

**ÉTUDE LONGITUDINALE DU DÉVELOPPEMENT  
DES ENFANTS DU QUÉBEC – (ELDEQ 1998-2020)**



**Guide de l'utilisateur de la  
banque de données du volet COVID 2020**

*Jeunes âgés d'environ 22 ans*

Document préparé par :

Mai Thanh Tu et Karine Tétreault, Direction principale des statistiques sociales et de santé  
Karine Dion et Luc Belleau, Direction de la méthodologie

**Institut de la statistique du Québec**

**Novembre 2020**

## **Présentation**

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 spécial<sup>1</sup> sur la COVID-19 (volet COVID)**, visant à étudier les perceptions, les expériences et les comportements des jeunes dans les mois qui ont suivi le début de la crise de la COVID-19 au Québec. La collecte a eu lieu durant l'été 2020.

Ce document, divisé en 6 sections, présente d'abord un aperçu de la collecte de données. 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, les remarques et mises en garde sont présentées à la section 4, les variables dérivées à la section 5 et, finalement, la pondération à la section 6.

### **1. Collecte de données : un aperçu**

Le volet 2020 de l'ELDEQ a été réalisé auprès de 1 182 jeunes<sup>2</sup> du début du mois de juillet à la fin août 2020. Ainsi, au moment de cette collecte, les jeunes<sup>3</sup> avaient environ 22 ans.

La banque de données du volet spécial sur la COVID-19 de 2020 inclut les informations tirées du Questionnaire en ligne au jeune (QELJ)<sup>4</sup>, qui a été rempli sans l'aide d'un intervieweur. Un seul instrument de mesure a été administré. Ce volet visait à recueillir des informations sur les perceptions, les expériences de vie et les comportements des jeunes dans les mois qui ont suivi le début de la crise de la COVID-19 au Québec.

### **2. Convention de désignation des variables**

Pour désigner et retrouver facilement chacune des variables du fichier principal de données, nous avons conçu une nomenclature propre à ce volet spécial.

#### **2.1 Nomenclature générale des variables**

Les noms de variables se présentent ainsi :

**COV \_ Q nnx**, où :

**COV** : les trois premières positions identifient le volet spécial sur la COVID-19 (COV).

- 
1. 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 financés par les partenaires. Plus spécifiquement, ce volet spécial a été réalisé auprès de l'ensemble de l'échantillon et a été financé par les partenaires financiers, mais n'était pas prévu initialement à la planification des volets réguliers de l'ELDEQ.
  2. 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.jesuisjeserai.stat.gouv.qc.ca>, sous l'onglet « Publications »).
  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 pour désigner l'enfant cible.
  4. Tous les instruments de collecte sont disponibles sur le site de l'ELDEQ ([www.jesuisjeserai.stat.gouv.qc.ca](http://www.jesuisjeserai.stat.gouv.qc.ca)) sous l'onglet « Outils de collecte ».

- 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 le questionnaire;
  - « 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.

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

**Exemple:** L'échelle construite à partir des variables COV\_Q21, COV\_Q22, COV\_Q23, COV\_Q24, COV\_Q25, COV\_Q26 et COV\_Q27 se nomme COV\_S21\_27.

Le tableau 1 présente le nom et le libellé des variables du fichier de données. Les lignes en grisé sont les variables dérivées à partir d'autres variables du fichier (échelles ou autre). La construction de ses variables dérivées est présentée à la section 5 de ce document. La mention Mx signifie qu'il s'agit d'une variable à choix multiple. Cette variable est donc présente plusieurs fois selon la réponse sélectionnée par le répondant.

