

**ÉTUDE LONGITUDINALE DU DÉVELOPPEMENT
DES ENFANTS DU QUÉBEC – (ELDEQ 1998-2017)**



**Guide de l'utilisateur de la
banque de données du volet 2017**

Jeunes âgés d'environ 19 ans

**Direction des enquêtes longitudinales et sociales
Institut de la statistique du Québec**

Septembre 2018

La documentation qui suit présente l'information nécessaire pour utiliser le plus efficacement possible les données de l'*Étude longitudinale du développement des enfants du Québec* (ELDEQ) du **volet 2017 (E20)**. Ce document, divisé en quatre sections, présente d'abord un bref aperçu de la collecte. La deuxième section traite, pour sa part, de la convention de désignation des variables présentes dans la banque de données. La structure et le contenu des fichiers sont abordés à la section 3 et, finalement, les remarques et mises en garde sont présentées à la section 4.

1. Collecte de données : Un aperçu

Le volet 2017 de l'*Étude longitudinale du développement des enfants du Québec* (ELDEQ) a été réalisé auprès de 1400 jeunes¹ du milieu du mois de février au début juin 2017. Ainsi, au moment de cette collecte, les jeunes² avaient environ 19 ans.

La banque de données du volet 2017 inclut les informations tirées du Questionnaire en ligne au jeune (QELJ)³, qui a été rempli en ligne et sans l'aide d'une intervieweuse. Un seul instrument de mesure a été administré en 2017.

2. Convention de désignation des variables

Pour désigner et retrouver facilement chacune des variables des différents fichiers de données, nous avons conçu une nomenclature qui s'appuie en grande partie sur celle utilisée pour l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) de Statistique Canada que nous avons adaptée pour l'ELDEQ. Généralement, toutes les variables contenues dans la banque de données ont au plus huit caractères.

2.1. Nomenclature générale des variables

Les noms de variables se présentent ainsi :

T S E C Q ppx, où :

T : la première position se rapporte au volet de l'étude. Ainsi,

- « A » correspond au volet 1998 (5 mois) ;
- « B » correspond au volet 1999 (17 mois) ;
- « C » correspond au volet 2000 (29 mois) ;
- « D » correspond au volet 2001 (41 mois) ;
- « E » correspond au volet 2002 (44 à 56 mois) ;
- « F » correspond au volet 2003 (environ 5 ans) ;
- « G » correspond au volet 2004 (environ 6 ans) ;
- « H » correspond au volet 2005 (environ 7 ans) ;
- « I » correspond au volet 2006 (environ 8 ans) ;

1 Pour obtenir plus de détails concernant la population et l'échantillon initial, consultez le numéro 1 des volumes 1 et 2 de la collection ELDEQ (disponible sur le site web de l'ELDEQ à l'adresse suivante : <http://www.iesuisieserai.stat.gouv.qc.ca>, sous l'onglet « Publications »).

2 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 pour désigner l'enfant-cible.

3 Tous les instruments de collecte sont disponibles sur le site de l'ELDEQ (www.iesuisieserai.stat.gouv.qc.ca) sous l'onglet « Outils de collecte ».

- « K » correspond au volet 2008 (environ 10 ans) ;
- « M » correspond au volet 2010 (environ 12 ans) ;
- « N » correspond au volet 2011 (environ 13 ans) ;
- « P » correspond au volet 2013 (environ 15 ans) ;
- « R » correspond au volet 2015 (environ 17 ans) ;
- « T » correspond au volet 2017 (environ 19 ans).

SE : les deuxième et troisième positions se rapportent à une section du Questionnaire en ligne du jeune (QELJ) (voir le tableau 2).

C : correspond généralement à la personne à laquelle la variable se rapporte. Ainsi,

- « N » signifie que la variable se rapporte à l'enfant **et qu'elle est tirée du questionnaire informatisé ou en ligne au jeune (QIE et QELJ)**;
- « E » signifie que la variable se rapporte à l'enfant;
- « R » signifie que la variable se rapporte à la fratrie;
- « M » signifie que la variable se rapporte à la mère/conjointe présente;
- « J » signifie que la variable se rapporte au père/conjoint présent;
- « F » signifie que la variable se rapporte au ménage;
- « A » signifie que la variable se rapporte au père biologique absent du ménage;
- « I » signifie que la variable se rapporte à l'enseignant/e de l'enfant cible;
- « 1 » signifie que la variable se rapporte au premier enseignant de l'enfant (volet 2011);
- « 2 » signifie que la variable se rapporte au deuxième enseignant de l'enfant (volet 2011).

