

Pondération et analyse de la non-réponse des données du volet 2023¹

Karine Dion et Luc Belleau
Direction de la méthodologie
Institut de la statistique du Québec
10 avril 2024

Le présent rapport a pour but de décrire la méthode de pondération utilisée pour les données de la première édition de l'Étude longitudinale du développement des enfants du Québec (ELDEQ 1) au volet 2023². Dans ce rapport, le lectorat est invité à consulter les sections 1 à 4 pour connaître les étapes ayant mené au calcul de la pondération et la section 5 pour savoir comment utiliser la pondération. Plus précisément,

- la section 1 propose une description des étapes ayant mené au choix de la stratégie de pondération;
- la section 2 présente l'analyse de la non-réponse totale ayant donné lieu à la création de la pondération;
- la section 3 présente l'analyse de la non-réponse partielle;
- la section 4 détaille le taux de réponse pondéré obtenu; et
- la section 5 renseigne l'utilisateur sur le fichier de pondération ainsi que sur la façon d'utiliser les poids échantillonnaires dans les analyses statistiques. Elle comporte également quelques mises en garde concernant l'utilisation de ces poids.

1. Stratégie de pondération

1.1 Admissibilité à l'enquête au volet 2023

Parmi les 2 120 répondants au volet initial, on compte 31 familles ayant quitté définitivement le Québec et 6 familles dont l'enfant cible est décédé entre les volets 1998 et 2023. Les enfants cibles de ces familles, considérés comme inadmissibles à l'enquête, ne sont plus visés par l'enquête en ce sens qu'ils ne font plus partie de la population sur laquelle porte l'inférence. La population visée est par conséquent composée des enfants survivants qui sont demeurés au Québec entre les volets 1998 et 2023 ou qui ont quitté la province temporairement.

¹ Il s'agit d'un document de travail qui n'a pas fait l'objet d'une révision linguistique.

² Les rapports de pondération des volets antérieurs sont disponibles sur le site de l'ELDEQ 1 (www.iesuisieserai.stat.gouv.qc.ca), sous l'onglet « Documentation technique ».

Les enfants cibles des familles n'ayant pu être retracés, ayant refusé de répondre ou ayant été dans l'impossibilité de le faire sont tous considérés comme admissibles à l'enquête. Bien que l'on sache que, parmi les familles n'ayant pu être retracées, certaines pourraient avoir déménagé définitivement hors du Québec, leur nombre est trop petit pour que l'on en tienne compte dans le calcul de la pondération. Sur cette base, l'échantillon admissible à l'enquête au volet 2023 est composé de 2 083 jeunes³. Leur répartition, selon la réponse à l'enquête à chacun des volets de 1998 à 2023, est présentée au tableau I.

Tableau I Nombre de répondants⁴ aux volets de 1998 à 2023.

Volets 1998 à 2002	Volets 2003 à 2021	Volet 2023	Nombre de répondants
Répondants aux 5 volets	Répondants à 12 volets	Oui	706
		Non	28
	Répondants à 8, 9, 10 ou 11 volets	Oui	459
		Non	164
	Répondants à 4, 5, 6 ou 7 volets	Oui	122
		Non	140
	Répondants à 1, 2, 3 volets	Oui	12
		Non	187
	Répondant à 0 volet	Oui	0
		Non	64
Répondants à 2, 3 ou 4 volets	Répondants à 12 volets	Oui	5
		Non	0
	Répondants à 8, 9, 10 ou 11 volets	Oui	29
		Non	6
	Répondants à 4, 5, 6 ou 7 volets	Oui	13
		Non	17
	Répondants à 1, 2 ou 3 volets	Oui	0
		Non	21
	Répondant à 0 volet	Oui	0
		Non	58
Répondants à 1 volet	Répondant à 0 volet	Oui	0
		Non	52
Nombre total de jeunes admissibles à l'enquête au volet 2023			2 083

Note : Il n'y a pas eu de collecte de données en 2007, 2009, 2012, 2014, 2016, 2018, 2020 et 2022.

³ À partir du volet 2013, les termes « jeune » et « enfant » sont tous deux utilisés dans la documentation technique et dans les bases de données de l'ELDEQ 1 pour désigner l'enfant cible.

