

Méthodes d'estimation du nombre de cas de COVID-19 dans la population générale au Québec

Date 27 juillet 2022

Sommaire

1. Mise en contexte	1
2. Méthodologie	2
3. Avantages et limites	5
4. Références	9
5. Annexe 1 (Questionnaire)	10

1. Mise en contexte

L'introduction des tests rapides dans la communauté (20 décembre 2021) et les changements dans la stratégie de dépistage (5 janvier 2022) affectent les indicateurs des nombres de cas dans la population générale. En effet, les tests d'amplification d'acides nucléiques (TAAN) ont été restreints à certains groupes spécifiques de la population, tels que les personnes en milieu de soins ou d'hébergement, et les travailleurs de la santé. Ces changements ont limité la capacité de suivi de l'évolution de la pandémie auprès de la population générale, notamment pour l'évolution de l'incidence (journalière/hebdomadaire) des nouveaux cas.

Dans ce contexte, la réalisation d'enquêtes en ligne a été proposée comme méthode complémentaire afin de suivre l'évolution des cas incidents de COVID-19 au Québec. Une première approche, désignée sous le nom de méthode directe, a été spécifiquement adoptée. Elle permet d'estimer le nombre de cas d'infection à l'échelle provinciale

auprès de la population âgée de 18 ans et plus. Parce que la méthode directe peut produire des estimations peu précises en période de basse circulation du virus, la méthode d'*amplificateur par réseau* (APR, en anglais *Network Scale-up*) ou méthode indirecte a été proposée au Québec pour surmonter les difficultés liées à l'estimation des personnes infectées par le SRAS-CoV-2⁽²⁾. La méthode APR est particulièrement adaptée pour étudier une population d'intérêt difficilement accessible⁽³⁾, à savoir dans notre cas les personnes infectées par le SRAS-CoV-2.

Ces deux approches ont été utilisées par le Centre interuniversitaire de recherche en analyse des organisations (CIRANO) pour suivre l'évolution des cas de COVID-19 au Québec de la mi-janvier jusqu'à la mi-mai 2022. L'INSPQ a été mandaté pour poursuivre le travail de ce groupe de chercheurs afin de produire des indicateurs complémentaires pour suivre l'évolution de la pandémie⁽²⁾.

- L'**OBJECTIF PRINCIPAL** de cette étude est de suivre l'évolution de l'incidence journalière/hebdomadaire des nouveaux cas d'infection chez les sujets âgés de 18 ans et plus (méthode directe) et dans la population générale (méthode indirecte).
- L'**OBJECTIF SECONDAIRE** est d'évaluer à fréquence régulière les biais des méthodes indirectes en fonction de la circulation ou de la transmission du virus.

Il est important de mentionner que ces deux approches serviront surtout à suivre l'évolution du nombre de cas de manière hebdomadaire et à détecter une reprise éventuelle des infections.

2. Méthodologie

2.1. Base de sondage

La plateforme *Clic Santé* de prise de rendez-vous pour la vaccination a servi de base de sondage pour l'étude *Évolution des cas de COVID-19 au Québec : approche par sondage populationnel*. Un peu plus de 6 millions d'adultes québécois y sont présents et 92 % de ceux-ci ont fourni une adresse courriel pour les joindre, une information nécessaire pour solliciter leur participation à cette étude. En utilisant une telle base de sondage, la population sondée n'est pas si loin de la population cible (pour plus de détails sur les avantages liés à l'utilisation de cette base de sondage cf. section 3.1.).

2.2. Plan d'échantillonnage et taille d'échantillon

La stratégie d'échantillonnage de base consiste à :

- ▶ trier la base de sondage selon la région sociosanitaire de résidence, le sexe et l'âge des répondants;
- ▶ procéder par sélection systématique de 30 000 courriels.

Cette méthode, totalement aléatoire, permet d'assurer une bonne représentativité de l'échantillon selon ces trois variables et d'obtenir un *échantillon probabiliste* au regard de la population cible des adultes québécois.