Q : la cinquième position se rapporte au genre de variable. Ainsi,

- « Q » signifie que la variable se rapporte à une question qui a été posée directement dans l'un des questionnaires;
- « S » signifie que la variable représente une note calculée pour une échelle utilisée dans le questionnaire;
- « D » signifie que la variable a été calculée à partir d'autres questions qui ont été posées dans le questionnaire;
- « T » signifie que la note calculée pour une échelle utilisée dans le questionnaire a été standardisée, c'est-à-dire ramenée à une valeur se situant entre 0 et 10;
- « L » signifie que la note calculée pour une échelle à un volet donné est basée uniquement sur les items longitudinaux, c'est-à-dire ceux communs à tous les volets;
- « M » signifie qu'il s'agit d'une information de type date (en mois);
- « A » signifie qu'il s'agit d'une information de type date (en année).
- « P » a plusieurs significations. Il importe de vérifier l'étiquette de la variable pour déterminer avec précision le genre de variable dont il est question.

ppx : les trois dernières positions se rapportent à l'identification de la question. De façon générale, **pp** est un numéro séquentiel attribué à la variable et **x** est un indicateur alphabétique séquentiel pour une série de variables similaires.

ATTENTION : Dans certains cas, les trois derniers caractères de la variable ne sont pas suffisants pour identifier correctement une question (p. ex. : questions permettant plus d'un choix de réponse). Dans de tels cas, les quatre derniers caractères ont alors été utilisés afin d'identifier la variable (le plus souvent en supprimant la lettre «Q»).

3. La structure et le contenu des fichiers

L'ensemble des questions posées durant la collecte se trouvent dans la base de données, exception faite des variables permettant d'identifier un ou plusieurs membres du ménage. La grande majorité des variables ont fait l'objet d'une validation transversale, alors que certaines ont également été validées de façon longitudinale⁴.

Un seul instrument de mesure a été utilisé au volet 2017 et il s'adressait seulement au jeune (QELJ). Afin de faciliter l'utilisation de la banque de données, le tableau 1 présente la structure des fichiers inclus dans la banque de données du volet 2017. Le tableau 2 décrit les sections du questionnaire QELJ.

Tableau 1. Fichiers inclus dans la banque de données du volet 2017 de l'ELDEQ (E20)

Nom du fichier	Nombre de lignes contenues dans le fichier	Données contenues dans le fichier
QELJ2001	1 400	Données colligées à l'aide du QELJ et concernant l'enfant-cible.
INDI2011	1 400	Variables dérivées construites par les partenaires de l'ELDEQ à partir des données issues du QELJ. <i>Elles sont présentées en détail dans la partie B du document des « Variables dérivées de l'ELDEQ – 1998-2017⁵ ».</i>
POIDS2001	1 400	Variable de poids échantillonnaux généraux transversaux relative aux données recueillies sur le ménage (PEGENT20 ⁶). <i>Le choix d'une pondération appropriée dépend des variables considérées dans le modèle statistique. Pour en connaître davantage sur la création et l'utilisation optimale de la variable de pondération, référez-vous au document « Pondération des données du volet 2017⁷ ».</i>

4 Pour un aperçu des méthodes de validation utilisées, consulter le numéro 12 (section 2) des volumes 1 et 2 de la collection ELDEQ disponible sur le site de l'ELDEQ (www.iesuisieserai.stat.gouv.qc.ca) sous l'onglet « Publications ».

5 Le document « Variables dérivées de l'ELDEQ – 1998-2017, Partie B » est disponible sur le site de l'ELDEQ, sous l'onglet « Documentation technique ».

6 Il est à noter que, comme au volet 2015, une seule pondération a été calculée en prévision de produire davantage de pondérations sur mesure aux diverses analyses. N'hésitez pas à communiquer avec l'équipe de l'ELDEQ si vous désirez obtenir de telles pondérations.