**Tableau 1. Liste des variables de la banque de données du volet 2020 de l'ELDEQ (COVID)**

Variable	Libellé
IDME	Numéro de ménage
COV_SEXE	Sexe du jeune
COV_LANG	Langue de l'entrevue
COV_MODE	Mode d'administration du questionnaire
COV_Q1	À la mi-mars 2020, combien de personnes en tout habitaient votre domicile?
COV_Q2_Mx	À la mi-mars 2020, avec qui habitez ( <i>Père ; Mère ; Beau-Parent ; Frère/Sœur ; Parenté ; Conjoint ; Ami/colocataire ; Enfant à charge ; Autre</i> )
COV_Q3	Depuis la mi-mars 2020, habitez-vous avec les mêmes personnes?
COV_Q4	Actuellement, combien de personnes en tout habitent votre domicile?
COV_Q5_Mx	Actuellement, avec qui habitez ( <i>Père ; Mère ; Beau-Parent ; Frère/Sœur ; Parenté ; Conjoint ; Ami/colocataire ; Enfant à charge ; Autre</i> )
COV_Q6	Depuis la mi-mars 2020, avez-vous déménagé?
COV_Q7	Aviez-vous prévu déménager durant cette période?
COV_Q8	En général, à quel point le fait de déménager vous a-t-il causé du stress ou de l'inquiétude?
COV_Q9	Depuis mi-mars 2020, qualité relations entre vous et votre conjoint(e) ou partenaire intime
COV_Q10	Depuis mi-mars 2020, qualité relations entre vous et votre ami(e) proche
COV_Q11	De façon générale, avant la crise liée à la COVID-19, est-ce que votre santé était...?
COV_Q12	De façon générale, actuellement, est-ce que votre santé est...?
COV_Q13_Mx	Un professionnel déjà dit aviez ( <i>Allergies saisonnières ; Asthme ou autres problèmes pulmonaires ; Problèmes cardiaques, rénaux, digestifs ; Trouble immunitaire ; Diabète ou hyperglycémie ; Cancer, Prob. santé émotionnelle ou mentale tels dépression ou anxiété ; Problèmes d'alcool ou de drogues ; Autre ; Aucun problème</i> )
COV_Q14	Avez-vous ressenti des symptômes qui ressemblent à ceux de la COVID-19?

<b>Variable</b>	<b>Libellé</b>
COV_Q15	Avez-vous subi un test de dépistage de la COVID-19?
COV_Q16	Combien de temps dû vous placer en quarantaine ou vous isoler à la maison?
COV_Q17A_Mx	Résultat positif test de dépistage COVID-19 ( <i>Oui, par un proche avec qui vous habitez ; Oui, par un proche avec qui n'habitez pas ; Non</i> )
COV_Q17B_Mx	Apparition symptômes ressemblent COVID-19 ( <i>Oui, par un proche avec qui vous habitez ; Oui, par un proche avec qui n'habitez pas ; Non</i> )
COV_Q17C_Mx	Hospitalisation ( <i>Oui, par un proche avec qui vous habitez ; Oui, par un proche avec qui n'habitez pas ; Non</i> )
COV_Q17D_Mx	Mise en quarantaine ou isolement ( <i>Oui, par un proche avec qui vous habitez ; Oui, par un proche avec qui n'habitez pas ; Non</i> )
COV_Q18A	Inquiétude à propos: la santé d'un de vos proches à risque?
COV_Q18B	Inquiétude à propos: la santé d'un(e) de vos proches n'étant pas à risque?
COV_Q18C	Inquiétude à propos: votre propre santé?
COV_Q19	De façon générale, avant la crise liée à la COVID-19, votre santé mentale était...?
COV_Q20	De façon générale, actuellement, votre santé mentale est...?
COV_Q21	Gêné par un sentiment de nervosité, d'anxiété ou de tension
COV_Q22	Gêné par une incapacité à arrêter de s'inquiéter ou à contrôler ses inquiétudes
COV_Q23	Gêné par une inquiétude excessive à propos de différentes choses
COV_Q24	Gêné par des difficultés à se détendre
COV_Q25	Gêné par une agitation telle qu'il est difficile à tenir en place
COV_Q26	Gêné par une tendance à être facilement contrarié ou irritable
COV_Q27	Gêné par un sentiment de peur comme si quelque chose de terrible risquait de se produire
COV_S21_27	Trouble d'anxiété généralisée
COV_D21_27	Trouble d'anxiété généralisée (4 catégories)
COV_Q28	À quelle fréquence avez-vous ressenti un manque de compagnie?
COV_Q29	À quelle fréquence vous êtes-vous senti laissé de côté?
COV_Q30	À quelle fréquence vous êtes-vous senti isolé des autres?
COV_T28_30	Solitude
COV_Q31_Mx	Entre août 2019 et décembre 2019, suiviez cours ( <i>Oui, en personne ; Oui à distance ; Non</i> )
COV_Q32_Mx	À la mi-mars 2020, suiviez cours ( <i>Oui, en personne ; Oui à distance ; Non</i> )
COV_Q33	À la mi-mars 2020, est-ce que vous étiez aux études...
COV_Q34	Quel type de formation suiviez-vous?
COV_Q35A	En raison crise liée à la COVID-19: cours ont été reportés automne 2020
COV_Q35B	En raison crise liée à la COVID-19: cours en personne ont été transformés en cours en ligne
COV_Q35C	En raison crise liée à la COVID-19: j'ai abandonné une partie ou l'ensemble de mes cours
COV_Q35D	En raison crise liée à la COVID-19: stage en milieu de travail a été reporté ou annulé
COV_Q35E	En raison crise liée à la COVID-19: plus long prévu terminer prog. ou obtenir grade, certificat ou diplôme
COV_Q35F	En raison crise liée à la COVID-19: j'ai décidé d'annuler mon inscription pour cet automne
COV_Q36A	De mi-mars 2020 à fin session hiv. ou année scol.: facil. accéder ressources nécessaires terminer cours en ligne
COV_Q36B	De mi-mars 2020 à fin session hiv. ou année scol.: enseignants fourni l'encadrement pour faire cours en ligne
COV_Q37	À la mi-mars 2020, occupiez-vous au moins un emploi rémunéré ou travaillé à votre compte?