⁴ Aux volets 1998 à 2002, les répondants ont tous rempli le QIRI; de 2003 à 2015, les répondants ont rempli au moins un instrument de collecte au volet concerné; à partir de 2017, seul le questionnaire en ligne du jeune était à remplir.

1.2 Répondants au volet 2023

La pondération est un outil qui permet d'inférer à la population visée les estimations produites à partir des données fournies par les répondants. Cette pondération est requise puisque, en plus d'avoir des probabilités de sélection initiales variables, les répondants diffèrent en général des non-répondants. Ainsi, pour une analyse donnée, toute la non-réponse observée devrait idéalement être traitée, c'est-à-dire que la pondération utilisée pour cette analyse devrait avoir fait l'objet d'un ajustement pour compenser toute perte de répondants.

Depuis 2017, il n'y a qu'un seul instrument de collecte, soit le questionnaire en ligne au jeune (QELJ). Il a été décidé de créer un poids qui refléterait donc le fait d'avoir complété cet instrument de collecte au volet 2023. Rappelons que pour les volets 2006 à 2015, la pondération devait refléter le fait d'avoir complété au moins un instrument de collecte, tandis que pour les volets 1998 à 2005, ce fut un poids spécifique au Questionnaire informatisé rempli par l'intervieweur (QIRI) qui a été construit. La stratégie de collecte est donc complètement axée sur le jeune au volet 2023. Le tableau II présente le nombre de répondants par volet.

Tableau II Nombre de répondants⁵ à certains volets de 1998 à 2023

Nb de répondants	volet 1998	volet 1999	volet 2000	volet 2001	volet 2002	volet 2003	volet 2004	volet 2005	volet 2006	volet 2008	volet 2010	volet 2011	volet 2013	volet 2015	volet 2017	volet 2019	volet 2021	volet 2023
pour un volet donné	2 120	2 045	1 997	1 950	1 944	1 775	1 529	1 537	1 526	1 402	1 415	1 312	1 466	1 348	1 400	1 323	1 368	1 346
longitudinaux (pour un volet donné et ses précédents)	2 120	2 045	1 985	1 924	1 894	1 723	1 462	1 355	1 286	1 186	1 121	1 035	991	917	839	760	734	706

1.3 Étapes pour l'ajustement pour la non-réponse

Rappelons qu'en 2018, un nouveau poids de base, soit un « poids enfant », a été créé afin de répondre aux besoins futurs de la phase 4 de l'ELDEQ 1. Pour ce faire, nous avons attribué un poids de base à tous les jeunes ayant répondu à au moins un volet entre 2011 et 2018, incluant le volet spécial⁶ sur la santé mentale. Ainsi, nous avons un poids de base qui servira comme volet de référence pour les prochaines pondérations transversales des questionnaires sur le jeune. Cette façon de faire permet de considérer des variables mesurées directement auprès du jeune, tout en considérant également les informations stables de 2002 qui ont été utilisées pour les volets précédents dans le modèle de la participation (ou non).

⁵ Aux volets 1998 à 2002, les répondants ont tous rempli le QIRI; de 2003 à 2015, les répondants ont rempli au moins un instrument de collecte au volet concerné; à partir de 2017, seul le questionnaire en ligne du jeune était à remplir.

⁶ Un volet spécial est une collecte de renseignements, généralement issue d'un projet de chercheur, auprès d'une partie ou de l'ensemble de l'échantillon, non prévue dans les volets réguliers. En 2018, un volet spécial sur la santé mentale fut mené auprès des jeunes de l'échantillon.

Un poids transversal général a ainsi été créé pour les 1 346 répondants au QELJ du volet 2023. La méthode de pondération sera décrite plus en détail à la section 2.