7 Le document « Pondération et analyse de la non-réponse des données du volet 2017 » est disponible sur le site de l'ELDEQ, sous l'onglet « Documentation technique ».

Tableau 2. Composition du fichier QELJ de la banque de données du volet 2017 de l'ELDEQ, jeunes d'environ 19 ans

Questionnaires et sections	Lettres indiquant la section du questionnaire informatisé ¹
Structure et mouvement familiaux	MF
Relations interpersonnelles	RI
Choix de carrière	CC
Parcours et engagement scolaires	SC
Parcours socioprofessionnels	TV
Situation financière	SI
Santé et bien-être	HL
Habitudes de vie	HD
Activités physiques et de loisirs	AP
Sommeil	SO
Contexte d'administration	QJ

1. Ces lettres se retrouvent à la 2^e et 3^e position du nom des variables.
Source : Institut de la statistique du Québec, ELDEQ 1998-2017.

Dans une optique longitudinale, il est conseillé de se référer aux questionnaires⁸ de chacun des volets pour s'assurer de la comparabilité des questions dans le temps. En effet, certaines questions ou certains choix de réponse peuvent avoir été modifiés, inversés ou supprimés d'un volet à l'autre.

Comme les volets précédents, chacun des fichiers contient une variable nommée IDME (numéro de ménage). Cette variable doit absolument être utilisée lors de la fusion de deux fichiers ou plus. Tous les fichiers sont constitués d'une ligne par sujet.

Finalement, tous les fichiers ont été développés avec le logiciel SAS 9.3 sur LINUX. Un fichier de formats SAS est également fourni. Le fichier de format permet d'obtenir le catalogue des formats sur n'importe quelle plateforme (par exemple sur WINDOWS ou LINUX). La commande suivante peut être utilisée pour créer le catalogue des formats en SAS :

```
LIBNAME t20 "le chemin du répertoire où se trouve vos fichiers";
PROC FORMAT cntlin=t20.sas_fmts library =t20.formats; RUN;
PROC CATALOG catalog=t20.formats; copy out=work.formats; RUN; QUIT;
```

Une version SPSS WINDOWS (.sav) du fichier est également fournie.

8 Les questionnaires de l'étude sont disponibles sur le site de l'ELDEQ, sous l'onglet « Outils de collecte ».

4. Remarques et mises en garde

4.1. Volet 2017

Mode de collecte – Fichier QELJ2001

La collecte de données a été réalisée en multimode (web, téléphonique ou les deux⁹). Le web était privilégié mais le jeune avait la possibilité de répondre aux questions au téléphone. Une variable (TQJNMODE) est disponible pour identifier le mode de collecte. Certaines questions pourraient présenter un biais de désirabilité sociale. Des analyses supplémentaires seront nécessaires dans certains cas.

Age du jeune et mois de passation – Fichier QELJ2001

La variable AGE20 identifie l'âge en années du jeune lors de la passation du questionnaire. Une variable qui identifie le mois de passation du questionnaire est aussi présente dans le fichier.

Indice de masse corporelle – Fichier QELJ2001

L'indice de masse corporelle a été dérivé à partir du poids et de la taille déclarés par le jeune. La classification standard internationale chez les adultes a été utilisée. Deux variables sont donc disponibles dans le fichier QELJ2001 (TQJNIMCA et TQJNIMCAC).

Nombre d'enfants – Fichier QELJ2001

En raison des petits effectifs, toutes les catégories de la variable THLNQ10 ont dû être regroupées en une seule (1 enfant ou plus).

Codification des emplois – Fichier QELJ2001

Une variable TTVND01 est disponible dans le fichier QELJ2001 et présente la codification de l'emploi principal du jeune. Cette variable est basée sur la classification nationale des professions (CNP-2016) en 10 catégories. En raison de petits effectifs, les catégories « Gestion » et « Sciences naturelles et appliquées et domaines apparentés » ont été regroupées avec la catégorie « Indéterminé ».

Nombre emplois – Fichier QELJ2001

TRVNQ04 est le nombre d'emplois au cours de dernier mois, TRVNQ34A est le nombre d'emplois depuis août 2015 en excluant le travail principal. On remarque quelques incohérences entre les deux variables. Ces incohérences ont été laissées telles quelles dans le fichier de données.