<b>Variable</b>	<b>Libellé</b>
COV_Q38H	À la mi-mars 2020, heures en moyenne par semaine à votre emploi principal? - Heures
COV_Q38M	À la mi-mars 2020, heures en moyenne par semaine à votre emploi principal? - Minutes
COV_Q39	À la mi-mars 2020, à quel endroit se trouvait votre lieu de travail habituel?
COV_Q40A	Après mi-mars 2020, comment crise affecté situation d'emploi? - J'ai perdu mon emploi
COV_Q40B	Après mi-mars 2020, comment crise affecté situation d'emploi? - J'ai fermé mon entreprise
COV_Q40C	Après mi-mars 2020, comment crise affecté situation d'emploi? - Mes heures de travail ont été réduites
COV_Q40D	Après mi-mars 2020, comment crise affecté situation d'emploi? - Affecté autres tâches
COV_Q40E	Après mi-mars 2020, comment crise affecté situation d'emploi? - J'ai dû faire mon travail à domicile
COV_Q40F	Après mi-mars 2020, comment crise affecté situation d'emploi? - J'ai obtenu un nouvel emploi
COV_Q40G	Après mi-mars 2020, comment crise affecté situation d'emploi? - Pas pu commencer nouveau travail
COV_Q41	Actuellement, occupez-vous au moins un emploi rémunéré ou travaillez-vous à votre compte?
COV_Q42H	Actuellement, heures consacrez en moyenne par semaine à votre emploi principal? - heures
COV_Q42M	Actuellement, heures consacrez en moyenne par semaine à votre emploi principal? - minutes
COV_Q43	Est-ce le même emploi principal que vous occupiez à la mi-mars?
COV_Q44A	Depuis quand occupez-vous votre emploi principal actuel? - année
COV_Q44M	Depuis quand occupez-vous votre emploi principal actuel? - mois
COV_Q45_Mx	Dans travail, contact avec personnes infectées COVID-19 ( <i>Oui ; Possiblement ; Non</i> )
COV_Q46	À quelle fréquence en contact avec personnes ayant été infectées par la COVID-19?
COV_Q47	À quel point êtes-vous inquiet d'être infecté par la COVID-19 sur votre lieu de travail?
COV_Q48	Dans contexte COVID-19, satisfait mesures employeur pour protéger employés?
COV_Q49A	Depuis mi-mars 2020, inquiété pas suffisamment de nourriture à cause manque d'argent?
COV_Q49B	Depuis mi-mars 2020, pas suffisamment de nourriture à cause manque d'argent?
COV_Q49C	Depuis mi-mars 2020, pas mangé aliments qualité ou variété désirée à cause manque argent?
COV_D49	Insécurité alimentaire
COV_Q50A	Depuis début crise COVID-19, demandé une diminution ou report paiements loyer ou hypothèque ?
COV_Q50B	Depuis début crise COVID-19, demandé diminution ou report paiements voiture ?
COV_Q50C	Depuis début crise COVID-19, demandé diminution ou report paiements prêt bancaire ?
COV_Q50D	Depuis début crise COVID-19, demandé une diminution ou report paiements carte de crédit ?
COV_Q50E	Avez-vous demandé d'autres diminutions ou reports des paiements?
COV_Q51_Mx	Depuis la mi-mars 2020, demande aide gouvernementale ( <i>Oui, PCU ; Oui, PCUE ; Oui, autre ; Non</i> )
COV_Q52A	Niveau préoccupation: détenir diplôme pas équivalent des personnes pas affectées par COVID-19