La modélisation de la non-réponse au volet 2023 comporte cinq étapes :

1. Ajustement de l'inverse des probabilités de sélection pour la non-réponse à l'enquête au volet 1998 → pondération QIRI du volet 1998.
2. Ajustement des poids QIRI du volet 1998 pour la non-réponse à l'enquête au volet 2000 parmi les répondants du volet 1998 toujours admissible à l'enquête au volet 2023 → pondération QIRI du volet 2000.
3. Ajustement des poids transversaux du volet 2000 pour la non-réponse à l'enquête au volet 2002 parmi les répondants du volet 2000 toujours admissibles à l'enquête au volet 2023 → pondération QIRI du volet 2002.
4. Ajustement des poids transversaux du volet 2002 pour la non-réponse à tous les volets entre 2011 et 2018, incluant le volet spécial sur la santé mentale, parmi les répondants du volet 2002 toujours admissibles à l'enquête au volet 2023 → pondération « enfant ».
5. Ajustement des poids « enfant » pour la non-réponse à l'enquête au volet 2023 parmi les répondants admissibles à l'enquête au volet 2023 → pondération générale transversale du volet 2023.

2. Analyse de la non-réponse totale

2.1 Pondération transversale des données du volet 2023

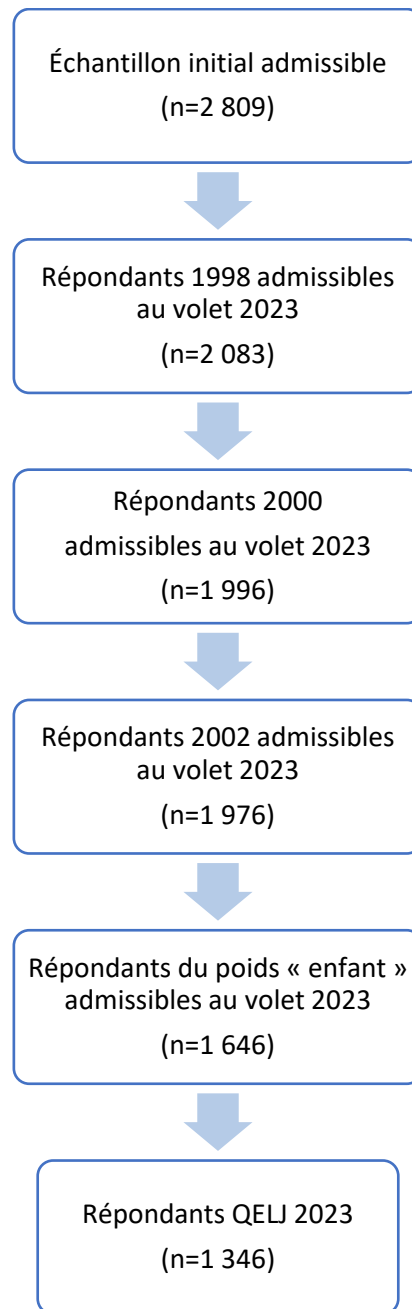
2.1.1 Démarche générale d'analyse

La création de pondérations ajustées pour la non-réponse est basée sur la formation de classes de pondération. C'est la méthode du score qui a été utilisée pour créer les classes de pondération (pour plus de détails sur la méthode, voir Haziza et Beaumont, 2007 et Eltinge et Yansaneh, 1997). Cette méthode crée des groupes homogènes selon la valeur d'un score, celui-ci étant issu d'un modèle de régression logistique. C'est la réponse au volet 2023 qui a été analysée à l'aide de ce modèle et la probabilité estimée de réponse constitue le score. Par la suite, la création des groupes s'effectue à l'aide d'une méthode de classification. Enfin, pour un jeune donné, l'ajustement de la pondération consiste à diviser le poids de référence par la proportion pondérée de jeunes répondants observée au sein du groupe auquel il appartient. Pour plus de détails concernant cette démarche d'analyse, consulter l'annexe A.

Pour tenir compte de la non-réponse au volet 2023, un ajustement a été fait à partir du poids de base « enfant » (voir la section 1.3). Cet ajustement est requis puisque les répondants au volet 2023 présentent des caractéristiques différentes des non-répondants. On minimise ainsi les risques de biais dus à la non-réponse dans les estimations qui seront produites.

La hiérarchie des répondants admissibles au volet 2023 est illustrée au schéma I.

