

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

L'attrition dans les enquêtes sociales longitudinales : le cas de l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ)

Par :
Julien Bérard-Chagnon

Département de démographie
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M. Sc.)
en démographie

Octobre, 2007

© Julien Bérard-Chagnon, 2007



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
L'attrition dans les enquêtes sociales longitudinales : le cas de l'Étude longitudinale du
développement des enfants du Québec (ÉLDEQ)

Présenté par :
Julien Bérard-Chagnon

A été évalué par un jury composé des personnes suivantes :

Thomas K. Legrand
Président-rapporteur

Simona Bignami
Directrice de recherche

Bertrand Perron
Membre du jury

Mémoire accepté le
17 octobre 2007

RÉSUMÉ

Les récents développements dans le domaine des enquêtes sociales longitudinales au cours de la seconde moitié du XX^e siècle permettent aux chercheurs des sciences sociales de compter sur des sources de données d'un potentiel incroyable. Cependant, un problème méthodologique majeur guette ces enquêtes : la perte prématurée de répondants au fil du temps. Ce phénomène, appelé attrition, donne bien des maux aux spécialistes des enquêtes longitudinales puisqu'il peut nuire à la qualité et à la comparabilité des données recueillies. Ce mémoire porte sur la problématique de l'attrition dans les grandes enquêtes longitudinales par l'entremise d'une analyse de l'attrition dans une enquête longitudinale québécoise actuellement en cours : l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ). Cette recherche a tout d'abord permis de montrer que les décrocheurs de l'ÉLDEQ présentent des caractéristiques très différentes des familles toujours dans l'enquête. Toutefois, au niveau multidimensionnel, parmi les variables étudiées, seuls le niveau d'éducation de la mère, son statut d'immigrante et le fait qu'elle ait eu au moins un avortement influencent significativement la probabilité de décrocher de l'enquête. En outre, le principal effet de l'attrition sur l'ÉLDEQ est l'érosion de l'échantillon puisqu'en contrôlant pour les caractéristiques des familles, seuls le statut d'immigrante de la mère et le fait qu'elle ait eu au moins un avortement sont biaisés de façon significative par l'attrition. Finalement, bien qu'ils ne soient pas parfaitement ajustés pour contrer l'attrition, les poids longitudinaux de l'ÉLDEQ sont néanmoins pertinents dans la correction des biais d'attrition.

Mots-clés : attrition, ÉLDEQ, enquêtes longitudinales, biais, poids longitudinaux

ABSTRACT

Longitudinal surveys are increasingly used in the social sciences to describe behavior dynamics, to identify the influence of past on current behaviors, and to make stronger causal inferences than it is possible with cross-sectional surveys. One of the main weaknesses of longitudinal surveys is that they are prone to attrition, that is, the loss of study subjects over time that is due to respondents leaving the study prematurely or being unable to participate in one or more of its follow-ups. Attrition is an important concern because it may seriously affect data quality if attritors are a non random subsample of the study population. This thesis addresses these issues by focusing on a large, ongoing panel survey: the Longitudinal Survey of Child Development in Québec (l'Étude longitudinale du développement des enfants du Québec, ÉLDEQ). The results indicate that attritors in the ÉLDEQ have different characteristics than non-attritors, since they are generally poorer and more likely to change residence. The level of education of the child's mother, her immigration status, and whether she had at least a prior abortion significantly influence the probability of attrition. Yet overall attrition does not bias estimated coefficients in multivariate analyses when controlling for the household's background characteristics, with the exception of two outcomes: the child's mother's immigration status, and whether she had at least a prior abortion. The main effect of attrition in the ÉLDEQ data is thus to reduce the sample size for analysis and, consequently, reduce the power of statistical inferences based on these data. Although the longitudinal weights included in the ÉLDEQ data do not perfectly correct for attrition in univariate analyses, they are nonetheless appropriate to correct for attrition bias.

Keywords : attrition, ÉLDEQ, longitudinal surveys, bias, longitudinal weights

TABLE DES MATIÈRES

TABLE DES MATIÈRES	V
LISTE DES TABLEAUX	VII
LISTE DES FIGURES	IX
LISTE DES ABBRÉVIATIONS	X
REMERCIEMENTS	XII
CHAPITRE 1 : PROBLÉMATIQUE, REVUE DE LA LITTÉRATURE ET OBJECTIFS VISÉS	1
1.1 DÉVELOPPEMENT DES ENQUÊTES LONGITUDINALES	1
1.1.1 <i>Principaux avantages des données longitudinales</i>	3
1.1.2 <i>Principaux inconvénients des données longitudinales</i>	4
1.2 DESCRIPTION DU PHÉNOMÈNE DE L'ATTRITION	4
1.2.1 <i>Données manquantes</i>	5
1.2.2 <i>Définition de l'attrition</i>	6
1.2.3 <i>Causes de l'attrition</i>	7
1.2.4 <i>Effets de l'attrition</i>	9
1.3 MÉTHODES D'ANALYSE DE L'ATTRITION.....	10
1.3.1 <i>Techniques pour détecter l'attrition</i>	11
1.3.2 <i>Techniques pour contrer l'attrition</i>	14
1.4 VARIABLES AFFECTÉES PAR L'ATTRITION	17
1.5 OBJECTIFS VISÉS	18
CHAPITRE 2 : PRÉSENTATION DE L'ENQUÊTE	21
2.1 DESCRIPTION DE L'ÉTUDE LONGITUDINALE DU DÉVELOPPEMENT DES ENFANTS DU QUÉBEC	21
2.1.1 <i>Présentation de l'Étude longitudinale du développement des enfants du Québec</i>	21
2.1.2 <i>Population visée, base de sondage et échantillonnage</i>	22
2.1.3 <i>Questionnaires</i>	23
2.1.4 <i>Prise en compte de l'attrition dans l'ÉLDEQ</i>	24
2.1.5 <i>L'accès aux données de l'ÉLDEQ</i>	27
2.2 CONCLUSION.....	27
CHAPITRE 3 : L'ATTRITION DANS L'ÉLDEQ	28
3.1. LA NON-RÉPONSE PAR UNITÉ DANS L'ÉLDEQ	28
3.1.1 <i>L'ampleur de la non-réponse par unité dans l'ÉLDEQ</i>	28
3.1.2 <i>Schémas de participation</i>	29
3.1.3 <i>Comparaison des deux définitions de non-réponse</i>	31
3.2 L'ATTRITION DANS L'ÉLDEQ	32
3.2.1 <i>Définition de l'attrition</i>	33
3.2.2 <i>Ampleur de l'attrition dans l'ÉLDEQ</i>	34
3.2.3 <i>Non-réponse versus attrition</i>	35

3.3	L'ATTRITION SELON LES CARACTÉRISTIQUES INDIVIDUELLES DES RÉPONDANTS	37
3.4	RATIOS D'ATTRITION	39
3.4.1	<i>Méthodologie</i>	40
3.4.2	<i>Résultats</i>	41
3.5	CONCLUSION.....	43
CHAPITRE 4 : LES CARACTÉRISTIQUES DES DÉCROCHEURS		44
4.1	MÉTHODOLOGIE.....	44
4.1.1	<i>Groupes d'attrition</i>	44
4.2	RÉSULTATS	45
4.3	CONCLUSION.....	48
CHAPITRE 5 : LES FACTEURS QUI INFLUENCENT LA PROBABILITÉ D'ATTRITION.....		49
5.1	MÉTHODOLOGIE.....	49
5.2	RÉSULTATS	51
5.3	CONCLUSION.....	56
CHAPITRE 6 : LES BIAIS CAUSÉS PAR L'ATTRITION.....		57
6.1	MÉTHODOLOGIE.....	57
6.2	RÉSULTATS	58
6.3	CONCLUSION.....	60
CHAPITRE 7 : VALIDATION DES POIDS LONGITUDINAUX		61
7.1	MÉTHODOLOGIE.....	61
7.1.1	<i>Description des poids longitudinaux</i>	61
7.1.2	<i>Méthodes d'analyse</i>	62
7.2	RÉSULTATS	63
7.2.1	<i>Analyse unidimensionnelle rétrospective</i>	63
7.2.2	<i>Analyse unidimensionnelle prospective</i>	66
7.3	CONCLUSION.....	68
DISCUSSION ET CONCLUSION.....		69
BIBLIOGRAPHIE.....		78
ANNEXES.....		XIII
ANNEXE MÉTHODOLOGIQUE		XXXIX

LISTE DES TABLEAUX

TABLEAU 3.1 DISTRIBUTION DES RÉPONDANTS À L'ÉLDEQ SELON LE NOMBRE DE VOLETS RÉPONDUS ET POURCENTAGE POUR CHAQUE VOLET	30
TABLEAU 3.2 SCHÉMAS DE PARTICIPATION DES RÉPONDANTS DE L'ÉLDEQ	31
TABLEAU 3.3 TAUX DE NON-RÉPONSE (%) DE L'ÉQUIPE DE L'ÉLDEQ ET DIFFÉRENCES (%) ENTRE LES TAUX DE NON-RÉPONSE CUMULATIFS (%) DE L'ÉQUIPE DE L'ÉLDEQ ET CEUX DE CE MÉMOIRE SELON LE VOLET DE L'ÉLDEQ	32
TABLEAU 3.4 TAUX D'ATTRITION CUMULATIFS (%) CHOISIS	38
TABLEAU 3.5 QUELQUES RATIOS D'ATTRITION CHOISIS CONSTRUITS À PARTIR DU VOLET 7 DE L'ÉLDEQ	42
TABLEAU 4.1 DIFFÉRENCES DANS LES CARACTÉRISTIQUES SELON LES TROIS GROUPES DE DÉCROCHEURS POUR QUELQUES VARIABLES CHOISIES (%)	47
TABLEAU 5.1 VARIABLES D'INTÉRÊT ET VARIABLES DE CONTRÔLE DES RÉGRESSIONS <i>PROBIT</i> MODÉLISANT LA PROBABILITÉ D'ATTRITION	50
TABLEAU 5.2 COEFFICIENTS DES RÉGRESSIONS <i>PROBIT</i> SELON CERTAINES VARIABLES D'INTÉRÊT ET DE CONTRÔLE SÉLECTIONNÉES	53
TABLEAU 5.3 PROBABILITÉ PRÉDITES DE DÉCROCHER DE L'ÉLDEQ ET VARIATION MARGINALE (Δp) DE CETTE PROBABILITÉ POUR LA VARIABLE DE CONTRÔLE SIGNIFICATIVE AINSI QUE POUR L'ENSEMBLE DES VARIABLES D'INTÉRÊT	55
TABLEAU 6.1 TESTS BGLW TRANSVERSAUX POUR DES VARIABLES D'INTÉRÊT ET DE CONTRÔLE SÉLECTIONNÉES	59
TABLEAU 7.1 FRÉQUENCES (%) DE QUELQUES VARIABLES FIXES POUR LES VOLETS UN ET HUIT, DONNÉES PONDÉRÉES	65
TABLEAU 7.2 CARACTÉRISTIQUES (%) DES DÉCROCHEURS ET DES NON-DÉCROCHEURS EN E5 (DONNÉES PONDÉRÉES)	67
TABLEAU A.3.1 TABLEAU COMPLET DES TAUX D'ATTRITION CUMULATIFS (%) PAR VOLET	XIV
TABLEAU A.3.2 TABLEAU COMPLET DES RATIOS D'ATTRITION	XVIII
TABLEAU A.4.1 DIFFÉRENCES DANS LES CARACTÉRISTIQUES SELON LES TROIS GROUPES DE DÉCROCHEURS (%)	XXI

TABLEAU A.5.1 TABLEAU COMPLET DES COEFFICIENTS DES RÉGRESSIONS <i>PROBIT</i> SELON LES VARIABLES D'INTÉRÊT.....	XXIV
TABLEAU A.6.1 TABLEAU COMPLET DES TESTS <i>BGLW</i> POUR DIVERSES VARIABLES D'INTÉRÊT.....	XXIX
TABLEAU A.7.1 FRÉQUENCES (%) POUR DES VARIABLES FIXES OU QUASI-FIXES POUR LES VOLETS UN ET HUIT, DONNÉES PONDÉRÉES.....	XXXIII
TABLEAU A.7.2 CARACTÉRISTIQUES (%) DES DÉCROCHEURS ET DES NON-DÉCROCHEURS EN E5 (DONNÉES PONDÉRÉES) – TESTS DU KHI-DEUX.....	XXXV
TABLEAU A.7.3 CARACTÉRISTIQUES (EN MOYENNE) DES DÉCROCHEURS ET DES NON- DÉCROCHEURS EN E5 (DONNÉES PONDÉRÉES) – TESTS T.....	XXXVIII

LISTE DES FIGURES

FIGURE 3.1 MÉNAGES RÉPONDANT À L'ÉLDEQ ET TAUX DE NON-RÉPONSE PAR UNITÉ CUMULATIF (%) SELON LE VOLET.....	29
FIGURE 3.2 EXEMPLES DES DIFFÉRENTS TYPES DE DÉCROCHEURS	34
FIGURE 3.3 TAUX D'ATTRITION CUMULATIF (%) DE L'ÉLDEQ SELON LE VOLET.....	35
FIGURE 3.4 NON-RÉPONDANTS PAR UNITÉ ET DÉCROCHEURS SELON LE VOLET DE L'ÉLDEQ	36
FIGURE 4.1 LES TROIS GROUPES D'ATTRITION ÉTUDIÉS SELON LES VOILETS RÉPONDUS PAR CHAQUE GROUPE	45
FIGURE AA.1 FONCTION DE RÉPARTITION DE LA LOI NORMALE (0,1)	XLIV

LISTE DES ABBRÉVIATIONS

ANOVA : Analyse de variance

BGLW : Beckett, Gould, Lillard et Welch

BHPS : British Household Panel Survey

CADRISQ : Centre d'accès aux données de recherche de l'Institut de la statistique du Québec

CLS : Centre for Longitudinal Studies

CPS : Current Population Survey

ECHP : European Community Household Panel

EDTR : Enquête sur la dynamique du travail et du revenu

ÉLDEQ : Étude longitudinale du développement des enfants du Québec

ELIC : Enquête longitudinale auprès des immigrants du Canada

ELNEJ : Enquête longitudinale nationale sur les enfants et les jeunes

ENSP : Enquête nationale sur la santé de la population

ISQ : Institut de la statistique du Québec

MCO : Moindres carrés ordinaires (régression)

MEQ : Ministère de l'Éducation du Québec (est devenu le MELS)

MSSS : Ministère de la Santé et des Services sociaux

NCDS : National Child Development Study

NELS : National Education Longitudinal Study

NLS : National Longitudinal Survey

NLSY : National Longitudinal Survey of Youth

NSFH : National Survey of Families and Households

PSID : Panel Survey of Income Dynamics

QIRI : Questionnaire informatique rempli par l'intervieweur

À mes grands-parents

REMERCIEMENTS

Puisque la réalisation d'un tel ouvrage ne peut se faire sans le constant support de son entourage, je tiens à remercier tous ceux qui furent associés, de près ou de loin, à la rédaction de ce mémoire de maîtrise en démographie.

J'aimerais tout d'abord remercier ma directrice de recherche, madame Simona Bignami, qui, par sa rigueur intellectuelle, sa patience, sa grande expertise et aussi sa gentillesse, fut une véritable source d'inspiration et de motivation pour moi dans la réalisation de ce mémoire.

J'aimerais également souligner la contribution financière du Fonds québécois de recherche sur la société et la culture (FQRSC) lors de la rédaction de ce mémoire.

J'aimerais aussi remercier mes amis, mes collègues étudiants, les membres de ma famille et le personnel du département de démographie de l'Université de Montréal, qui ont toujours été présents afin de me supporter et de m'encourager dans les bons moments tout comme dans les moments plus difficiles. J'aimerais particulièrement remercier Guillaume Marois, mon camarade dans les cours de maîtrise, et Ariane Lanoue, ma colocataire de bureau !

J'aimerais également souligner l'apport de madame Lucie Gingras, responsable du CADRISQ, pour son aide méthodologique précieuse avec la manipulation des différents fichiers de l'ÉLDEQ.

Finalement, je ne peux passer sous silence la contribution de ma copine Cindy, dont les suggestions furent à la fois très utiles et très pertinentes, et qui a même lu ce mémoire dans sa totalité sans décrocher prématurément !

CHAPITRE 1 : PROBLÉMATIQUE, REVUE DE LA LITTÉRATURE ET OBJECTIFS VISÉS

Les développements des enquêtes longitudinales réalisés au cours de la seconde moitié du XX^e siècle permettent aux chercheurs des sciences sociales de compter sur des sources de données possédant un potentiel énorme. Cependant, ces enquêtes sont sujettes à un problème méthodologique majeur : l'érosion graduelle de l'échantillon. Ce phénomène, appelé attrition, peut grandement nuire à la qualité des données longitudinales. Afin de bien saisir la totalité de la problématique de l'attrition, il ne faut pas seulement s'attarder à ses conséquences mais également à ses causes et aux techniques permettant d'identifier l'attrition et de la contrer. Ce chapitre sera donc divisé en cinq sections. Pour débiter, les caractéristiques des enquêtes longitudinales, leurs atouts et leurs faiblesses seront brièvement décrits. Par la suite, l'attrition sera définie en mettant l'accent sur les causes et les conséquences de ce phénomène. Les variables qui sont généralement reconnues comme étant liées à l'attrition dans la littérature seront ensuite introduites, en mettant l'emphase sur les enquêtes longitudinales portant sur les ménages et sur les enfants. Finalement, les objectifs de ce mémoire seront présentés et décrits.

1.1 DÉVELOPPEMENT DES ENQUÊTES LONGITUDINALES

Depuis le milieu du dernier siècle, on assiste à d'énormes développements dans l'utilisation des données quantitatives en sciences sociales grâce aux grandes enquêtes longitudinales. Ces enquêtes se distinguent des enquêtes transversales par le suivi d'un même échantillon à travers le temps, permettant ainsi d'obtenir des informations sur les mêmes répondants à des moments différents. Elles se distinguent également des systèmes de surveillance démographique comme le projet Matlab alors qu'elles permettent de répondre à des questions très spécifiques et très variées en plus d'être fondées sur une méthodologie très différente. Les deux premières enquêtes longitudinales de grande envergure qui furent mises sur pied sont deux enquêtes américaines datant de la décennie 1960 : le Panel Survey of Income Dynamics (PSID) et le National Longitudinal Survey (NLS). Le PSID est une enquête qui étudie

principalement les aspects dynamiques des comportements économiques des Américains alors que le NLS est une série d'enquêtes qui s'intéressent aux transitions sociales et économiques dans la vie des Américains en ciblant initialement les jeunes et les gens plus âgés. Ces deux enquêtes pionnières furent d'ailleurs utilisées pour développer diverses méthodes longitudinales et ont servi d'exemples pour plusieurs autres enquêtes. Vers la fin de la décennie 1980, des enquêtes ciblant des problématiques plus précises furent aussi réalisées, complétant ainsi les informations des enquêtes en cours à l'époque mais permettant aussi d'offrir davantage de renseignements sur des enjeux plus particuliers. C'est notamment le cas du National Survey of Families and Households (NSFH), une enquête américaine consacrée à l'étude des transitions familiales et du National Educational longitudinal Survey (NELS), une enquête ciblant les étudiants américains de 8^e année. Parallèlement à ces nouvelles enquêtes, des grandes enquêtes longitudinales furent aussi réalisées à l'extérieur des États-Unis. Ainsi, en Angleterre, le Centre for Longitudinal Studies (CLS) réalisa notamment le British Household Panel Survey (BHPS) et le National Child Development Study (NCDS) afin de mieux comprendre les dynamiques économiques et sociales des ménages anglais et du développement des enfants britanniques. Ces enquêtes se répandirent ensuite à travers plusieurs autres pays européens lors de la création du European Community Household Panel (ECHP).

Au Canada, l'aventure des enquêtes longitudinales débuta durant la dernière décennie du 20^e siècle avec une série d'enquêtes abordant diverses thématiques très variées. Par exemple, l'Enquête longitudinale auprès des immigrants du Canada (ELIC) cible une cohorte d'immigrants venus s'installer au Canada alors que l'Enquête sur la dynamique du travail et du revenu (EDTR) s'intéresse aux activités économiques des ménages canadiens et l'Enquête nationale sur la santé de la population (ENSP) étudie les divers éléments entourant la santé des Canadiens. En outre, l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ) et l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) abordent la thématique du développement des enfants dans le contexte québécois et canadien. Ces enquêtes laissèrent entrevoir une véritable révolution dans la recherche en sciences sociales au Canada. Elles étaient tellement

populaires à cette époque qu'il fallait même presque justifier le fait de ne pas mener une enquête longitudinale lors de la tenue d'une nouvelle enquête (Picot, 2005).

Néanmoins, comme toutes les autres sources de données, les enquêtes longitudinales possèdent des forces et des faiblesses qu'il est important de connaître avant d'envisager leur utilisation.

1.1.1 Principaux avantages des données longitudinales

Les données longitudinales possèdent évidemment des atouts substantiels pour les chercheurs des sciences sociales et pour les décideurs des sphères publiques car elles permettent d'obtenir une meilleure compréhension des phénomènes se déroulant dans le temps. En effet, les données longitudinales se distinguent particulièrement des données transversales dans leur grande utilité pour décrire les dynamiques des comportements, pour identifier l'influence de comportements passés sur les comportements actuels ainsi que pour établir des liens de causalité plus solides.

Ces nombreux atouts font en sorte que l'utilisation de données longitudinales est encouragée dans plusieurs domaines très variés, de la démographie et l'économie à la nutrition et à l'éducation physique. Par exemple, des enquêtes longitudinales canadiennes furent utilisées afin de démontrer le lien entre le manque de nourriture et l'obésité chez les enfants d'âge préscolaire (Dubois, 2006), l'impact de la pauvreté à long terme sur la santé physique des enfants (Jones et *al.*, 2002) ou bien le rôle de l'environnement familial et économique dans l'apprentissage du vocabulaire chez les enfants québécois (Neill, 2005b). En démographie, l'utilisation de données longitudinales est particulièrement encouragée dans l'étude des transitions familiales puisque ces données permettent de mieux saisir l'évolution de la structure familiale au fil du temps (Miller et *al.*, 1995) ainsi que dans la mesure de l'impact du passé familial sur le futur des enfants (Marcil-Gratton et *al.*, 2003). Par exemple, le recours aux données du National Longitudinal Survey of Youth (NLSY), une enquête longitudinale américaine pionnière, a mis en lumière l'impact longitudinal du mariage et du divorce sur l'état financier des individus (Zagorsky, 2005). Ces exemples illustrent donc la

pertinence de l'utilisation des données longitudinales dans la résolution de problématiques de recherche actuelles en démographie et dans plusieurs autres disciplines différentes. En somme, les enquêtes longitudinales possèdent des qualités indéniables pour l'avancement de la recherche scientifique, notamment dans les sciences sociales.

1.1.2 Principaux inconvénients des données longitudinales

Malgré leurs avantages certains, les enquêtes longitudinales possèdent également plusieurs faiblesses. Ce type d'enquêtes demande tout d'abord un investissement initial très élevé de la part des chercheurs et des organismes subventionnaires. De plus, puisque l'intérêt de l'analyse longitudinale réside dans le suivi des individus au fil du temps, les chercheurs et les organismes subventionnaires ne pourront voir le fruit de leur investissement que plusieurs années après le début de l'enquête. Ceci fait en sorte qu'il est commun de voir plusieurs organismes subventionnaires investir dans la même enquête, entraînant souvent une forte augmentation du nombre de questions, alourdissant ainsi le fardeau du répondant. Ceci peut même aller jusqu'à faire dévier l'enquête de ses objectifs initiaux alors que les organismes subventionnaires désirent obtenir des données sur les problématiques qui les intéressent. En outre, plusieurs questions de recherche nécessitent des données longitudinales s'étalant sur plusieurs années, ce qui demande alors un financement très élevé pour maintenir l'enquête sur une longue période. Finalement, ces enquêtes nécessitent de relever plusieurs défis méthodologiques majeurs, parmi lesquels figure l'attrition (Picot, 2005).

1.2 DESCRIPTION DU PHÉNOMÈNE DE L'ATTRITION

L'attrition dans les enquêtes longitudinales est essentiellement un problème de données manquantes. Cependant, l'attrition se distingue des autres types de données manquantes non seulement par ses conséquences mais aussi par ses causes. Pour cette raison, cette section débutera en distinguant l'attrition des autres types de données manquantes pour ensuite définir l'attrition de façon théorique. Par la suite, les causes et les conséquences de l'attrition seront présentées.

1.2.1 Données manquantes

Les données manquantes sont un problème très fréquent et très important dans la collecte et dans l'utilisation de données d'enquête, autant pour les enquêtes longitudinales que pour les enquêtes transversales. Il est possible de diviser la non-réponse en trois catégories, dont les deux premières sont communes aux enquêtes longitudinales et transversales et la troisième est unique aux enquêtes longitudinales. Tout d'abord, la non-réponse par unité, qui survient lorsque le sujet rate un passage d'une enquête longitudinale mais répond à au moins un des passages suivants ou refuse simplement de répondre à une certaine enquête, longitudinale ou transversale. Ce type de non-réponse peut être causé par un déplacement temporaire (comme un voyage), suite à un manque de temps ou d'intérêt de la part du sujet, ou bien à un manque de flexibilité dans les vagues de collecte de l'enquête. Alternativement, surtout dans le cas longitudinal, ce type de non-réponse peut être causé par une période d'instabilité au cours de la vie de l'individu suite à un événement important (divorce, cessation d'emploi, changement d'emploi, déménagement, etc.). Cette dernière raison est particulièrement préoccupante pour la qualité des données provenant d'enquêtes longitudinales étant donné le désir de capter les transitions dans la vie des individus à travers ces enquêtes. Le second type de non-réponse est la non-réponse par item, qui se produit lorsqu'une valeur est manquante pour un sujet même si celui-ci a participé à l'enquête. Ceci peut souvent se produire pour des questions sensibles (comme le revenu du sujet ou l'adresse postale du sujet), par manque de connaissance du sujet ou simplement suite à la fatigue dans le cas d'enquêtes nécessitant beaucoup de temps de la part du répondant. Il s'agit du type de non-réponse le plus étudié puisque beaucoup de chercheurs travaillent seulement avec quelques variables d'intérêt et non sur la totalité des variables d'une enquête, ce qui entraîna alors un développement de méthodes pour traiter ce type de non-réponse (comme les différentes formes d'imputation). Le dernier type de non-réponse, qui est un cas particulier de non-réponse par unité unique aux enquêtes longitudinales, est l'attrition. Ce dernier type de non-réponse sera défini dans la section suivante.

1.2.2 Définition de l'attrition

De façon générale, l'attrition peut être définie comme étant la fin prématurée de la participation d'un sujet à une enquête longitudinale (Schmitz, 2002). Généralement, les répondants qui ne sont plus ciblés par l'enquête sont inadmissibles et ne sont pas comptabilisés comme étant des décrocheurs. Ceci engendre alors des données manquantes, lesquelles peuvent être aléatoires (la probabilité d'attrition du sujet est indépendante de la réponse qui aurait été observée), complètement aléatoires (la donnée manquante est aléatoire en plus d'être indépendante des mesures précédentes) ou pas aléatoires. Les biais causés par l'attrition se produisent lorsque les données manquantes dues à l'attrition ne sont pas aléatoires. En effet, si l'attrition se produit aléatoirement, on peut alors supposer que les décrocheurs forment un sous-groupe aléatoire de l'échantillon et que l'attrition ne cause pas de biais systématique dans l'échantillon. Dans le cas contraire, il est possible que l'attrition engendre des biais dans l'échantillon, nuisant alors à la qualité des données longitudinales recueillies. Or, l'hypothèse que des données soient manquantes de façon aléatoire ou complètement aléatoire est très irréaliste (Rubin, 1976). Une définition plus mathématique de l'attrition se trouve en annexe (voir annexe méthodologique 1).

Dans la pratique, la définition exacte du concept de l'attrition varie régulièrement de chercheur en chercheur (Rendtel, 2002). Par exemple, dans leur analyse de l'attrition du PSID, Fitzgerald et *al.* (1998) considèrent l'effet des décès et des émigrants quittant la population de base de l'enquête sur l'érosion de l'échantillon dans leur analyse descriptive même si ces répondants ne font plus partie de la population ciblée par le PSID alors que Chen (2005) distingue les répondants du PSID qui sont toujours dans l'enquête de ceux qui ont décroché à au moins un passage ou qui sont décédés¹. D'un autre côté, Watson (2003) ne considère pas les gens exclus de la population ciblée (qui sont principalement les décédés et les émigrants) comme des décrocheurs, bien qu'elle considère les ménages ayant raté au moins un passage du ECHP comme étant des

¹ Il utilise la terminologie « always in » pour les répondants qui ont répondu à tous les passages de l'enquête et « ever out » pour définir les répondants qui sont décédés ou qui ont décroché de l'échantillon à au moins un passage du PSID.

décrocheurs. Ces différentes définitions, qui peuvent parfois faire varier considérablement les résultats, démontrent non seulement qu'il n'y a pas de consensus sur la définition de l'attrition mais aussi l'importance de bien définir ce phénomène avant l'analyse.

1.2.3 Causes de l'attrition

Il est possible qu'un sujet décroche d'une enquête longitudinale pour plusieurs raisons. Cependant, il est possible de classer ces raisons en deux grands groupes : l'attitude du répondant par rapport aux enquêtes et la collecte sur le terrain. Ces deux causes de l'attrition seront décrites ici.

Premièrement, l'attitude du répondant face aux enquêtes, pouvant être vue comme étant la probabilité du sujet de refuser de participer à une enquête, peut causer divers niveaux d'attrition chez les sujets qui possèdent différents niveaux de collaboration aux enquêtes. Cette attitude est communément mesurée de façon indirecte par des caractéristiques socioéconomiques du répondant comme le niveau de scolarité, l'état matrimonial et l'âge du répondant (Behr et *al.*, 2005). Contrairement à la seconde cause majeure de l'attrition, qui possède une grande partie aléatoire (les gens qui changent d'adresse n'ont pas, par exemple, tous les mêmes caractéristiques), l'attrition provenant de cette cause est plus dangereuse car elle implique directement que les caractéristiques des décrocheurs ne sont probablement pas les mêmes que celles des répondants toujours dans l'enquête.

Deuxièmement, les procédures mises en application dans la collecte des données peuvent aussi favoriser l'attrition. En effet, la capacité des enquêteurs sur le terrain à retenir les sujets est un facteur déterminant de l'ampleur de l'attrition dans une enquête longitudinale. Ceci peut être réalisé de plusieurs façons, comme en offrant une grande flexibilité aux répondants pour les entrevues, en tentant de rejoindre plusieurs fois les répondants, en minimisant le plus possible la durée des entrevues ou bien en tentant de retracer les répondants qui ont changé d'adresse dans le but de continuer leur suivi longitudinal. Effectivement, Behr et *al.* (2005) ont remarqué l'importance de la collecte

de données sur le terrain dans leur analyse de l'attrition dans l'ECHP alors que le changement d'intervieweur et le changement de résidence du sujet furent les deux variables ayant l'effet le plus important sur l'attrition. De plus, Rendtel (2002) parle de l'importance de la collecte de données sur la rétention des répondants en montrant l'effet que le changement du mode d'entrevue ainsi que les changements dans les questionnaires peuvent avoir sur la rétention des répondants. L'attrition provenant de cette cause peut naturellement biaiser l'échantillon alors que les gens très mobiles ou très occupés risquent de quitter plus facilement le suivi longitudinal. Néanmoins, il s'agit d'une cause qui pourra graduellement être contrôlée, du moins en partie, dans le futur alors que la littérature sur l'attrition détaillera des méthodes de collecte de données qui furent efficaces pour limiter l'attrition dans des enquêtes particulières.