<b>Variable</b>	<b>Libellé</b>
COV_Q52B	Niveau préoccupation: pas avoir perspective emploi avenir rapproché
COV_Q52C	Niveau préoccupation: pas avoir suffisamment argent pour besoins fondamentaux
COV_Q53	À quel point optimiste pandémie COVID-19 sera maîtrisée au Québec prochaine année?
COV_Q54H	Avant la mi-mars, combien heures sommeil besoin pour sentir en forme? - heures
COV_Q54M	Avant la mi-mars, combien heures sommeil besoin pour sentir en forme? - minutes
COV_Q55H	Avant la mi-mars, durée totale de votre sommeil de nuit (en moyenne)? - heures
COV_Q55M	Avant la mi-mars, durée totale de votre sommeil de nuit (en moyenne)? - minutes
COV_Q56H	Depuis la mi-mars, durée totale de votre sommeil de nuit (en moyenne) - heures
COV_Q56M	Depuis la mi-mars, durée totale de votre sommeil de nuit (en moyenne) - minutes
COV_Q57A	Au cours dernière semaine: pas eu envie de manger, peu d'appétit
COV_Q57B	Au cours dernière semaine: sentiment pas débarrasser du cafard, même avec aide famille ou amis
COV_Q57C	Au cours dernière semaine: difficulté à concentrer
COV_Q57D	Au cours dernière semaine: senti déprimé
COV_Q57E	Au cours dernière semaine: sentiment que tout demandait un effort
COV_Q57F	Au cours dernière semaine: plein d'espoir face à l'avenir
COV_Q57G	Au cours dernière semaine: sommeil agité
COV_Q57H	Au cours dernière semaine: heureux
COV_Q57I	Au cours dernière semaine: senti seul
COV_Q57J	Au cours dernière semaine: joui de la vie
COV_Q57K	Au cours dernière semaine: pleuré
COV_Q57L	Au cours dernière semaine: sentiment que les gens aimaient pas
COV_Q57M	Au cours dernière semaine: ressenti peur ou panique sans avoir une très bonne raison
COV_T57	Niveau symptomatique dépressif
COV_Q58	Au cours des deux dernières semaines, penser au suicide?
COV_Q59	Entre mi-décembre 2019 et mi-mars 2020, fréquence consommé de l'alcool?
COV_Q60	Entre mi-décembre 2019 et mi-mars 2020, combien verres d'alcool consommés, par semaine?
COV_Q61	Entre mi-décembre 2019 et mi-mars 2020, combien fois bu 4-5 verres alcool ou plus même occasion?
COV_Q62	Entre mi-décembre 2019 et mi-mars 2020, fréq. consommé du cannabis, mari, pot ou hachisch?
COV_Q63	Entre mi-mars et mi-juin 2020, fréquence avez-vous bu de l'alcool?
COV_Q64	Entre mi-mars et mi-juin 2020, combien verres alcool consommés, par semaine?
COV_Q65	Entre mi-mars et mi-juin 2020, combien fois bu 4-5 verres alcool ou plus même occasion?
COV_Q66	Entre mi-mars et mi-juin 2020, fréquence consommé cannabis, mari, pot ou hachisch?
COV_Q67	Entre mi-mars et mi-juin 2020, consommation cigarettes comparativement avant cette période?
COV_Q68	Entre mi-mars et mi-juin 2020, consommation vapotage comparativement avant cette période?
COV_Q69	Entre mi-mars et mi-juin 2020, jeux sur Internet argent comparativement avant cette période?
COV_Q70A	Entre mi-décembre 2019 et mi-mars 2020, temps par jour contenus télévisuels/médias numériques/ Internet?
COV_Q70B	Entre mi-décembre 2019 et mi-mars 2020, temps par jour réseaux sociaux ?
COV_Q70C	Entre mi-décembre 2019 et mi-mars 2020, temps par jour jouer à des jeux vidéo?