Schéma I Hiérarchie des répondants aux différents instruments de collecte, admissibles au volet 2023



Note : Aux volets 2000 et 2002, le nombre indiqué inclut les pseudo-répondants. Aux fins de la pondération transversale du volet pour lequel un poids transversal devait être construit, les enfants dont la famille a participé à un volet donné mais qui n'étaient pas participants au volet 2002 ont été considérés comme « répondants » (pseudo-répondants) au volet 2002. De cette manière, il est possible de recalculer un nouveau poids pour l'ensemble des participants au volet 2002, incluant ces derniers.

2.1.2 Variables considérées et résultats

Les variables considérées pour la modélisation sont principalement de nature socioéconomique. Elles portent sur la mère de l'enfant cible ou sur sa famille et sont tirées du QIRI du volet 2002. Des variables dites longitudinales ont également été étudiées en créant un indice à partir de la même mesure prise de 1998 à 2002. Ces variables sont :

- le revenu du ménage (revenu faible à au moins un des 5 volets (soit moins de 10 000 \$), versus autres; revenu faible à au moins un des 5 volets (soit moins de 15 000 \$), versus autres);
- le type de famille (monoparentalité à au moins un volet versus autres; monoparentalité ou nouveau conjoint à au moins un volet versus autres);
- la présence du père biologique (le père biologique est absent du ménage à au moins un volet versus autres);
- le niveau de suffisance du revenu du ménage (insuffisance du revenu à au moins un volet versus autres);
- le travail de la mère au cours des douze derniers mois (n'a pas travaillé au cours des douze mois précédant l'enquête pour plus d'un volet versus autres);
- la principale source de revenu du ménage (aide sociale comme principale source de revenus à aucun volet, à 1 ou 2 volets, à 3 volets ou plus); et
- la situation en emploi des parents (aucun parent en emploi à aucun volet, à 1 ou 2 volets, à 3 volets ou plus).

Compte tenu de la disponibilité de microdonnées administratives⁷ provenant du ministère de l'Éducation du Québec (MEQ) et le ministère de l'Enseignement supérieur (MES), des variables supplémentaires ont été considérées pour la pondération du volet 2023. Ces variables sont exclusivement reliées à la diplomation. Il est important de mentionner que la majorité des jeunes de ce volet avaient terminé leurs secondaires en 2015 ou 2016, ainsi il y avait très peu d'informations disponibles provenant du MEQ et du MES. Pour certaines de ces variables, une imputation avec les microdonnées d'années antérieures était possible, telles que les variables sur la diplomation ou l'obtention du DES.

Également, des variables du questionnaire au jeune de 2013, 2015, 2017, 2019 et 2021 ont été considérées pour la pondération du volet 2023. Ces variables sont principalement liées à la motivation scolaire et aux occupations du jeune : a déjà doublé une année scolaire, l'obtention d'un diplôme, s'il travaille ou non, etc. Des variables dites longitudinales ont également été étudiées en créant un indice à partir de la même mesure prise avec le questionnaire du jeune pendant cette période. Par exemple :

- le jeune habitant le foyer familial (0 volet, à 1 volet, à 2 volets, à 3 ou 4 volets);
- la fréquentation scolaire (0 ou 1 volet, 2 ou 3 volets, 4 ou 5 volets) ;
- la participation du jeune (0 ou 1 volet, 2 ou 3 volets, 4 ou 5 volets);
- le jeune ayant travaillé le dernier mois (0 ou 1 volet, 2 ou 3 volets, 4 ou 5 volets).

⁷ Microdonnées provenant du fichier daté du mois d'avril 2019 pour les jeunes inscrits à l'école au 30 septembre 2018.

Ces nouvelles variables permettent d’avoir une information directement mesurée auprès du jeune, et pouvant être liées ou non à la participation de celui-ci au volet 2023.