En terminant, il faut préciser que les pertes démographiques, qui peuvent par exemple se produire suite au décès du sujet ou bien suite à une migration amenant le sujet à l'extérieur de la population étudiée par l'enquête, ne sont pas toujours considérées comme de l'attrition. Bien que ces pertes engendrent une diminution de l'échantillon, puisque ces répondants ne font parfois plus partie de la population ciblée par l'enquête (selon les spécificités de l'enquête), ils ne sont pas considérés comme étant des décrocheurs dans plusieurs enquêtes. C'est notamment le cas du ECHP, qui considère que les sujets qui meurent ou qui quittent l'Union européenne ne sont pas des décrocheurs (Behr et *al.*, 2005). Cependant, il faut se rappeler que dans une enquête sévèrement touchée par les pertes démographiques, comme dans le cas d'une enquête se déroulant sur plusieurs dizaines d'années ou bien une enquête ciblant une population particulièrement à risque d'émigrer ou de décéder (tels que les immigrants ou les personnes âgées), la diminution de la taille de l'échantillon peut quand même nuire considérablement à la qualité des données même si les pertes démographiques ne sont pas considérées comme de l'attrition. Finalement, il faut préciser que les pertes associées à une stratégie d'échantillonnage particulière, comme dans le cas de groupes rotatifs, ne sont pas considérés comme des décrocheurs.

1.2.4 Effets de l'attrition

L'attrition peut avoir deux types d'effets sur un échantillon longitudinal : un effet de quantité et un effet de distribution (Watson, 2003). Ces deux effets, qui peuvent évidemment nuire à la qualité des données longitudinales, seront décrits ici.

La quantité d'attrition se réfère à la diminution de la taille échantillonnale provenant de l'érosion graduelle de l'échantillon. Cette perte de sujets contribue directement à réduire l'efficacité des différentes techniques statistiques car la précision des différents estimateurs statistiques (coefficients de régression, intervalles de confiance, etc.) et la puissance statistique des tests d'hypothèses sont fonction de la taille de l'échantillon. En outre, l'érosion de l'échantillon causée par l'attrition peut rendre l'étude de petites sous-populations particulièrement ardue, surtout pour les sous-populations très affectées par l'attrition. Finalement, bien que cet effet soit facilement identifiable, il demeure toutefois un problème grave dans la qualité des données des enquêtes longitudinales.

La distribution de l'attrition est le caractère non-aléatoire de la perte de répondants. Effectivement, si les décrocheurs présentent des caractéristiques différentes des sujets qui demeurent dans l'échantillon, la représentativité de l'échantillon par rapport à la population étudiée sera graduellement réduite alors que certains sujets décrocheront de l'enquête. Par exemple, si les gens peu scolarisés sont plus enclins à abandonner une enquête longitudinale, il s'en suivra qu'à la longue, cette catégorie de répondants sera sous-représentée. En outre, la distribution de l'attrition peut modifier les liens existant entre plusieurs variables. En effet, si les gens présentant certaines caractéristiques sont plus enclins à décrocher d'une enquête longitudinale, ceci peut avoir des répercussions sur d'autres variables liées à ces caractéristiques. Ceci aura donc comme effet de modifier la force de la covariance entre plusieurs variables de l'enquête, créant alors des biais dans les données. Par exemple, les liens entre la dépression et la durée du mariage peuvent être biaisés si les gens dépressifs sont plus enclins à abandonner le suivi longitudinal. Alors que l'effet de la quantité d'attrition est bien visible, l'impact de la distribution de l'attrition est beaucoup plus difficile à vérifier et à

étudier (Watson, 2003). Les biais de l'attrition peuvent également avoir de grandes conséquences sur la validité des résultats obtenus avec les données longitudinales.

Néanmoins, plusieurs auteurs (Alderman et *al.*, 2000, Hoddinott et *al.*, 2002, Fitzgerald et *al.*, 1998) arrivent à la conclusion que l'attrition, bien qu'importante dans certaines enquêtes, ne biaise pas de façon significative les coefficients obtenus à l'aide des analyses multidimensionnelles, malgré la présence de biais parfois très importants au niveau unidimensionnel. Cependant, il faut rappeler que l'effet de la réduction de l'échantillon est toujours présent. De plus, puisque chaque enquête longitudinale est différente, il demeure nécessaire d'effectuer une analyse de l'attrition pour chaque enquête afin de s'assurer de la qualité des données utilisées.

En somme, l'attrition peut affecter un échantillon longitudinal par sa quantité et par sa distribution. En outre, malgré des résultats rassurants dans la littérature, il demeure toutefois nécessaire de développer des outils pour identifier et contrer ces problèmes.

1.3 MÉTHODES D'ANALYSE DE L'ATTRITION

Deux philosophies existent dans l'étude de l'attrition. La première philosophie consiste à prédire l'attrition à l'aide de différentes variables explicatives dans le but de minimiser l'attrition. Cette approche de minimisation est surtout utilisée par les producteurs de données, qui veulent réduire le plus possible l'ampleur de l'attrition. La seconde philosophie, qui est une approche d'évaluation du biais de l'attrition, est principalement utilisée par les chercheurs utilisant des données provenant d'enquêtes longitudinales dans leurs travaux de recherche. Ces chercheurs sont donc surtout intéressés à savoir si leurs variables d'intérêt sont biaisées par l'attrition. Ces deux approches ne sont pas du tout opposées, mais se distinguent plutôt par rapport aux objectifs de l'utilisateur des données longitudinales. Cette section présente d'abord les techniques les plus connues pour détecter l'attrition pour ensuite introduire les techniques permettant de contrer les effets de l'attrition. Il faut cependant noter que la définition de l'attrition qui est choisie en pratique peut grandement influencer l'efficacité des diverses techniques présentées

dans cette section, d'où l'importance de prendre le temps nécessaire pour bien définir ce phénomène.

1.3.1 Techniques pour détecter l'attrition

Les techniques pour détecter les effets causés par l'attrition sont multiples (Rendtel, 2002). Elles visent cependant toutes à identifier l'ampleur de l'impact de la quantité d'attrition et surtout de la distribution de l'attrition. Les techniques les plus connues et les plus populaires sont introduites dans cette section.

L'approche classique consiste à comparer les caractéristiques des sujets ayant décroché de l'enquête et des sujets qui sont demeurés dans l'enquête pour certaines variables d'intérêt (Miller, 1995). Si on parvient à repérer des différences statistiquement significatives entre les décrocheurs et les non-décrocheurs, ceci voudra donc dire que l'attrition ne se produit pas de façon aléatoire. Ceci peut être réalisé à l'aide d'outils statistiques de comparaison de moyennes et de distribution comme des tests de Student (tests t), des tests d'indépendance du khi-deux ou bien des analyses de variance (ANOVA) à un facteur. Par exemple, Hiskey et Troop (2002) utilisent cette technique dans leur étude de l'attrition dans une enquête longitudinale sur Internet alors qu'Hoddinot *et al.* (2002) débutent leur analyse de l'attrition en montrant les différences dans les caractéristiques des sujets à l'aide de tests t .

De plus, une approche descriptive peut être réalisée en calculant des taux d'attrition (ou des taux de réponse) pour différentes variables (Behr *et al.*, 2005). Cette méthode relativement simple à appliquer permet de montrer l'ampleur de l'érosion de l'échantillon en plus de permettre l'identification rapide des variables et des catégories au sein des différentes variables où l'attrition est particulièrement élevée. Par exemple, grâce à cette méthode, Behr *et al.* (2005) parviennent à montrer que le statut matrimonial du répondant influence le taux d'attrition dans l'ECHP alors que les individus mariés tendent à décrocher moins souvent d'une enquête longitudinale que les célibataires. Cette méthode est cependant limitée par son caractère essentiellement transversal

puisque, pour une variable donnée, les sujets peuvent changer de catégorie au fil du temps.

Alternativement, Watson (2003) a développé un indicateur simple à calculer, qu'elle nomme ratio d'attrition, permettant d'évaluer l'ampleur du biais de distribution de l'attrition pour les différentes catégories des variables discrètes. Ce ratio est défini, pour une certaine catégorie d'une variable, comme étant le rapport entre la proportion de cette catégorie au premier passage et cette même proportion au dernier passage pour le sous-groupe de l'échantillon qui a répondu à tous les passages de l'enquête. À l'aide des ratios d'attrition, Watson a ainsi pu montrer que les couples avec des jeunes enfants et que les individus de 30 ans ou moins vivant seuls sont plus sujets à décrocher du suivi longitudinal de l'ECHP.

Il est aussi possible d'étudier l'attrition de façon multidimensionnelle avec des régressions logistiques et *probit*. Cette stratégie permet de modéliser l'attrition en tant que variable dépendante qui est expliquée par une série de variables indépendantes pour mesurer l'impact de différentes variables sur la probabilité d'attrition d'un sujet tout en contrôlant pour certaines caractéristiques. Behr et *al.* (2005) utilisent plusieurs régressions logistiques afin de comparer l'impact de l'attrition selon le passage de l'enquête et selon le pays pour l'ECHP. Fitzgerald et *al.* (1998) ainsi qu'Alderman et *al.* (2000) utilisent des régressions *probit* à partir des données du premier passage d'une enquête longitudinale afin de prédire la probabilité d'attrition d'un sujet. Cette dernière stratégie cadre bien avec les objectifs de l'approche de minimisation de l'attrition énoncée précédemment puisqu'elle permet de modéliser la probabilité d'attrition d'un sujet selon une gamme de variables indépendantes dans le but d'identifier les variables qui sont pertinentes dans l'explication de l'attrition.

Étant un phénomène longitudinal de nature, l'attrition peut également être évaluée à l'aide des modèles d'analyse de survie. Dans ces modèles, on modélise la durée de séjour du sujet dans l'échantillon ainsi que le risque de quitter celui-ci. Par exemple, Beckett et *al.* (1988) utilisent un modèle de risques proportionnels et un

modèle de Weibull dans leur analyse de l'attrition dans le PSID. Néanmoins, selon la définition de l'attrition qui est considérée, il peut être ardu de modéliser ce phénomène avec l'analyse de survie si plusieurs répondants effectuent un retour (ou même plusieurs retours) dans l'échantillon. De plus, les modèles de survie sont généralement plus complexes à appliquer que les modèles de régression classiques en plus de soulever des questions méthodologiques par rapport la façon dont le modèle traite l'hétérogénéité non observée.

Des procédures particulières furent également développées pour détecter les biais d'attrition. Une procédure bien connue est le test de Beckett, Gould, Lillard et Welch (BGLW) (Beckett *et al.*, 1988). Cette procédure consiste à estimer la variable d'intérêt, dont on désire évaluer le biais d'attrition, par les caractéristiques des répondants ainsi que par une variable dichotomique représentant l'attrition du sujet. L'intérêt du test est de vérifier si la variable dichotomique d'attrition est statistiquement significative ou non. Cette procédure cadre bien avec l'approche d'évaluation du biais d'attrition puisqu'elle permet de bien identifier si l'attrition joue un rôle significatif dans l'explication de la variable d'intérêt. Si c'est le cas, on peut ainsi avancer que la variable dépendante est biaisée par l'attrition. Finalement, Fitzgerald *et al.* (1998) ont montré que cette procédure est l'inverse de la procédure d'estimation directe de l'attrition qui fut présentée précédemment. La distinction entre les deux procédures se trouve ainsi dans les objectifs de l'utilisateur alors qu'un utilisateur intéressé à vérifier si une variable d'intérêt est biaisée par l'attrition aura intérêt à utiliser cette technique. Alderman *et al.* (2000) ont utilisé cette technique afin de mesurer le biais d'attrition dans trois enquêtes longitudinales effectuées dans des pays différents.

Alors que l'approche classique compare les caractéristiques des décrocheurs avec celles des non-décrocheurs, il peut également être possible de comparer les caractéristiques des répondants de l'enquête longitudinale avec les données provenant d'une autre source comme un registre de population ou un recensement. Cependant, cette approche est très limitée dans la mesure où son efficacité dépend grandement de la qualité de la source de données utilisée comme comparaison ainsi que des informations

qui sont contenues dans cette source. Or, dans un pays qui ne possède pas de registre de population comme le Canada, il faut s'en remettre aux données du recensement quinquennal canadien, dont le questionnaire de base (le questionnaire 2A) ne contient que quelques questions, ou bien à une autre grande enquête transversale et représentative de la population. C'est la stratégie adoptée par Chen (2005) et Beckett et *al.* (1988) alors qu'ils comparent les données du PSID avec celles du CPS (Current Population Survey). Cependant, l'utilisation de données censitaires pourrait s'avérer être un puissant outil dans l'ajustement des poids échantillonnaires longitudinaux (qui seront abordés à la prochaine section) pour l'attrition.

En somme, il existe un grand nombre de techniques pour détecter les effets de l'attrition. Mais aucune technique ne fait présentement l'unanimité, faisant en sorte que l'utilisation simultanée de plusieurs de ces techniques peut s'avérer comme la meilleure façon de bien détecter les biais causés par l'attrition.

1.3.2 Techniques pour contrer l'attrition

Tout comme pour la détection de l'attrition, plusieurs techniques existent afin de compenser pour les effets de l'attrition. La littérature met ainsi en lumière un certain nombre de techniques possédant des forces et des faiblesses uniques à chaque technique. Les techniques les plus connues et les plus utilisées sont présentées ici.

Tout d'abord, une ancienne procédure pour traiter l'attrition consistait à supprimer les sujets manquants. Cependant, Little et Rubin (1987) ont montré les grands dangers de procéder de cette façon. En effet, en plus de dangereusement réduire la taille de l'échantillon, cette méthode n'élimine aucunement les biais encourus par l'érosion de l'échantillon puisqu'elle se base sur l'hypothèse (très rarement respectée en pratique) que les décrocheurs quittent l'échantillon de façon aléatoire. Les problèmes de cette stratégie ont néanmoins mené à plusieurs développements dans la gestion des données manquantes, notamment dans le développement des méthodes d'imputation, beaucoup plus efficaces que de simplement les omettre.

Présentement, l'application de poids échantillonnaires longitudinaux ajustés pour l'attrition est la méthode la plus recommandée pour contrer les biais d'attrition (Hill, 1997). Les poids échantillonnaires sont un facteur multiplicatif lié à chaque sujet dans un plan d'échantillonnage probabiliste où la probabilité d'inclusion dans l'échantillon n'est pas la même pour chaque sujet. Ces poids permettent de pouvoir faire des inférences sans biais de l'échantillon à la population étudiée. De façon classique, les poids sont l'inverse de la probabilité d'inclusion dans l'échantillon² bien qu'il soit possible d'ajuster les poids pour la non-réponse ainsi que pour d'autres facteurs pouvant affecter l'inférence. Fitzgerald et *al.* (1998) proposent d'ailleurs une pondération particulière pour corriger les biais. Cependant, l'utilisation efficace des poids, particulièrement ceux de Fitzgerald et *al.*, exige une excellente connaissance du processus d'attrition au sein de l'enquête utilisée. En outre, Hoddinot et *al.* (2002) constatent que des différences importantes demeurent entre les caractéristiques des répondants demeurant dans l'ELNEJ et ceux quittant l'ELNEJ malgré l'application des poids longitudinaux. Il est également très important de noter que les poids longitudinaux sont souvent recommandés pour contrer les biais liés à l'attrition même si leur efficacité n'est testée que très rarement (Rendtel, 2002).

Une autre technique permettant de contrer l'attrition est l'imputation de données pour remplacer les données manquantes causées par l'attrition. L'imputation consiste à générer des valeurs simulées pour remplacer les valeurs manquantes. Il existe deux grandes catégories d'imputation : l'imputation déterministe (comme l'imputation par la moyenne), qui ne comporte pas de composante aléatoire et atténue ainsi la structure de la covariance du jeu de données, et l'imputation aléatoire (comme l'imputation multiple), qui comprend une composante aléatoire permettant de maintenir la covariance sous certaines hypothèses. L'analyse longitudinale pose un problème particulier aux méthodes d'imputation puisque ces méthodes sont généralement basées pour traiter la non-réponse partielle en créant une valeur simulée grâce aux autres valeurs disponibles pour le sujet. Dans le cas où les valeurs manquantes sont causées par l'attrition du sujet,

² Ces poids sont basés sur le célèbre estimateur d'Horwitz-Thompson, dont on peut trouver des informations dans Särndal et *al.* (1992).

on ne dispose pas d'information actuelle sur le sujet pour la création des valeurs simulées. Conséquemment, plusieurs méthodes d'imputation ne sont pas bien adaptées pour contrer l'attrition. De nouvelles méthodes d'imputation tentent de contourner ce problème en utilisant, par exemple, les dernières valeurs observées mais elles ne sont pas encore au point (Baskin, 2005).

Pour compenser pour l'érosion de l'échantillon qui se produit avec l'attrition, il est aussi possible d'ajouter des nouveaux sujets à l'enquête. Ceci permet de maintenir une taille de l'échantillon satisfaisante mais n'élimine pas les biais pouvant être occasionnés par l'attrition. De plus, cette technique soulève plusieurs questions, notamment par rapport à la représentativité de l'échantillon face à la population étudiée ainsi que sur le suivi longitudinal individuel. L'ELNEJ et l'EDTR, de Statistique Canada, utilisent notamment cette technique pour renouveler son échantillon longitudinal.

Miller et Wright (1995) présentent une adaptation de la procédure d'Heckman pour la correction du biais d'attrition, le modélisant comme étant un biais de sélection. Cette procédure, divisée en deux étapes, considère le biais de sélection comme étant un problème d'omission d'une variable pertinente. La première étape consiste à utiliser un modèle *probit* pour obtenir un score λ , représentant la probabilité d'attrition de chaque sujet selon une gamme de caractéristiques. Par la suite, les scores sont ajoutés aux régressions pour corriger le problème d'omission de variables. Ceci permet alors d'éliminer le biais causé par l'attrition en plus d'obtenir des coefficients de régression qui ne sont plus biaisés. Néanmoins, cette méthode, basée sur la procédure d'Heckman, demeure contestée suite aux hypothèses parfois très fortes sur lesquelles reposent la procédure (Puhani, 1997).

En résumé, bien qu'il existe différentes procédures pour contrer l'attrition, ces procédures ne font généralement pas l'unanimité suite aux limitations causées par les hypothèses intrinsèques de chaque méthode. De plus, bien peu de ces techniques furent

véritablement testées afin de vérifier si elles permettent de contrer efficacement les effets de l'attrition.

1.4 VARIABLES AFFECTÉES PAR L'ATTRITION

La littérature fait état de plusieurs types de variables sujettes à être affectées par l'attrition. Cependant, l'effet de l'attrition dans les enquêtes longitudinales dépend considérablement du type d'étude, du contexte de l'enquête et de la population étudiée (Ahern, 2005). Néanmoins, certains groupes de variables sont reconnus comme étant spécialement touchés par l'attrition dans le cadre des enquêtes longitudinales ciblant la famille et les ménages. Cette section détaille donc les groupes de variables qui sont reconnus comme étant liés à l'attrition en mettant l'accent sur les enquêtes longitudinales de ménages ou portant sur la famille.

Tout d'abord, il est généralement reconnu que les variables associées à la pauvreté, à l'exclusion sociale et au statut socioéconomique sont particulièrement touchées par l'attrition. Par exemple, Behr et *al.* (2005) montrent que les gens moins scolarisés ainsi que ceux qui sont sans emploi sont plus à risque de décrocher de l'ECHP alors que Fitzgerald et *al.* (1998) arrivent aux mêmes conclusions avec les données du PSID. Ces caractéristiques sont également associées à l'attrition dans l'enquête Cocon (Razafindratsima et *al.*, 2004). En outre, Bor et *al.* (1993) montrent que l'attrition du Mater-University of Queensland Study of Pregnancy (MUSP), une enquête australienne, se concentre surtout chez les jeunes mères avec peu de scolarité et des revenus faibles.

En outre, les variables liées à l'instabilité dans la vie des individus sont aussi touchées par l'attrition. Watson (2003) montre l'importance de l'attrition chez les répondants qui changent de résidence dans l'ECHP alors qu'Alderman et *al.* (2000) arrivent aux mêmes conclusions pour une enquête bolivienne portant sur le développement des enfants. De son côté, Hill (1997) montre que l'attrition dans le Survey of Income and Program Participation (SIPP) fut particulièrement problématique chez les sujets qui venaient de divorcer.

Finalement, des variables liées à l'état de santé physique et mental des répondants, comme l'âge, la solitude, la dépression et le mauvais état de santé, sont également liées à l'attrition (Ahern, 2005). Cependant, dans des enquêtes longitudinales étudiant la famille, ces variables sont souvent plus difficiles à identifier car peu de questions sont posées sur ces facteurs dans ce type d'enquêtes longitudinales.

Il faut cependant noter que les variables potentiellement biaisées par l'attrition ne permettent pas d'expliquer l'attrition dans sa totalité. En effet, Watson (2003) a montré que bien que l'attrition ne soit pas distribuée aléatoirement dans l'ECHP, les variables d'intérêt pour l'analyse de la pauvreté et de l'exclusion sociale n'expliquent qu'une petite partie de la variation de l'attrition. Ceci n'est pas surprenant compte tenu du caractère partiellement aléatoire du phénomène et de la difficulté à trouver des variables *proxy* pour les phénomènes causant l'attrition.

1.5 OBJECTIFS VISÉS

Malgré plusieurs preuves du caractère non-aléatoire de l'attrition dans un grand nombre d'enquêtes longitudinales, l'attrition demeure un sujet relativement peu étudié par les scientifiques des sciences sociales (Ahern, 2005). Pourtant, avec le développement des enquêtes longitudinales canadiennes au cours de la dernière décennie, l'attrition devient une problématique très importante de l'analyse longitudinale en sciences sociales et mérite donc l'attention des chercheurs utilisant ce type de données dans leurs travaux de recherche. En effet, bien que plusieurs articles portent sur les techniques existant afin de vérifier la présence du biais d'attrition et d'y remédier (Miller et *al.*, 1995, Schmitz et *al.*, 2002, Marini et *al.*, 1980) ainsi que sur l'importance de l'attrition dans diverses enquêtes longitudinales pour quelques variables d'intérêt particulières (Behr et *al.*, 2005, Watson, 2003, Hiskey et *al.*, 2002), peu d'études portent sur des analyses approfondies de l'attrition dans une enquête précise ainsi que sur l'évaluation de l'efficacité des techniques connues pour contrer l'attrition malgré le grand nombre d'enquêtes longitudinales au Québec et au Canada. En outre, une identification des variables pouvant être reliées à l'attrition est nécessaire pour chaque enquête longitudinale (Ahern, 2005).

Ce mémoire portera donc sur ces thèmes pour une enquête longitudinale québécoise toujours en cours et couramment utilisée par les chercheurs s'intéressant à la démographie mais aussi à d'autres champs d'étude : l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ). Cette enquête est particulièrement pertinente en démographie puisque la population ciblée par l'enquête (une cohorte d'enfants nés au Québec) permet d'aborder plusieurs thèmes de nature démographique et au grand nombre de passages au cours desquels des données furent collectées. Ainsi, en démographie, les chercheurs utilisent notamment cette enquête dans l'étude approfondie de l'évolution des familles québécoises. Par exemple, les données longitudinales de l'ÉLDEQ ont servi dans l'étude des ménages multigénérationnels (Neill et *al.*, 2005a) ainsi que dans l'étude des impacts des systèmes de service de garde au Québec (Neill, 2004). Bien évidemment, à l'instar des autres enquêtes longitudinales, peu d'analyses de l'attrition furent réalisées pour l'ÉLDEQ, un manque que ce mémoire vise à combler.

De façon plus formelle, ce mémoire tentera de répondre aux questions suivantes :

1. Quelles sont les caractéristiques des familles décrochant prématurément de l'ÉLDEQ ?
2. Quels sont les facteurs influençant la probabilité d'attrition dans l'ÉLDEQ ?
3. Quelles variables d'intérêt sont biaisées par l'attrition ?
4. Est-ce que les poids longitudinaux permettent de corriger les biais causés par l'attrition dans l'ÉLDEQ ?

L'atteinte du premier objectif permettra de tracer un portrait des familles décrocheuses en mettant en lumière leurs caractéristiques par rapport aux familles toujours dans l'échantillon. Par la suite, à partir des résultats du premier objectif, la réalisation du second objectif permettra d'identifier les caractéristiques qui sont associées à la probabilité d'attrition d'un répondant. Cette modélisation du processus d'attrition permettra donc de pouvoir estimer l'ampleur de l'effet de la quantité d'attrition. Ces deux premiers objectifs permettront d'orienter l'équipe de l'ÉLDEQ et les chercheurs par rapport aux caractéristiques des décrocheurs et aux caractéristiques favorisant l'attrition afin de pouvoir développer des stratégies afin de limiter et de

corriger pour l'attrition. Par la suite, l'atteinte du troisième objectif permettra d'évaluer le second effet de l'attrition : les biais pouvant être causés par l'attrition. Il permettra de vérifier si certaines variables susceptibles d'intéresser des chercheurs utilisant l'ÉLDEQ sont biaisées par l'attrition et d'évaluer l'ampleur de ce biais. Finalement, le dernier objectif permettra de vérifier une méthode fort populaire pour contrer l'attrition, l'utilisation de poids longitudinaux, est réellement efficace dans la correction des biais causés par l'attrition dans le cas de l'ÉLDEQ.

Afin d'atteindre ces quatre objectifs, ce mémoire sera structuré de la façon suivante. Le prochain chapitre introduira l'enquête utilisée pour ce mémoire : l'ÉLDEQ. Le troisième chapitre présentera l'analyse descriptive de l'attrition afin de pouvoir évaluer l'ampleur globale du phénomène pour l'ÉLDEQ. Par la suite, les quatre chapitres suivants sont consacrés aux quatre objectifs principaux du mémoire. Conséquemment, le quatrième chapitre présentera un profil des familles décrochant prématurément de l'ÉLDEQ en comparant les familles décrocheuses de celles toujours dans l'enquête. Le cinquième chapitre présentera les probabilités pour une famille de quitter l'échantillon longitudinal de l'ÉLDEQ, une des deux conséquences majeures de l'attrition, grâce à des régressions *probit*. L'étude du deuxième effet de l'attrition, les biais, suivra dans le sixième chapitre avec des tests de Beckett, Gould, Lillard et Welch (BGLW). Finalement, le septième et dernier chapitre évaluera l'efficacité des poids longitudinaux de l'ÉLDEQ pour contrer les biais d'attrition.

Dans l'ensemble, ce mémoire visera donc à contribuer à l'essor de la connaissance sur l'attrition dans les enquêtes longitudinales dans le contexte d'une enquête longitudinale québécoise sur les enfants en plus de sensibiliser la communauté scientifique aux risques pouvant être occasionnés par l'attrition pour la qualité des données d'enquêtes longitudinales.

CHAPITRE 2 : PRÉSENTATION DE L'ENQUÊTE

Dans ce second chapitre, l'enquête qui sera utilisée dans ce mémoire, l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ), sera introduite en mettant l'emphase sur les stratégies mises en place par l'équipe de l'ÉLDEQ pour minimiser l'attrition et favoriser le retour des décrocheurs dans l'enquête.

2.1 DESCRIPTION DE L'ÉTUDE LONGITUDINALE DU DÉVELOPPEMENT DES ENFANTS DU QUÉBEC

2.1.1 Présentation de l'Étude longitudinale du développement des enfants du Québec

L'Étude longitudinale du développement des enfants du Québec (ÉLDEQ)³ est l'une des bases de données les plus utilisées pour étudier les facteurs influençant le développement d'une cohorte d'enfants québécois. Issue du fruit d'une collaboration entre plusieurs ministères québécois et étant une suite logique à l'ELNEJ de Statistique Canada, l'ÉLDEQ suit une cohorte d'enfants nés au Québec dans le but de comprendre les nombreux facteurs socioéconomiques ou autres impliqués dans le développement des enfants. À cette fin, l'enquête cible quatre grandes thématiques : le développement cognitif, l'évolution du comportement, l'adaptation sociale et la réussite scolaire. Il s'agit de la première enquête québécoise de la sorte basée sur un échantillon représentatif provenant d'une cohorte de nouveau-nés qui est suivie annuellement depuis la naissance jusqu'à l'entrée dans le milieu scolaire. Suite au succès des cinq passages⁴ de la première phase, l'ÉLDEQ a amorcé une deuxième phase en 2003, qui doit se terminer en 2010. Cette seconde phase, se déroulant en parallèle avec l'arrivée des enfants dans le milieu scolaire, permettra d'apporter plusieurs renseignements essentiels sur l'environnement scolaire des enfants cibles. Afin de pouvoir mettre sur pied une telle enquête, de cerner pleinement la complexité des facteurs entourant le développement des enfants (famille, service de garde, environnement social, sommeil, alimentation, milieu

³ Il est possible de consulter plusieurs documents portant sur l'ÉLDEQ en ligne à l'adresse suivante : <http://www.jesuisj'esperai.stat.gouv.qc.ca>

⁴ Les passages de l'ÉLDEQ sont codés EX où X est le numéro du passage. Ainsi, E1 représente le passage initial de l'enquête. Cette notation sera parfois utilisée dans le cadre de ce mémoire.

scolaire, etc.) et de faire face aux nombreux défis méthodologiques et logistiques engendrés par une enquête longitudinale de cette importance, une équipe regroupant des spécialistes de plusieurs disciplines variées comme la démographie, la sociologie et la psychologie fut mise à contribution. Cette équipe interdisciplinaire a également permis d'innover sur plusieurs points. Par exemple, l'ÉLDEQ a innové en collectant des informations sur le père biologique de l'enfant même si celui-ci ne réside pas avec l'enfant. Ceci jette ainsi un tout nouveau regard sur l'environnement social de l'enfant mais apporte également plusieurs nouveaux problèmes méthodologiques (comment rejoindre efficacement ces pères ?). Bien évidemment, l'un des défis méthodologiques majeurs auquel fait face l'équipe de l'ÉLDEQ est l'érosion de l'échantillon. Décrire l'ampleur de l'attrition dans l'ÉLDEQ sera d'ailleurs la phase préliminaire avant d'atteindre les objectifs de ce mémoire.

2.1.2 Population visée, base de sondage et échantillonnage

La population ciblée par l'ÉLDEQ est constituée des bébés issus de naissances simples, qui avaient 59 ou 60 semaines d'âge gestationnel⁵ au début de la période de collecte et dont la mère vit au Québec. Les naissances de mères vivant dans les régions sociosanitaires 10 (Nord-du-Québec), 17 (territoire cri) et 18 (territoire inuit) ainsi que dans les réserves indiennes ont été exclues de la population visée par l'enquête.

La base de sondage de l'ÉLDEQ a été créée à l'aide du Fichier maître des naissances vivantes du MSSS, qui contient les enregistrements des formulaires de naissance pour chaque année civile. Ce fichier comprend plusieurs informations sur la naissance de l'enfant telles que le lieu et la date de la naissance, la durée de la grossesse ainsi que quelques renseignements sur les parents. Certains cas, comme les naissances s'étant produites après plus de 42 semaines de gestation, ne purent être retenus, faisant en sorte que la population échantillonnée pour l'ÉLDEQ a été estimée à 94,5% de la population visée (Jetté et *al.*, 2000)

⁵ On définit l'âge gestationnel comme étant la somme de la durée de gestation (grossesse) et de l'âge chronologique de l'enfant.