Variable	Libellé
COV_Q71A	Entre mi-mars et mi-juin 2020, temps par jour contenus télévisuels/médias numériques/Internet?
COV_Q71B	Entre mi-mars et mi-juin 2020, temps par jour réseaux sociaux?
COV_Q71C	Entre mi-mars et mi-juin 2020, temps par jour jouer à des jeux vidéo?
COV_Q72	Actuellement, suivez actualité en lien avec la crise de la COVID-19?
COV_Q73A	De mi-mars à mi-juin 2020, temps par jour consulter médias évolution COVID-19? - Médias traditionnels
COV_Q73B	De mi-mars à mi-juin 2020, temps par jour consulter médias évolution COVID-19 - Médias sociaux

### 3. Structure et 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 le jeune. Toutes les variables ont fait l'objet d'une validation<sup>5</sup>.

Un seul instrument de mesure a été utilisé au volet 2020 et il s'adressait seulement au jeune (QELJ). Afin de faciliter l'utilisation de la banque de données, le tableau 2 présente la structure des fichiers inclus dans la banque de données du volet 2020.

**Tableau 2. Fichiers inclus dans la banque de données du volet 2020 de l'ELDEQ (COVID)**

Nom du fichier	Nombre de lignes contenues dans le fichier	Données contenues dans le fichier
COV_2020	1 182	Données colligées à l'aide du QELJ et concernant l'enfant cible.
PDSCOVID	1 182	Variable de poids échantillonnaires généraux transversaux relative aux données recueillies sur le ménage (PDSCOVID19 <sup>6</sup> ). <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 à la section 6 de ce document.</i>

***Dans une optique longitudinale, il est conseillé de se référer aux questionnaires<sup>7</sup> 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 pour 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 9.4 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

5. 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.jesuisjeserai.stat.gouv.qc.ca](http://www.jesuisjeserai.stat.gouv.qc.ca)) sous l'onglet « Publications ».

6. Il est à noter que, comme au volet 2019, 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. Les questionnaires de l'étude sont disponibles sur le site de l'ELDEQ, sous l'onglet « Outils de collecte ».

plateforme (par exemple sur WINDOWS ou LINUX). La commande suivante peut être utilisée pour créer le catalogue des formats en SAS :

```
LIBNAME COV "le chemin du répertoire où se trouvent vos fichiers";

PROC FORMAT cntlin=COV.sas_fmfs library =COV.formats; RUN;
PROC CATALOG catalog=COV.formats; copy out=work.formats; RUN; QUIT;
```

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

#### 4. Remarques et mises en garde

##### Mode de collecte – Fichier COV\_2020

La collecte de données a été réalisée en multimode (web ou téléphonique). Le web était privilégié, mais le jeune avait la possibilité de répondre aux questions au téléphone. Une variable (COV\_MODE) 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.