Parmi l’ensemble des variables considérées, voici celles qui ont été retenues pour le modèle final de régression logistique :

- Le sexe du jeune (sexe)
- Obtention du diplôme d’études secondaires (DES)
- Nombre de volets où le jeune a habité au foyer familial au cours des quatre derniers volets, modalités regroupées pour former deux catégories (residFam_2c)
- Le jeune a travaillé le dernier mois pendant au moins un volet au cours des cinq derniers volets (trav_2422201816)

Une méthode de classification hiérarchique de Ward a permis de regrouper les probabilités estimées en cinq classes de pondération. Le tableau III présente les proportions pondérées de répondants au volet 2023 parmi les répondants du poids « enfant » pour ces cinq groupes. De plus, il présente le nombre de répondants, parmi les 1 346 répondants, à qui la proportion pondérée sera appliquée en guise de correction de la non-réponse. Par exemple : il y a 133 répondants au volet 2023 dont le poids de référence sera ajusté par l’inverse de la proportion pondérée de la quatrième classe de pondération, qui est de 67,8 %.

Tableau III Proportions pondérées de répondants et nombre de répondants par classe de pondération (transversal)

Classe de pondération	Proportions pondérées de répondants au volet 2023 (en %)	Nombre de répondants
1	85,3	508
2	20,3	7
3	93,6	608
4	67,8	133
5	58,5	90

Au sein des différentes classes d’ajustement de la pondération, la proportion de répondants varie de 20,3 % à 93,6 % (relativement à une proportion globale de 81,8 %). La proportion la plus faible est observée dans une classe où l’on retrouve, en proportion, un plus grand nombre des jeunes de sexe masculin, un plus grand nombre de jeunes n’ayant pas obtenu leur diplôme d’études secondaires, un plus grand nombre de jeunes n’ayant pas habité au foyer familial entre 2015 et 2021 (aucun volet sur les quatre) et un plus grand nombre de jeunes n’ayant pas travaillé au moins un volet entre 2013 et 2021 (0 volet parmi les cinq).

Comme au volet précédent, il n’y a pas eu d’ajustement de la pondération à l’aide de données administratives. En effet, les statistiques agrégées habituellement utilisées pour effectuer cet ajustement aux volets précédents provenaient du ministère de l’Éducation et de l’Enseignement supérieur (MEES). En 2023, les jeunes ont environ 25 ans et la majorité d’entre eux, parmi ceux qui fréquentent un établissement scolaire, vont au collégial ou à l’université. Il serait intéressant de réfléchir à des sources alternatives de données administratives afin de continuer à évaluer s’il faut faire un redressement des poids d’enquête afin que les estimations s’ajustent à des totaux connus (Lavallée et Durning, 1993).

3. Analyse de la non-réponse partielle

3.1 Introduction

La pondération transversale a été produite afin de tenir compte de la non-réponse totale, mais celle-ci n'a pas été ajustée pour la non-réponse partielle à une question. Une non-réponse partielle importante peut entraîner certains biais dans les estimations, au même titre que la non-réponse totale. Cette section relève les questions (ou items) présentant un taux de non-réponse partielle supérieur à 5 %. Le taux de non-réponse partielle est défini ici comme le rapport entre le nombre pondéré de non-répondants à une question d'un instrument donnée et le nombre pondéré de répondants à cet instrument admissibles à y répondre.

À noter qu'avant de choisir une variable pour une analyse, il importe de considérer non seulement l'ampleur de la non-réponse partielle pour celle-ci, mais également la non-réponse totale à l'instrument considéré, s'il y a lieu⁸. Il faut également considérer la non-réponse partielle cumulative si l'analyse comprend plusieurs variables simultanément.

3.2 Variables présentant une non-réponse partielle supérieure à 5%

Le tableau IV présente les variables d'analyse pour lesquelles le taux de non-réponse partielle est supérieur à 5 %. Pour chacune des variables identifiées, le taux pondéré est présenté, ainsi que le nombre de personnes (non pondéré) qui avaient à répondre à la question correspondante⁹. Pour le seul instrument de collecte du volet 2023 (QELJ), ce sont les questions qui touchent les études et le travail qui sont davantage entachés de non-réponse partielle.

Tableau IV Questions ayant une non-réponse partielle, questionnaire en ligne du jeune (QELJ)

Fichier QELJ2601 (n=1 346)	Description	Taux de non-rép. partielle pondéré (%)
ZSCNQ04A_AN ZSCNQ04A_MO	Date obtention diplôme d'études secondaires (DES) le plus récent?	16,0 (n=23)
ZTVNQ31CBA	Depuis août 2021, troisième emploi, depuis combien de temps : AN(s)	10,5 (n=13)

⁸ Il est prévu que des pondérations « sur mesure » soient produites pour des analyses spécifiques lorsque la non-réponse cumulative est importante.