À partir de la deuxième semaine de collecte, la stratégie d'échantillonnage a été améliorée, en tenant compte des taux de réponse différents selon le sexe et l'âge. Ainsi, des strates d'échantillonnage ont été formées à partir de ces deux variables. Par conséquent, la répartition de la taille d'échantillon dans ces strates d'échantillonnage présentée dans le **tableau 1** reflète à la fois la répartition proportionnelle aux effectifs de population québécoise ainsi que les taux de réponse. La vérification de ces ajustements aux taux de réponse est réalisée régulièrement.

Tableau 1 Répartition de l'échantillon initial selon l'âge et le sexe

Sexe à la naissance	18-34 ans	35-54 ans	55-64 ans	65 ans et +
Hommes	18,2 %	17,9 %	9,0 %	9,2 %
Femmes	13,3 %	14,9 %	8,1 %	9,4 %

2.3. Collecte

Le questionnaire comprend une série de questions concernant l'expérience du répondant avec la COVID-19 durant les 7 derniers jours (présence de symptômes apparentés à la COVID-19, obtention de tests positifs à l'infection par le SRAS-CoV-2, isolement dû à la COVID-19), en lien avec les méthodes d'estimations APR (nombre de cas de COVID-19 dans le réseau de connaissances, nombre de connaissances parmi 8 sous-populations) et enfin des questions de nature sociodémographique (niveau d'éducation, lieu de naissance, temps vécu au Canada, langue couramment utilisée à la maison, nombre de personnes formant le ménage, l'âge des personnes vivant dans le même ménage). Pour de plus amples détails, le questionnaire est présenté en **Annexe 1** à ce document.

Chaque collecte est prévue sur une période de 7 jours commençant le lundi à 12 h et se terminant le dimanche à minuit. Pour augmenter le taux de participation, un courriel de rappel de l'étude est lancé le jeudi à 12 h.

2.4. Traitements

Les méthodes d'estimation APR s'appuient essentiellement sur des questions posées à propos des personnes formant le réseau de connaissances des répondants à l'étude. Pour éviter que quelques répondants aient des impacts importants sur ces estimations en raison de leurs nombres de connaissances trop importants, ces nombres sont ramenés aux valeurs des 99^{es} percentiles. De plus, les individus ayant des nombres de connaissances supérieurs aux 99^{es} percentiles pour au moins trois sous-populations sont retirés des calculs des méthodes APR. La répartition de l'échantillon de répondants peut être fort différente de celle de la population générale des adultes (18 ans et plus) québécois. Pour inférer les résultats à cette population, les rendre représentatifs, on applique une pondération qui tient compte de la probabilité initiale de sélection, du mécanisme de non-réponse et du calage aux marges sur différentes caractéristiques de la population [sexe, groupe d'âge (18-34 ans, 35-54 ans, 55-64 ans et 65 ans et plus), regroupement de régions sociosanitaires¹, niveau de scolarité et fait de vivre seul]. La méthode pour le calage aux marges a été vérifiée pour ne pas produire de poids extrêmes. Toutes les estimations sont ainsi pondérées par ces poids d'enquête.

¹ En quatre catégories : Montréal, la couronne de Montréal (Montérégie, Lanaudière, Laurentides, Laval), autres régions urbaines (Capitale-Nationale, Mauricie-Centre-du-Québec, Estrie, Outaouais, Chaudières-Appalaches, Saguenay-Lac-St-Jean) et petites régions (Bas-St-Laurent, Abitibi-Témiscamingue, Côte-Nord, Nord-du-Québec, Gaspésie-Îles-de-la-Madeleine, Terres-Cries-de-la-Baie-James et Nunavik).

2.5. Diffusion des résultats

Comme les questions se réfèrent aux 7 derniers jours, les résultats pour une semaine donnée de collecte peuvent se rapporter sur une période de deux semaines (la semaine de la collecte de données et celle qui la précède). Il faut en tenir compte lors de l'interprétation.