Variables ayant un taux de non-réponse partielle supérieur à 5 %

Le nombre de variables comportant un taux de non-réponse partielle supérieur à 5 % varie en fonction de l'instrument de collecte et de la section observée. Ces variables doivent être traitées avec prudence. Le document « Pondération et analyse de la non-réponse des données du volet 2017 » analyse la non-réponse partielle pour les variables qui posent problème.

⁹ Certains questionnaires incomplets ont un mode de collecte indéterminé car il n'a pas été possible de vérifier si une partie du questionnaire aurait pu être rempli au téléphone.

4.2. Modifications aux variables des années antérieures

Des corrections ont été apportées à certains fichiers de la banque de données depuis la dernière livraison. Voici la liste exhaustive de ces modifications.

Il est de la responsabilité des utilisateurs d'utiliser les données les plus à jour.

Volet	Fichier	Nom des variables visées par la modification, l'ajout ou la suppression	Commentaires ou justifications
2010 (E13)	QAAM	MQMMQ50 à MQMMQ55	Invalider les réponses pour les pères décédés
	PÈRE	MAGJD02	Correction du libellé
	INDI	MQEED35	Ajouter une nouvelle variable dérivée sur l'activité physique
2011 (E14)	QAAM	NQMMQ55B, NQMMQ56B	Rétablir les réponses du répondant qui avaient été invalidées lorsqu'il y avait plusieurs événements à la question précédente
	INDI	NQEED35	Ajouter une nouvelle variable dérivée sur l'activité physique
2013 (E16)	Enfant	PEEEQ18	Ajouter 4 nouvelles catégories (rendez-vous médical, fatigue, absence non-motivée et décès) aux choix de réponses qui viennent des cas « Autres » de cette question
	INDI	PQEED26A et PQEED26B	Ajouter deux nouvelles variables dérivées sur la consommation problématique d'alcool ou de drogues
		PQEED35	Ajouter une nouvelle variable dérivée sur l'activité physique
	INDIMIA	PMINT09	Supprimer la variable parce que son alpha est faible (0,22)
	QELJ	PAPNQ11B	Corriger le libellé de la variable
		PTRN39BH et PTRN39BM	Corriger le filtre entre les deux variables
		PQJNIMC	Associer la variable à un format
		PAPNQ8A	Corriger le format de la variable
	QAAM	PQMMQ43B, PQMMQ44B	Rétablir les réponses du répondant qui avaient été invalidées lorsqu'il y avait plusieurs événements à la question précédente
		PQMMQ45 à PQMMQ49B	Invalider les réponses pour les pères décédés

Volet	Fichier	Nom des variables visées par la modification, l'ajout ou la suppression	Commentaires ou justifications
2015 (E18)	MÈRE	RCNMQ1	Cette variable a été déplacée dans le fichier PCM
	PÈRE	RCNJQ1	Cette variable a été déplacée dans le fichier PCM
	PCM	RCNFQ1	Ajouter la variable dans le bon fichier
	QELJ	RQJNIMC	Associer la variable à un format
		RAPNQ11B	Corriger le libellé de la variable
		RAPNQ8A	Corriger le format de la variable
	INDI	RQEED26A et RQEED26B	Ajouter deux nouvelles variables dérivées sur la consommation problématique d'alcool ou de drogues
		RQEED35	Ajouter une nouvelle variable dérivée sur l'activité physique
	INDIMIA	RMINT09 et RMINT05D	Supprimer les deux variables parce que les alphas sont faibles (0,22 et 0,30)
	QAAM	Éliminer une ligne	Supprimer la ligne du ménage 5372 car la personne qui a répondu au QAAM n'est pas une figure maternelle du jeune
RQMMQ33B et RQMMQ34B		Rétablir les réponses du répondant qui avaient été invalidées lorsqu'il y avait plusieurs événements à la question précédente	
2011, 2013, 2015	TRAJ_FAM	NC2ED10, NC2ED10A PC2ED10, PC2ED10A RC2ED10, RC2ED10A	Corrections apportées aux variables dérivées suite à des validations faites pour des analyses