⁹ Lorsque le nombre de personnes qui avaient à répondre à une question est inférieur à 100, seules les variables ayant un taux de non-réponse partielle supérieur à 10 % sont présentées. Par contre, lorsque le nombre de personnes qui avaient à répondre à une question est inférieur à 10, même si le taux de non-réponse partielle est supérieur à 10 %, les résultats ne sont pas présentés.

4. Taux de réponse

Le tableau V présente le taux de réponse pondéré transversal obtenu au volet 2023. Ce taux est obtenu en multipliant les taux obtenus aux différentes étapes de pondération, selon le cas. Mentionnons que nous avons obtenu un taux de réponse pondéré transversal légèrement moins élevé au volet 2023, soit 47,7% comparativement à 48,2% au volet 2021.

Tableau V Taux de réponse pondéré transversal au volet 2023

Taux de réponse au volet 1998	75,3 % (n=2 809)
Proportion de répondants (incluant les pseudo-répondants) au volet 2000 parmi les répondants au volet 1998 admissibles au volet 2023	95,0 % (n=2 083)
Proportion de répondants au volet 2002 parmi les répondants au volet 2000 admissibles au volet 2023 (incluant les pseudo-répondants)	98,9 % (n=1 996)
Proportion de répondants du poids « enfant » parmi les répondants au volet 2002 admissibles au volet 2023 (incluant les pseudo-répondants)	82,4 % (n=1 976)
Proportion de répondants au volet 2023 parmi les répondants du poids « enfant » admissibles au volet 2023 (incluant les pseudo-répondants)	81,8 % (n=1 646)
Taux de réponse transversal au volet 2023	47,7 %

Note : le chiffre présenté entre parenthèses représente le dénominateur à partir duquel le calcul est effectué.

5. Utilisation de la pondération par les utilisateurs des données du volet 2023

5.1 L'importance de la pondération

La pondération est le résultat du traitement de la non-réponse. Elle permet d'inférer les résultats à la population visée tout en minimisant les biais dans les estimations. Elle doit ainsi être utilisée afin d'inférer l'analyse des données à la population de l'ELDEQ 1.

Le taux de réponse au volet 2023 est de l'ordre de 47,7 % (voir tableau V). Ce taux confirme l'importance du traitement effectué lors de la pondération.

5.2. Tests statistiques

Le fichier POIDS2601 contient la variable de pondération PEGENT26 (poids général transversal du volet 2023). C'est un poids échantillonnal, c'est-à-dire un poids qui a été multiplié par une constante de sorte que la somme des poids soit égale à la taille de l'échantillon. Ce poids doit faire partie de toute analyse des données du volet 2023, comme indiqué à la section 5.1. Des logiciels statistiques, tels que SUDAAN, SAS ou STATA, permettent l'intégration de la pondération dans les différentes procédures offertes. En plus d'intégrer la pondération afin d'obtenir le bon calcul de la variabilité des estimations, le plan de sondage peut aussi être pris en compte lors des analyses. Le logiciel SUDAAN le permet, ainsi que

certaines procédures du logiciel SAS.

Si les logiciels utilisés ne tiennent pas compte du plan de sondage complexe, le poids PEGENT26 peut être utilisé pour faire des tests approximatifs.

Afin de pallier le caractère approximatif des tests statistiques réalisés à l'aide de poids échantillonnaires, il est recommandé d'adopter une approche conservatrice en abaissant le seuil théorique des tests. Par exemple, si l'on souhaite faire des tests au seuil théorique de 0,05, on peut choisir de n'interpréter que les résultats significatifs au seuil 0,01. Il serait ainsi possible de conclure, avec un seuil observé de 0,005 obtenu d'un test statistique, que l'hypothèse nulle du test est rejetée au seuil théorique de 0,05 (étant donné que 0,005 est inférieur à 0,01).