Le plan de sondage retenu pour l'ÉLDEQ est un plan stratifié à trois degrés (Institut de la statistique du Québec, 2000). Tout d'abord, le territoire visé par l'enquête est scindé en unités primaires d'échantillonnage (UPE), qui sont formées d'une région sociosanitaire. Ces régions sont ensuite partagées entre les « régions éloignées⁶ » et les « régions non éloignées », afin de maximiser le rendement financier de l'enquête. Au second degré, les régions sociosanitaires sont divisées en unités secondaire d'échantillonnage (USE), qui sont constituées d'une ou de deux municipalités régionales de comté (MRC). Ces unités secondaires d'échantillonnage sont par la suite divisées en deux groupes selon le nombre de naissances enregistrées en 1999 afin de contenir l'effet de plan global. Finalement, au troisième degré, on sélectionne des unités tertiaires d'échantillonnage (UTE), représentant les enfants, à l'intérieur des USE sélectionnées.

Afin de pouvoir obtenir des statistiques longitudinales de bonne qualité pour la durée de l'enquête, de tenir compte des contraintes budgétaires et d'inclure un échantillon supplémentaire de naissances venant de la région sociosanitaire de la Montérégie suite au passage de la tempête verglaçante, un total de 2 940 naissances furent choisies pour constituer l'échantillon initial. De cet échantillon, 2 223 ont répondu au QIRI, l'instrument principal de l'enquête (voir la section suivante), lors du premier volet de l'enquête, ce qui représente un taux de réponse initial de 75,8%. De ces répondants, 2 120 furent éligibles au suivi longitudinal (les autres étant les familles montérégiennes qui furent interviewées suite à la tempête verglaçante et qui ne furent pas interviewées aux passages suivants).

2.1.3 Questionnaires

Étant une enquête très complexe, l'ÉLDEQ possède plusieurs questionnaires et instruments de mesure. Au total, pour la première phase, un total de quatorze instruments ont été utilisés afin de collecter des informations ou d'assurer le suivi longitudinal des enfants. Par la suite, d'autres instruments ont été intégrés, notamment pour recueillir des informations sur le milieu scolaire fréquenté par l'enfant.

⁶ Les régions du Bas-Saint-Laurent, de l'Abitibi-Témiscamingue, de la Côte-Nord et de la Gaspésie-Îles-de-la-Madeleine forment les « régions éloignées » par leur faible densité de population, leur vaste étendue géographique et leur éloignement des grands centres urbains.

Parmi ces différents instruments, on distingue notamment le questionnaire informatisé rempli par l'intervieweur (QIRI), qui est en quelque sorte l'instrument maître de l'ÉLDEQ. Ce questionnaire contient environ 600 variables couvrant une multitude de thèmes comme les caractéristiques socioéconomiques et sociodémographiques du ménage, l'état de santé et les habitudes de vie des individus du ménage ainsi que le fonctionnement familial. De plus, une section est consacrée aux personnes de référence afin de faciliter le contact avec la famille lors des prochaines collectes.

Il est également pertinent de mentionner les principaux questionnaires gravitant autour du QIRI. Le questionnaire papier rempli par l'intervieweur (QPRI) vient compléter l'entrevue avec l'intervieweur et contient notamment des informations sur les familles maternelle et paternelle de l'enfant. Le questionnaire autoadministré de la mère (QAAM) permet la saisie d'informations sur les relations entre la mère et l'enfant ainsi que la collecte de données plus sensibles comme les grossesses antérieures. Le questionnaire autoadministré du père (père présent : QAAP, père biologique absent : QAAPABS) est l'équivalent du QAAM pour le père de l'enfant. Ce questionnaire, en plus d'être unique à l'ÉLDEQ, permet ainsi de collecter de l'information sur le père biologique de l'enfant même si celui-ci ne vit plus avec l'enfant.

Pour la première phase de la collecte, tous les instruments et les questionnaires étaient remplis en même temps, lors d'une seule et même visite (avec l'exception du QAAP, qui était généralement envoyé par la poste). Par la suite, dans la seconde phase de l'enquête, lorsque les enfants furent en âge d'aller à l'école, les enfants répondent aux instruments et aux tests qui les concernent à l'école alors que les parents répondent à leurs instruments à la maison.

2.1.4 Prise en compte de l'attrition dans l'ÉLDEQ

Dès le premier passage de l'enquête, l'équipe de l'ÉLDEQ a fait de grands efforts afin de maximiser la participation des ménages répondant à l'enquête. Ces démarches ont été

mises à exécution dans une optique de mise en confiance des répondants face à l'enquête en expliquant les motivations et le contexte derrière les décisions prises par l'équipe de l'ÉLDEQ.

Afin de mieux comprendre le contexte entourant chaque passage de l'enquête ainsi que les différentes mesures qui ont été prises pour limiter l'attrition et favoriser le retour des décrocheurs dans l'enquête, une entrevue fut effectuée avec deux membres de l'équipe de l'ÉLDEQ : Delphine Provençal et Josette Thibault. Cette entrevue a été réalisée jeudi le 5 juillet 2007 aux bureaux de l'ISQ de Montréal. Ces informations qualitatives viendront compléter avantageusement les données quantitatives afin de mieux comprendre la dynamique de l'attrition dans l'ÉLDEQ.

Première phase de l'ÉLDEQ : L'entente initiale avec les répondants était d'une durée de 5 ans, soit pour les cinq premiers passages. Dès cette première phase, divers moyens ont été utilisés afin de fidéliser les répondants. En plus de sensibiliser les parents à l'importance de l'ÉLDEQ, des cartes de fête et des lettres ont été envoyées aux enfants. De plus, un album souvenir a été graduellement réalisé pour chaque enfant et remis après le cinquième passage.

Seconde phase de l'ÉLDEQ : À la fin de la première phase, après une période d'incertitude quant à l'avenir de l'enquête et au financement de celle-ci, une seconde phase a finalement été mise sur pied. Cependant, c'est au début de cette seconde phase, lors des sixième et septième passages, que l'attrition fut particulièrement élevée (voir le chapitre suivant). À ce moment, il était impossible pour l'équipe de l'ÉLDEQ de demander un consentement de la part de la famille pour plus d'un an. Afin de faire face à cette problématique, l'équipe de l'ÉLDEQ a choisi de diversifier ses stratégies de fidélisation.

Premièrement, on a tenté de préserver un bon noyau d'intervieweurs afin d'établir et de préserver un lien de confiance entre les intervieweurs et les répondants. À ce titre, la coordonnatrice chargée de la prise des entrevues est demeurée la même depuis

le premier passage de l'enquête, lui permettant ainsi de bien connaître chaque famille de l'échantillon. Un appel téléphonique supplémentaire a été également effectué pour chaque famille afin de favoriser la communication entre l'équipe de l'enquête et les répondants.

Deuxièmement, l'équipe de l'ÉLDEQ tenta de changer l'image de l'enquête. Effectivement, le site Internet (<http://www.jesuisjeserai.stat.gouv.qc.ca>) a été créé en plus de changer le slogan de l'enquête. Celui-ci est passé de : « En 2002... J'aurai 5 ans ! » pour la phase I à : « En 2002... J'avais 5 ans ! » afin de bien montrer la nouvelle direction prise par l'enquête. Par la suite, le slogan a été modifié une seconde fois à partir de E8 pour devenir : « Je suis, je serai », faisant alors référence aux buts de l'enquête. À la demande des parents, davantage de résultats généraux leur ont été envoyés pour attester de l'utilité des données collectées. Suite à la participation de l'organisme Québec en forme, un diplôme d'activité physique a également été remis à chaque enfant cible en plus de leur adresser directement des lettres.

Troisièmement, l'équipe de l'ÉLDEQ a tenté d'offrir davantage de flexibilité aux répondants. Ceci fut tout d'abord réalisé en réduisant la durée des entrevues et des tests. En effet, la durée moyenne des entrevues en E7 en personne était de 210 minutes, comparativement à 90 minutes au passage suivant. Il faut spécifier ici que suite à l'incertitude de la fin de la première phase, plusieurs questions et tests avaient été ajoutés aux questions et aux tests existants afin de tirer le plus de renseignements possibles au cas où on mettrait un terme à l'enquête. Ceci alourdissait évidemment le fardeau des répondants, nuisant ainsi à leur niveau de participation. En outre, on a proposé aux parents réticents d'aller seulement voir l'enfant cible. De plus, la compensation monétaire a été augmentée (passant de 20\$ à 25\$) et un cadeau, un livre, fut offert à l'enfant. Finalement, à partir de E9, la collecte de données ne sera plus annuelle mais s'effectuera aux deux ans. Ceci réduira évidemment le fardeau des répondants et les coûts de l'enquête.

Quatrièmement, l'équipe de l'ÉLDEQ a tenté de retracer tous les décrocheurs en plus de changer les critères de rappel des familles ayant refusé de répondre à un certain volet. Par exemple, alors que la stratégie initiale de prendre en note deux références pour retrouver les familles « perdues » s'avérait ne pas être efficace dans tous les cas, la Commission d'accès à l'information permit à l'équipe de l'ÉLDEQ d'utiliser les fichiers administratifs du Ministère de l'Éducation (MEQ) afin de retracer les enfants décrocheurs. Cette technique connut un succès significatif alors que sur 41 familles disparues identifiées, 31 ont été retracés dans les fichiers administratifs du MEQ et 23 de ces familles sont retournées dans l'enquête.

En somme, la seconde phase de l'ÉLDEQ fut marquée par une diversification des stratégies de fidélisation des répondants. L'efficacité de cette stratégie pourra être bien évaluée lors des passages futurs de l'ÉLDEQ.

2.1.5 L'accès aux données de l'ÉLDEQ

Les données de l'ÉLDEQ sont accessibles au Centre d'accès aux données de recherche de l'Institut de la statistique du Québec (CADRISQ). Tout chercheur désirant utiliser les données de cette enquête doit en faire la demande auprès de l'ISQ. Une fois la demande acceptée, le chercheur peut alors avoir accès aux données en se rendant au CADRISQ.

2.2 CONCLUSION

L'ÉLDEQ est donc une enquête longitudinale québécoise innovatrice et toujours en cours ciblant une cohorte d'enfants québécois dans le but de mieux comprendre les nombreux facteurs qui influencent le développement des enfants québécois ainsi que leur adaptation sociale et leur réussite scolaire. Il est très intéressant de constater tous les efforts de l'équipe de l'ÉLDEQ pour minimiser l'attrition et favoriser le retour des décrocheurs dans l'enquête, ce qui montre bien les préoccupations face à l'effet de l'attrition sur la qualité des données. Toutefois, bien qu'elle ait été utilisée par plusieurs chercheurs afin de répondre à diverses questions de recherche, peu d'analyses ont été effectuées sur l'attrition. Ce mémoire vise ainsi à combler ce vide.

CHAPITRE 3 : L'ATTRITION DANS L'ÉLDEQ

Avant de débiter une analyse plus systématique de l'attrition, il convient d'étudier la non-réponse par unité et l'attrition de façon descriptive pour mieux saisir l'ampleur de ces phénomènes dans l'ÉLDEQ. Puisque l'attrition est essentiellement un cas particulier de non-réponse par unité, une brève analyse de ce type de non-réponse sera présentée en premier. Par la suite, toujours de façon descriptive, l'attrition dans l'ÉLDEQ sera introduite, principalement à l'aide de deux outils complémentaires : les taux d'attrition, permettant particulièrement de mesurer l'effet de l'érosion de l'échantillon, et les ratios d'attrition, dont l'utilité principale réside en la mesure du biais causé par l'attrition.

3.1. LA NON-RÉPONSE PAR UNITÉ DANS L'ÉLDEQ

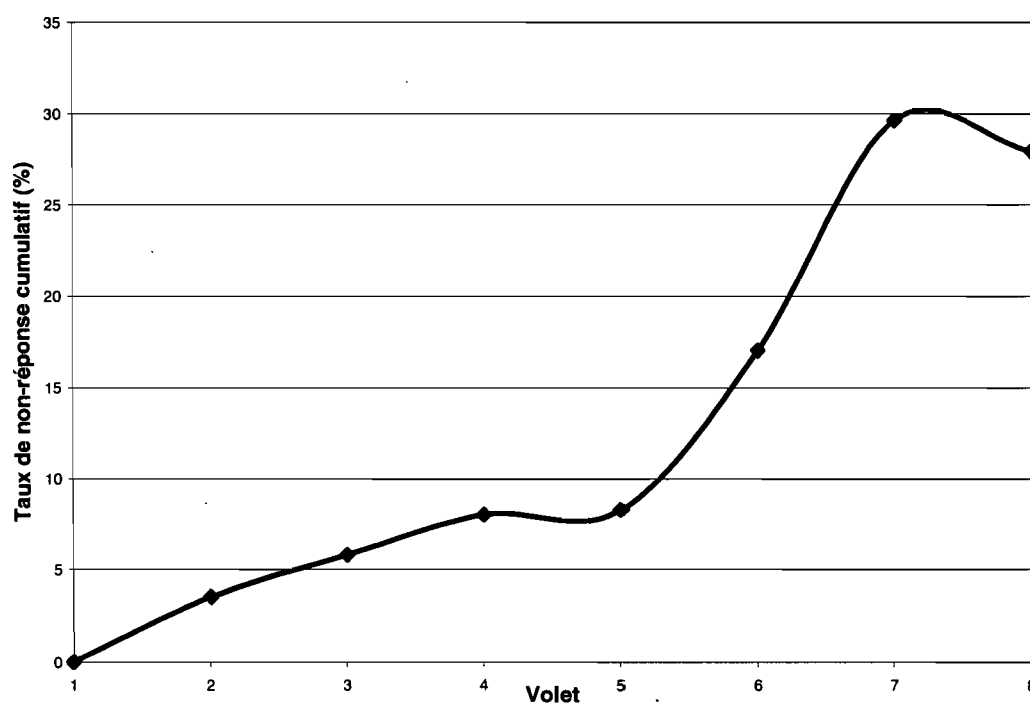
À l'instar des autres enquêtes longitudinales, l'ÉLDEQ est évidemment sujette à l'érosion de son échantillon. L'effet de la non-réponse par unité sur l'échantillon de l'ÉLDEQ sera étudié de manière descriptive en s'attardant au nombre de répondants présents dans l'échantillon à chaque volet. Ceci est réalisable grâce au fichier des dates d'entrevue du QIRI, où une valeur manquante signifie qu'aucune entrevue n'a été réalisée. Par la suite, les différents parcours de participation par unité des familles de l'ÉLDEQ seront étudiés.

3.1.1 L'ampleur de la non-réponse par unité dans l'ÉLDEQ

La figure 3.1 montre le nombre de répondants, le taux de non-réponse par unité et le taux de non-réponse par unité cumulatif pour chaque volet de l'ÉLDEQ. L'érosion graduelle de l'échantillon est bien visible puisque des 2 120 ménages répondant au premier volet de l'ÉLDEQ en 1998, il n'en reste plus que 1 528 après le huitième passage de l'enquête (72,1%). Pendant la première phase de l'ÉLDEQ (volets 1 à 5), l'érosion de l'échantillon est très faible alors que 91,7% de l'échantillon est présent au cinquième volet et que le taux de non-réponse par unité pour chaque volet est toujours inférieur à 4,0%. Cependant, l'érosion de l'échantillon augmente considérablement aux sixième et septième volets alors qu'environ 200 sujets sont perdus à chacun de ces volets, pour des taux de non-réponse supérieurs à 9,0%. Le sixième volet marque

d'ailleurs le passage à la deuxième phase de l'enquête. Par la suite, le taux de non-réponse devient négatif au huitième passage (-2,4%), alors que le nombre de répondants est plus élevé qu'au passage précédent. Ceci signifie donc que plusieurs familles sont de retour dans l'enquête, ce qui peut être imputable aux grands efforts réalisés par l'équipe de l'ÉLDEQ pour minimiser la non-réponse et l'attrition en plus d'encourager le retour des familles dans l'enquête.

Figure 3.1
Ménages répondant à l'ÉLDEQ et taux de non-réponse par unité cumulatif (%) selon le volet



Volet	1	2	3	4	5	6	7	8
<i>n</i>	2 120	2 045	1 997	1 950	1 944	1 759	1 492	1 528

Source : ÉLDEQ, ISQ.

3.1.2 Schémas de participation

Après avoir constaté de façon générale la diminution de la taille échantillonnale de l'ÉLDEQ suite à la non-réponse (et indirectement à l'attrition), il est intéressant de s'attarder aux schémas de participation des familles de l'ÉLDEQ. Ainsi, avec des ménages décrochant de l'enquête (et revenant parfois dans l'enquête), on se trouve dans la situation où les répondants ont répondu à un nombre différent de volets en plus d'avoir une multitude de schémas de réponse différents. Le tableau 3.1 présente la

distribution des volets répondus pour les familles de l'enquête. Bien que la majorité des 2 120 répondants aient répondu aux huit volets, environ le quart des ménages ont raté au moins un des huit passages de l'ÉLDEQ. En effet, un peu plus d'un ménage sur cinq a raté un ou deux volets de l'enquête alors que près de 5% des familles n'ont répondu qu'à un ou deux volets. De plus, il faut noter que les 63 répondants qui n'ont répondu qu'à un seul volet ont tous répondu au premier volet.

Tableau 3.1
Distribution des répondants à l'ÉLDEQ selon le nombre de volets répondus et pourcentage pour chaque volet

Nombre de volets répondus	<i>n</i>	%
1	63	3,0
2	45	2,1
3	22	1,0
4	39	1,8
5	140	6,6
6	253	11,9
7	222	10,5
8	1 336	63,0
Total	2 120	100

Source : ÉLDEQ, ISQ.

Ceci mène donc à plusieurs schémas de participation différents, qui sont décrits dans le tableau 3.2. À ce stade, il faut se rappeler que la réponse d'une famille est définie par la réponse au QIRI, l'instrument principal de l'enquête. Mis à part le parcours où les ménages ont répondu aux huit volets de l'enquête, les parcours les plus fréquents révèlent une attrition, se produisant lorsqu'un répondant quitte l'échantillon longitudinal de manière définitive, particulièrement lors de la deuxième phase de l'enquête (aux volets 6, 7 et 8). Ces parcours regroupent donc des individus ayant commencé à répondre aux premiers passages de l'enquête pour cesser définitivement (du moins jusqu'au volet 8) leur participation à l'ÉLDEQ à partir d'un certain passage de la seconde phase. Il faut toutefois noter qu'il soit possible que certains décrocheurs au volet 8 ne puissent simplement pas avoir répondu lors de ce volet et répondront aux volets suivants. En outre, il est intéressant de noter que plusieurs des parcours présentés ici sont constitués de ménages ayant quitté l'échantillon et qui sont retournés lors du

huitième volet. Ceci donne un indice du succès des méthodes utilisées par l'équipe de l'ÉLDEQ pour ramener des familles décrocheuses dans l'échantillon.

Tableau 3.2
Schémas de participation des répondants de l'ÉLDEQ

Volet 1	Volet 2	Volet 3	Volet 4	Volet 5	Volet 6	Volet 7	Volet 8	<i>n</i>
Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	1 336
Oui	Oui	Oui	Oui	Oui	Oui	Oui	---	98
Oui	Oui	Oui	---	Oui	Oui	Oui	Oui	8
Oui	Oui	Oui	Oui	Oui	---	Oui	Oui	24
Oui	Oui	Oui	Oui	Oui	Oui	---	Oui	79
Oui	Oui	Oui	Oui	Oui	Oui	---	---	199
Oui	Oui	Oui	Oui	Oui	---	---	Oui	32
Oui	Oui	Oui	Oui	Oui	---	---	---	122
Oui	Oui	Oui	Oui	---	---	---	---	14
Oui	Oui	Oui	---	---	---	---	Oui	10
Oui	Oui	Oui	---	---	---	---	---	21
Oui	Oui	---	---	---	---	---	---	43
Oui	---	---	---	---	---	---	---	63
Oui	Oui	Oui	---	Oui	---	---	---	9
Oui	Oui	Oui	Oui	---	Oui	---	Oui	5

Note : seuls les parcours comprenant 5 répondants ou plus sont présentés ici afin de préserver la confidentialité des répondants.

Source : ÉLDEQ, ISQ.

3.1.3 Comparaison des deux définitions de non-réponse

Avant de poursuivre dans l'analyse descriptive de l'attrition, il est pertinent de s'arrêter et de considérer la définition de la non-réponse par unité choisie par l'équipe de l'ÉLDEQ.

Contrairement à la définition choisie pour ce mémoire, en prenant la réponse au QIRI comme gage de la participation du sujet, l'équipe de l'ÉLDEQ considère qu'une famille a répondu à un passage donné si elle a complété au moins un des instruments de l'ÉLDEQ. La différence entre les deux définitions risque évidemment d'entraîner une différence dans les taux de non-réponse et donc, dans les taux d'attrition. Ces différences, pour chaque passage, sont illustrées dans le tableau 3.3. Les nombres obtenus de E1 à E6 sont les mêmes puisque pour ces passages, les instruments pour les parents et pour les enfants étaient remplis en même temps. Il était donc impossible, lors

de ces passages, qu'un enfant remplisse les instruments préparés pour lui sans que les parents répondent à leurs propres instruments. Par la suite, de petites différences apparaissent pour E7 (1,7%) et E8 (0,4%). Il est vraisemblable de penser que ces différences s'accroîtront lors des passages futurs suite à la stratégie de l'équipe de l'ÉLDEQ de séparer l'entrevue des enfants de celle des parents. Ces écarts provenant de la différence dans la définition du phénomène, bien que marginaux pour le moment, montrent toutefois l'importance de la définition de phénomènes comme la non-réponse et de la façon dont ils sont considérés en pratique dans l'analyse du phénomène de l'attrition.

Tableau 3.3

Taux de non-réponse (%) de l'équipe de l'ÉLDEQ et différences (%) entre les taux de non-réponse cumulatifs (%) de l'équipe de l'ÉLDEQ et ceux de ce mémoire selon le volet de l'ÉLDEQ

Volet	<i>n</i> (Équipe ÉLDEQ)	Taux de non- réponse (Équipe ÉLDEQ)	Taux de non- réponse cumulatif (Équipe ÉLDEQ)	Taux de non- réponse cumulatif (Bérard- Chagnon)	Différence
Volet 1	2120	0,0	0,0	0,0	0,0
Volet 2	2045	3,5	3,5	3,5	0,0
Volet 3	1997	2,4	5,8	5,8	0,0
Volet 4	1950	2,4	8,0	8,0	0,0
Volet 5	1944	0,3	8,3	8,3	0,0
Volet 6	1759	9,5	17,0	17,0	0,0
Volet 7	1529	13,1	27,9	29,6	1,7
Volet 8	1537	-0,5	27,5	27,9	0,4

Note : La colonne intitulée « Différence » représente, pour chaque passage, la différence entre la méthode de ce mémoire et la méthode de l'équipe de l'ÉLDEQ.

Source : ÉLDEQ, ISQ.

3.2 L'ATTRITION DANS L'ÉLDEQ

Suite à la présentation de la non-réponse par unité dans l'ÉLDEQ, il est maintenant en possible de présenter l'ampleur de l'attrition dans cette enquête longitudinale québécoise. Cette section débutera par la présentation des taux d'attrition et par une comparaison entre les effectifs de non-répondants et de décrocheurs. Ensuite, la

présentation des taux d'attrition calculés selon les caractéristiques des répondants et des ratios d'attrition, deux outils permettant d'évaluer rapidement les deux conséquences de l'attrition sur un échantillon longitudinal, viendront compléter l'analyse descriptive.

3.2.1 Définition de l'attrition

Bien que la définition théorique de l'attrition, la perte prématurée et définitive de répondants, soit reconnue et acceptée par la communauté scientifique, il existe de nombreuses différences dans l'application de ce concept aux diverses enquêtes longitudinales passées et présentes. Il est donc hautement pertinent de préciser la définition de l'attrition qui sera utilisée dans le cadre de ce mémoire.

Ainsi, pour ce mémoire, l'attrition sera définie comme étant la cessation prématurée (et de façon permanente) de la participation d'une famille à l'ÉLDEQ en date du huitième passage. Conséquemment, les sujets effectuant un retour dans l'enquête après avoir raté un ou plusieurs passages ne seront pas considérés comme étant des décrocheurs. Cette définition longitudinale de l'attrition permet de distinguer les gens ratant quelques passages de l'enquête mais qui reviennent par la suite (il s'agit alors de non-réponse par unité) de ceux qui cessent leur participation de façon définitive (les décrocheurs). Toutefois, dans le cas d'une enquête comme l'ÉLDEQ, avec un grand nombre de passages et offrant la possibilité aux décrocheurs de revenir plus tard dans l'enquête, il faut noter qu'au fur et à la mesure que l'enquête progresse, plusieurs décrocheurs changent évidemment de groupe suite à un retour à l'enquête. La figure 3.2 montre les trois statuts possibles pour un sujet se trouvant dans la population cible à chaque passage de l'enquête. Ayant répondu à tous les passages de l'enquête, le sujet A est considéré comme un non décrocheur. Le sujet B, ayant cessé de participer à l'enquête longitudinale lors de trois volets mais étant revenu par la suite, n'est pas considéré comme un décrocheur même s'il n'a pas participé à tous les passages. En effet, le sujet B entre dans la catégorie de la non-réponse par unité. Finalement, le sujet C a cessé sa participation à l'enquête de façon définitive (en date du huitième passage), faisant de lui un décrocheur. Cependant, il n'est pas exclu qu'il reprenne sa participation dans le futur, ce qui lui enlèverait le statut de décrocheur.

Figure 3.2
Exemples des différents types de décrocheurs

Sujet	1	2	3	4	5	6	7	8	Statut
A	✓	✓	✓	✓	✓	✓	✓	✓	Non décrocheur
B	✓	✓	✓				✓	✓	Non-réponse par unité
C	✓	✓	✓	✓	✓				Attrition

De plus, il est important de noter que les familles dont l'enfant-cible est décédé ou qui ont quitté le Québec ne seront pas exclues des analyses même si, théoriquement, ils ne font plus partie de la population-cible de l'ÉLDEQ. Ceci se justifie tout d'abord par le faible nombre de ces familles par rapport à l'échantillon total (environ 30 familles en date de E8 dont la vaste majorité ont quitté le Québec) mais surtout car l'équipe de l'ÉLDEQ a continué de collecter des informations sur les familles qui ont quitté le Québec suite au caractère prospectif de l'enquête. Par exemple, il est possible que certains enfants ayant déménagé hors du Québec reviennent éventuellement au Québec pour leurs études ou pour y vivre de façon définitive, justifiant ainsi la poursuite de la collecte de données avec ces familles. Il est toutefois important de spécifier que ces familles sont évidemment beaucoup plus difficiles à rejoindre et à interviewer étant donné leur nouveau lieu de résidence.

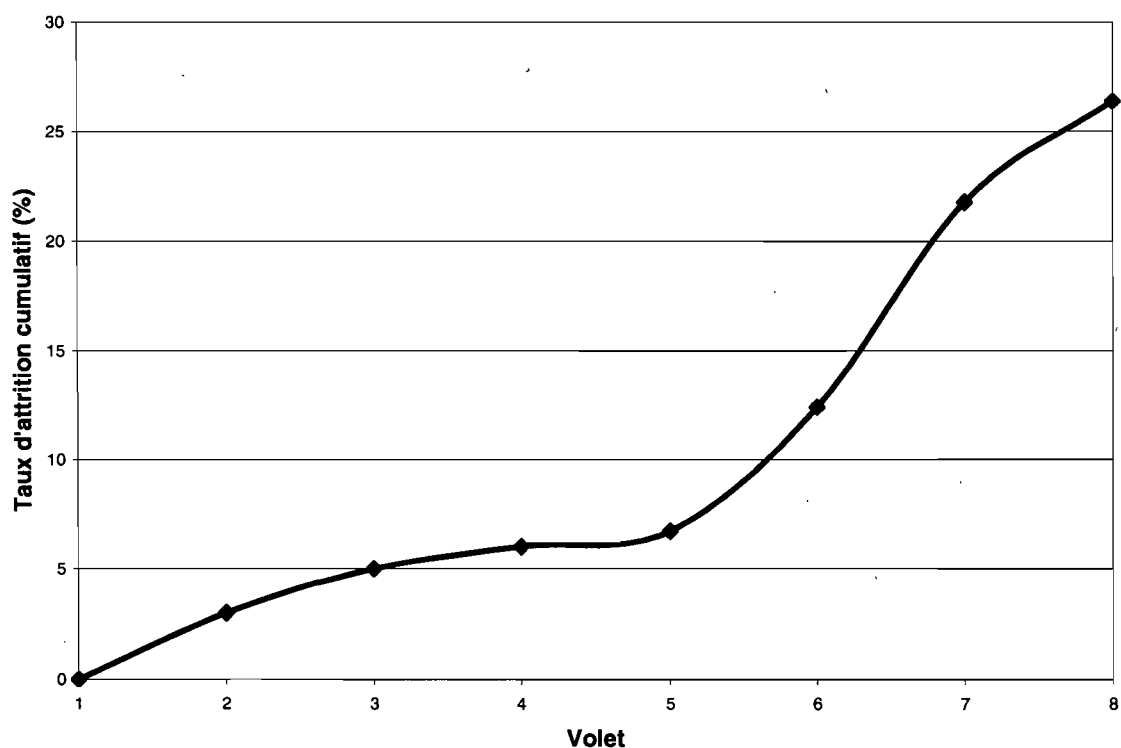
3.2.2 Ampleur de l'attrition dans l'ÉLDEQ

Tout comme pour la non-réponse, il est possible de calculer des taux et des taux cumulatifs pour l'attrition. Ces taux, pour chaque volet de l'ÉLDEQ, sont présentés dans la figure 3.3. Dans l'ensemble, le taux d'attrition cumulatif de l'ÉLDEQ en date de E8 est de 26,4%, ce qui est légèrement inférieur au taux de non-réponse cumulatif de ce volet (27,9%). De plus, tout comme pour la non-réponse, on voit que l'attrition est plutôt faible dans la première phase de l'ÉLDEQ, constituée des cinq premiers volets, alors que les taux d'attrition instantanés sont toujours inférieurs à 3,0%. Cependant, à partir de la seconde phase de l'enquête, les taux d'attrition instantanés augmentent considérablement, avec un sommet à 11,3% en E7. L'incertitude qui régnait quant à l'avenir de l'enquête lors de la fin de la première phase ainsi que le sentiment par les familles d'avoir rempli leur mandat initial (qui était seulement de répondre à la première phase) ont contribué à cette forte augmentation de l'attrition. Il est intéressant de voir le

taux d'attrition élevé de E8 (6,6%), qui se contraste avec le taux de non-réponse particulier de E8 (-2,4%). Ceci s'explique par le retour d'un grand nombre de familles, masquant ainsi le départ de certaines familles lors de ce volet. Cependant, il ne faut pas oublier qu'il soit possible que certains des décrocheurs des derniers volets (particulièrement en E8, le dernier volet étudié), reviennent dans l'enquête lors d'un volet futur.