Les résultats sont en fait des estimations obtenues à l'aide d'un échantillon de participants et ne doivent pas être perçus avec une précision absolue. Ainsi, les résultats portant sur le nombre de personnes ayant contracté la COVID-19 sont arrondis au millier près. De plus, pour apprécier la précision des estimations présentées, elles sont toujours accompagnées de leurs intervalles de confiance à un niveau de 95 %. Ces intervalles de confiance ont été obtenus en utilisant l'approximation normale sur la transformation « *logit* » des proportions estimées. Les erreurs types proviennent, quant à elles, des méthodes de rééchantillonnage de type « *bootstrap* » (avec 500 itérations).

Les résultats doivent respecter le critère d'un coefficient de variation (CV) inférieur à 33,3 %⁽⁴⁾, sans quoi ils sont jugés comme ayant une variabilité trop importante pour être diffusés.

2.6. Méthodes d'estimation et analyses statistiques

Les méthodes d'estimation présentées sont de deux types : directes (avec ou sans confirmation de tests positifs) et indirectes (les méthodes dites « amplificateur par réseau », APR). Il faut préciser que ces deux méthodes portent sur les nouveaux cas de COVID-19 au cours des 7 derniers jours et incluent les cas de réinfections.

- ▶ Les méthodes directes se rapportent directement aux répondants et donc à la population des adultes québécois. Elles permettent de ventiler les résultats selon différentes variables sociodémographiques (sexe, groupe d'âge, régions). Les estimations directes ont été calculées en se basant uniquement sur les réponses fournies à la question 7 (voir l'**Annexe 1**). Il y a deux types de résultats produits : le premier comprenant uniquement une confirmation par un test positif (PCR ou rapide) et le second, incluant l'autodiagnostic des participants sans résultat de test. Le premier type d'estimation est ainsi inclus dans le deuxième. Il faut noter cependant que l'estimation incluant l'autodiagnostic peut surestimer le nombre réel de cas de COVID-19, car l'autodiagnostic peut inclure des symptômes liés à d'autres virus respiratoires (consulter *cf.* section 3.2.1 portant sur les limites des méthodes d'estimation).
- ▶ Les méthodes indirectes se réfèrent plutôt à l'ensemble de la population. Elles s'appuient sur les nombres de connaissances des répondants pour 8 sous-populations (voir l'**Annexe 1**). La méthode de *Killworth*⁽⁵⁾ retenue par l'équipe CIRANO (présentée dans leur [rapport de base](#) du 28 janvier 2022) a été utilisée⁽²⁾. Les estimations indirectes ont été calculées en se basant uniquement sur les réponses fournies aux questions 8 A à 8 I (voir l'**Annexe 1**). Il est à noter que durant les premières semaines, nous avons utilisé, en plus de la méthode de *Killworth*⁽⁵⁾, la méthode inspirée de *Habecker*⁽⁶⁾, en accord avec l'approche de l'équipe CIRANO, dans un but de tester la concordance des deux méthodes. Toutefois, considérant que celles-ci fournissaient des estimations similaires, il a été décidé de garder uniquement la méthode de *Killworth* (méthode de base de l'APR) pour simplifier l'interprétation des résultats.

Afin de suivre l'évolution des cas par caractéristiques sociodémographiques, des taux sont présentés. Ces taux sont calculés en divisant les incidences hebdomadaires, calculées avec la méthode directe avec autodiagnostic, par la taille de la population visée.

Des tests statistiques de comparaison sont présentés entre les estimations d'une semaine avec la précédente pour juger des changements statistiquement significatifs dans le temps. Des tests statistiques sont aussi effectués entre les taux hebdomadaires estimés pour les différentes caractéristiques

sociodémographiques (groupes de régions, d'âge et sexe). Comme pour la construction des intervalles de confiance, ces tests statistiques utilisent l'approximation normale sur la transformation « *logit* » des proportions estimées. Le seuil des tests retenu pour conclure à une différence statistiquement significative est de 5 %.

3. Avantages et limites

3.1. Avantages et limites associés à l'échantillonnage et à la méthode de collecte

La base de sondage *Clic Santé* (incluant ≈ 5,5 millions d'individus) est jugée plus représentative de la population du Québec qu'un panel Web. En utilisant une telle base de sondage, les caractéristiques de la population sondée se rapprochent de la population cible. Le **tableau 2** décrit en détail les avantages et inconvénients de considérer *Clic Santé* comme base de sondage.