Dans le cas particulier de tests du khi-deux sur un tableau de fréquences, l'utilisation des poids échantillonnaires divisés par un effet de plan moyen égal à 1,3 demeure appropriée pour obtenir un test approximatif. Il n'est alors pas nécessaire d'abaisser le seuil des tests. Un résultat pour lequel le seuil observé est près de 0,05 devrait néanmoins être interprété avec nuances.

5.3 Choix de la pondération

Les possibilités d'analyse incluant des données du volet 2023 sont innombrables. Ainsi, en raison de la non-réponse qui varie selon les instruments de collecte et les volets considérés, le choix d'une pondération adéquate nécessite un examen cas par cas. **En précisant la population visée, de même que les instruments et les volets considérés pour l'analyse, l'Institut peut évaluer si une pondération appropriée est disponible. Dans le cas contraire, une pondération sur mesure peut être requise.** Il s'agirait alors pour l'Institut de faire un ajustement de la pondération existante, de manière à minimiser les biais potentiels qui pourraient être induits par une non-réponse non prise en compte.

Le choix d'une pondération appropriée doit tenir compte de deux éléments importants. Il faut d'abord minimiser le risque de biais lorsqu'il y a de la non-réponse (totale ou partielle) non traitée par la pondération. Le second élément dont il faut tenir compte est la diminution de puissance statistique lorsqu'il y a une perte d'unités d'analyse due à l'absence de poids pour ces unités. Dans ce cas, l'unité a répondu à l'ensemble des variables incluses dans l'analyse, mais elle n'a pas de poids pour être retenue. Au moyen d'un ajustement adéquat, une pondération devrait généralement tenir compte de ces deux aspects. Le lecteur est invité à consulter des exemples qui illustrent la démarche à suivre pour évaluer la situation. Ceux-ci se retrouvent dans les rapports de pondération des volets antérieurs.

6. Références bibliographiques

Eltinge, J. L. et Yansaneh, I.S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey, *Techniques d'enquête*, vol. 23, no. 1, pages 33-40.

Haziza, D. et Beaumont, J.-F. (2007). On the construction of imputation classes in surveys. *International Statistical Review*, 75, 25-43

Lavallée, P. et Durning. A. (1993). Estimateur jackknife de la variance pour l'estimation par calage sur marges, article extrait de la présentation faite dans le cadre du congrès de l'Association canadienne français pour l'avancement des sciences (ACFAS) en 1993.

ANNEXE A

Les étapes de la création d'une pondération générale

Voici la description de la séquence des étapes de création de la pondération transversale pour les participants au volet 2023.

Étape 1 :

Analyses bivariées pour réduire le nombre de variables considérées pour la modélisation (environ 90 variables). Les variables ayant les seuils observés les plus faibles sont conservées.

Étape 2 :

Modélisation préliminaire avec la régression logistique afin d'identifier les variables retenues à l'étape 1 qui présentent un problème de multicolinéarité. Plusieurs essais de modélisation ont été effectués afin de ne retenir qu'un sous-ensemble de variables. Celles-ci ne présentent pas de problème de multicolinéarité entre elles, ni de taux de non-réponse partielle élevée, ni de seuils observés très élevés.

Étape 3 :

Estimation de la taille du modèle par la minimisation du critère d'Akaike (à titre indicatif).

Étape 4 :

Détermination d'un modèle de régression logistique avec SUDAAN pour prédire la probabilité de réponse, en excluant les jeunes pour lesquels il y a présence de non-réponse partielle combinée

Étape 5 :

Création d'une catégorie de valeurs manquantes pour les variables du modèle retenu à l'étape 4. À noter qu'à ce volet-ci, les valeurs manquantes de certaines variables ont été imputées à une catégorie aléatoirement de façon à conserver les proportions observées. La validation de ce modèle est effectuée et un modèle final est retenu.

Étape 6 :

Création des classes de pondération effectuée à l'aide de la méthode du score, ce dernier étant la probabilité de réponse estimée à l'aide du modèle. La détermination du nombre de classes et le regroupement sont effectués à l'aide d'une méthode de classification hiérarchique ou non hiérarchique. Ceci étant fait, les poids de base sont ajustés selon la proportion pondérée de répondants par classe. Ainsi, la pondération 2023 est constituée.