#### 5. Variables dérivées

##### 5.1 Construction des variables

Le tableau 3 présente une brève description de la construction des différentes variables<sup>8</sup>.

**Tableau 3. Description des variables dérivées du volet COVID de l'ELDEQ**

Nom de la variable	Items utilisés	Description
COV_S21_27 <i>Trouble d'anxiété généralisé (somme)</i>	COV_Q21 COV_Q22 COV_Q23 COV_Q24 COV_Q25 COV_Q26 COV_Q27	Variable calculée pour le volet spécial santé mentale (SM_S30). On soustrait 1 de chacun des items afin d'obtenir une valeur minimale de 0 pour chacun d'entre eux. Si le nombre de réponses valides aux x items retenus est égal ou supérieur à x-y (où y = le nombre possible de variables manquantes déterminé par l'équipe de recherche responsable de la construction de l'échelle), alors : COV_S21_27= Moyenne des items * nombre d'items puis on arrondit le score à deux décimales.
COV_D21_27 <i>Trouble d'anxiété généralisé (catégories)</i>	COV_S21_27	Variable calculée pour le volet spécial santé mentale (SM_D30). Si COV_S21_27 est valide, alors : COV_D21_27 correspond aux scores de COV_S21_27 catégorisés de la façon suivante: <ol style="list-style-type: none"><li>1. Score 0-4: Anxiété minimale</li><li>2. Score 5-9: Anxiété légère</li><li>3. Score 10-14: Anxiété modérée</li><li>4. Score 15 et plus: Anxiété grave</li></ol>
COV_T28_30 <i>Solitude</i>	COV_Q28	On soustrait 1 de chacun des items afin d'obtenir une valeur minimale de 0 pour chacun d'entre eux.

8. N'hésitez pas à communiquer avec l'équipe de l'ELDEQ pour plus de détails.

Nom de la variable	Items utilisés	Description
	COV_Q29 COV_Q30	Si le nombre de réponses valides aux x items retenus est égal ou supérieur à x-y (où y = le nombre possible de variables manquantes déterminé par l'équipe de recherche responsable de la construction de l'échelle), alors :  COV_T28_30 = Moyenne des items * nombre d'items puis on arrondit le score à deux décimales.  Les scores sont ramenés à une échelle variant entre 0 et 10.  Sinon, COV_T28_30 est indéterminée.
COV_D49 <i>Insécurité alimentaire</i>	COV_Q49A COV_Q49B COV_Q49C	Variable calculée pour les volets 2017 et 2019.  Cette variable sert à identifier s'il y a présence d'insécurité alimentaire dans la vie du jeune.
COV_T57 <i>Niveau symptomatique dépressif</i>	COV_Q57A, COV_Q57B, COV_Q57C, COV_Q57D, COV_Q57E, COV_Q57F, COV_Q57G, COV_Q57H, COV_Q57I, COV_Q57J, COV_Q57K, COV_Q57L, COV_Q57M	Variable calculée pour le volet spécial santé mentale. Voir méthode présentée pour la variable COV_T28_30.  ⇒ On inverse la valeur de certains des items (Q57F, Q57H, Q57J).

## **6. Pondération<sup>9</sup>**

Les prochaines pages de ce document présentent la description de la méthode de pondération utilisée pour pondérer les données de l'Étude longitudinale du développement des enfants du Québec (ELDEQ) au volet spécial sur la COVID-19, mené en 2020 auprès des jeunes de l'échantillon ainsi que l'analyse de la non-réponse.

Plus précisément,

- la section 6.1 propose une description des étapes ayant mené au choix de la stratégie de pondération;
- la section 6.2 présente l'analyse de la non-réponse totale et partielle ayant donné lieu à la création de la pondération;
- la section 6.3 présente l'analyse de la non-réponse partielle;
- la section 6.4 détaille le taux de réponse pondéré obtenu; et
- la section 6.5 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.