Tableau 2 Avantages et inconvénients de l'utilisation de *Clic Santé* comme base de sondage

Paramètre	Avantages	Désavantages
Population à l'étude	Plus large et considérée comme plus représentative de la population visée, comparativement à un panel Web qui est en partie composé de volontaires et de gens intéressés à remplir des sondages en ligne.	Les personnes non vaccinées (en date du 26 juin, 7,9 % pour les 12 ans et plus) sont peu présentes dans le bassin des répondants, car il faut s'être créé un compte sur le site <i>Clic Santé</i> . Toutefois, le biais pourrait être perçu comme négligeable, car il faudrait que le taux d'infection des non-vaccinés soit nettement plus élevé que celui des vaccinés. Dans le cas où il y a présence d'un biais, il y aurait une sous-estimation du nombre de cas de COVID-19.
Sélection des participants	<ul style="list-style-type: none"> Sélection d'un échantillon probabiliste. Ceci permet de mesurer l'erreur d'échantillonnage liée aux estimations obtenues et de faire des inférences au sujet de la population. Plus grand bassin de répondants, ce qui réduit les chances d'être échantillonné plusieurs fois. Ceci permet théoriquement de maintenir un taux de réponse raisonnable et de réduire le biais d'attrition. 	Taux de réponse relativement faibles, surtout chez les jeunes adultes et les hommes.
Précision des estimations	Aucune contrainte imposée par les quotas. Pondération tient compte de la probabilité de sélection des individus par strate et du mécanisme de non-réponse (voir la section Méthodologie).	

Le sondage en ligne est une méthode peu dispendieuse et rapide par rapport aux enquêtes de terrain ou avec entrevues.

Nous avons adopté deux approches complémentaires : i) une *méthode directe* qui permet d'estimer le nombre de sujets âgés de 18 ans et plus infectés par le SRAS-CoV-2 dans la province du Québec; ii) une *méthode indirecte* qui permet d'estimer le nombre total de sujets infectés dans la province.

La méthode d'amplification par réseau (APR) permet de parvenir à des sous-groupes difficilement accessibles pour diverses raisons⁽³⁾. Cette méthode est d'autant plus intéressante sachant que la population d'intérêt est très volatile⁽⁶⁾ : autrement dit, le délai de temps qui s'écoule entre le moment où la personne est décelée positive (ou symptomatique) et le moment où elle guérit est très court (une dizaine de jours environ). La réalisation d'enquêtes épidémiologiques utilisant la méthode APR de manière récurrente sur de courts laps de temps (hebdomadaire) permet de pallier ce problème spécifiquement.

Contrairement à la méthode indirecte, les méthodes directes permettent de suivre l'incidence hebdomadaire des cas en fonction des variables sociodémographiques (âge, sexe et région de résidence). Toutefois, la méthode indirecte permet de couvrir l'ensemble de la population générale incluant les moins de 18 ans.

Mentionnons finalement que, pour une taille d'échantillon donnée, les estimations indirectes permettent d'obtenir de meilleures mesures de précision (variances des estimations plus petites et intervalles de confiance plus étroits) que celles des estimations directes.

3.2. Limites associées aux différentes méthodes d'estimation

3.2.1. MÉTHODE DIRECTE

La méthode directe repose sur la collecte de données auprès des répondants en se basant sur leur propre expérience. En ce sens, un plan d'échantillonnage bien déterminé est préétabli pour assurer une représentativité des estimations obtenues à l'échelle provinciale. Sachant que le taux de réponse peut être associé aux caractéristiques sociodémographiques des répondants, la représentativité des estimés peut être altérée par les faibles taux de participation. Pour remédier à ce problème, la stratégie d'échantillonnage a été raffinée en tenant compte des taux de réponse (c.f. section 2.2. Plan d'échantillonnage et taille d'échantillon). Dans ce contexte, un biais d'attrition peut émerger lorsque le taux de réponse est faible. Ce biais a été minimisé non seulement par l'ajustement de l'échantillonnage, mais aussi par l'adoption de différentes stratégies de pondération qui incluent entre autres le mécanisme de non-réponse. De plus, un contrôle a été effectué lors de l'étape du calage aux marges pour s'assurer de minimiser l'occurrence possible de poids ayant des valeurs extrêmes.