Figure 3.3

Taux d'attrition cumulatif (%) de l'ÉLDEQ selon le volet



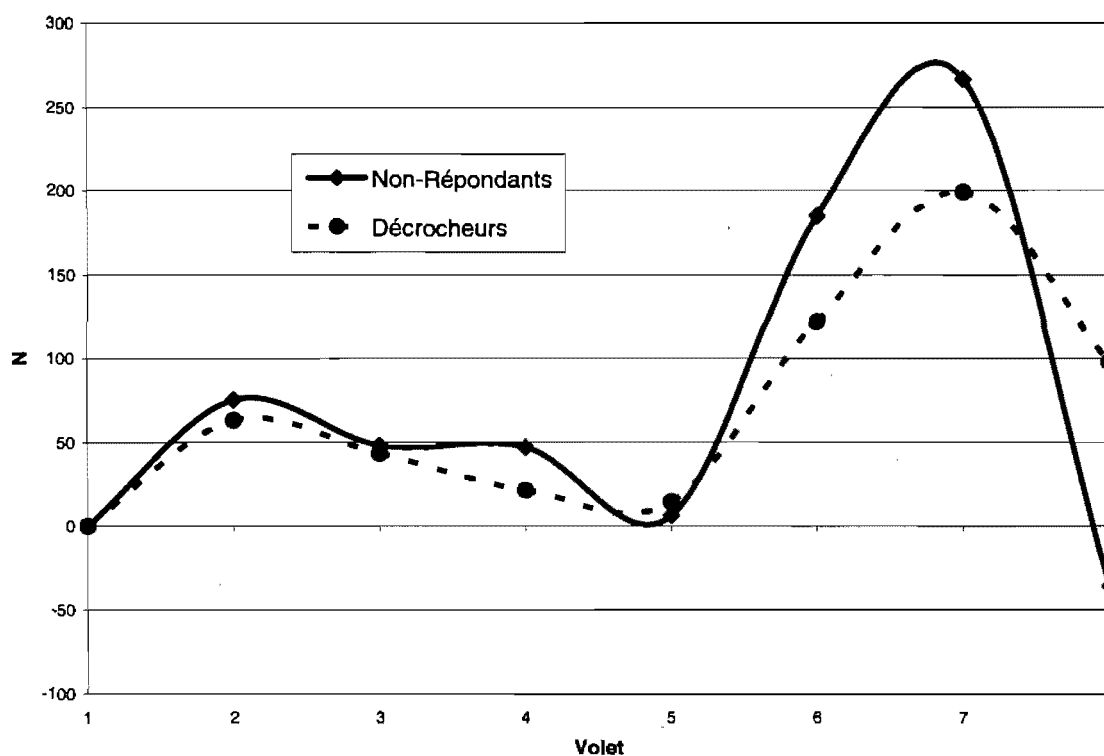
Source : ÉLDEQ, ISQ.

3.2.3 Non-réponse versus attrition

Étant donné la distinction faite entre les non-répondants et les décrocheurs, il est important de prendre un peu de temps afin de vérifier si cette distinction influence beaucoup le processus d'érosion de l'échantillon. Ceci est d'autant plus pertinent que, dans plusieurs enquêtes longitudinales, les définitions de la non-réponse et de l'attrition sont fréquemment confondues, amenant donc un problème de comparabilité de l'attrition entre différentes enquêtes. La figure 3.4 montre les différences, en effectifs absolus,

entre les non-répondants (qui comprennent naturellement les décrocheurs) et les décrocheurs pour chaque volet.

Figure 3.4
Non-répondants par unité et décrocheurs selon le volet de l'ÉLDEQ



Source : ÉLDEQ, ISQ.

Des 2 120 répondants initiaux de l'ÉLDEQ, en date de E8, 560 sont devenus des décrocheurs, ce qui signifie qu'ils ont cessé de façon définitive leur participation à l'enquête à partir d'un certain volet. En outre, toujours en date de E8, l'échantillon longitudinal a perdu 592 répondants par rapport au premier volet. Au total, à E8, les décrocheurs représentent donc 32 familles de moins que les non-répondants par unité. Ces deux catégories de répondants sont évidemment très corrélées puisque les décrocheurs sont naturellement inclus dans les non-répondants. Il est particulièrement intéressant de comparer les deux types de non-réponse pour E8, le troisième volet avec le plus de décrocheurs, alors que les nombreuses familles de retour dans l'enquête masquent les 98 familles ayant décroché à ce passage. Il s'agit d'ailleurs du volet où l'écart entre les non-répondants par unité et les décrocheurs est le plus élevé suite au

retour de nombreuses familles dans l'enquête. Dans l'ensemble, la faible ampleur des différences d'effectifs entre ces deux catégories de décrocheurs indique que, dans le cas de l'ÉLDEQ, la non-réponse par unité est constituée en très grande partie des décrocheurs, signifiant donc que lorsqu'une famille ne répond pas à un volet, les chances pour que cette famille ne réponde pas aux volets suivants sont très fortes.

3.3 L'ATTRITION SELON LES CARACTÉRISTIQUES INDIVIDUELLES DES RÉPONDANTS

Après avoir constaté l'ampleur de l'érosion de l'échantillon de l'ÉLDEQ, il est pertinent de s'intéresser aux taux d'attrition selon certaines caractéristiques des répondants au premier volet. Le tableau 3.4 présente donc les taux d'attrition cumulatifs pour chaque catégorie de quelques variables de nature socioéconomique spécialement sélectionnées suite à leur lien avec l'attrition dans la littérature.

Tableau 3.4
Taux d'attrition cumulatifs (%) choisis

Variables	n	Volet						
		2	3	4	5	6	7	8
<i>Présence du père biologique</i>								
Présent	1937	2,7	4,6	5,3	6,0	11,7	21,1	25,6
Absent	177	5,7	9,6	13,0	13,6	19,8	29,4	35,6
<i>Type de famille (3 cat.)</i>								
Intacte	1706	2,6	4,5	5,7	5,7	11,7	21,0	25,7
Recomposée	235	3,4	5,5	6,4	7,7	11,5	21,7	24,7
Monoparentale	171	5,9	9,9	13,5	14,0	21,1	30,4	36,3
<i>Statut d'immigrante (mère)</i>								
N'est pas immigrante	1865	1,7	3,3	4,1	4,7	10,3	19,6	24,3
Immigrante européenne	62	12,9	16,1	16,1	17,7	24,2	35,5	38,7
Immigrante non européenne	191	12,6	17,8	21,5	22,0	28,8	38,2	42,4
<i>Langue(s) parlée(s) à la maison (mère)</i>								
Français	1711	1,8	3,3	4,0	4,6	10,3	19,8	24,4
Anglais	208	4,3	8,2	10,6	11,5	17,8	25,5	28,9
Ni français ni anglais	128	15,6	22,7	25,0	25,8	32,8	43,0	48,4
Français, anglais et autre	37	5,4	5,4	5,4	5,4	5,4	16,2	21,6
Français ou anglais et autre	34	2,9	5,9	8,8	8,8	11,8	26,5	32,4
<i>Diplôme le plus élevé (mère)</i>								
Pas de DES	385	5,2	8,6	9,9	11,2	16,4	26,5	32,5
DES	555	2,4	4,4	5,9	6,5	13,5	20,6	27,7
Post-sec. Sauf univ.	611	2,9	4,0	4,0	4,0	9,5	20,7	25,1
Universitaire	566	2,3	4,1	4,9	5,7	9,6	19,0	21,9
<i>Niveau de suffisance du revenu</i>								
Suffisant	1571	1,4	2,8	3,3	3,8	9,3	18,5	23,1
Insuffisant	235	4,7	6,8	8,5	9,4	14,0	25,1	28,1
Très insuffisant	276	8,3	13,0	16,3	17,8	25,4	34,4	40,2
<i>Propriétaire du logement</i>								
Oui	1327	1,9	3,2	3,5	3,9	9,7	19,1	23,4
Non	793	4,8	7,9	10,1	11,2	16,9	26,4	31,5
Attrition cumulative	2120	3,0	5,0	6,0	6,7	12,4	21,8	26,4

Note : le tableau complet se trouve en annexe (tableau A.3.1).

Source : ÉLDEQ, ISQ.

L'examen des taux d'attrition cumulatifs pour les variables affectées par l'attrition selon la littérature montre l'effet différentiel de l'attrition sur les différentes catégories des variables. En effet, pour une même variable, il y a souvent de fortes

différences dans les taux d'attrition cumulatifs pour chaque catégorie, suggérant dès lors que l'attrition dans l'ÉLDEQ ne se produirait pas de façon aléatoire. Par exemple, les familles qui sont propriétaires de leur logement ont un taux d'attrition cumulatif plus faible que les familles qui sont locataires et ce, aux huit passages de l'enquête. Effectivement, 31,5% des familles n'étant pas propriétaires de leur logement en E1 ont quitté l'échantillon définitivement comparativement à 23,4% chez les familles propriétaires. De plus, ces résultats sont convergents avec ceux de la littérature alors que les variables liées à la pauvreté, à l'exclusion sociale et à l'instabilité présentent des taux d'attrition cumulatifs très différents d'une catégorie à l'autre. L'effet de l'attrition sur la taille de l'échantillon pour certains sous-groupes est également visible. Ainsi, en E1, 128 familles comprennent une mère biologique qui parle une autre langue que le français ou l'anglais à la maison. Toutefois, en E8, près de la moitié de ces familles ont décroché de l'enquête. Si on ajoute les familles qui ont raté quelques volets pour revenir par la suite dans l'enquête (non-réponse par unité), on se rend compte que seulement une minorité de ces familles ont répondu à tous les volets de l'ÉLDEQ. L'analyse longitudinale de ces familles est bien évidemment rendue plus ardue à réaliser suite à l'érosion particulièrement forte de ces familles.

Il faut cependant garder à l'esprit que ces taux sont calculés à partir des données du premier passage de l'ÉLDEQ. Il est donc fort possible, pour certaines variables, que plusieurs sujets changent de catégorie aux passages suivants (comme une famille initialement locataire qui devient propriétaire de leur logement ou bien une mère décrochant un nouveau diplôme ou un nouvel emploi). Néanmoins, à la lumière de ces résultats, il est vraisemblable de penser que l'attrition dans l'ÉLDEQ ne se produit pas de façon aléatoire et que ce phénomène réduit parfois considérablement la taille échantillonnale de certains sous-groupes au sein de l'échantillon de l'ÉLDEQ.

3.4 RATIOS D'ATTRITION

Après avoir présenté les taux d'attrition cumulatifs, ce chapitre se termine par la présentation des ratios d'attrition, indicateurs développés par Watson (2003) dans son analyse de l'attrition dans l'European Community Household Panel (ECHP). Ces ratios,

représentant en quelque sorte la pondération devant être utilisée au dernier volet (ici le volet 7) pour compenser pour l'attrition, permettent aussi de donner une bonne idée des catégories de variables qui pourraient être biaisées par l'attrition.

3.4.1 Méthodologie

Les ratios d'attrition (Watson, 2003), sont définis comme suit, pour une variable catégorielle donnée :

$$R = P_{f_i} / P_{r_i}$$

P_{f_i} = proportion i au premier volet

P_{r_i} = proportion i au dernier volet du sous-groupe ayant participé à tous les volets

Une catégorie d'une variable où le ratio sera supérieur à l'unité ($R > 1$) signifie que cette catégorie de la variable étudiée est particulièrement affectée par l'attrition. Par exemple, pour le sexe du répondant, si la proportion d'hommes dans l'enquête au premier volet (P_{f_i}) est plus élevée que celle au dernier volet (P_{r_i}) chez les sujets ayant participé à tous les volets, le ratio d'attrition sera supérieur à l'unité, signifiant alors que les hommes sont plus touchés par l'attrition que les femmes.

Ceci permet d'avoir rapidement une bonne idée des catégories de répondants qui sont plus sujets à quitter prématurément l'enquête. Ces ratios peuvent ainsi être vus comme étant la proportion par laquelle les répondants présents au dernier volet devraient être pondérés afin de compenser pour l'attrition. En outre, cette méthode permet de comparer des catégories de plusieurs variables différentes par rapport à leur impact sur le biais d'attrition.

Cependant, il faut noter que cette méthode est basée pour des variables discrètes. De plus, puisque cette méthode compare la distribution d'une variable catégorielle donnée à deux moments différents, le ratio d'attrition est sujet à être lui-même biaisé par les changements dans la vie des répondants. Par exemple, il est impossible que le niveau de scolarité d'un individu diminue dans le temps, ce qui fait en sorte qu'il soit normal d'assister à une baisse de la proportion d'individus ayant des niveaux de scolarité plus

faibles alors que l'enquête longitudinale progresse. Il est également possible que certaines catégories ne comprennent qu'un très petit nombre de répondants à un des deux volets utilisés pour le calcul du ratio d'attrition, ce qui peut évidemment mener à des résultats extrêmes étant donné le faible nombre d'observations.

Contrairement à la méthodologie de Watson, les numérateurs des ratios d'attrition seront constitués de l'ensemble des répondants au volet 1, en n'excluant pas les répondants qui sont devenus inéligibles au fil de l'enquête (par exemple suite à un déménagement hors Québec ou au décès de l'enfant-cible). Ce choix est convergent avec la définition de l'attrition qui est utilisée dans le cadre de ce mémoire.

3.4.2. Résultats

Le tableau 3.5 présente les ratios d'attrition pour quelques variables de nature sociodémographique.

Tableau 3.5
Quelques ratios d'attrition choisis construits à partir du volet 7 de l'ÉLDEQ

Variabes	Ratio d'attrition
<i>Statut d'immigrante (mère)</i>	
N'est pas immigrante	0,96
Immigrante européenne	1,32
Immigrante non européenne	1,58
<i>Statut d'immigrant (père)</i>	
N'est pas immigrant	0,97
Immigrant européen	0,93
Immigrant non européen	1,43
<i>Plus haut diplôme (mère)</i>	
Pas de diplôme secondaire	1,52
Diplôme secondaire	1,22
Diplôme post-secondaire (sauf univ.)	0,85
Diplôme universitaire	0,82
<i>Plus haut diplôme (père)</i>	
Pas de diplôme secondaire	1,30
Diplôme secondaire	1,20
Diplôme post-secondaire(sauf univ.)	0,83
Diplôme universitaire	0,90
<i>Principale source de revenus du ménage</i>	
Salaire et traitement	0,95
Travail autonome	0,75
Aide sociale	2,20
Assurance-emploi	2,57
Autres	1,22
<i>Propriétaire du logement</i>	
Oui	0,80
Non	1,70

Note : le tableau complet se trouve en annexe (tableau A.3.2).

Source : ÉLDEQ, ISQ.

L'examen des ratios d'attrition sélectionnés ici permet notamment de constater que les enfants provenant de familles au revenu très insuffisant ainsi que de celles qui ne sont pas propriétaires de leur logement tendent à décrocher davantage de l'échantillon. En outre, les mères peu scolarisées et qui sont immigrantes tendent à décrocher davantage de l'enquête. Ces conclusions vont donc dans le même sens que celles tirées avec la littérature sur l'attrition ainsi qu'avec les taux d'attrition cumulatifs.

3.5 CONCLUSION

L'analyse descriptive de l'attrition dans l'échantillon longitudinal de l'ÉLDEQ a tout d'abord permis de constater que 592 répondants du premier volet de l'ÉLDEQ n'ont pas répondu au huitième passage de l'enquête, ce qui représente 27,9% de l'échantillon initial de l'ÉLDEQ. De plus, après les huit premiers volets de l'ÉLDEQ, 560 répondants ont décroché de l'enquête en cessant de façon définitive leur participation à l'enquête. Dans les deux cas, la non-réponse par unité et l'attrition varient considérablement d'un volet à l'autre, en étant plutôt faibles lors de la première phase de l'ÉLDEQ et très fortes au début de la seconde phase. En outre, suite aux grands efforts réalisés par l'équipe de l'ÉLDEQ pour fidéliser les répondants et convaincre les décrocheurs de revenir dans l'enquête, un taux de non-réponse par unité négatif a été observé lors du huitième passage. Dans l'ensemble, pour l'ÉLDEQ, la non-réponse par unité et l'attrition sont fortement liées, suggérant ainsi que lorsqu'une famille ne répond pas à un volet de l'enquête, elle risque de ne pas répondre aux volets suivants aussi.

En créant des taux d'attrition cumulatifs et des ratios d'attrition pour quelques variables spécialement choisies, on a pu constater que l'attrition n'est visiblement pas aléatoire puisque certaines catégories sont nettement plus touchées par l'attrition que les autres, autant pour la quantité d'attrition que pour la distribution de l'attrition. Dans les deux cas, les résultats obtenus convergent avec ceux de la littérature alors que l'attrition touche principalement les sujets dont les caractéristiques sont liées à la pauvreté, à l'exclusion sociale et à l'instabilité semblent particulièrement affectées par l'attrition.

CHAPITRE 4 : LES CARACTÉRISTIQUES DES DÉCROCHEURS

Le chapitre précédent a permis constaté l'ampleur de l'attrition dans l'ÉLDEQ ainsi que son caractère déterministe pour différentes caractéristiques des familles. Sachant cela, il est donc possible d'atteindre le premier objectif de ce mémoire : tracer un portrait des caractéristiques des décrocheurs. Pour se faire, des tests d'indépendance du khi-deux et des analyses de variance à un facteur ont été effectuées pour comparer les caractéristiques des décrocheurs initiaux (après le premier passage de l'ÉLDEQ) de celles des décrocheurs de la seconde phase (ayant décroché à partir du sixième passage) et des familles ayant répondu à tous les volets, en date du huitième passage.

4.1 MÉTHODOLOGIE

Afin d'atteindre le premier objectif de ce mémoire, qui est l'identification des caractéristiques des familles décrochant prématurément de l'ÉLDEQ, l'approche classique dans l'analyse de l'attrition sera utilisée. Cette approche consiste à effectuer des tests d'indépendance du khi-deux (variables discrètes) et des analyses de variance à un facteur (variables continues) afin de comparer les caractéristiques des décrocheurs à celles des familles demeurant dans l'enquête. La façon dont cet objectif sera atteint sera introduite ici.

4.1.1 Groupes d'attrition

Afin de pouvoir comparer les caractéristiques des décrocheurs de celles des familles demeurant dans l'enquête tout en tenant compte de l'évolution particulière de l'attrition dans l'ÉLDEQ, trois groupes d'attrition seront formés :

- Le groupe A est constitué des ménages ayant seulement répondu au premier volet ;
- Le groupe B est constitué des ménages ayant répondu à tous les volets de la première phase de l'enquête (volets 1 à 5) et ayant arrêté de répondre à partir du début de la deuxième phase (volets 6 à 8) ;
- Le groupe C est constitué des ménages ayant répondu aux huit volets.

Cette procédure permet ainsi de distinguer les décrocheurs de la première phase de ceux de la seconde phase en plus d'être appropriée dans le cas de l'ÉLDEQ suite au grand nombre de passages de l'enquête. De plus, mentionnons que ces trois groupes comprennent 71,7% de l'ensemble des répondants du premier volet de l'enquête. La figure 4.1 illustre schématiquement les trois groupes d'attrition qui seront considérés dans les analyses.

Figure 4.1

Les trois groupes d'attrition étudiés selon les volets répondus par chaque groupe

	1	2	3	4	5	6	7	8
A (n = 63)	✓							
B (n = 122)	✓	✓	✓	✓	✓			
C (n = 1 336)	✓	✓	✓	✓	✓	✓	✓	✓

Ces trois groupes seront alors utilisés pour les analyses de variance et les tests d'indépendance du khi-deux. Dans le cas des analyses de variance, le test de Levene sera utilisé pour tester l'homogénéité de la variance entre les trois groupes d'attrition et le test post-hoc de Tukey sera utilisé pour faire des comparaisons deux-à-deux. En outre, il est important de mentionner qu'afin de respecter les règles de confidentialité du CADRISQ et les hypothèses du test du khi-deux, certaines catégories de plusieurs variables durent ont du fusionnées. Ceci réduit la sensibilité des résultats mais permet de respecter l'aspect confidentiel des données ainsi que les hypothèses du test. L'annexe méthodologique contient des détails sur le test du khi-deux (note 2), l'analyse de variance à un facteur et le test de Levene (note 3).

4.2 RÉSULTATS

Le tableau 4.1 présente les résultats pour quelques variables socioéconomiques spécialement choisies. Les décrocheurs initiaux constituent le groupe A alors que les décrocheurs de la seconde phase de l'ÉLDEQ sont regroupés au sein du groupe B et les familles toujours dans l'enquête forment le groupe C.

Il existe d'importantes différences dans les caractéristiques socioéconomiques des décrocheurs par rapport à celles des non-décrocheurs. Ces différences vont dans le même sens que celles obtenues pour les autres grandes enquêtes longitudinales portant sur la famille alors que les variables associées à l'exclusion sociale, à la pauvreté et à l'instabilité sont particulièrement affectées par l'attrition. Ainsi, les familles qui décrochent de l'ÉLDEQ sont notamment plus nombreuses à compter une mère dont la langue maternelle n'est pas le français ou l'anglais ($p < 0,01$) ou qui n'a pas complété ses études secondaires ($p < 0,01$). En outre, les familles quittant prématurément l'ÉLDEQ sont plus nombreuses à avoir un revenu familial très insuffisant ($p < 0,01$), une mère qui n'a pas travaillé au cours de la dernière année ($p < 0,01$) et à être locataire de leur logement ($p < 0,01$). Il est également fort intéressant d'observer que les familles du groupe B, ayant donc décroché de l'enquête au début de la seconde phase, possèdent des caractéristiques intermédiaires entre les groupes A et C. Par exemple, les familles du groupe B sont plus nombreuses à être propriétaires de leur logement que les familles du groupe A (63,10% contre 39,70%) mais moins nombreuses que les familles du groupe C (67,10%). Cette gradation dans les caractéristiques des familles de ces trois groupes montre bien le phénomène de sélection qui se produit pour l'attrition, renforçant l'idée que l'attrition n'est pas aléatoire dans l'ÉLDEQ.

Tableau 4.1
Différences dans les caractéristiques selon les trois groupes de décrocheurs pour quelques variables choisies (%)

Variables	A	B	C
<i>Groupe d'âge (mère)</i>			
24 ans et moins	33,3	18,2	21,4
25-29 ans	25,4	30,6	32,8
30 ans et plus	41,3	51,2	45,8
<i>Présence du père biologique*</i>			
Oui	84,1	91,0	93,2
Non	15,9	9,0	6,8
<i>Type de famille à l'enquête (3 cat.)*</i>			
Famille intacte	71,4	82,8	82,8
Famille recomposée	12,7	7,4	10,7
Famille monoparentale	15,9	9,8	6,5
<i>Première langue apprise (mère)***</i>			
Français	50,8	81,0	84,9
Anglais	7,9	7,4	7,4
Ni français ni anglais	41,3	11,6	7,7
<i>Plus au diplôme (mère)***</i>			
Pas dip. second.	32,3	16,5	15,1
Dip. études sec.	25,8	27,3	25,4
Dip. post-sec. (sauf univ.)	21,0	35,5	29,7
Dip. universitaire	21,0	20,7	29,8
<i>Travaille 12 derniers mois (mère)***</i>			
Non	53,2	29,2	27,1
Oui	46,8	70,8	72,9
<i>Niveau de suffisance du revenu (3 cat.)***</i>			
Revenu suffisant	39,3	73,1	80,6
Revenu insuffisant	19,6	9,2	10,4
Revenu très insuffisant	41,1	17,6	8,9
<i>Propriétaire du logement***</i>			
Oui	39,7	63,1	67,1
Non	60,3	36,9	32,9

* p < 0,1 ** p < 0,05 *** p < 0,01

A : familles ayant décroché dès le second volet.

B : familles ayant décroché au début de la seconde phase de l'ÉLDEQ.

C : familles ayant participé à tous les volets de l'ÉLDEQ.

Note : le tableau complet se trouve en annexe (tableau A.4.1).

Source : ÉLDEQ, ISQ.

4.3 CONCLUSION

Suite aux analyses effectuées ici, il est possible de conclure que des différences importantes existent entre les familles décrocheuses et celles demeurant dans l'échantillon de l'ÉLDEQ. Ces différences vont dans le même sens que celles observées dans les autres enquêtes longitudinales sur la famille et les ménages alors que les familles décrochant de l'ÉLDEQ possèdent des caractéristiques liées à la pauvreté, à l'exclusion sociale et à l'instabilité. Finalement, les nombreuses différences entre les décrocheurs et les non-décrocheurs confirment donc que l'attrition ne se produit pas de façon aléatoire dans l'ÉLDEQ, ce qui justifie la pertinence de l'analyse des biais pouvant être causés par l'attrition.

CHAPITRE 5 : LES FACTEURS QUI INFLUENCENT LA PROBABILITÉ D'ATTRITION

Après avoir réalisé que les décrocheurs possèdent des caractéristiques différentes du reste des répondants dans l'ÉLDEQ, il est pertinent de vérifier si les caractéristiques des familles influencent la probabilité de décrocher de l'enquête. Ce chapitre présente les résultats des régressions *probit* qui ont été effectuées afin de réaliser le second objectif du mémoire : modéliser la probabilité d'attrition d'une famille selon ses caractéristiques pour différentes variables d'intérêt afin de pouvoir estimer la quantité d'attrition. Les résultats des régressions *probit*, qui permettront notamment de vérifier la significativité des variables d'intérêt choisies pour prédire l'attrition, seront tout d'abord présentés. Par la suite, il sera alors possible de répondre au second objectif en présentant les probabilités d'attrition des familles de l'ÉLDEQ selon leurs caractéristiques. Cette stratégie fut notamment utilisée par Hoddinot et *al.* (2002) et Fitzgerald et *al.* (1998) pour estimer la probabilité d'attrition d'un sujet en fonction de ses caractéristiques.

5.1 MÉTHODOLOGIE

Afin de modéliser la probabilité d'attrition d'une famille, une série de régressions *probit* pour une gamme de variables d'intérêt sera réalisée à l'aide des données de E1, le seul passage où il n'y a pas encore d'attrition. La transformation *probit* est utile pour modéliser une variable dichotomique (comme une probabilité) en permettant notamment de traiter une variable dépendante dont les catégories ne sont pas indépendantes entre elles. Pour chaque variable d'intérêt, une régression *probit* unidimensionnelle sera tout d'abord exécutée dans le but d'évaluer l'effet brut de la variable d'intérêt sur la probabilité de décrocher de l'ÉLDEQ. Par la suite, des variables de contrôle seront ajoutées simultanément, afin de déterminer l'effet net de chaque variable d'intérêt. Une variable dont l'effet est statistiquement significatif voudrait alors dire que cette variable a un effet significatif sur la probabilité de décrocher de l'ÉLDEQ. Ceci peut être vérifié en examinant directement le coefficient et la valeur-p de la variable en question ou bien à l'aide d'un test conjoint, qui permet de vérifier si l'ajout d'une nouvelle variable à un modèle est significatif ou non. Un modèle contenant seulement les variables de contrôle

sera également réalisé afin d'évaluer l'effet des caractéristiques des familles sur la probabilité d'attrition sans aucune variable d'intérêt. L'annexe méthodologique contient des renseignements supplémentaires concernant la régression *probit* (note 2) et le test du rapport de vraisemblance (note 4).

Étant donné le caractère multidisciplinaire de l'ÉLDEQ et le très grand nombre de variables se retrouvant dans l'enquête, les régressions *probit* seront réalisées pour une gamme de variables non seulement sujettes à être affectées par l'attrition dans la littérature mais aussi susceptibles d'intéresser les chercheurs provenant de différents domaines, allant de la démographie de la famille à la psychologie du développement. Le tableau 5.1 présente les variables d'intérêt qui seront étudiées ainsi que les variables de contrôle qui seront prises en compte dans chaque régression. De surcroît, la sélection des variables de contrôle se base sur l'analyse de l'attrition dans l'ELNEJ réalisée par Hoddinott et *al.* (2002) suite aux similitudes entre les deux enquêtes, qui ciblent toutes les deux des cohortes d'enfants canadiens.

Tableau 5.1
Variables d'intérêt et variables de contrôle des régressions *probit* modélisant la probabilité d'attrition

Variables de contrôle	Variables d'intérêt
Caractéristiques de la mère :	Mère immigrante
- Âge	Autre union de la mère (sauf avec le père)
- Plus haut diplôme obtenu	Surprotection parentale (mère)
- Type d'emploi	Perception d'impact parental (mère)
Nombre de frères/sœurs	Présence du père biologique
Propriétaire du logement	Famille monoparentale à la naissance de l'enfant
Revenus du ménage	Statut socioéconomique du ménage
Revenus du ménage au carré	Insuffisance du revenu
	Avortement précédant la naissance
	Naissance de rang 1
	Naissance prématurée
	Retard de croissance de l'enfant

Note : les régressions sont effectuées avec les données de E1.

5.2 RÉSULTATS

Premièrement, le tableau 5.2 présente les résultats de ces régressions pour des variables d'intérêt sélectionnées et pour l'ensemble des variables de contrôle. Il faut rappeler ici que les paramètres de régression représentent un changement dans la probabilité cumulée que la famille décroche de l'enquête et non directement dans la probabilité de décrocher. Dans le modèle contenant uniquement les variables de contrôle, seul le plus haut diplôme obtenu par la mère prédit de façon significative l'attrition. La relation va dans le sens attendu alors que moins la mère est scolarisée par rapport aux mères détenant un diplôme universitaire, plus ses chances de décrocher de l'ÉLDEQ sont élevées (Catégorie pas de diplôme d'études secondaire : $\hat{\beta} = 0,454$, $p < 0,01$). Toutefois, il faut préciser ici que la forte variabilité due aux petits effectifs des deux modalités supérieures de la variable représentant le nombre de frères et de sœurs de l'enfant rend les coefficients de ces deux modalités non significatives malgré la forte valeur de l'estimation ponctuelle. Pour les variables d'intérêt choisies, l'analyse au niveau unidimensionnel révèle que de toutes les variables d'intérêt étudiées, seules les variables indiquant si la naissance de l'enfant était prématurée ($\hat{\beta} = 0,021$, $p = 0,697$), si l'enfant est une naissance de premier rang ($\hat{\beta} = 0,037$, $p = 0,428$) ou bien si la mère avait déjà eu une union avant celle avec le père ($\hat{\beta} = -0,056$, $p = 0,374$) et si la mère avait eu au moins un avortement ($\hat{\beta} = -0,037$, $p = 0,440$) ne sont pas fortement significatives pour prédire l'attrition. Cependant, lorsque les variables de contrôle sont ajoutées au modèle, seuls le statut d'immigrante de la mère ($\hat{\beta} = 0,387$, $p < 0,01$) et le fait que la mère ait eu au moins un avortement ($\hat{\beta} = -0,164$, $p < 0,05$) sont significatifs. Ainsi, par exemple, le fait que la mère de l'enfant soit immigrante augmente la cote Z^7 , haussant alors la probabilité de décrocher de l'ÉLDEQ. Il en est de même pour l'avortement alors que les mères ayant eu au moins un avortement ont moins de chances de quitter prématurément l'ÉLDEQ que les femmes n'ayant jamais eu d'avortement. Sans surprise, les seuls tests conjoints significatifs sont ceux pour le statut d'immigrante

⁷ La régression *probit* permet de calculer des cotes Z, lesquelles servent ensuite à calculer des probabilités.

de la mère de l'enfant ($\chi^2_1 = 7,30, p < 0,01$) et le fait d'avoir eu un avortement ($\chi^2_1 = 3,84, p < 0,05$), signifiant donc que l'ajout de ces variables est pertinent dans la prédiction de l'attrition.