### **6.1 Stratégie de pondération**

#### **6.1.1 Admissibilité à l'enquête au volet spécial sur la COVID-19**

Parmi les 2 120 répondants au volet initial, on compte 31 familles ayant quitté définitivement le Québec et 5 familles dont l'enfant cible est décédé entre les volets 1998 et 2019. 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 2019 ou qui ont quitté la province temporairement.

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 spécial sur la COVID-19 est composé de 2 084 jeunes<sup>10</sup>.

#### **6.1.2 Choix du volet de référence 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. 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 spécial sur la santé mentale. Ainsi, nous avons un poids de base qui sert de 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 l'ajustement pour la non-réponse.

---

9. Cette section a été produite par Karine Dion et Luc Belleau de la Direction de la méthodologie de l'Institut de la statistique du Québec.

10. 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.

## **6.2 Analyse de la non-réponse totale**

### **6.2.1 Pondération transversale des données du volet spécial sur la COVID-19**

#### *6.2.1.1 Démarche générale d'analyse*

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.

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 spécial sur la COVID-19 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, un ajustement a été fait à partir du poids de base « enfant » (voir la section 6.1.2). Cet ajustement est requis puisque les répondants au volet spécial sur la COVID-19 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.

#### *6.2.1.2 Variables considérées et résultats*

Comme le questionnaire du volet spécial sur la COVID-19 est rempli par le jeune et non la famille, il est intéressant de considérer des variables qui le caractérisent pour la pondération. Ainsi, des variables du questionnaire au jeune de 2013, 2015, 2017 et 2019 ont été considérées : s'il aime l'école, s'il a déjà doublé une année scolaire, plus haut niveau de scolarité qu'il désire atteindre, s'il travaille ou non, s'il passe 15 heures ou plus à l'ordinateur dans une semaine, s'il passe 15 heures ou plus à jouer à des jeux vidéo dans une semaine, s'il est en couple, s'il habite à la même adresse que ses parents ou tuteurs, la perception de son état de santé, la relation avec sa mère, la relation avec son père, etc. Ces 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 spécial sur la COVID-19.

D'autres variables, principalement de nature socioéconomique, qui ont été utilisées dans la pondération des volets précédents de l'ELDEQ ont aussi été considérées. Certaines 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.

Enfin, compte tenu de la disponibilité de microdonnées administratives<sup>11</sup> provenant du ministère de l'Éducation et de l'Enseignement supérieur (MEES), des variables supplémentaires ont été considérées pour la pondération du volet spécial sur la COVID-19. Ces variables sont reliées au cheminement scolaire du jeune : sexe, langue maternelle et diplomation. Ce fichier permet aussi de créer une variable qui

---

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

exprime la mobilité du jeune entre 2017 et 2019. Cette variable a été créée en comparant les adresses de la résidence du jeune pour les années 2017 et 2019. Il est important de mentionner que la majorité des jeunes de ces volets avaient terminé leurs secondaires en 2015 ou 2016, ainsi il y avait très peu d'informations disponibles provenant du MEES. 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.

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

- le sexe du jeune;
- l'obtention ou non de son diplôme d'études secondaires (DES);
- le jeune habite le foyer familial (VMFNQ01A, TMFNQ04 et RCNNQ02);
- le jeune passe 15h ou plus par semaine sur l'ordinateur (RAPNQ8, PAPNQ8 et NAPNQ8).

Une méthode de classification hiérarchique de Ward a permis de regrouper les probabilités estimées en cinq classes de pondération. Le tableau 4 présente les proportions pondérées de répondants au volet spécial sur la COVID-19 parmi les répondants du poids « enfant » pour ces 5 groupes. De plus, il présente le nombre de répondants, parmi les 1 182 répondants, à qui la proportion pondérée sera appliquée en guise de correction de la non-réponse. Par exemple : il y a 191 répondants dont le poids de référence sera ajusté par l'inverse de la proportion pondérée de la première classe de pondération, qui est de 74,2 %.