Tableau 5.2
Coefficients des régressions *probit* selon certaines variables d'intérêt et de contrôle sélectionnées

	Variables de contrôle uniquement	Présence du père biologique	Naissance prématurée	Avortement	Mère immigrante	Statut Socio-économique du ménage	Perception d'impact Parental (mère)	Naissance de rang 1
<i>Modèle unidimensionnel</i>								
Constante		-0,120 (0,104)	-0,442*** (0,030)	-0,430*** (0,037)	-0,497*** (0,031)	-0,443*** (0,030)	-0,069 (0,135)	-0,458*** (0,040)
Variable d'intérêt		-0,347*** (0,108)	0,021 (0,138)	-0,037 (0,062)	0,509*** (0,093)	-0,150*** (0,031)	-0,044*** (0,016)	0,037 (0,059)
<i>Modèle multidimensionnel</i>								
Constante	-0,968*** (0,311)	-0,959*** (0,368)	-0,962*** (0,311)	-1,095*** (0,308)	-0,903*** (0,313)	-0,976*** (0,314)	-0,822** (0,380)	-1,269*** (0,375)
Plus haut diplôme (mère) (Réf : Dip. Universitaire) Pas dip. Second.	0,454*** (0,167)	0,453*** (0,167)	0,452*** (0,167)	0,478*** (0,166)	0,487*** (0,167)	0,431** (0,194)	0,444*** (0,167)	0,472*** (0,167)
Locataire du logement	0,135 (0,092)	0,138 (0,092)	0,135 (0,092)	0,168** (0,090)	0,093 (0,093)	0,135 (0,092)	0,132 (0,092)	0,124 (0,092)
Variable d'intérêt		-0,016 (0,207)	-0,196 (0,199)	-0,164** (0,084)	0,387*** (0,143)	-0,024 (0,103)	-0,016 (0,024)	0,283 (0,199)
Test χ^2 pour la relation globale (d.l.) [valeur p]	31,755 (15) [0,007]	32,005 (16) [0,010]	32,753 (16) [0,008]	36,683 (16) [0,002]	39,055 (16) [0,001]	31,812 (16) [0,011]	32,190 (16) [0,009]	33,770 (16) [0,006]
Test χ^2 conjoint de la Variable d'intérêt (d.l.) [valeur p]		0,881 (1) [0,348]	0,998 (1) [0,318]	3,84 (1) [0,050]	7,3 (1) [0,007]	0,057 (1) [0,811]	0,435 (1) [0,509]	2,015 (1) [0,156]

Notes : Les erreurs standard des paramètres de régression se trouvent entre parenthèses.

* p < 0,1 ** p < 0,05 *** p < 0,01

Le tableau complet se trouve en annexe (tableau A.5.1).

Source : ÉLDEQ, ISQ.

Deuxièmement, afin de prédire la probabilité d'attrition pour chaque variable d'intérêt, une probabilité d'attrition de base est tout d'abord générée à l'aide des variables de contrôle de chaque modèle. Pour les variables continues, la valeur moyenne est prise alors que pour la valeur modale est choisie pour les variables discrètes. Il est important de noter que dans ce modèle, la variable d'intérêt est soit à sa valeur moyenne (variable continue) ou nulle (variable discrète). Par la suite, on change la valeur de la variable d'intérêt en ajoutant un écart-type (variable continue) ou en la mettant à un (variable discrète). Les composantes de cette probabilité de base pour la variable de contrôle significative, le plus haut diplôme obtenu par la mère, seront tout d'abord analysées. Ensuite, le changement dans la probabilité de décrocher des variables d'intérêt à partir de la probabilité de base en changeant la modalité si la variable d'intérêt est discrète ou en augmentant sa valeur d'un écart-type par rapport à sa moyenne si la variable d'intérêt est continue sera étudié. Les résultats de ces calculs se trouvent au tableau 5.3. Il est important de mentionner que les cas de base varient légèrement d'une variable à l'autre étant donné la faible variation (généralement non significative) des coefficients des variables de contrôle et de la constante lors de l'ajout d'une variable d'intérêt dans chaque modèle.

La probabilité de quitter prématurément l'échantillon longitudinal de l'ÉLDEQ est beaucoup plus forte chez les familles dont la mère n'a pas de diplôme d'études secondaires (0,398) que dans les familles où la mère est universitaire (0,238). C'est également le cas si la mère possède un diplôme d'études secondaires (0,323) ou un diplôme d'études post-secondaires (sauf université) (0,306). Ceci montre la direction du lien entre scolarité et attrition : plus la mère de l'enfant est scolarisée, plus les chances que la famille décroche de l'ÉLDEQ sont faibles. En outre, les familles dont la mère est immigrante (0,343) sont également beaucoup plus sujettes à quitter l'échantillon que les familles où la mère n'est pas immigrante (0,214), tout comme les familles dont la mère a eu au moins un avortement (0,209) par rapport à celles dont la mère n'a pas eu d'avortement (0,257). Il faut également souligner la probabilité de décrocher si l'enfant est une naissance de rang un (0,233) qui est plus élevée que la probabilité de base (0,156) même si ceci n'est pas statistiquement significatif. Les probabilités de décrocher

de l'ÉLDEQ pour les autres variables d'intérêt sont très rapprochées du cas de base étant donné le caractère non significatif de ces variables dans la prédiction de la probabilité d'attrition.

Tableau 5.3

Probabilité prédites de décrocher de l'ÉLDEQ et variation marginale (Δp) de cette probabilité pour la variable de contrôle significative ainsi que pour l'ensemble des variables d'intérêt

Caractéristiques des familles répondantes	Probabilité de décrocher de base	Probabilité de décrocher selon la caractéristique	Δp
<i>Plus haut diplôme (mère)</i>			
Mère sans diplôme d'études secondaires	0,238	0,398***	0,160
Mère avec diplôme d'études secondaires	0,238	0,323**	0,085
Mère avec diplôme post-secondaire (sauf univ.)	0,238	0,306*	0,068
Père biologique présent	0,243	0,238	-0,005
Enfant né prématurément	0,242	0,185	-0,057
Revenu familial insuffisant	0,233	0,257	0,024
Mère immigrante	0,214	0,343***	0,129
Statut socioéconomique (+ σ)	0,214	0,234	0,020
Surprotection parentale (mère) (+ σ)	0,236	0,253	0,017
Perception d'impact parental (mère) (+ σ)	0,239	0,230	-0,009
Enfant étant une naissance de rang 1	0,156	0,233	0,077
Autre union (sauf avec père)	0,234	0,238	0,004
Famille monoparentale à la naissance	0,237	0,249	0,012
Retard de croissance	0,241	0,212	-0,029
Avortement	0,257	0,209**	-0,048

* $p < 0,1$

** $p < 0,05$

*** $p < 0,01$

Note : pour les variables continues, la variation de la probabilité de décrocher de l'enquête est mesurée en ajoutant un écart-type (+ σ) à la moyenne de cette variable.

Source : ÉLDEQ, ISQ.

5.3 CONCLUSION

L'analyse des probabilités d'attrition a permis de constater que, bien que la plupart des variables d'intérêt sont significatives dans la prédiction de la probabilité d'attrition au niveau unidimensionnel, une fois les variables de contrôle ajoutées, seuls le statut d'immigrante de la mère et le fait que la mère ait eu au moins un avortement demeurent significativement liés à l'attrition. Ceci fait en sorte que, sauf pour ces deux variables, la probabilité de décrocher de l'ÉLDEQ varie faiblement pour les variables d'intérêt. De surcroît, il est intéressant de constater que parmi les variables de contrôle, seul le plus haut diplôme obtenu par la mère de l'enfant prédit de façon significative l'attrition au niveau multidimensionnel. Dans ce cas, moins la mère est scolarisée, plus la probabilité de décrocher est forte. En somme, sauf pour le cas du statut d'immigrante de la mère et du plus haut diplôme obtenu par la mère, la probabilité d'un sujet de décrocher de l'échantillon de l'ÉLDEQ n'est pas liée avec les caractéristiques de la famille. Ceci montre le caractère partiellement aléatoire de l'attrition et la difficulté de bien modéliser la probabilité d'attrition d'une famille.

CHAPITRE 6 : LES BIAIS CAUSÉS PAR L'ATTRITION

Après avoir analysé le premier effet de l'attrition dans le chapitre précédent en modélisant la probabilité d'attrition d'une famille selon ses caractéristiques, le second effet de l'attrition, les biais, sera étudié ici. Dans ce chapitre, les résultats pour atteindre le troisième objectif de ce mémoire, l'évaluation des biais engendrés par l'attrition pour certaines variables d'intérêt, seront présentés. À cette fin, des tests classiques de Beckett, Gould, Lillard et Welch (BGLW) (Beckett et *al.*, 1988) seront effectués pour quelques variables susceptibles d'être utilisées par les chercheurs travaillant avec les données de l'ÉLDEQ.

6.1 MÉTHODOLOGIE

Dans le but de vérifier si une certaine variable d'intérêt est biaisée par l'attrition, la procédure développée par Beckett, Gould, Lillard et Welch (Beckett et *al.*, 1988) sera utilisée. Cette procédure permet de modéliser la variable d'intérêt au premier passage de l'enquête par les caractéristiques des répondants ainsi que par une variable dichotomique représentant l'attrition du sujet de E2 à E8 où la valeur un est donnée si le sujet a quitté l'étude entre E2 et E8. Le choix du premier passage s'explique par le fait qu'il s'agit du seul passage où tous les répondants sont présents. L'objectif du test est de vérifier si l'ajout de la variable dichotomique d'attrition influence les coefficients des autres variables de façon significative. Ceci est testé à l'aide d'un test de Fisher partiel pour la régression par les moindres carrés (variables dépendantes continues) et par un test de rapport de vraisemblance pour la régression *probit* (variables dépendantes dichotomiques). Si l'ajout de la variable d'attrition est significatif, la variable d'intérêt étudiée s'avère ainsi être biaisée par l'attrition et les analyses faites avec cette variable d'intérêt devront tenir compte de ce biais. Les variables d'intérêt et les variables de contrôle qui seront utilisées pour les tests BGLW sont les mêmes que pour les régressions faites lors du second objectif, pour modéliser la probabilité d'attrition (voir tableau 5.1). L'annexe méthodologique contient des renseignements supplémentaires concernant le test F partiel (note 5) et le test du rapport de vraisemblance (note 6).

6.2 RÉSULTATS

Le tableau 6.1 montre les résultats pour quelques variables d'intérêt choisies. Parmi l'éventail de variables d'intérêt présentées, seuls le statut d'immigrante de la mère ($\hat{\beta}_{attrition} = 0,375, p < 0,01$) et le fait que la mère ait eu au moins un avortement avant la naissance de l'enfant ($\hat{\beta}_{attrition} = -0,170, p < 0,05$) s'avèrent être biaisés de façon significative par l'attrition. Ceci signifie par exemple que la probabilité pour la mère biologique de l'enfant d'avoir eu au moins un avortement est significativement plus faible chez les familles décrocheuses que chez les familles toujours dans l'enquête. Cette différence significative cause donc un biais pour cette variable. Ces résultats s'observent également avec le test conjoint ($\chi^2 = 4,030, p = 0,045$), signifiant que l'ajout de la variable d'attrition pour expliquer la probabilité que la mère ait eu au moins un avortement est pertinent au modèle.

Tableau 6.1
Tests BGLW transversaux pour des variables d'intérêt et de contrôle sélectionnées

	Présence du père biologique	Naissance prématurée	Avortement	Mère immigrante	Statut socio-économique du ménage	Perception d'impact parental (mère)	Naissance de rang 1
Modèle utilisé	<i>Probit</i>	<i>Probit</i>	<i>Probit</i>	<i>Probit</i>	MCO	MCO	<i>Probit</i>
Constante	1,945*** (0,603)	-1,783*** (0,504)	-2,420*** (0,308)	-3,277*** (0,453)	-,578*** (0,159)	9,042*** (0,667)	3,357*** (0,615)
<i>Plus haut diplôme (mère)</i> (Réf : Dip. Universitaire)							
Pas dip. Second.	-0,416 (0,341)	-0,238 (0,276)	0,227 0,165	-0,405* (0,240)	-0,960*** (0,047)	-0,648*** (0,198)	-0,792** (0,317)
Locataire du logement	-0,192 (0,165)	0,067 (0,147)	0,317*** (0,090)	0,698*** (0,124)	0,007 (0,026)	-0,245** (0,109)	0,564*** (0,175)
Effet de l'attrition	-0,068 (0,163)	-0,089 (0,147)	-0,170** (0,085)	0,375*** (0,116)	-0,008 (0,024)	-0,065 (0,102)	0,160 (0,157)
Test χ^2 /Test F pour la relation globale avec l'attrition (d.l.) [valeur p]	91,097 (16) [<,001]	17,863 (16) [0,332]	70,306 (16) [<,001]	118,963 (16) [<,001]	328,290 (16, 127) [<,001]	7,820 (16, 125) [<,001]	1429,854 (16) [<,001]
Test χ^2 conjoint/Test F partiel de l'attrition (d.l.) [valeur p]	0,169 (1) [0,681]	0,376 (1) [0,540]	4,030 (1) [0,045]	10,091 (1) [0,001]	0,107 (1,127) [0,743]	0,392 (1,125) [0,531]	1,040 (1) [0,308]

Notes : Les erreurs standard des paramètres de régression se trouvent entre parenthèses.

* p < 0,1 ** p < 0,05 *** p < 0,01

Le tableau complet se trouve en annexe (tableau A.6.1).

Source : ÉLDEQ, ISQ.

6.3 CONCLUSION

Les tests BGLW indiquent que, parmi les variables d'intérêt choisies, seuls le statut d'immigrante de la mère et le fait que la mère ait eu au moins un avortement sont biaisés de façon significative par l'attrition. Conséquemment, il est possible d'avancer que, de façon générale, la principale conséquence de l'attrition au niveau multidimensionnel sur les variables d'intérêt de l'ÉLDEQ est la réduction graduelle de la taille de l'échantillon. Toutefois, plusieurs variables d'intérêt ont une certaine tendance, plus ou moins forte selon la variable, à être biaisées par l'attrition. Ainsi, il se pourrait qu'un biais de très faible envergure soit présent pour certaines variables d'intérêt, que des biais significatifs pourraient apparaître dans les volets futurs ou bien que des variables qui ne furent pas étudiées ici pourraient être biaisées par l'attrition. Ceci suggère alors la nécessité pour les chercheurs de vérifier si l'attrition biaise leurs variables d'intérêt. En terminant, les résultats obtenus ici ne sont pas surprenants étant donné les résultats obtenus précédemment pour la probabilité d'attrition (voir chapitre 5). Effectivement, puisque les deux techniques sont liées, on pouvait s'attendre à obtenir des résultats statistiques similaires (bien que les conclusions tirées à partir des résultats obtenus soient différentes étant donné que la probabilité d'attrition, et non les variables d'intérêt, est modélisée au chapitre).

CHAPITRE 7 : VALIDATION DES POIDS LONGITUDINAUX

Après avoir étudié les conséquences de l'attrition sur l'échantillon de l'ÉLDEQ dans les chapitres précédents, il est maintenant pertinent de vérifier si l'une des méthodes les plus utilisées pour contrer l'attrition, l'utilisation de poids longitudinaux corrigés pour tenir compte de la non-réponse, est une technique réellement efficace pour contrer les biais causés par l'attrition. Cet objectif est réalisé en deux étapes permettant une analyse unidimensionnelle rétrospective et une analyse unidimensionnelle prospective. Ce chapitre présente tout d'abord la méthodologie pour chaque étape pour ensuite en montrer les résultats.

7.1 MÉTHODOLOGIE

7.1.1 Description des poids longitudinaux

À l'instar des poids échantillonnaires classiques ou transversaux, les poids longitudinaux permettent une inférence longitudinale de l'échantillon à la population visée par l'enquête qui est sans biais. Ils diffèrent des poids transversaux puisqu'ils sont notamment construits à partir de la population pour laquelle l'échantillon longitudinal initial a été sélectionné. En outre, les poids longitudinaux sont généralement ajustés pour la non-réponse et pour l'attrition, faisant en sorte qu'ils ont des valeurs différentes des poids transversaux alors que l'enquête longitudinale progresse.

Dans le cadre de l'ÉLDEQ, l'ajustement de la non-réponse et de l'attrition est basé sur la création de classes de pondération à l'aide de la modélisation par segmentation exécutée avec l'algorithme CHAID (« *Chi-square automatic interaction detection* ») (Étude longitudinale du développement des enfants du Québec, 2003). Pour chaque famille, l'ajustement du poids longitudinal pour les données du QIRI consiste à diviser le poids de référence par la proportion des familles répondantes observée au sein de la classe à laquelle la famille appartient. Cet ajustement a été effectué pour les variables suivantes : 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.

7.1.2 Méthodes d'analyse

La validation des poids longitudinaux se fera en deux étapes. Ceci s'explique par la grande difficulté à trouver une seule méthode vraiment efficace dans l'évaluation des poids longitudinaux pour la correction des biais d'attrition. L'absence d'une telle méthode et les nombreuses difficultés rencontrées pour atteindre cet objectif justifient alors la réalisation de cet objectif en deux étapes.

Tout d'abord, l'effet des poids longitudinaux au niveau unidimensionnel sera étudié en comparant les valeurs des différentes modalités de quelques variables pour E1 et E8 en utilisant les données pondérées. Puisque les variables qui seront choisies ici sont des variables qui ne devraient pas varier de façon significative dans le temps, on ne devrait pas retrouver de différence significative entre les données pondérées de E1 et de E8. Ceci permet donc d'isoler l'effet de l'attrition des autres variations de la variable se déroulant au fil du temps. Par exemple, dans le cas de l'ÉLDEQ, il est possible de constater une érosion considérable de la proportion des familles intactes simplement suite aux dynamiques matrimoniales des couples québécois. Cette analyse rétrospective de l'efficacité des poids longitudinaux ne sera toutefois que descriptive étant donné la difficulté à trouver des modèles statistiques appropriés pour tester les différences dans les modalités entre E1 et E8. Effectivement, une telle analyse requiert l'utilisation d'un test de proportions pour données appariées. Or, les tests de ce genre qui existent ne sont pas construits pour tester les variations dans les proportions des modalités d'une variable qui proviennent de la diminution de la taille échantillonnale. Ainsi, par exemple, le test le plus souvent utilisé dans ce genre de situations, le test de non-paramétrique de McNemar, permet de tester, à l'aide de données appariées, les variations dans la proportion d'une variable binaire à l'aide d'un test du khi-deux. Toutefois, ce test se base sur les changements d'état des sujets et non sur les différences dans la taille échantillonnale entre les deux mesures (Rice, 1995).

Par la suite, toujours au niveau unidimensionnel, on comparera les caractéristiques des décrocheurs de la seconde phase de l'ÉLDEQ (à partir de E6) par

rapport à celles des familles étant dans l'enquête à chaque passage en utilisant les poids. Contrairement au premier objectif, qui était d'obtenir un portrait des décrocheurs, on considère ici l'ensemble des décrocheurs des volets six, sept et huit et non seulement les décrocheurs du volet six. Cette décision est prise avec l'idée, non pas d'obtenir le portrait des décrocheurs, mais plutôt d'évaluer l'efficacité des poids longitudinaux dans la correction des biais unidimensionnels causés par l'attrition. À cette fin, il a été jugé pertinent de rassembler l'ensemble des décrocheurs de la deuxième phase de l'ÉLDEQ pour obtenir un nombre de décrocheurs plus élevé. En outre, étant le dernier volet de la première phase de l'ÉLDEQ, le volet E5 est un choix stratégique offrant à la fois un horizon longitudinal intéressant (E1 à E5) et un bon nombre de familles décrocheuses après ce volet. Les comparaisons seront effectuées à l'aide de tests du khi-deux (pour les variables discrètes) et de tests de Student pour échantillons indépendants (pour les variables continues) en prenant soin d'utiliser les poids longitudinaux de E5. De plus, pour tenir compte de l'effet de plan de l'ÉLDEQ, les poids longitudinaux furent divisés par l'effet de plan moyen de l'ÉLDEQ (1,3) pour faire les tests du khi-deux. Si les poids longitudinaux corrigent efficacement l'attrition, on ne devrait pas retrouver de différence significative entre les décrocheurs et les familles restant dans l'enquête. Le test de Student est détaillé dans la note 7 de l'annexe méthodologique.

Pour la première étape, les variables de pondération longitudinale PEPCML8 (volet 8) et PEPCM1M (volet 1) seront utilisées alors que pour la seconde étape, la variable PEPCM5 (volet 5) sera utilisée.

7.2 RÉSULTATS

7.2.1 Analyse unidimensionnelle rétrospective

La première étape pour valider l'efficacité des poids longitudinaux consiste à comparer la distribution des fréquences pour des variables fixes ou quasi-fixes aux volets un et huit en utilisant les poids longitudinaux de chaque volet. Les résultats, pour quelques variables choisies, se trouvent dans le tableau 7.1. Si les poids longitudinaux corrigent bien pour l'attrition, on ne devrait pas retrouver des différences importantes entre les distributions pour ces variables puisque celles-ci sont fixes ou quasi-fixes.

Malgré l'application des poids longitudinaux, il existe encore des petites différences dans la distribution des fréquences de quelques variables entre les volets 1 et 8 de l'ÉLDEQ. Par exemple, la situation conjugale de la famille à la naissance de l'enfant varie considérablement où la proportion des couples mariés au moment où l'enfant est né passe de 18,7% à 15,5% alors que la proportion des couples mariés après une union libre passe de 25,2% au volet 1 à 27,4% au volet 8. Compte tenu du nombre élevé d'observations aux deux volets (rappelons que les poids longitudinaux en E8 sont basés sur 1 336 familles), il est raisonnable de supposer que quelques-unes de ces différences sont statistiquement significatives. Toutefois, la possibilité d'appliquer des tests de proportions pour données appariées qui seraient adaptés à cette situation permettrait de tirer des conclusions beaucoup plus rigoureuses.

Tableau 7.1
Fréquences (%) de quelques variables fixes pour les volets un et huit, données pondérées

Variables	Volet 1	Volet 8
<i>Statut immigrant (mère)</i>		
N'est pas immigrante	84,6	86,8
Immigrante européenne	3,3	2,7
Immigrante non européenne	12,2	10,4
<i>Statut immigrant (père)</i>		
N'est pas immigrant	84,0	85,6
Immigrant européen	3,2	3,3
Immigrant non européen	12,7	11,1
<i>Langue maternelle (mère)</i>		
Français	76,4	78,9
Anglais	8,7	8,2
Ni français ni anglais	14,9	13,0
<i>Langue maternelle (père)</i>		
Français	76,0	77,9
Anglais	8,3	8,2
Ni français ni anglais	15,7	13,9
<i>Rang de naissance de l'enfant*</i>		
1	43,6	44,4
2	39,6	40,1
3	11,3	10,8
4	3,4	3,0
5+	2,1	1,6
<i>Type de famille à la naissance de l'enfant*</i>		
Intacte	78,0	78,8
Intacte (eua)	2,8	2,4
Recomposée : mère	7,0	6,8
Recomposée : père seul.	3,8	4,2
Monoparentale	8,4	7,8
<i>Situation conjugale à la naissance de l'enfant*</i>		
Mariés	18,7	15,5
Mariage précédé d'une union libre	25,2	27,4
Union libre	47,6	49,2
Séparés	3,8	3,8
Parents n'ont jamais vécu ensemble	4,7	4,0

* Ces variables ne sont disponibles qu'en E1. Les différences dans la distribution de celles-ci provient alors que de l'érosion graduelle de l'échantillon.

Note : le tableau complet se trouve en annexe (tableau A.7.1).

Source : ÉLDEQ, ISQ.

7.2.2 Analyse unidimensionnelle prospective

La seconde étape pour valider l'efficacité des poids longitudinaux consiste à comparer les caractéristiques des décrocheurs et celles des non-décrocheurs pour la seconde phase de l'ÉLDEQ de façon prospective. Ceci est fait en utilisant les données et les poids longitudinaux de E5 en distinguant les décrocheurs de la seconde phase aux familles qui répondent toujours à l'ÉLDEQ. Si les poids longitudinaux corrigent l'attrition de façon efficace, il ne devrait pas y avoir de différence statistiquement significative entre ces deux groupes. Le tableau 7.2 résume les tests statistiques effectués pour quelques variables choisies.

Ce tableau indique qu'il existe de nombreuses différences entre les familles décrocheuses à la seconde phase de l'enquête et celles demeurant dans l'ÉLDEQ même avec l'application des poids longitudinaux. En ayant notamment une mère généralement moins instruite (Aucun diplôme d'études secondaires : 14,4% chez les non décrocheurs contre 20,4 chez les décrocheurs) et un père biologique plus souvent absent (décrocheurs : 22,0%, non décrocheurs : 16,5%), les familles décrocheuses se distinguent toujours par leurs caractéristiques associées à la pauvreté, à l'exclusion sociale et à l'instabilité.

Tableau 7.2
Caractéristiques (%) des décrocheurs et des non-décrocheurs en E5 (données pondérées)

Catégorie	Non décrocheurs	Décrocheurs
<i>Diplôme le plus élevé (mère/conjointe)***</i>		
Pas dip. second.	14,4	20,4
Dip. d'études sec.	20,9	23,3
Dip. post-sec. (sauf univ.)	35,1	36,8
Dip. universitaire	29,5	19,5
<i>Présence du père biologique***</i>		
Oui	83,5	78,0
Non	16,5	22,0
<i>Statut d'immigrante (mère)***</i>		
Pas immigrante	87,9	80,5
Immigrante européenne	3,0	4,1
Immigrante non-européenne	9,1	15,4
<i>Première langue apprise (mère)***</i>		
Français	79,3	71,6
Anglais	8,5	9,4
Ni français ni anglais	12,1	19,0
<i>Propriétaire du logement***</i>		
Oui	71,7	64,3
Non	28,3	35,7
<i>Statut civil (mère)***</i>		
Mariée	50,3	43,0
Célibataire	43,2	49,9
Séparée légalement ou veuve	6,5	7,2
<i>Principale source de revenu du ménage*</i>		
Salaire et traitement	80,7	75,7
Travail autonome	9,5	12,3
Aide sociale	6,5	7,3
Assurance-emploi	1,8	1,8
Autres	1,5	3,0

* p < 0,1 ** p < 0,05 *** p < 0,01

Note : le tableau complet se trouve en annexe (tableau A.7.2).

Source : ÉLDEQ, ISQ.

7.3 CONCLUSION

Malgré l'application des poids longitudinaux, il existe encore de faibles différences dans la distribution de certaines variables fixes ou quasi-fixes dans le temps. L'ampleur limitée des différences observées suggère alors que l'utilisation des poids longitudinaux est pertinente bien que légèrement imparfaite. De plus, il existe encore des différences importantes entre les familles qui resteront dans l'enquête et celles qui décrocheront. Bien que les poids ne soient évidemment pas faits pour corriger pour les futurs décrocheurs, ceci montre néanmoins les limites de la correction faite par les poids longitudinaux. Toutefois, une évaluation longitudinale multidimensionnelle plus systématique, pouvant par exemple être basée sur l'analyse de survie ou l'analyse longitudinale de trajectoires, pourrait s'avérer très utile dans l'évaluation de l'efficacité des poids longitudinaux.

DISCUSSION ET CONCLUSION

L'attrition est un problème inévitable dans la qualité des données provenant d'enquêtes longitudinales. Cependant, les enquêtes longitudinales sont trop riches en information pour limiter leur utilisation simplement à cause de l'attrition. Ainsi, les chercheurs utilisant des données longitudinales devront tenir compte de l'effet que l'attrition peut avoir sur la qualité de leurs résultats, autant pour la diffusion des résultats que dans la comparabilité entre différentes enquêtes longitudinales. Ce mémoire portait donc sur une analyse de l'attrition pour une enquête longitudinale québécoise traitant sur le développement des enfants : l'Étude longitudinale du développement des enfants du Québec (ÉLDEQ). Cette analyse de l'attrition se divisait en quatre objectifs principaux : tracer un profil des décrocheurs, estimer la probabilité d'attrition des familles, estimer les biais causés par l'attrition et évaluer l'efficacité des poids longitudinaux dans la correction des biais occasionnés par l'attrition.

Tout d'abord, on a vu qu'à l'instar des autres enquêtes longitudinales, l'ÉLDEQ est affectée par l'attrition. En effet, en date du huitième volet, l'attrition touche près du quart de l'échantillon longitudinal de l'ÉLDEQ. De plus, l'attrition ne se produit pas de façon uniforme, survenant principalement lors de la seconde phase de l'enquête. Il est également très intéressant de remarquer que les efforts de l'équipe de l'ÉLDEQ afin de fidéliser les répondants et de favoriser le retour des familles ayant quitté l'enquête semblent porter fruit alors que la taille échantillonnale augmente au huitième volet. Puisque les résultats de ce mémoire indiquent que le principal effet de l'attrition sur l'échantillon longitudinal de l'ÉLDEQ est la diminution de celui-ci et que l'équipe de l'ÉLDEQ est en mesure de préserver une bonne taille échantillonnale en appliquant diverses méthodes de fidélisation des répondants pour minimiser l'érosion de l'échantillon, on s'assure ainsi d'avoir des données longitudinales de bonne qualité.

La création du portrait des décrocheurs, le premier objectif de ce mémoire, a mis en lumière les caractéristiques très différentes des décrocheurs par rapport aux répondants qui demeurent dans l'ÉLDEQ. Ainsi, à l'instar des autres enquêtes longitudinales étudiant la famille et les enfants (Behr et *al.*, 2005, Hoddinott et *al.*, 2002,

Razafindratsima et Kishimba, 2004), Watson, 2003), les familles décrochant prématurément de l'ÉLDEQ présentent certains traits particuliers liés à la pauvreté, à l'instabilité et à l'exclusion sociale, notamment des parents moins scolarisés, des parents immigrants, des revenus plus faibles, une famille qui n'est pas intacte, un père biologique absent et une famille locataire de son logement. Les caractéristiques des décrocheurs de l'ÉLDEQ sont similaires à celles répertoriées dans la littérature sur les enquêtes longitudinales, montrant ainsi que les défis causés par l'attrition dans l'ÉLDEQ peuvent se rapprocher de ceux pour les autres grandes enquêtes longitudinales.

L'atteinte du second objectif, l'estimation de la probabilité d'attrition des familles, a principalement montré la difficulté à prédire l'attrition d'une famille selon ses caractéristiques. En effet, parmi les variables utilisées, seuls le niveau de scolarité de la mère de l'enfant, le statut d'immigrante de la mère et le fait que la mère ait eu au moins un avortement permettent de prédire de façon significative la probabilité de la famille à quitter l'ÉLDEQ. Ces résultats montrent tout d'abord l'importance de l'attitude du répondant face aux enquêtes (régulièrement représentée dans la littérature sur les enquêtes par le niveau de scolarité du répondant) alors que les individus moins scolarisés décrochent plus facilement de l'ÉLDEQ. Par la suite, le statut d'immigrante de la mère permet de constater l'impact de l'intégration des immigrants (les entrevues de l'ÉLDEQ se déroulant en français ou en anglais) et de la mobilité particulière des immigrants (ils sont habituellement plus nombreux à effectuer une migration interprovinciale ou internationale) sur la rétention des familles. La variable portant sur l'avortement montre que certaines variables d'intérêt peuvent être biaisées par l'attrition, bien que ce résultat soit quelque peu surprenant. Toutefois, dans le cas de cette dernière variable, il serait pertinent de vérifier les interactions possibles, notamment entre l'âge de la mère, son statut d'immigrante et le fait qu'elle se soit fait avorter au moins une fois. Finalement, la difficulté à retracer les répondants très mobiles, mesurée ici par le statut de propriétaire du logement, ne joue pas dans la prédiction de l'attrition, signifiant alors que l'équipe de l'ÉLDEQ a pu rejoindre efficacement les familles qui déménageaient. Il faut toutefois préciser qu'il aurait pu être fort intéressant de mesurer la mobilité des répondants par une variable indiquant si la famille a déménagé (et à quel endroit) ou non entre deux

volets plutôt que par le statut de propriétaire. Cette variable serait évidemment un meilleur *proxy* de la mobilité géographique des familles.

Les tests de BGLW, réalisés pour atteindre le troisième objectif, ont montré que seuls le statut d'immigrante de la mère et le fait que la mère ait eu au moins un avortement sont biaisés de façon significative par l'attrition. Cependant, le statut d'immigrante de la mère soulève les mêmes interrogations dans l'évaluation des biais d'attrition que pour la prédiction de la probabilité d'attrition. De plus, plusieurs des variables d'intérêt utilisées ici ont des tendances plus ou moins fortes à être biaisées. Il se pourrait alors que des biais apparaissent de façon significative pour des variables d'intérêt lors des prochains passages de l'enquête ou simplement pour des variables d'intérêt qui ne furent pas étudiées ici. Ceci met donc en lumière l'importance pour les chercheurs utilisant des données longitudinales de prendre le temps nécessaire afin de vérifier la présence des biais d'attrition pour s'assurer de la qualité des données longitudinales utilisées. Toutefois, il faut rappeler que dans la plupart des cas, les variables d'intérêt ne sont pas biaisées significativement par l'attrition, montrant ainsi que l'attrition n'affecte que très peu la qualité des données de l'ÉLDEQ. En outre, ces conclusions se basent seulement sur les variables étudiées ici et non sur l'ensemble des variables contenues dans l'ÉLDEQ.

Le quatrième et dernier objectif de ce mémoire, la validation de l'efficacité des poids longitudinaux dans la correction des biais d'attrition, est une analyse très pertinente compte tenu de la popularité grandissante de cette technique pour corriger l'attrition, spécialement dans les enquêtes longitudinales canadiennes et québécoises où la représentativité transversale et longitudinale est encouragée. Les résultats de ce mémoire ont montré que les poids longitudinaux sont pertinents dans la mesure où ils permettent de diminuer considérablement les différences dans les caractéristiques fixes pour les répondants présents au volet initial et au huitième volet. Il serait intéressant de trouver une procédure afin d'étendre cette comparaison aux variables qui varient dans le temps mais une telle procédure nécessite de relever plusieurs défis méthodologiques, particulièrement la modélisation des changements d'état des décrocheurs qui devrait

probablement se faire avec des hypothèses très fortes. Les poids longitudinaux ne permettent cependant pas de corriger les différences dans les caractéristiques des futurs décrocheurs par rapport aux familles qui resteront dans l'enquête, montrant ainsi une certaine limitation à leur efficacité même si le rôle premier des poids est de corriger de façon rétrospective. Ces constatations vont dans le même sens que la littérature sur l'attrition, qui montre aussi que les biais causés par l'attrition se retrouvent principalement dans l'analyse unidimensionnelle (Hoddinott, 2002).

Toutefois, les résultats de ce mémoire génèrent un certain nombre d'interrogations. Tout d'abord, il a été vu que la façon de tenir compte du concept d'attrition diffère selon les enquêtes suite au développement de diverses enquêtes longitudinales ayant des méthodologies dissemblables et ciblant des populations différentes. Il serait donc nécessaire, à l'avenir, d'uniformiser la façon de traiter l'attrition afin d'améliorer la comparabilité entre les différentes enquêtes et pour développer des nouvelles techniques pour contrer les effets de l'attrition. Étant donné les nombreuses particularités propres à chaque enquête longitudinale (allant de la population ciblée par l'enquête aux techniques pour rejoindre les sujets), cette uniformisation méthodologique risque d'être une tâche très ardue à réaliser. Il importera donc que les chercheurs s'intéressant à des problèmes méthodologiques comme l'attrition prennent le temps de bien définir les concepts qu'ils utilisent.

Une réflexion critique sur l'attrition nous amène à répartir les répondants des enquêtes longitudinales en trois grandes catégories selon leur persévérance dans le suivi longitudinal : les fidèles, les fragiles et les réticents. Alors que les fidèles tendent à demeurer facilement dans le suivi longitudinal, des efforts croissants doivent être réalisés pour garder les fragiles dans l'enquête alors que celle-ci progresse. Les mesures visant à réduire l'attrition sont donc principalement utiles pour garder ce type de répondants dans le suivi longitudinal. Malheureusement, un certain nombre de sujets seront toujours très enclins à quitter l'échantillon (ils sont ici nommés réticents), bien que les faibles taux d'attrition lors de la première phase de l'ÉLDEQ suggèrent que l'engagement de cinq ans pris par les parents fut sérieux. Ceci suggère donc que

l'approche de l'ÉLDEQ pour la première phase serait très efficace pour garder les répondants appartenant aux trois types. Toutefois, il ne faut pas oublier le biais de sélection initial de l'enquête (l'ÉLDEQ affichait un taux de réponse initial de 75,8%), qui peut également nuire à la qualité des données d'enquête en plus de réduire considérablement la proportion de répondants futurs décrocheurs dans l'échantillon. Ainsi, il est pratiquement impossible de limiter totalement la non-réponse par unité et l'attrition (même des sources de données exhaustives comme les recensements sont touchées par un certain sous-dénombrement).

Une piste de recherche très intéressante serait d'étendre l'analyse selon les types de non réponse (refus, incapacité à rejoindre la famille, pertes démographiques, etc.). Ceci permettrait de distinguer les deux causes principales de l'attrition : l'attitude du répondant face aux enquêtes (refus) et la collecte de données (incapacité à rejoindre la famille). Une telle analyse permettrait notamment d'évaluer les différentes mesures prises par l'équipe de l'ÉLDEQ pour diminuer l'attrition en plus de pouvoir être très intéressante dans une perspective comparative avec l'ELNEJ ou avec une autre enquête longitudinale portant sur les enfants. En outre, cette piste est très prometteuse dans la mesure où ce genre d'informations peut être disponible auprès des producteurs de données.

Une façon alternative de mesurer les biais causés par l'attrition serait de comparer les données de l'ÉLDEQ avec une base de données jugée non biaisée. Cette autre source pourrait être une enquête transversale (comme l'Enquête sociale générale de Statistique Canada), les données censitaires de 2001 ou de 2006 (ajustées pour le sous-dénombrement) ou des données administratives (fichiers scolaires, déclarations de revenus, etc.). Évidemment, une telle comparaison possède malheureusement certaines limitations, notamment par rapport au nombre de variables de nature sociodémographiques sur lesquelles axer l'analyse. Néanmoins, avec l'accès prochain aux données censitaires dans les centres de recherche canadiens, une telle comparaison pourrait être éventuellement possible, non seulement pour évaluer l'impact de l'attrition, mais également pour évaluer la représentativité de toutes les enquêtes, longitudinales et

transversales. De plus, une enquête fondée sur un plan d'échantillonnage rotatif, comme l'EDTR, pourrait être utilisée afin d'obtenir une mesure de l'attrition en comparant les caractéristiques des répondants pour les passages où deux échantillons longitudinaux se superposent.

En outre, une perspective longitudinale pourrait aussi être utilisée afin d'étudier l'attrition. Ceci pourrait tout d'abord être réalisé avec un modèle d'analyse de survie où la durée de survie de l'individu jusqu'au moment de l'attrition serait étudiée. Ceci permettrait alors de modéliser le risque de décrocher de façon prématurée au lieu de modéliser la probabilité de décrocher. Il faut par contre noter que dans une enquête comme l'ÉLDEQ, la pertinence d'un tel modèle est questionnable. Effectivement, puisque l'ÉLDEQ a conservé près de 70% de son échantillon, un modèle de survie n'apporterait probablement pas beaucoup de renseignements supplémentaires par rapport aux tests transversaux, en plus d'être plus compliqué à mettre en œuvre. Cependant, un modèle de ce genre pourrait être très pertinent pour une enquête de très longue durée où l'attrition touche une grande partie de l'échantillon (comme le PSID) ou bien comme une façon détournée d'évaluer de façon longitudinale le biais d'attrition pour une certaine variable d'intérêt. Une seconde approche longitudinale prometteuse serait l'analyse de trajectoires. À l'aide de cette technique, il pourrait être possible de modéliser le développement d'une variable d'intérêt au fil du temps tout en contrôlant pour l'attrition comme il a été réalisé au niveau transversal avec les tests BGLW. Ceci permettrait notamment de corriger le caractère transversal des tests BGLW, l'une des principales faiblesses de cette approche, tout en permettant aux chercheurs de contrôler pour l'attrition dans leurs analyses. Toutefois, cette approche soulève plusieurs interrogations par rapport à l'horizon utilisé puisqu'il est nettement préférable d'utiliser des répondants ayant répondu à tous les volets faisant partie de l'horizon choisi. Pour cette raison, l'approche par l'analyse de survie semble plus prometteuse, même si elle estime le biais d'attrition sur une variable d'intérêt de façon indirecte.

L'analyse de l'efficacité des poids longitudinaux a bien illustré les difficultés existant pour bien évaluer l'efficacité de cette technique pour corriger les biais

d'attrition. Effectivement, puisque les décrocheurs n'ont évidemment pas de pondération longitudinale à partir du moment de leur départ (même s'ils reviennent par la suite dans l'échantillon) et que les familles toujours dans l'enquête reçoivent une correction dans leur pondération pour compenser pour le départ des décrocheurs, l'évaluation de l'efficacité des poids longitudinaux doit se faire de façon indirecte. De plus, étant donné le caractère longitudinal des données, les caractéristiques des familles peuvent évidemment changer au fil du temps, compliquant davantage la validation des poids longitudinaux. Effectivement, ceci crée en sorte une dualité dans les proportions pour chaque variable variant dans le temps où s'opposent le risque de décrocher de l'enquête et le risque de changer de catégorie. Il faut alors se concentrer sur des variables fixes dans le temps (comme le rang de naissance de l'enfant) ou dont on connaît la variation (comme l'âge) pour bien isoler l'effet de l'attrition ou alors de tenter de mesurer les variations au sein de chaque variable étudiée pour les décrocheurs et pour les familles demeurant dans l'enquête. En plus de limiter les analyses à seulement quelques variables, ceci soulève alors le problème de trouver un test statistique approprié pour traiter ces données appariées. Or, les tests de proportions pour données appariées fréquemment utilisés, comme le test de McNemar, se basent sur les transitions entre les deux temps de mesure, ce qui n'est évidemment pas approprié pour les présentes analyses, étant donné le caractère fixe des variables qu'il convient d'étudier. Cependant, avec le nombre grandissant d'enquêtes longitudinales qui utilisent cette technique pour contrer les biais d'attrition (Vandecasteele et *al.*, 2006), il serait très important de développer une solide méthodologie afin d'étudier l'efficacité des poids longitudinaux dans la correction des biais d'attrition.

Évidemment, les analyses effectuées dans ce mémoire sont essentiellement des analyses basées sur la sélection sur les observables. Cependant, il est bien possible que des différences qui ne sont pas observables dans l'échantillon pourraient également causer l'attrition, comme dans la situation où un changement (divorce, déménagement, voyage, perte d'emploi, maladie, accident, etc.) ayant lieu entre deux passages provoque le décrochage du répondant. Malheureusement, rares sont les techniques permettant de tester efficacement les biais de sélection sur des variables non observables. Le

développement de telles techniques seraient un apport formidable non seulement à l'analyse de l'attrition et dans la validation des poids longitudinaux mais aussi dans plusieurs autres domaines où des données de toute sorte sont utilisées. Le Canada compte d'ailleurs sur plusieurs dossiers administratifs, comme le registre de l'état civil ou les fichiers de déclaration fiscale, qui pourraient permettre de saisir une partie de cet effet qui n'est pas observable directement dans l'enquête.

Un autre point d'interrogation important demeure les pertes démographiques. Ainsi, les sujets qui décèdent et ceux qui migrent, advenant la situation où une enquête ne considère pas les pertes démographiques dues à la migration comme étant de l'attrition, peuvent alors devenir une autre source potentielle de biais. Effectivement, ces répondants pourraient naturellement avoir des caractéristiques particulières résultant d'un certain processus de sélection. Par exemple, des variables comme l'âge et l'état de santé peuvent évidemment être corrélées avec la probabilité de décès (qui est très faible chez les enfants canadiens) alors que des variables telles que les langues comprises, l'âge, le statut d'emploi et le niveau de scolarité pourraient être liées à la probabilité d'émigrer, autant au niveau international qu'aux niveaux interprovincial ou interrégional. Par exemple, il est reconnu que les Québécois francophones émigrent moins vers d'autres provinces canadiennes que les Québécois anglophones ou allophones ou encore que les immigrants internationaux possèdent des comportements migratoires très différents de ceux des natifs. Les conséquences des pertes démographiques pourraient même s'avérer aussi dommageables sur la qualité des données longitudinales que celles causées par l'attrition. Toutefois, une enquête longitudinale comme l'ÉLDEQ n'est pas propice à l'étude de ces biais étant donné le petit nombre d'enfants qui sont morts ou qui ont migré de façon permanente à l'extérieur du Québec.

En somme, bien que l'ÉLDEQ soit affectée par l'attrition de près du quart de ses répondants initiaux, la principale conséquence de l'attrition sur la qualité des données de l'ÉLDEQ est la réduction de la taille échantillonnale. Il est également très intéressant de remarquer que les efforts de l'équipe de l'ÉLDEQ afin de fidéliser les répondants et de

favoriser le retour des familles ayant quitté l'enquête semblent porter fruit alors que la taille échantillonnale augmente en E8. Puisque le principal effet de l'attrition sur l'échantillon longitudinal de l'ÉLDEQ est la diminution de celui-ci et que l'équipe de l'ÉLDEQ est en mesure de préserver une bonne taille échantillonnale en appliquant diverses méthodes de fidélisation des répondants pour minimiser l'érosion de l'échantillon, on s'assure ainsi d'avoir des données longitudinales de très bonne qualité. Toutefois, il demeure très pertinent que les utilisateurs de données longitudinales prennent le temps de vérifier si les variables qu'ils utilisent sont biaisées par l'attrition. À cette fin, la méthode BGLW longitudinale présentée dans ce mémoire possède un bon potentiel afin de tenir compte du caractère longitudinal des variables étudiées.

En terminant, la grande pertinence des analyses de l'attrition dans les enquêtes longitudinales suggère que bien que les décrocheurs soient perdus, ils sont loin d'être oubliés !

BIBLIOGRAPHIE

AHERN, Kathy et Robyne LE BROCCQUE, 2005, « Methodological Issues in the Effects of Attrition: Simple Solutions for Social Scientists », *Field Methods*, Vol. 17, no 1, février, p. 53-69.

ALDERMAN, Harold, Jere R. BEHRMAN, Hans-Peter KOHLER, John A. MALUCCIO et Susan Cotts WATKINS, 2000, « Attrition in Longitudinal Household Survey Data: Some Tests for Three Developing-country Samples », *FCND Discussion Paper No. 96*, Food Consumption and Nutrition Division, International Food Policy Research Institute.

BASKIN, Robert M., 2005, « Traitement des données d'enquête manquantes dans le cadre d'analyse longitudinale », *Recueil du Symposium 2005 de Statistique Canada : Défis méthodologiques pour les besoins futurs d'information*, Ottawa, Statistique Canada, catalogue No 11-522-XIF.

BECKETTI, S., W. GOULD, L. LILLARD et F. WELCH, 1988, « The Panel Study of Income Dynamics after Fourteen Years: An Evaluation », *Journal of Labor Economics*, no. 6, p. 472-492.

BEHR, Andreas, Egon BELLGARDT et Ulrich RENDTEL, 2005, « Extent and Determinants of Panel Attrition in the European Community », *European Sociological Review*, Vol. 21, no. 5, pp. 489-512.

BÉRARD-CHAGNON, Julien, 2006, « L'attrition: un problème majeur dans les enquêtes sociales longitudinales », *Dire*, vol. 16 n°1, pp.28-31.

BIGNAMI-VAN ASSCHE, Simona, 2003, « Are we measuring what we want to Measure? Individual consistency in survey response in rural Malawi », *Demographic Research*, Collection spéciale 1, article 3, pp 78-108.

BOR, W., J. M. NAJMAN, M. ANDERSEN, J. MORRISON et G. WILLIAMS, 1993, « Socioeconomic Disadvantage and Child Morbidity: An Australian Longitudinal Study », *Social Science & Medicine*, Vol. 36, no. 8, pp. 1053-1061.

CASELLA, George et Roger L. BERGER, 2002, *Statistical Inference. Second Edition*, Duxbury Advanced Series, Duxbury Press, Pacific Grove, 660 p.

CHEN, Yuanyuan, 2005, « The Disparity of Black-White Male Wage Gaps between the CPS and PSID— An Analysis of Sample Attrition in Panel Data »

DUBOIS, Lise, Anna FARMER, Manon GIRARD et Marion PORCHERIE, 2006, « Family food insufficiency is related to overweight among preschoolers », *Social Science & Medicine*, Vol. 63, no 6, pp. 1503-1516.

ÉTUDE LONGITUDINALE DU DÉVELOPPEMENT DES ENFANTS DU QUÉBEC (ÉLDEQ), 2003, *Pondération longitudinale : volets 1998 à 2002*, Direction Santé Québec, Institut de la statistique du Québec, 7 p.

FITZGERALD, John, Peter GOTTSCHALK et Robert MOFFITT, 1998, « An Analysis of Sample Attrition in Panel Data : The Michigan Panel Study of Income Dynamics », *The Journal of Human Resources*, Vol. 33, no 2, pp. 300-344.

INSTITUT DE LA STATISTIQUE DU QUÉBEC, 2000, « Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2002), Les nourrissons de 5 mois : Enquête : description et méthodologie. (Volume 1, numéro 1) », *Collection la santé et le bien-être*, Institut de la statistique du Québec, 60 p.

HAHN, Eugene D. et Refik SOYER, 2005, « Probit and Logit Models: Differences in the Multivariate Realm », soumis au *The Journal of the Royal Statistical Society*, Series B.

HILL, Daniel H., 1997, « Adjusting for Attrition in Event-History Analysis », *Sociological Methodology*, Vol. 27, no. 1, pp. 393-416.

HISKEY, Syd et Nicholas A. TROOP, 2002, « Online Longitudinal Survey Research : Viability and Participation », *Social Science Computer Review*, Vol. 20, no. 3, pp. 250-259.

HODDINOTT, John, Lynn LETHBRIDGE et Shelley PHIPPS, 2002, « Is History Destiny ? Resources, Transitions and Child Education Attainments in Canada », Human Resources Development Canada, Catalogue no.RH63-1/551-12-02E, 71 p.

HOSMER, David W. et Stanley LEMESHOW, 1989, *Applied Logistic Regression*, Wiley Series in Probabilities and Mathematical Statistics, John Wiley & Sons, 307 p.

JETTÉ, Mireille, et Lyne DES GROSEILLIERS, 2000, « L'enquête : description et méthodologie » dans *Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2002)*, Québec, Institut de la statistique du Québec, vol. 1, no 1, 26 p.

JONES, Charles, Linn CLARK, Joan GRUSEC, Randle HART, Gabriele PLICKERT et Lorne TEPPERMAN, 2002, « La pauvreté, le capital social, les compétences parentales et les résultats des enfants au Canada », Développement des ressources humaines Canada, Direction générale de la recherche appliquée, catalogue no. SP-557-01-03F, 84 p.

KARNEY, Benjamin R. et Thomas N. Bradbury, 1995, « Assessing Longitudinal Change in Marriage : An Introduction to the Analysis of Growth Curves », *Journal of Marriage and the Family*, Vol. 57, pp. 1091-1108.

KENNEDY, Peter, 2003, *A Guide to Econometrics – Fifth Edition*, The MIT Press, Cambridge, Massachusetts, 623 p.

LÉGER, Christian, 2002, *STT 1700. Introduction à la statistique*, Université de Montréal, 245 p.

LEHNEN, Robert G. et Gary G. KOCH, 1974, « Analyzing Panel Data with Uncontrolled Attrition », *The Public Opinion Quarterly*, Vol. 38, no. 1, pp. 40-56.

LITTLE, Roderick J.A. et Donald B. RUBIN, 1987, *Statistical Analysis with Missing Data*, John Wiley and Sons, New York, 278 p.

MARCIL-GRATTON, Nicole, Heather JUBY et Céline LE BOURDAIS, 2003, « Du passé conjugal des parents au devenir familial des enfants : un exemple de la nécessité d'une approche longitudinale », *Sociologie et sociétés*, Vol. 35, no. 1, pp. 143-164.

MARINI, Margaret Mooney, Anthony R. OLSEN et Donald B. RUBEN, 1980, « Maximum-Likelihood Estimation in Panel Studies with Missing Data », *Sociological Methodology*, Vol. 11, pp. 314-357.

MILLER, Richard B. et David W. WRIGHT, 1995, « Detecting and Correcting Attrition Bias in Longitudinal Family Research », *Journal of Marriage and the Family*, Vol. 57, no. 4, p. 921-929.

MOORE, David S. et MCCABE, George P., 1999, *Introduction to the Practice of Statistics, Third Edition*, W. H. Freeman and Company, New York, 825 p.

NEILL, Ghyslaine, 2004, *Conditions économiques, travail des parents et garde des enfants au Québec : les familles dans un monde en mouvement*, Communication présentée à la conférence « Des services de garde pour tous. Un enjeu pour le XXI^e siècle », organisé par le Conseil canadien de développement social, Winnipeg, novembre 2004.

NEILL, Ghyslaine, Hélène DESROSIERS et Évelyne LAPIERRE-ADAMCYK, 2005a, « Les transformations de l'environnement familial des jeunes enfants au Québec : la place des grands-parents », Réflexions proposées en vue de la table ronde du colloque *La famille et les relations intergénérationnelles : quelle place dans l'intervention*, mai 2005.

NEILL, Ghyslaine, Hélène DESROSIERS, Amélie DUCHARME et Lucie GINGRAS, 2005b, « L'acquisition du vocabulaire chez les jeunes enfants au Québec : le rôle de l'environnement familial et économique », Communication présentée au colloque *Études longitudinales et défis démographiques du XXI^e siècle*, Université de Montréal, 18-19 novembre 2005.

NEILL, Ghyslaine, 2006, « Le suivi longitudinal des enfants et des familles: l'expérience québécoise et canadienne », Communication présentée dans le cadre de la

Journée scientifique « Les enfants et leur famille », organisée par l'Institut national des études démographiques (INED), Paris, 24 mars 2006.

NETER, John, Michael H. KUTNER, Christopher J. NACHTSHEIM et William WASSERMAN, 1996, *Applied Linear Statistical Models. Fourth Edition*, McGraw-Hill, 1408 p.

PICOT, Garnet et Maryanne WEBBER, 2005, « Faire le point : l'avenir des enquêtes longitudinales », *Recueil du Symposium 2005 de Statistique Canada : Défis méthodologiques pour les besoins futurs d'information*, Ottawa, Statistique Canada, Catalogue No. 11-522-XIF.

PLANTE, Nathalie, Catherine FONTAINE et Robert COURTEMANCHE, 2005, *Pondération des données du volet 2004*. Direction de la méthodologie, de la démographie et des enquêtes spéciales, Institut de la statistique du Québec.

PUHANI, Patrick A., 1997, « Foul or Fair? The Heckman Correction for Sample Selection and Its Critique : A Short Survey », Discussion Paper No. 97-07 E, Centre for European Economic Research, SELAPO, Université de Munich.

RAFTERY, Adrian E., 2001, « Statistics in Sociology, 1950-2000 : A Selective Review », *Sociological Methodology*, Vol. 31, pp. 1-45.

RAZAFINDRATSIMA, Nicolas et Ngoy KISHIMBA, 2004, « La déperdition dans la cohorte Cocon entre 2000 et 2002 », *Population*, Institut national d'études démographiques (INED), Vol. 59, no 3-4, pp. 419-448.

RENDTEL, Ulrich, 2002, « Attrition in Household Panels : A Survey », Communication présentée dans le cadre du colloque *The Future of Social Surveys in Europe*, Helsinki, 32 p.

RICE, John A., 1995, *Mathematical Statistics and Data Analysis. Second Edition*, Duxbury Press, Belmont, 602 p.

RUBIN, Donald, 1976, « Inference and Missing Data », *Biometrika*, Vol. 63, no. 3, pp. 581-592.

SÄRNDAL, Carl-Erik, Bengt SWENSSON et Jan WRETMAN, 1992, *Model Assisted Survey Sampling*, Springer Series in Statistics, New York, 694 p.

SCHMITZ, Norbert et Matthias FRANZ, 2002, « A Bootstrap Method To Test If Study Dropouts Are Missing Randomly », *Quality and Quantity*, Vol. 36, no. 1, pp. 1-16.

VANDECASTEELE, Leen et Annelies Debels, 2006, « Attrition in Panel Data : The Effectiveness of Weighting », *European Sociological Review*, Publication avec accès avancé (<http://www.esr.oxfordjournals.org>), pp. 1-17.

WATSON, Dorothy, 2003, « Sample Attrition between Waves 1 and 5 in the European Community Household Panel », *European Sociological Review*, Vol. 19, no. 4, pp. 361-378.

WONNACOTT, Thomas H. et Ronald J. WONNACOTT, 1991, *Introductory Statistics for Business and Economics, fourth edition*, John Wiley & Sons, New York, 920 p.

ZAGORSKY, Jay L., 2005, « Marriage and divorce's impact on wealth », *Journal of Sociology*, Vol. 41, no 4, p. 406-424.

ANNEXES

Tableau A.3.1

Tableau complet des taux d'attrition cumulatifs (%) par volet

Variables	n	Volet						
		2	3	4	5	6	7	8
<i>Présence du père biologique</i>								
Présent	1937	2,74	4,59	5,32	5,99	11,72	21,12	25,55
Absent	177	5,65	9,60	12,99	13,56	19,77	29,38	35,59
<i>Type de famille (3 cat.)</i>								
Intacte	1706	2,64	4,45	5,16	5,74	11,66	20,98	25,67
Recomposée	235	3,40	5,53	6,38	7,66	11,49	21,70	24,68
Monoparentale	171	5,85	9,94	13,45	14,04	21,05	30,41	36,26
<i>Statut d'immigrante (mère)</i>								
Pas immigrante	1865	1,66	3,32	4,08	4,72	10,29	19,62	24,34
Immigrante européenne	62	12,90	16,13	16,13	17,74	24,19	35,48	38,71
Immigrante non européenne	191	12,57	17,80	21,47	21,99	28,80	38,22	42,41
<i>Nombre d'années depuis l'immigration (mère)</i>								
Moins de 5 ans	81	14,81	19,75	24,69	24,69	34,57	40,74	46,91
5-9 ans	76	9,21	17,11	19,74	21,05	25,00	39,47	40,79
10 ans et plus	96	13,54	15,63	16,67	17,71	23,96	33,33	37,50
<i>Langue(s) parlée(s) à la maison (mère)</i>								
Français	1711	1,81	3,27	3,97	4,62	10,34	19,75	24,43
Anglais	208	4,33	8,17	10,58	11,54	17,79	25,48	28,85
Ni français ni anglais	128	15,63	22,66	25,00	25,78	32,81	42,97	48,44
Français, anglais et autre	37	5,41	5,41	5,41	5,41	5,41	16,22	21,62
Français ou anglais et autre	34	2,94	5,88	8,82	8,82	11,76	26,47	32,35
<i>Diplôme le plus élevé (mère)</i>								
Pas de DES	385	5,19	8,57	9,87	11,17	16,36	26,49	32,47
DES	555	2,42	4,44	5,86	6,46	13,54	20,61	27,68
Post-sec. Sauf univ.	611	2,91	4,00	4,00	4,00	9,45	20,73	25,09
Universitaire	566	2,25	4,09	4,91	5,73	9,61	19,02	21,88
<i>Niveau de suffisance du revenu</i>								
Suffisant	1571	1,40	2,80	3,25	3,76	9,29	18,52	23,11
Insuffisant	235	4,68	6,81	8,51	9,36	14,04	25,11	28,09
Très insuffisant	276	8,33	13,04	16,30	17,75	25,36	34,42	40,22
<i>Statut socioéconomique du ménage</i>								
Premier quartile	529	4,16	8,13	10,78	12,10	18,71	27,22	31,95
Deuxième quartile	526	3,42	4,56	5,32	5,70	13,31	22,62	29,47
Troisième quartile	523	1,91	3,44	3,82	4,21	9,56	20,65	25,05
Quatrième quartile	517	1,55	3,09	3,09	3,68	7,35	16,25	18,96
<i>Propriétaire du logement</i>								
Oui	1327	1,88	3,24	3,54	3,92	9,72	19,07	23,36
Non	793	4,79	7,94	10,09	11,22	16,90	26,36	31,53

Source : ÉLDEQ, ISQ.

Tableau A.3.1

Tableau complet des taux d'attrition cumulatifs (%) par volet (suite)

Variables	n	Volet						
		2	3	4	5	6	7	8
<i>Groupe d'âge (mère)</i>								
Moins de 20 ans	59	5,08	6,78	8,47	8,47	10,17	15,25	20,34
20-24 ans	414	4,35	7,25	8,45	9,42	14,49	24,64	29,95
25-29 ans	609	2,40	4,50	5,41	6,16	11,71	20,42	25,68
30-34 ans	695	3,17	5,04	6,04	6,47	12,66	22,59	26,91
35-39 ans	239	1,67	2,93	3,77	4,60	11,72	21,34	24,27
40+ ans	46	0,00	0,00	0,00	0,00	4,35	13,04	15,22
<i>Groupe d'âge (père)</i>								
Moins de 25 ans	165	3,64	7,27	7,88	9,09	13,94	23,03	27,27
25-29 ans	536	1,87	2,99	4,10	5,04	11,75	20,34	25,75
30-34 ans	669	2,39	4,33	4,93	5,38	10,91	20,18	24,36
35-39 ans	422	2,61	4,27	4,98	5,45	10,66	21,33	26,07
40+ ans	158	6,33	8,86	9,49	10,13	15,19	24,68	27,22
<i>Nombre de frères/sœurs</i>								
0	887	2,82	4,28	5,30	5,98	12,18	22,77	27,73
1	850	2,82	5,29	6,12	6,94	12,35	21,06	25,65
2	262	3,82	4,58	5,73	6,11	12,21	20,23	24,05
3	84	2,38	5,95	8,33	8,33	13,10	20,24	26,19
4+	37	5,41	16,22	16,22	16,22	18,92	29,73	29,73
<i>Nombre de personnes dans le ménage</i>								
2	42	7,14	14,29	16,67	19,05	28,57	45,24	47,62
3	803	2,37	3,74	4,48	5,11	11,08	21,30	26,53
4	854	2,93	5,04	5,97	6,79	12,30	21,19	25,88
5	281	3,20	4,27	5,34	5,69	12,46	20,28	23,49
6	94	4,26	8,51	11,70	11,70	13,83	21,28	27,66
7+	46	6,52	15,22	15,22	15,22	19,57	30,43	30,43
<i>Rang de naissance de l'enfant</i>								
1	936	2,56	3,95	4,91	5,66	11,75	22,44	27,24
2	841	3,21	5,71	6,66	7,37	12,60	20,81	25,56
3	237	3,38	4,22	5,06	5,49	12,24	21,10	24,89
4	67	1,49	4,48	7,46	7,46	13,43	22,39	28,36
5+	39	7,69	20,51	20,51	20,51	23,08	30,77	30,77
<i>Deux parents biologiques présents</i>								
Oui	1936	2,74	4,60	5,32	5,99	11,67	21,07	25,52
Non	178	5,62	9,55	12,92	13,48	20,22	29,78	35,96
<i>Statut d'immigrant (père)</i>								
Pas immigrant	1706	1,70	3,11	3,69	4,40	10,20	19,46	23,97
Immigrant européen	55	3,64	7,27	9,09	9,09	14,55	25,45	30,91
Immigrant non européen	186	11,29	16,67	18,82	19,35	24,19	34,41	38,71
<i>Nombre d'années depuis l'immigration (père)</i>								
Moins de 5 ans	63	14,29	14,29	14,29	14,29	14,29	14,29	14,29
5-9 ans	56	7,14	14,29	19,64	19,64	26,79	37,50	42,86
10+ ans	118	8,47	13,56	14,41	14,41	16,95	27,12	30,51

Source : ÉLDEQ, ISQ.