**Tableau 4 Proportions pondérées de répondants et nombre de répondants par classe de pondération**

Classe de pondération	Proportions pondérées de répondants (en %)	Nombre de répondants
1	74,2	191
2	61,0	279
3	82,3	398
4	38,5	60
5	87,1	254

Au sein des différentes classes d'ajustement de la pondération, la proportion de répondants varie de 38 % à 87 % (relativement à une proportion globale de 72 %). La proportion la plus faible est observée dans une classe où l'on retrouve, en proportion, un plus grand nombre de jeunes de sexe masculin; de jeunes n'ayant pas obtenu leur DES; de jeunes vivant dans le foyer familial; et de jeunes passant moins de 15h par semaine à l'ordinateur.

### **6.3 Analyse de la non-réponse partielle**

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. 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 lieu<sup>12</sup>. Il faut également considérer la non-réponse partielle cumulative si l'analyse comprend plusieurs variables simultanément.

Pour le seul instrument de collecte du volet spécial sur la COVID, aucune question n'est entachée de non-réponse partielle supérieure à 5 %.

#### **6.4 Taux de réponse**

Le tableau 5 présente le taux de réponse pondéré obtenu au volet spécial sur la COVID-19. Ce taux est obtenu en multipliant les taux obtenus aux différentes étapes de pondération, selon le cas.

**Tableau 5 Taux de réponse pondéré au volet spécial sur la COVID-19**

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 spécial sur la COVID-19	95,0 % (n=2 084)
Proportion de répondants au volet 2002 parmi les répondants au volet 2000 admissibles au volet spécial sur la COVID-19 (incluant les pseudo-répondants)	98,9 % (n=1 997)
Proportion de répondants du poids « enfant » parmi les répondants au volet 2002 admissibles au volet spécial sur la COVID-19 (incluant les pseudo-répondants)	82,4 % (n=1 977)
Proportion de répondants au volet spécial sur la COVID-19 parmi les répondants du poids « enfant » admissibles au volet spécial sur la COVID-19 (incluant les pseudo-répondants)	71,6 % (n=1 647)
Taux de réponse au volet spécial sur la COVID-19	39,7 %

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

#### **6.5 Utilisation de la pondération par les utilisateurs des données du volet spécial sur la COVID-19**

##### **6.5.1 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.

Le taux de réponse au volet spécial sur la COVID-19 est de l'ordre de 39,7 % (voir tableau 5). Ce taux confirme l'importance du traitement effectué lors de la pondération.

##### **6.5.2. Tests statistiques**

Le fichier PDSCOV19 contient la variable de pondération PDSCOV19 (poids général transversal du volet spécial sur la COVID-19). 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

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

de toute analyse des données du volet spécial sur la COVID-19. 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 PDSCOV19 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. Par exemple, il serait 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 prudence.

### **6.5.3 Choix de la pondération**

Les possibilités d'analyse incluant des données du volet spécial sur la COVID-19 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<sup>13</sup>.

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 suivie pour évaluer la situation. Ceux-ci se retrouvent dans les rapports de pondération des volets antérieurs<sup>14</sup>.

## **6.6. 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

---

13. Des frais pourraient s'appliquer pour la réalisation de ces travaux.

14. Les rapports de pondération des volets antérieurs sont disponibles sur le site de l'ELDEQ ([www.jesuisjeserai.stat.gouv.qc.ca](http://www.jesuisjeserai.stat.gouv.qc.ca)), sous l'onglet « Documentation technique ».

## **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 spécial sur la COVID-19.

**Étape 1 :**

Analyses bivariées pour réduire le nombre de variables considérées pour la modélisation (environ 60 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 :**

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 4 :**

Création d'une catégorie de valeurs manquantes pour les variables du modèle retenu à l'étape 3. À noter qu'à ce volet-ci, les valeurs manquantes de certaines variables ont été imputées aléatoirement à une catégorie de façon à conserver les proportions observées. La validation de ce modèle est effectuée et un modèle final est retenu.

**Étape 5 :**

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.