Tableau A.3.1
Tableau complet des taux d'attrition cumulatifs (%) par volet (suite)

Variables	n	Volet						
		2	3	4	5	6	7	8
<i>Langue(s) parlée(s) à la maison (père)</i>								
Français	1603	1,68	2,93	3,49	4,12	9,92	19,15	23,77
Anglais	182	3,30	6,59	8,24	9,34	15,38	24,18	26,92
Ni français ni anglais	104	16,35	25,00	25,96	26,92	33,65	41,35	48,08
Français, anglais et autre	26	3,85	3,85	3,85	3,85	3,85	15,38	19,23
Français ou anglais et autre	32	3,13	6,25	12,50	12,50	12,50	37,50	40,63
<i>Première langue apprise (mère)</i>								
Français	1718	1,86	3,32	4,02	4,60	10,30	19,62	24,21
Anglais	166	3,01	6,02	9,04	10,24	15,66	24,70	28,92
Ni français ni anglais	234	11,11	16,67	18,38	19,23	25,21	35,47	40,60
<i>Première langue apprise (père)</i>								
Français	1567	1,79	3,19	3,77	4,40	10,08	19,27	23,99
Anglais	149	1,34	3,36	4,70	6,04	12,08	23,49	26,17
Ni français ni anglais	231	9,52	14,29	16,02	16,45	22,08	31,60	35,93
<i>Langue(s) de conversation (mère)</i>								
Français ou anglais	1036	2,22	3,67	4,63	5,12	11,49	21,24	26,74
Français et anglais	779	1,54	3,34	3,98	4,49	9,50	18,61	21,82
Français, anglais et autre	174	5,17	9,20	10,34	12,64	18,39	28,16	33,33
Français ou anglais et autre	129	14,73	20,16	23,26	24,03	28,68	36,43	41,86
<i>Langue(s) de conversation (père)</i>								
Français ou anglais	857	1,87	3,50	4,08	4,67	11,32	21,82	26,95
Français et anglais	816	1,47	2,94	3,55	4,29	9,07	16,79	20,71
Français, anglais et autre	189	5,82	8,99	9,52	10,05	14,81	27,51	31,75
Français ou anglais et autre	85	15,29	20,00	24,71	25,88	32,94	40,00	44,71
<i>Diplôme le plus élevé (père)</i>								
Pas de DES	398	3,27	6,03	6,78	8,29	14,57	24,12	27,89
DES	495	2,42	4,44	5,86	6,46	13,54	20,61	27,68
Post-sec. Sauf univ.	550	2,91	4,00	4,00	4,00	9,45	20,73	25,09
Universitaire	489	2,25	4,09	4,91	5,73	9,61	19,02	21,88
<i>Principale source de revenus du ménage</i>								
Salaire et traitement	1625	2,46	3,94	4,62	5,29	10,52	19,45	24,25
Travail autonome	166	2,41	5,42	6,02	6,02	13,86	26,51	29,52
Aide sociale	207	6,76	11,59	15,46	16,91	24,64	33,33	37,68
Assurance-emploi	38	5,26	7,89	7,89	7,89	15,79	31,58	34,21
Autre	58	3,45	5,17	6,90	6,90	10,34	24,14	31,03
<i>Travaille à l'enquête (mère)</i>								
Non	1716	3,21	5,13	6,29	7,05	13,00	22,67	27,45
Oui	371	2,16	4,31	4,58	4,85	9,70	18,33	22,37
<i>Travaille 12 derniers mois (mère)</i>								
Non	630	5,24	8,10	10,32	11,11	16,67	23,65	28,41
Oui	1475	1,97	3,59	4,07	4,68	10,44	20,81	25,42

Source : ÉLDEQ, ISQ.

Tableau A.3.1
Tableau complet des taux d'attrition cumulatifs (%) par volet (suite)

Variables	n	Volet						
		2	3	4	5	6	7	8
<i>Statut d'emploi principal (mère)</i>								
Temps partiel	358	2,23	3,63	3,91	4,75	11,73	22,63	25,98
Temps plein	1109	1,89	3,61	4,15	4,69	10,10	20,38	25,34
<i>Travaille à l'enquête (père)</i>								
Non	223	5,38	8,97	10,76	11,21	16,14	26,46	31,39
Oui	1723	2,38	4,00	4,59	5,28	11,09	20,37	24,84
<i>Travaille 12 derniers mois (père)</i>								
Non	96	8,33	12,50	15,63	16,67	25,00	35,42	40,63
Oui	1847	2,38	4,11	4,71	5,36	10,94	20,30	24,80
<i>Statut d'emploi principal (père)</i>								
Temps partiel	70	5,71	7,14	7,14	7,14	12,86	20,00	25,71
Temps plein	1769	2,09	3,84	4,41	5,09	10,68	20,18	24,65
<i>Situation en emploi des parents</i>								
Biparentale 2 emplois	1393	1,79	3,37	3,66	4,24	9,98	20,17	24,69
Biparentale 1 emploi	475	4,21	6,11	7,58	8,42	13,47	20,84	25,26
Biparentale 0 emploi	69	10,14	15,94	20,29	21,74	28,99	36,23	42,03
Monoparentale 1 emploi	55	5,45	9,09	14,55	16,36	21,82	32,73	40,00
Monoparentale 0 emploi	110	6,36	10,91	13,64	13,64	21,82	30,00	35,45
<i>Revenus du ménage</i>								
Quartile 1	494	5,26	8,30	10,73	11,94	18,42	28,74	33,00
Quartile 2	495	1,62	3,43	4,04	4,65	10,30	17,58	23,43
Quartile 3	511	1,37	2,94	3,72	4,11	10,57	21,33	26,03
Quartile 4	500	1,20	2,40	2,40	3,00	6,80	16,40	19,40
Attrition cumulative	2120	2,97	5,00	5,99	6,65	12,41	21,79	26,42

Source : ÉLDEQ, ISQ.

Tableau A.3.2
Tableau complet des ratios d'attrition

Variables	% Volet 1	% Volet 7	R
<i>Nombre de frères/sœurs</i>			
0	41,8	13,5	3,1
1	40,1	53,1	0,76
2	12,4	25,4	0,49
3	4,0	6,3	0,63
4+	1,7	1,8	0,94
<i>Nombre de personnes dans le ménage</i>			
2	2	4,1	0,49
3	37,9	14,5	2,61
4	40,3	46,9	0,86
5	13,3	25,1	0,53
6	4,4	6,6	0,67
7+	2,2	2,7	0,81
<i>Rang de naissance de l'enfant</i>			
1	44,2	45,1	0,98
2	39,7	40	0,99
3	11,2	10,7	1,05
4	3,2	2,8	1,14
5+	1,8	1,4	1,29
<i>Présence du père biologique</i>			
Oui	91,6	78,6	1,17
Non	8,4	21,4	0,39
<i>Type de famille (3 cat.)</i>			
Famille intacte	80,8	71,4	1,13
Famille recomposée	11,1	14,1	0,79
Famille monoparentale	8,1	14,5	0,56
<i>Statut d'immigrante (mère)</i>			
N'est pas immigrante	88,1	92,1	0,96
Immigrante européenne	2,9	2,2	1,32
Immigrante non européenne	9,0	5,7	1,58
<i>Nombre d'années depuis l'immigration (mère)</i>			
N'est pas immigrante	88,1	92,1	0,96
Moins de 10 ans	7,4	1,7	4,35
10+ ans	4,5	6,3	0,71
<i>Statut d'immigrant (père)</i>			
N'est pas immigrant	87,6	90,3	0,97
Immigrant européen	2,8	3,0	0,93
Immigrant non européen	9,6	6,7	1,43
<i>Nombre d'années depuis l'immigration (père)</i>			
N'est pas immigrant	87,8	90,6	0,97
Moins de 10 ans	6,1	1,9	3,21
10+ ans	6,1	7,5	0,81

Source : ÉLDEQ, ISQ.

Tableau A.3.2
Tableau complet des ratios d'attrition (suite)

Variables	% Volet 1	% Volet 7	R
<i>Langue(s) parlée(s) à la maison (mère)</i>			
Français seul.	80,8	84,6	0,96
Anglais	9,8	8,8	1,11
Ni français ni anglais	6,0	2,9	2,07
Français et anglais	1,7	2,5	0,68
Français ou anglais + autre	1,6	1,2	1,33
<i>Langue(s) parlée(s) à la maison (père)</i>			
Français seul.	82,3	84,4	0,98
Anglais	9,3	10	0,93
Ni français ni anglais	5,3	3,3	1,61
Français et anglais	1,3	1,3	1,00
Français ou anglais + autre	1,6	1,0	1,60
<i>Première langue apprise (mère)</i>			
Français	81,1	84,8	0,96
Anglais	7,8	7,4	1,05
Ni français ni anglais	11,0	7,8	1,41
<i>Première langue apprise (père)</i>			
Français	80,5	83,3	0,97
Anglais	7,7	7,6	1,01
Ni français ni anglais	11,9	9,1	1,31
<i>Langue(s) de conversation (mère)</i>			
Français ou anglais	48,9	49,1	1,00
Français et anglais	36,8	39,8	0,92
Français et anglais + autre	8,2	7,1	1,15
Français ou anglais + autre	6,1	4,0	1,53
<i>Langue(s) de conversation (père)</i>			
Français ou anglais	44,0	44,0	1,00
Français et anglais	41,9	45,2	0,93
Français et anglais + autre	9,7	8,2	1,18
Français ou anglais + autre	4,4	2,6	1,69
<i>Plus haut diplôme (mère)</i>			
Pas de diplôme secondaire	18,2	12,0	1,52
Diplôme secondaire	26,2	21,5	1,22
Diplôme post-secondaire (sauf univ.)	28,9	34,1	0,85
Diplôme universitaire	26,7	32,4	0,82
<i>Plus haut diplôme (père)</i>			
Pas de diplôme secondaire	20,6	15,9	1,30
Diplôme secondaire	25,6	21,4	1,20
Diplôme post-secondaire (sauf univ.)	28,5	34,5	0,83
Diplôme universitaire	25,3	28,1	0,90

Source : ÉLDEQ, ISQ.

Tableau A.3.2
Tableau complet des ratios d'attrition (suite)

Variables	% Volet 1	% Volet 7	R
<i>Principale source de revenus du ménage</i>			
Salaire et traitement	77,6	82,0	0,95
Travail autonome	7,9	10,5	0,75
Aide sociale	9,9	4,5	2,20
Assurance-emploi	1,8	0,7	2,57
Autres	2,8	2,3	1,22
<i>Travaille à l'enquête (mère)</i>			
Non	82,2	26,3	3,13
Oui	17,8	73,7	0,24
<i>Travaille 12 derniers mois (mère)</i>			
Non	29,9	18,3	1,63
Oui	70,1	81,7	0,86
<i>Statut d'emploi (mère)</i>			
Temps partiel	17,1	19,0	0,90
Temps plein	52,9	62,7	0,84
<i>Travaille à l'enquête (père)</i>			
Non	11,5	6,7	1,72
Oui	88,5	93,2	0,95
<i>Travaille 12 derniers mois (père)</i>			
Non	4,9	3,1	1,58
Oui	95,1	96,9	0,98
<i>Statut d'emploi (père)</i>			
Temps partiel	3,6	2,7	1,33
Temps plein	91,4	94,2	0,97
<i>Niveau de suffisance du revenu (3 cat.)</i>			
Revenu suffisant	75,5	84,6	0,89
Revenu insuffisant	11,3	9,7	1,16
Revenu très insuffisant	13,3	5,7	2,33
<i>Propriétaire du logement</i>			
Oui	62,9	78,2	0,80
Non	37,1	21,8	1,70

Source : ÉLDEQ, ISQ.

Tableau A.4.1
Différences dans les caractéristiques selon les trois groupes de décrocheurs (%)

Variables	A	B	C
<i>Nombre de frères/sœurs</i>			
0	39,7	45,1	42,6
1	38,1	37,7	40,3
2+	22,2	17,2	17,1
<i>Nombre de personnes dans le ménage</i>			
3-	34,9	42,6	40,4
4	39,7	38,5	40,4
5+	25,4	18,9	19,2
<i>Groupe d'âge (mère)</i>			
24 ans et moins	33,3	18,2	21,4
25-29 ans	25,4	30,6	32,8
30 ans et plus	41,3	51,2	45,8
<i>Type de famille à l'enquête (3 cat.)**</i>			
Famille intacte	71,4	82,8	82,8
Famille recomposée	12,7	7,4	10,7
Famille monoparentale	15,9	9,8	6,5
<i>Langue(s) parlée(s) à la maison (mère) ***</i>			
Français	51,7	81,7	86,9
Anglais	15,0	10,8	9,7
Ni français ni anglais	33,3	7,5	3,4
<i>Première langue apprise (mère) ***</i>			
Français	50,8	81,0	84,9
Anglais	7,9	7,4	7,4
Ni français ni anglais	41,3	11,6	7,7
<i>Langue(s) de conversation (mère) ***</i>			
Français ou anglais	36,5	54,5	49,2
Français et anglais	19,0	32,2	39,8
Français, anglais et autre	14,3	8,3	7,0
Français ou anglais et autre	30,2	5,0	4,0
<i>Plus au diplôme (mère) ***</i>			
Pas dip. second.	32,3	16,5	15,1
Dip. études sec.	25,8	27,3	25,4
Dip. post-sec. (sauf univ.)	21,0	35,5	29,7
Dip. universitaire	21,0	20,7	29,8
<i>Travaille à l'enquête (mère)</i>			
Non	87,3	85,0	80,9
Oui	12,7	15,0	19,1
<i>Travaille 12 derniers mois (mère) ***</i>			
Non	53,2	29,2	27,1
Oui	46,8	70,8	72,9

* $p < 0,1$ ** $p < 0,05$ *** $p < 0,01$

A : familles ayant décroché dès le second volet.

B : familles ayant décroché au début de la seconde phase de l'ÉLDEQ.

C : familles ayant participé à tous les volets de l'ÉLDEQ.

Source : ÉLDEQ, ISQ.

Tableau A.4.1
Différences dans les caractéristiques selon les trois groupes de décrocheurs (%)
(suite)

Variables	A	B	C
<i>Niveau de suffisance du revenu (3 cat.) ***</i>			
Revenu suffisant	39,3	73,1	80,6
Revenu insuffisant	19,6	9,2	10,4
Revenu très insuffisant	41,1	17,6	8,9
<i>Nombre d'années depuis immigration (mère)</i>			
Moins de 5 ans	37,5	47,1	26,4
5+ ans	62,5	52,9	73,6
<i>Statut civil (mère)</i>			
Mariée	50,8	38,1	45,5
Célibataire, séparée et divorcée	49,2	61,9	54,5
<i>Propriétaire du logement***</i>			
Oui	39,7	63,1	67,1
Non	60,3	36,9	32,9
<i>Groupe d'âge (père)</i>			
Moins de 25 ans	11,3	7,2	8,4
25-29 ans	18,9	32,4	27,7
30-34 ans	30,2	33,3	35,9
35-39 ans	20,8	19,8	20,5
40 ans et plus	18,9	7,2	7,5
<i>Présence du père biologique**</i>			
Oui	84,1	91,0	93,2
Non	15,9	9,0	6,8
<i>Statut d'immigrant (père) ***</i>			
N'est pas immigrant	55,8	89,2	90,9
Immigrant	44,2	10,8	9,1
<i>Nombre d'années depuis immigration (père)</i>			
Moins de 5 ans	39,1	41,7	22,5
5 ans ou plus	60,9	58,3	77,5
<i>Langue(s) parlée(s) à la maison (père) ***</i>			
Français	54,0	83,8	87,7
Anglais	12,0	9,9	9,4
Ni français ni anglais	34,0	6,3	3,0

* $p < 0,1$ ** $p < 0,05$ *** $p < 0,01$

A : familles ayant décroché dès le second volet.

B : familles ayant décroché au début de la seconde phase de l'ÉLDEQ.

C : familles ayant participé à tous les volets de l'ÉLDEQ.

Source : ÉLDEQ, ISQ.

Tableau A.4.1
Différences dans les caractéristiques selon les trois groupes de décrocheurs (%)
(suite)

Variables	A	B	C
<i>Première langue apprise (père) ***</i>			
Français ou anglais	57,7	88,3	91,3
Ni français ni anglais	42,3	11,7	8,7
<i>Langue(s) de conversation (père) ***</i>			
Franç. ou ang. seul.	30,8	51,4	43,9
Franç. et ang. seul.	23,1	35,1	45,6
Fr. et ang.+ aut. lang.	21,2	8,1	7,8
Fr. ou ang.+ aut. lang.	25,0	5,4	2,6
<i>Plus au diplôme (père)</i>			
Pas dip. second.	25,0	22,9	19,2
Dip. études sec.	23,1	32,1	24,9
Dip. post-sec. (sauf univ.)	30,8	27,5	29,1
Dip. universitaire	21,2	17,4	26,8
<i>Travaille à l'enquête (père) **</i>			
Non	22,6	9,9	10,3
Oui	77,4	90,1	89,7
<i>Travaille 12 derniers mois (père) ***</i>			
Non	15,4	7,2	3,5
Oui	84,6	92,8	96,5
<i>Statut d'emploi principal (père) ***</i>			
Temps partiel et sans objet	24,5	10,8	7,2
Temps plein	75,5	89,2	92,8

* $p < 0,1$ ** $p < 0,05$ *** $p < 0,01$

A : familles ayant décroché dès le second volet.

B : familles ayant décroché au début de la seconde phase de l'ÉLDEQ.

C : familles ayant participé à tous les volets de l'ÉLDEQ.

Source : ÉLDEQ, ISQ.

Tableau A.5.1

Tableau complet des coefficients des régressions *probit* selon les variables d'intérêt

	Variables de contrôle uniquement		Présence du père biologique		Naissance prématurée	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Modèle unidimensionnel</i>						
Constante			-0,121	0,104	-0,442***	0,030
Variable d'intérêt			-0,347***	0,1084	0,021	0,138
Test χ^2 pour la relation globale (d.l.) [valeur p]			10,182 (1) [0,001]		0,023 (1) [0,881]	
<i>Modèle multidimensionnel</i>						
Constante	-0,968***	0,311	-0,959***	0,368	-0,962***	0,311
Âge de la mère (en années)	0,007	0,010	0,008	0,0097	0,008	0,0097
Plus haut diplôme (mère) (Réf : Dip. Universitaire)						
Pas dip. Second.	0,454***	0,167	0,453***	0,167	0,452***	0,167
Dip. études sec.	0,255**	0,128	0,258**	0,128	0,254**	0,128
Dip. post-sec. (sauf univ.)	0,206*	0,110	0,205*	0,110	0,202*	0,111
Nombre de frères/sœurs (Réf : aucun)						
1	0,014	0,086	0,017	0,086	0,009	0,086
2	-0,015	0,136	-0,015	0,136	-0,022	0,136
3	-0,363	0,285	-0,364	0,285	-0,375	0,286
4+	-0,202	0,471	-0,204	0,471	-0,214	0,471
Locataire du logement	0,135	0,092	0,138	0,092	0,135	0,092
Regroupement des professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	0,027	0,129	0,021	0,129	0,022	0,129
Cadre inter.tech	0,020	0,109	0,019	0,109	0,016	0,109
Contrem.ouv qua.	0,226	0,152	0,223	0,152	0,228	0,152
Ouv.non-qualifié	0,143	0,183	0,141	0,183	0,153	0,183
Revenus du ménage (en dollars)	-2,17E-06	1,97E-06	-2,18E-06	1,98E-06	-2,23E-06	1,97E-06
Revenus du ménage au Carré (en dollars ²)	1,86E-12	3,87E-12	1,88E-12	3,88E-12	1,93E-12	3,86E-12
Variable d'intérêt			-0,016	0,207	-0,196	0,199
Test χ^2 pour la relation globale (d.l.) [valeur p]	31,755 (15) [0,007]		32,005 (16) [0,010]		32,753 (16) [0,008]	
Test χ^2 conjoint de la variable d'intérêt (d.l.) [valeur p]			0,881 (1) [0,348]		0,998 (1) [0,318]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.5.1

Tableau complet des coefficients des régressions *probit* selon les variables d'intérêt (suite)

	Insuffisance du revenu		Mère immigrante		Statut socioéconomique du ménage	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Modèle unidimensionnel</i>						
Constante	-0,531***	0,034	-0,497***	0,031	-0,443***	0,030
Variable d'intérêt	0,332***	0,0704	0,509***	0,093	-0,150***	0,031
Test χ^2 pour la relation globale (d.l.) [valeur p]	22,183 (1) [<,0001]		29,849 (1) [<,0001]		23,994 (1) [<,0001]	
<i>Modèle multidimensionnel</i>						
Constante	-0,997***	0,316	-0,903***	0,313	-0,976***	0,314
Âge de la mère (en années)	0,007	0,010	0,003	0,010	0,008	0,010
Plus haut diplôme (mère) (Réf : Djp. Universitaire)						
Pas dip. Second.	0,453***	0,167	0,487***	0,167	0,431**	0,194
Dip. études sec.	0,257**	0,128	0,284**	0,129	0,238	0,147
Dip. post-sec. (sauf univ.)	0,207*	0,111	0,222**	0,111	0,196*	0,117
Nombre de frères/sœurs (Réf : aucun)						
1	0,015	0,086	0,009	0,086	0,014	0,086
2	-0,018	0,136	-0,003	0,136	-0,015	0,136
3	-0,373	0,286	-0,380	0,287	-0,365	0,286
4+	-0,210	0,472	-0,161	0,472	-0,204	0,471
Locataire du logement	0,128	0,092	0,093	0,093	0,135	0,092
Regroupement des professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	0,023	0,129	0,035	0,129	0,033	0,137
Cadre inter.tech	0,018	0,109	0,020	0,109	0,025	0,111
Contrem.ouv. qua.	0,224	0,152	0,210	0,153	0,224	0,152
Ouv.non-qualifié	0,137	0,183	0,080	0,186	0,139	0,184
Revenus du ménage (en dollars)	-1,71E-06	2,12E-06	-1,56E-06	1,98E-06	-1,79E-06	2,55E-06
Revenus du ménage au carré (en dollars ²)	1,18E-12	4,16E-12	1,09E-12	4,01E-12	1,44E-12	4,26E-12
Variable d'intérêt	0,077	0,136	0,387***	0,143	-0,024	0,103
Test χ^2 pour la relation globale (d.l.) [valeur p]	32,068 (16) [0,010]		39,055 (16) [0,001]		31,812 (16) [0,01]	
Test χ^2 conjoint de la variable d'intérêt (d.l.) [valeur p]	0,313 (1) [0,576]		7,3 (1) [0,007]		0,057 (1) [0,811]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.5.1

Tableau complet des coefficients des régressions *probit* selon les variables d'intérêt (suite)

	Surprotection Parentale (mère)		Perception d'impact parental (mère)		Naissance de rang 1	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Modèle unidimensionnel</i>						
Constante	-0,636***	0,073	-0,069	0,135	-0,458***	0,040
Variable d'intérêt	0,036***	0,012	-0,044***	0,016	0,037	0,059
Test χ^2 pour la relation globale (d.l.) [valeur p]	8,575 (1) [0,003]		7,888 (1) [0,005]		0,396 (1) [0,529]	
<i>Modèle multidimensionnel</i>						
Constante	-1,098***	0,324	-0,822**	0,380	-1,27***	0,375
Âge de la mère (en années)	0,007	0,010	0,007	0,010	0,008	0,010
Plus haut diplôme (mère) (Réf : Dip. Universitaire)						
Pas dip. Second.	0,442***	0,167	0,444***	0,167	0,472***	0,167
Dip. études sec.	0,249*	0,129	0,251*	0,128	0,264**	0,128
Dip. post-sec. (sauf univ.)	0,208*	0,111	0,203*	0,111	0,214*	0,111
Nombre de frères/sœurs (Réf : aucun)						
1	0,030	0,087	0,014	0,086	0,275	0,202
2	-0,005	0,136	-0,016	0,136	0,233	0,222
3	-0,339	0,286	-0,371	0,286	-0,111	0,338
4+	-0,174	0,472	-0,195	0,471	0,048	0,502
Locataire du logement	0,139	0,092	0,132	0,092	0,124	0,092
Regroupement des Professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	0,045	0,130	0,027	0,129	0,035	0,129
Cadre inter.tech	0,029	0,109	0,022	0,109	0,027	0,109
Contrem.ouv qua.	0,210	0,153	0,217	0,153	0,232	0,152
Ouv.non-qualifié	0,141	0,183	0,134	0,183	0,149	0,183
Revenus du ménage (en dollars)	-2,05E-06	1,97E-06	-2,07E-06	1,97E-06	-2,29E-06	1,97E-06
Revenus du ménage au carré (en dollars ²)	1,91E-12	3,85E-12	1,76E-12	3,88E-12	2,02E-12	3,85E-12
Variable d'intérêt	0,022	0,017	-0,016	0,024	0,283	0,199
Test χ^2 pour la relation Globale (d.l.) [valeur p]	33,478 (16) [0,006]		32,190 (16) [0,009]		33,770 (16) [0,006]	
Test χ^2 conjoint de la variable d'intérêt (d.l.) [valeur p]	2,354 (1) [0,125]		0,435 (1) [0,509]		2,015 (1) [0,156]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.5.1

Tableau complet des coefficients des régressions *probit* selon les variables d'intérêt (suite)

	Autre union (sauf avec père)		Famille monoparentale	
	Estimation	Erreur standard	Estimation	Erreur standard
<i>Modèle unidimensionnel</i>				
Constante	-0,456***	0,035	-0,469***	0,031
Variable d'intérêt	-0,056	0,071	0,384***	0,113
Test χ^2 pour la relation globale (d.l.) [valeur p]	0,619 (1) [0,432]		11,470 (1) [0,001]	
<i>Modèle multidimensionnel</i>				
Constante	-0,948***	0,321	-0,969***	0,311
Âge de la mère (en années)	0,008	0,010	0,007	0,010
Plus haut diplôme (mère) (Réf : Dip. Universitaire)				
Pas dip. Second.	0,383**	0,172	0,453***	0,167
Dip. études sec.	0,254*	0,130	0,254**	0,128
Dip. post-sec. (sauf univ.)	0,202*	0,111	0,206*	0,110
Nombre de frères/sœurs (Réf : aucun)				
1	-0,009	0,087	0,015	0,086
2	-0,052	0,139	-0,015	0,136
3	-0,313	0,290	-0,364	0,285
4+	-0,217	0,472	-0,200	0,471
Locataire du logement	0,085	0,094	0,135	0,092
Regroupement des Professions Pineo (5 cat.) (Réf : Empl.bureau.service)				
Prof. cadre sup.	0,008	0,129	0,022	0,129
Cadre inter.tech	0,005	0,110	0,020	0,109
Contrem.ouv qua.	0,244	0,158	0,224	0,152
Ouv.non-qualifié	0,114	0,191	0,142	0,183
Revenus du ménage (en dollars)	-2,09E-06	2,01E-06	-2,11E-6	1,987E-6
Revenus du ménage au carré (en dollars ²)	1,62E-12	4,00E-12	1,78E-12	3,9E-12
Variable d'intérêt	0,015	0,090	0,039	0,219
Test χ^2 pour la relation globale (d.l.) [valeur p]	24,494 (16) [0,0793]		31,785 (16) [0,01]	
Test χ^2 conjoint de la variable d'intérêt (d.l.) [valeur p]	0,026 (1) [0,872]		0,030 (1) [0,862]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.5.1

Tableau complet des coefficients des régressions *probit* selon les variables d'intérêt (suite)

	Retard de croissance		Avortement	
	Estimation	Erreur standard	Estimation	Erreur standard
<i>Modèle unidimensionnel</i>				
Constante	-0,455***	0,031	-0,430***	0,037
Variable d'intérêt	0,178*	0,106	-0,037	0,062
Test χ^2 pour la relation globale (d.l.) [valeur p]	2,791 (1) [0,095]		0,367 (1) [0,545]	
<i>Modèle multidimensionnel</i>				
Constante	-0,962***	0,311	-1,095***	0,308
Âge de la mère (en années)	0,0076	0,010	0,013	0,010
Plus haut diplôme (mère) (Réf : Dip. Universitaire)				
Pas dip. Second.	0,461***	0,167	0,478***	0,166
Dip. études sec.	0,259**	0,128	0,298**	0,127
Dip. post-sec. (sauf univ.)	0,207*	0,110	0,231**	0,110
Nombre de frères/sœurs (Réf : aucun)				
1	0,010	0,086	0,005	0,085
2	-0,022	0,136	0,036	0,133
3	-0,365	0,285	-0,280	0,272
4+	-0,202	0,471	-0,228	0,471
Locataire du logement	0,136	0,092	0,168**	0,090
Regroupement des Professions Pineo (5 cat.) (Réf : Empl.bureau.service)				
Prof. cadre sup.	0,021	0,129	0,046	0,127
Cadre inter.tech	0,020	0,109	-0,006	0,108
Contrem.ouv qua.	0,223	0,152	0,172	0,150
Ouv.non-qualifié	0,137	0,183	0,155	0,181
Revenus du ménage (en dollars)	-2,18E-6	1,966E-6	-2,05E-6	1,937E-6
Revenus du ménage au carré (en dollars ²)	1,85E-12	3,88E-12	1,65E-12	3,85E-12
Variable d'intérêt	-0,094	0,145	-0,164**	0,084
Test χ^2 pour la relation globale (d.l.) [valeur p]	32,175 (16) [0,010]		36,683 (16) [0,002]	
Test χ^2 conjoint de la variable d'intérêt (d.l.) [valeur p]	0,42 (1) [0,520]		3,84 (1) [0,050]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.6.1

Tableau complet des tests BGLW pour diverses variables d'intérêt

	Présence du père biologique (Probit)		Naissance prématurée (Probit)		Insuffisance du revenu (Probit)	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Variables de contrôle</i>						
Constante	1,945***	0,603	-1,783***	0,504	5,045*	2,729
Âge de la mère (en années)	-0,04**	0,017	0,021	0,015	0,048*	0,026
Plus haut diplôme (mère) (Réf : Dip. universitaire)						
Pas dip. second.	-0,416	0,341	-0,238	0,276	0,280	0,469
Dip. études sec.	-0,408	0,300	-0,221	0,209	0,290	0,424
Dip. post-sec. (sauf univ.)	-0,382	0,284	-0,337*	0,189	0,383	0,389
Nombre de frères/sœurs (Réf : aucun)						
1	0,774***	0,217	-0,254*	0,144	1,252***	0,297
2	0,067	0,244	-0,372	0,244	1,747***	0,471
3	0,142	0,429	-0,458	0,485	2,225***	0,598
4+	-0,701	0,544	-3,699	141,600	2,761***	0,917
Locataire du logement	-0,192	0,165	0,067	0,147	0,476**	0,240
Regroupement des professions Pineo (5 cat.) (Réf : Empl. bureau.service)						
Prof. cadre sup.	0,152	0,320	-0,008	0,205	-0,125	0,558
Cadre inter.tech	0,210	0,252	-0,228	0,198	0,684*	0,353
Contrem.ouv qua.	0,189	0,257	0,050	0,244	-0,348	0,370
Ouv.non-qualifié	-0,278	0,262	0,354	0,264	0,027	0,422
Revenus du ménage (en dollars)	2,7E-05***	5,63E-06	-4,90E-06	3,44E-06	-3,10E-04	2,10E-04
Revenus du ménage au carré (en dollars ²)	-3E-11***	7,62E-12	4,68E-12	7,24E-12	3,47E-11	4,07E-09
Effet de l'attrition	-0,068	0,163	-0,089	0,147	0,145	0,252
Test χ^2 / Test F pour la relation globale avec l'attrition (d.l.) [valeur p]	91,097 (16) [<,0001]		17,863 (16) [0,332]		772,283 (16) [<,0001]	
Test c2 conjoint / Test F partiel de l'attrition (d.l.) [valeur p]	0,169 (1) [0,681]		0,376 (1) [0,540]		0,353 (1) [0,552]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.6.1
Tableau complet des tests BGLW pour diverses variables d'intérêt (suite)

	Mère immigrante (<i>Probit</i>)		Statut socioéconomique du ménage (MCO)		Surprotection parentale (mère) (MCO)	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Variables de contrôle</i>						
Constante	-3,277***	0,453	-0,578***	0,159	4,149***	0,955
Âge de la mère (en années)	0,073***	0,013	0,005*	0,003	0,023	0,016
Plus haut diplôme (mère) (Réf : Dip. universitaire)						
Pas dip. second.	-0,405*	0,240	-0,960***	0,047	0,867***	0,283
Dip. études sec.	-0,384**	0,186	-0,711***	0,035	0,521***	0,213
Dip. post-sec. (sauf univ.)	-0,120	0,155	-0,373***	0,030	0,173	0,181
Nombre de frères/sœurs (Réf : aucun)						
1	0,080	0,122	-0,023	0,024	-0,677***	0,143
2	-0,151	0,204	0,008	0,037	-0,546**	0,226
3	0,182	0,299	-0,089	0,071	-1,013**	0,427
4+	-4,517	137,200	-0,069	0,128	-0,973	0,764
Locataire du logement	0,698***	0,124	0,007	0,026	-0,146	0,155
Regroupement des professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	-0,062	0,187	0,475***	0,035	-0,705***	0,211
Cadre inter.tech	0,070	0,157	0,230***	0,030	-0,325*	0,181
Contrem.ouv qua.	0,240	0,199	-0,074*	0,044	0,682**	0,265
Ouv.non-qualifié	0,862***	0,217	-0,155***	0,053	0,194	0,320
Revenus du ménage (en dollars)	-1E-05***	3,08E-06	1,6E-05***	5,2E-07	-6,1E-06*	3,1E-06
Revenus du ménage au carré (en dollars ²)	1,3E-11***	4,74E-12	-1E-11***	< 1E-11	< 1E-11	< 1E-11
Effet de l'attrition	0,375***	0,116	-0,008	0,024	0,202	0,145
Test χ^2 / Test F pour la relation globale avec l'attrition (d.l.) [valeur p]	118,963 (16) [<,0001]		328,290 (16, 1274) [<,0001]		7,820 (16, 1246) [<,0001]	
Test χ^2 conjoint / Test F partiel de l'attrition (d.l.) [valeur p]	10,091 (1) [0,001]		0,107 (1, 1274) [0,743]		0,392 (1, 1246) [0,531]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.6.1
Tableau complet des tests BGLW pour diverses variables d'intérêt (suite)

	Perception d'impact parental (mère) (MCO)		Naissance de rang 1 (Probit)		Autre union (sauf avec père) (Probit)	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Variables de contrôle</i>						
Constante	9,042***	0,667	3,357***	0,615	-2,578***	0,331
Âge de la mère (en années)	-0,023**	0,011	-0,032*	0,018	0,054***	0,010
Plus haut diplôme (mère) (Réf : Dip. universitaire)						
Pas dip. second.	-0,648***	0,198	-0,792**	0,317	0,478***	0,174
Dip. études sec.	-0,271*	0,148	-0,446*	0,232	0,402***	0,129
Dip. post-sec. (sauf univ.)	-0,199	0,126	-0,345*	0,201	0,199*	0,111
Nombre de frères/sœurs (Réf : aucun)						
1	0,020	0,100	-4,042***	0,198	-0,033	0,088
2	-0,068	0,158	-3,716***	0,235	0,192	0,132
3	-0,495	0,303	-3,876***	0,408	0,140	0,251
4+	0,551	0,533	-3,785***	0,619	0,447	0,431
Locataire du logement	-0,245**	0,109	0,564***	0,175	0,162*	0,095
Regroupement des professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	0,361**	0,147	-0,593**	0,244	-0,091	0,129
Cadre inter.tech	0,133	0,126	-0,361*	0,205	-0,012	0,109
Contrem.ouv qua.	-0,554***	0,186	-0,181	0,312	-0,063	0,162
Ouv.non-qualifié	-0,491**	0,224	-0,311	0,349	-0,322	0,212
Revenus du ménage (en dollars)	4,8E-06**	2,2E-06	7,3E-06**	3,30E-06	2,46E-06	3,37E-06
Revenus du ménage au carré (en dollars ²)	0	0	-9,1E-12	6,07E-12	-1,35E-11	1,56E-11
Effet de l'attrition	-0,065	0,102	0,160	0,157	0,007	0,088
Test χ^2 / Test F pour la relation globale avec l'attrition (d.l.) [valeur p]	6,87 (16, 1244) [<.0001]		1429,854 (16) [<.0001]		60,716 (16) [<.0001]	
Test χ^2 conjoint / Test F partiel de l'attrition (d.l.) [valeur p]	0,431 (1, 1244) [0,512]		1,040 (1) [0,308]		0,007 (1) [0,933]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.6.1

Tableau complet des tests BGLW pour diverses variables d'intérêt (suite)

	Famille monoparentale (<i>Probit</i>)		Retard de croissance (<i>Probit</i>)		Avortement (<i>Probit</i>)	
	Estimation	Erreur standard	Estimation	Erreur standard	Estimation	Erreur standard
<i>Variables de contrôle</i>						
Constante	-1,926***	0,442	-1,926***	0,442	-2,420***	0,308
Âge de la mère (en années)	0,017	0,013	0,017	0,013	0,064***	0,010
Plus haut diplôme (mère) (Réf : Dip. universitaire)						
Pas dip. second.	0,128	0,355	0,536**	0,225	0,227	0,165
Dip. études sec.	0,281	0,303	0,316*	0,177	0,048	0,124
Dip. post-sec. (sauf univ.)	0,049	0,297	0,126	0,156	0,070	0,106
Nombre de frères/sœurs (Réf : aucun)						
1	-0,548***	0,208	-0,339***	0,119	0,052	0,083
2	-0,078	0,272	-0,621***	0,222	0,151	0,128
3	-0,056	0,442	-0,158	0,337	0,021	0,244
4+	-3,743	221,200	-0,057	0,582	-0,377	0,475
Locataire du logement	0,003	0,1765	0,066	0,123	0,317***	0,090
Regroupement des professions Pineo (5 cat.) (Réf : Empl.bureau.service)						
Prof. cadre sup.	-0,274	0,363	-0,011	0,178	1,0E-04	0,122
Cadre inter.tech	-0,049	0,262	-0,003	0,148	0,014	0,104
Contrem.ouv qua.	0,118	0,240	-0,237	0,216	-0,382**	0,163
Ouv.non-qualifié	0,171	0,290	-0,513	0,318	0,077	0,185
Revenus du ménage (en dollars)	-4,0E-05***	6,683E-6	3,915E-6	6,928E-6	-9,25E-7	1,877E-6
Revenus du ménage au carré (en dollars ²)	4,5E-11***	9,00E-12	-4,64E-11	4,79E-11	-1,02E-12	4,3E-12
Effet de l'attrition	0,088	0,172	-0,078	0,119	-0,170**	0,085
Test χ^2 / Test F pour la relation globale avec l'attrition (d.l.) [valeur p]	99,354 (16) [<0,0001]		28,606 (16) [0,027]		70,306 (16) [<0,0001]	
Test χ^2 conjoint / Test F partiel de l'attrition (d.l.) [valeur p]	0,247 (1) [0,619]		0,433 (1) [0,511]		4,03 (1) [0,045]	

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.7.1

Fréquences (%) pour des variables fixes ou quasi-fixes pour les volets un et huit, données pondérées

Variables	Volet 1	Volet 8
<i>Statut d'immigrante (mère)</i>		
N'est pas immigrante	84,6	86,8
Immigrante européenne	3,3	2,7
Immigrante non européenne	12,2	10,4
<i>Statut immigrant (père)</i>		
N'est pas immigrant	84,0	85,6
Immigrant européen	3,2	3,3
Immigrant non européen	12,7	11,1
<i>Langue maternelle (mère)</i>		
Français	76,4	78,9
Anglais	8,7	8,2
Ni français ni anglais	14,9	13,0
<i>Langue maternelle (père)</i>		
Français	76,0	77,9
Anglais	8,3	8,2
Ni français ni anglais	15,7	13,9
<i>Plus haut diplôme (mère)</i>		
Pas dip. second.	20,1	14,2
Dip. études sec.	26,8	23,5
Dip. post-sec. (sauf univ.)	28,2	34,8
Dip. universitaire	24,8	27,6
<i>Plus haut diplôme (père)</i>		
Pas dip. second.	20,6	18,2
Dip. études sec.	26,3	22,5
Dip. post-sec. (sauf univ.)	28,3	33,0
Dip. universitaire	24,8	26,3
<i>Rang de naissance de l'enfant</i>		
1	43,6	44,4
2	39,6	40,1
3	11,3	10,8
4	3,4	3,0
5+	2,1	1,6
<i>Naissance prématurée</i>		
Oui	6,2	6,1
Non	93,8	93,9
<i>Union antérieure (mère)</i>		
Aucune union	74,1	74,5
Oui, union	25,9	25,5

Source : ÉLDEQ, ISQ.

Tableau A.7.1
Fréquences (%) pour des variables fixes ou quasi-fixes pour les volets un et huit,
données pondérées (suite)

Variables	Volet 1	Volet 8
<i>Union antérieure (père)</i>		
Aucune union	73,0	71,9
Oui, union	27,0	28,1
<i>Type de famille à la naissance</i>		
Intacte	78,0	78,8
Intacte (eua)	2,8	2,4
Recomposée : mère	7,0	6,8
Recomposée : père seulement	3,8	4,2
Monoparentale	8,4	7,8
<i>Situation conjugale à la naissance</i>		
Mariés	18,7	15,5
Mariage précédé d'union libre	25,2	27,4
Union libre	47,6	49,2
Séparés	3,8	3,8
Parents n'ont jamais vécu ensemble	4,7	4,0

Source : ÉLDEQ, ISQ.

Tableau A.7.2
Caractéristiques (%) des décrocheurs et des non-décrocheurs en E5 (données pondérées) – Tests du khi-deux

Variables	Non-décrocheurs	Décrocheurs
<i>Groupe d'âge du père/conjoint</i>		
Moins de 25 ans	0,8	1,9
25-29 ans	12,7	10,1
30-34 ans	29,8	32,2
35-39 ans	33,9	32,2
40 ans et plus	22,8	23,5
<i>Groupe d'âge du mère/conjointe</i>		
Moins de 25 ans	6,0	6,4
25-29 ans	22,7	24,3
30-34 ans	32,7	29,1
35-39 ans	28,7	30,0
40 ans et plus	9,9	10,1
<i>Diplôme le plus élevé (père/conjoint)**</i>		
Pas dip. second.	17,2	19,2
Dip. d'études sec.	20,7	23,6
Dip. post-sec. (sauf univ.)	34,7	37,3
Dip. universitaire	27,4	20,0
<i>Diplôme le plus élevé (mère/conjointe)***</i>		
Pas dip. second.	14,4	20,4
Dip. d'études sec.	20,9	23,3
Dip. post-sec. (sauf univ.)	35,1	36,8
Dip. universitaire	29,5	19,5
<i>Travaille à l'enquête (mère/conjointe)</i>		
Non	34,6	35,8
Oui	65,4	64,2
<i>Travaille à l'enquête (père/conjoint)</i>		
Non	9,2	10,1
Oui	90,8	89,9
<i>Nombre de personnes dans le ménage</i>		
2	4,0	5,2
3	19,0	20,2
4	48,0	49,0
5	19,6	17,0
6	6,1	5,7
7+	3,2	2,9
<i>Nombre de frères/sœurs</i>		
0	18,5	20,7
1	53,8	53,6
2	20,2	18,0
3	5,7	5,2
4+	1,8	2,5

* $p < 0,1$

** $p < 0,05$

*** $p < 0,01$

Source : ÉLDEQ, ISQ.

Tableau A.7.2

Caractéristiques (%) des décrocheurs et des non-décrocheurs en E5 (données pondérées) – Tests du khi-deux (suite)

Variables	Non-décrocheurs	Décrocheurs
<i>Type de famille (3 cat.)**</i>		
Famille intacte	75,0	70,9
Famille recomposée	12,1	11,4
Famille monoparentale	13,0	17,7
<i>Présence du père biologique***</i>		
Oui	83,5	78,0
Non	16,5	22,0
<i>Statut d'immigrante (mère)***</i>		
Pas immigrante	87,9	80,5
Immigrante européenne	3,0	4,1
Immigrante non-européenne	9,1	15,4
<i>Nombre d'années depuis immigration (mère)***</i>		
N'est pas immigrante	87,8	80,5
Moins de 10 ans	4,3	7,4
10 ans et plus	7,9	12,2
<i>Langue(s) de conversation (mère)***</i>		
Franç. ou ang. seul.	46,2	49,5
Franç. et ang. seul.	38,2	29,4
Fr. et ang.+ aut. lang.	9,1	11,9
Fr. ou ang.+ aut. lang.	6,5	9,2
<i>Première langue apprise (mère)***</i>		
Français	79,3	71,6
Anglais	8,5	9,4
Ni français ni anglais	12,1	19,0
<i>Langue(s) parlée(s) à la maison (mère)***</i>		
Français	78,3	72,6
Anglais	11,6	11,3
Ni français ni anglais	6,0	11,7
Français et anglais	2,3	1,6
Fr. ou ang. + aut. Lang.	1,8	2,8
<i>Statut d'immigrant (père)</i>		
Pas immigrant	86,5	82,4
Immigrant européen	3,4	4,3
Immigrant non-européen	10,1	13,4
<i>Nombre d'années depuis immigration (père)</i>		
N'est pas immigrante	86,9	83,3
Moins de 10 ans	4,2	5,7
10 ans et plus	8,9	10,9

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.7.2

Caractéristiques (%) des décrocheurs et des non-décrocheurs en E5 (données pondérées) – Tests du khi-deux (suite)

Variables	Non-décrocheurs	Décrocheurs
<i>Langue(s) de conversation (père)***</i>		
Franç. ou ang. seul.	40,9	46,4
Franç. et ang. seul.	44,1	32,5
Fr. et ang.+ aut. lang.	10,3	15,1
Fr. ou ang.+ aut. lang.	4,7	6,0
<i>Première langue apprise (père)*</i>		
Français	77,6	72,4
Anglais	9,3	9,4
Ni français ni anglais	13,1	18,2
<i>Langue(s) parlée(s) à la maison (père)*</i>		
Français	78,1	73,8
Anglais	12,3	11,8
Ni français ni anglais	6,1	9,9
Français ou/et anglais + autre langue	3,4	4,5
<i>Principale source de revenu du ménage*</i>		
Salaire et traitement	80,7	75,7
Travail autonome	9,5	12,3
Aide sociale	6,5	7,3
Assurance-emploi	1,8	1,8
Autres	1,5	3,0
<i>Statut civil (mère)**</i>		
Mariée	50,3	43,0
Célibataire	43,2	49,9
Séparée légalement ou veuve	6,5	7,2
<i>Propriétaire du logement***</i>		
Oui	71,7	64,3
Non	28,3	35,7

* p < 0,1

** p < 0,05

*** p < 0,01

Source : ÉLDEQ, ISQ.

Tableau A.7.3

**Caractéristiques (en moyenne) des décrocheurs et des non-décrocheurs en E5
(données pondérées) – Tests t**

Variables	Non-décrocheurs	Décrocheurs
Satisfaction conjugale (mère) **	31,6	32,3
Satisfaction conjugale (père) † *	32,3	32,9
Hyperactivité (selon PCM)	4,0	3,8
Troubles émotifs (selon PCM) *	1,7	1,6
Anxiété (selon PCM) **	2,1	1,9
Agressivité proactive (selon PCM)	1,1	1,1
Agressivité physique globale (selon PCM) ***	2,1	1,9

† Hypothèse d'homogénéité des variances qui n'est pas respectée. La correction de Welch fut appliquée.

Source : ÉLDEQ, ISQ.

ANNEXE MÉTHODOLOGIQUE

1) Définition mathématique de l'attrition :

Si on considère l'attrition de façon plus mathématique, on peut définir ce concept de la façon suivante⁸ :

Soient un échantillon observé aux temps 1, 2, ..., T, une variable d'intérêt $Y = f(X)$ et une variable dichotomique A_t , qui est défini comme suit pour un certain sujet :

$$A_t = \begin{cases} 0 & \text{si le sujet fut interviewé} \\ 1 & \text{sinon} \end{cases}$$

On est intéressé par $f(y_t | x_t)$, qui représente la fonction de densité de Y sachant un certain vecteur de variables indépendantes X . Y_t est donc observé si $A_t = 0$. Ceci fait en sorte qu'on obtient les deux équations suivantes :

$$(1) Y_t = bX_t + \varepsilon_t$$

$$(2) Y_t \text{ est observée si } A_t = 0$$

On suppose alors que X_t est connue pour toutes les unités, comme s'il s'agissait de caractéristiques fixes ($X_t = c \forall t = 1, \dots, T$) ou d'une valeur passée de Y ($X_t = Y_{t-n}$ avec $n > 0$). Puisque l'équation (1) peut être estimée avec les sujets interviewés, on peut tirer des équations (1) et (2) que nous connaissons $f(y_t | x_t, A_t = 0)$. Cependant, notre but étant d'estimer $f(y_t | x_t)$ sur l'ensemble de l'échantillon, il nous manque des informations sur A_t . On fait donc l'hypothèse que A_t est induite par une variable latente A_t^* :

$$(3) A_t^* = \delta_0 + \delta_1 X_t + \delta_2 Z_t + v_t$$

Z_t est un vecteur de variables indépendantes distinctes de X_t mais observables pour toutes les unités. Z_t pourrait par exemple être des caractéristiques fixes du répondant ($Z_t = c \forall t = 1, \dots, T$), une valeur passée ($Z_t = W_{t-n}$ avec $n > 0$) ou bien ne pas dépendre de l'entrevue ($Cov(A_t, Z_t) = 0$). De l'équation (3), on peut redéfinir la variable dichotomique A_t comme suit :

⁸ Ces concepts et cette notation proviennent de Fitzgerald et al. (1998) et furent notamment repris par Alderman et al. (2001) et Razafindratsima et al. (2004). Le discours présenté ici reprend principalement l'excellente synthèse faite par Razafindratsima et al. (2004).

$$A_t = \begin{cases} 0 & \text{si } A_t^* < 0 \\ 1 & \text{sinon} \end{cases}$$

On a alors les deux types de sélection suivants :

- Sélection sur des variables inobservables si $P(A_t = 0 | y_t, x_t, z_t) \neq P(A_t = 0 | x_t, z_t)$. Ceci signifie alors que des variables qui ne sont pas observées agissent sur Y_t et sur A_t .
- Sélection sur des variables observables si $P(A_t = 0 | y_t, x_t, z_t) \neq P(A_t = 0 | x_t, z_t)$. Ainsi, la probabilité de décrocher de l'échantillon, sachant x_t et z_t , est indépendante de y_t . Conséquemment, la probabilité d'attrition est indépendante des facteurs qui ne sont pas observables.

Par la suite, on peut distinguer trois types de sélection sur des variables observables :

- Sélection complètement aléatoire si $P(A_t = 0 | y_t, x_t, z_t) = P(A_t = 0)$. Dans ce cas, la probabilité de décrocher de l'échantillon est totalement indépendante des autres variables, situation évidemment très irréaliste.
- Sélection aléatoire⁹ si on se retrouve dans une des deux situations suivantes :
 - a) y_t et z_t sont indépendantes conditionnellement à x_t et à $A_t = 0$.
 - b) $P(A_t = 0 | y_t, x_t, z_t) = P(A_t = 0 | x_t)$, donc que la probabilité d'attrition est indépendante de z_t ou encore que $\delta_2 = 0$ dans l'équation 3.
- Sélection non aléatoire¹⁰ si aucune des deux conditions précédentes n'est respectée. Ceci veut donc dire que les variables y_t et z_t sont endogènes.

Si l'attrition est aléatoire ou complètement aléatoire, le vecteur b de l'équation 1 n'est pas biaisé par l'attrition. Dans le cas où l'attrition n'est pas aléatoire, le vecteur b pourrait être biaisé et il faudra apporter une correction pour rectifier le vecteur. Fitzgerald et *al.* (1998) suggèrent une pondération spéciale permettant de corriger pour les biais d'attrition. Le test de Beckett, Gould, Lillard et Welch (BGLW) (Beckett et *al.*, 1988) permet de repérer les biais d'attrition et est basé sur cette définition de l'attrition.

⁹ Le terme originel pour cette sélection est « ignorable ».

¹⁰ Le terme originel pour cette sélection est « non ignorable ».

2) Test du khi-deux

Le test d'indépendance du khi-deux de Pearson permet de tester l'indépendance de deux variables discrètes provenant, par exemple, d'un tableau croisé. Ce test permet ainsi de tester l'hypothèse nulle (H_0) où les deux variables sont indépendantes contre l'hypothèse alternative (H_1) où elles ne sont pas indépendantes. De façon générale, ce test s'exprime comme suit :

$$\sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_k$$

Pour chaque cellule, O_i représente la fréquence observée et E_i représente la fréquence attendue (ou théorique). De plus, le nombre de degrés de liberté de la statistique du χ^2 est déterminé par les modalités des deux variables discrètes alors que $k = (m_1 - 1) \times (m_2 - 1)$ où m_1 et m_2 représentent le nombre de catégories des deux variables.

3) Analyse de variance à 1 facteur

L'analyse de variance à 1 facteur est un modèle statistique dont l'objectif est la comparaison, en moyenne, de plusieurs groupes pour une certaine variable continue donnée. De façon statistique, il s'agit d'un modèle linéaire de la forme suivante :

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ avec } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

où Y_{ij} représente la valeur de l'observation j du traitement i , μ représente la grande moyenne et α_i l'effet différentiel du groupe i .

À l'aide de la table ANOVA, il est possible de faire un test de Fisher testant l'hypothèse nulle : $H_0 : \alpha_i = \alpha_j \forall i, j$ contre l'hypothèse alternative : $H_1 : \alpha_i \neq \alpha_j$ pour au moins une paire (i, j) . Cependant, bien que le test F nous renseigne sur l'ensemble des groupes, il est souvent fort utile de faire des comparaisons deux-à-deux entre les différents groupes étudiés. Afin de faire ces comparaisons dites multiples, il existe plusieurs méthodes, dont les plus connues sont les méthodes de Tukey, Bonferroni et Scheffé. Ici, on utilisera la méthode de Tukey, qui consiste à construire des intervalles de confiance de façon à ce que chaque intervalle possède une probabilité de couverture connue pour ensuite appliquer la dualité entre les tests d'hypothèses et les intervalles de confiance pour

identifier les paires de groupes où il existe des différences significatives. Ceci nous permettra d'identifier les groupes d'attrition dont les caractéristiques sont significativement différentes.

En outre, pour tester l'hypothèse d'homogénéité de la variance nécessaire à l'utilisation de l'analyse de variance, le test de Levene sera utilisé. En comparant l'écart à la moyenne de chaque groupe, ce test permet de tester l'hypothèse nulle suivante : $H_0 : \sigma_i = \sigma_j \forall i, j$ contre l'hypothèse alternative : $H_1 : \sigma_i \neq \sigma_j$ pour au moins une paire (i, j) . Dans la situation où l'hypothèse nulle serait rejetée, donc que les variances des trois groupes ne sont pas homogènes, une transformation sera utilisée sur la variable dépendante du modèle. Ces transformations sont dans l'ordre : le logarithme ($Y' = \ln(Y)$), la racine carrée ($Y' = Y^{1/2}$) et l'inverse ($Y' = Y^{-1}$). Dans la situation où ces trois transformations s'avèrent inefficaces, la transformation donnant la valeur-p la plus élevée pour le test de Levene sera utilisée.

4) Régression *probit*

La transformation *probit*, inventée par le biométriste Bliss, est une méthode très populaire dans l'analyse des données binomiales. Cette transformation utilise les propriétés de la fonction de répartition de la loi normale (voir figure 4.1 pour visualiser cette fonction) pour formuler le modèle suivant :

$$P(Y = 1 | X) = \Phi(Xb)$$

où $\Phi(x)$ représente la fonction de répartition d'une loi normale $(0,1)$, X une matrice d'observations et b un vecteur de coefficients de régression.

On est en mesure d'obtenir un modèle satisfaisant les hypothèses de linéarité et de normalité du terme d'erreur des modèles linéaires en inversant l'équation précédente pour obtenir :

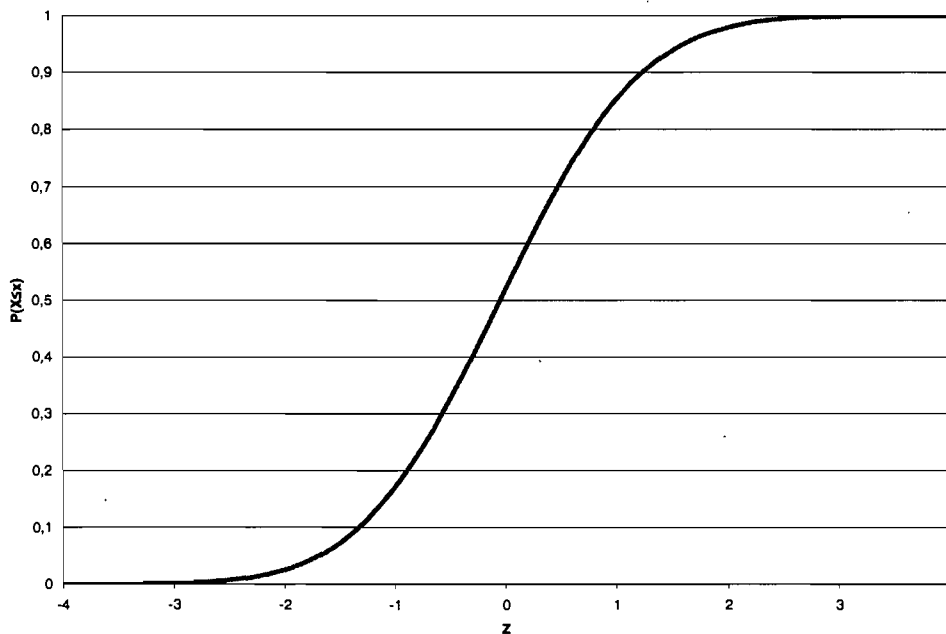
$$\Phi^{-1}(Y) = Xb \text{ avec } \Phi^{-1} : [0,1] \rightarrow (-\infty, +\infty)$$

Dans ce modèle, les coefficients de régression s'estiment par maximum de vraisemblance à l'aide de la fonction de vraisemblance suivante :

$$\ln L = \sum \ln(\Phi(x, b)) + \sum \ln(1 - \Phi(x, b))$$

Les modèles *logit* et *probit* donnent des résultats très comparables mais ils se distinguent sur quelques points. Tout d'abord, outre les fonctions de vraisemblance qui sont évidemment différentes, les coefficients b du modèle *probit* expriment un changement

Figure AA.1
Fonction de répartition de la loi normale (0,1)



dans la probabilité cumulée de la loi normale, mieux connu sous le nom de cote Z , qu'un sujet subisse l'événement étudié. Ceci distingue le modèle *probit* du modèle logistique où les coefficients b représentent un changement dans le logarithme de la cote (*log odds*) de la probabilité de subir l'événement. Conséquemment, il est impossible de créer des rapports de cote avec le modèle *probit*. En outre, la régression *probit* suppose que les catégories de la variable dépendante reflètent une distribution normale, contrairement au modèle logistique où les catégories de la variable dépendante sont distribuées uniformément. En général, sauf dans la situation où la variable indépendante possède une grande variabilité, les coefficients logistiques sont environ 1,7 fois plus grands que les coefficients de la régression *probit*.

5) Test F partiel

Dans la réalisation d'un modèle statistique, il est très intéressant d'obtenir le modèle le plus parcimonieux possible en gardant dans le modèle que les variables dont l'apport au modèle est significatif. Dans le cas de la régression par les moindres carrés ordinaires (MCO), il existe plusieurs méthodes de sélection de modèles (comme les méthodes pas-à-pas, par en avant, par en arrière ou bien le Cp de Mallows). Une des façons de tester si l'ajout d'une variable à un certain modèle est significatif dans une régression par les moindres carrés, qui est d'ailleurs utilisée dans plusieurs méthodes de sélection de modèles plus complexes comme la méthode pas-à-pas, est d'utiliser un test de Fisher partiel. Ce test permet de tester $H_0 : \beta_k = 0$ contre $H_1 : \beta_k \neq 0$ pour un i donné à l'aide de la statistique suivante :

$$F = \text{MSR} (X_k | X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_{p-1}) / \text{MSE} (X_1, \dots, X_k, \dots, X_{p-1}) \sim F(1, n - p)$$

Où MSR est la somme des carrés moyens du modèle de régression (« *regression mean square* »), MSE la somme des carrés moyens des erreurs (« *error mean square* »), n le nombre d'observations et p le nombre de paramètres (incluant l'ordonnée à l'origine) .

6) Test du rapport de vraisemblance

Pour les cas où la régression s'estime par la vraisemblance (comme c'est le cas pour la régression logistique ou *probit*), il est possible de vérifier si l'ajout d'une nouvelle variable à un modèle est significatif à l'aide d'un test de rapport de vraisemblance. On peut ainsi tester $H_0 : \beta_i = 0$ contre $H_1 : \beta_i \neq 0$ pour un i donné à l'aide de la statistique :

$$G = D(\text{sous-modèle}) - D(\text{modèle complet}) \sim \chi^2(1)$$

Où $D(x)$ représente la statistique de déviance d'un modèle donné.

7) Test de Student

Le test de Student (ou test-t) est un outil statistique dont l'objectif est la comparaison, en moyenne, de deux groupes indépendants pour une certaine variable continue (distribuée normalement). Il permet de tester $H_0 : \bar{X} = \bar{Y}$ contre $H_1 : \bar{X} \neq \bar{Y}$ à l'aide de la statistique suivante :

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\text{Var}(\bar{X} - \bar{Y})}} \sim t(n-1)$$

De plus, comme pour les analyses de variance, le test d'homogénéité des variances de Levene sera utilisé pour les tests de Student. Si on doit rejeter l'hypothèse des variances homogènes, la correction de Welch sera utilisée.