Cependant, il est généralement difficile d'évaluer la qualité des réponses obtenues dans le cadre des enquêtes en ligne. Les principaux biais associés à cette approche sont : i) le biais de mémoire, qui est peu probable à cause de la courte période de rappel. Toutefois, il est attendu que les sujets les plus sensibilisés au sujet de la COVID-19 tendent à rapporter plus précisément l'information surtout liée aux symptômes et à l'autodiagnostic; ii) le biais de désirabilité; iii) le biais de mesure lié à l'utilisation correcte des tests rapides et aussi à leur performance moins importante que le test d'amplification des acides nucléiques. À noter qu'avec cette approche, les personnes asymptomatiques ne sont pas identifiées, ce qui peut causer une sous-estimation du vrai nombre de cas de COVID-19. À l'inverse, une surestimation du nombre réel de cas de COVID-19 pourrait être observée puisque l'autodiagnostic peut inclure des symptômes liés à d'autres virus respiratoires.

Finalement, mentionnons que lorsque l'incidence des cas de COVID-19 sera faible, le nombre de personnes sondées ne sera pas suffisant pour obtenir des estimations précises (avec un coefficient de variation inférieur à 33,3 %) par la méthode directe.

3.2.2. MÉTHODE INDIRECTE « AMPLIFICATEUR PAR RÉSEAU »

La méthode APR repose essentiellement sur le fait que le réseau de connaissances des répondants est représentatif de la population générale⁽⁶⁾. En plus de cette hypothèse initiale liée à la représentativité du réseau des répondants, trois hypothèses sont à priori admises pour faire les approximations nécessaires⁽⁶⁾, à

savoir : *i*) les liens sociaux se forment complètement au hasard entre le répondant et son réseau. Donc les répondants ont la même probabilité de connaître un individu de la population d'intérêt (sujets infectés par le SRAS-CoV-2); *ii*) les répondants connaissent parfaitement leurs réseaux, autrement dit, ils sont au courant de l'occurrence de toutes infections au sein de leurs réseaux respectifs; et *iii*) les répondants sont capables de se remémorer et de fournir des réponses précises sur leurs réseaux.

Or, sur le plan pratique, ces hypothèses ne sont pas toujours respectées⁽⁹⁾. De ce fait, leurs violations aboutissent respectivement à l'émergence d'un certain nombre de biais, à savoir l'*effet barrière*, l'*effet de transmission* et le *biais de mémoire*^(6,10). L'*effet barrière* indique que les répondants peuvent avoir une connaissance plus ou moins différente d'individus appartenant à la population d'intérêt ou à l'une des 8 sous-populations⁽⁶⁾. La probabilité de connaître un individu issu de la population d'intérêt peut être associée aux caractéristiques sociodémographiques du répondant (par exemple, âge, statut professionnel)⁽⁶⁾. L'*effet de transmission* est un biais associé à la violation de la seconde hypothèse, à savoir que les répondants ne sont pas au courant de tous les cas d'infection survenus dans leur réseau de connaissances⁽⁶⁾. Enfin, le *biais de mémoire* est associé à la troisième hypothèse⁽⁶⁾. En effet, les répondants peuvent oublier de rapporter avec précision les individus infectés par le SRAS-CoV-2 ou de décrire avec une précision suffisante leur réseau de connaissances. Si un répondant connaît un faible nombre d'individus d'un sous-groupe, il a plus de chance de s'en souvenir que s'il connaissait plusieurs individus d'un sous-groupe et dans ce cas, il risquerait d'en omettre un certain nombre⁽¹¹⁾. Par exemple, si un répondant appartient à la catégorie des travailleurs de la santé (spécifiquement les infirmiers), il y a de fortes chances qu'il oublie d'énumérer tous les individus appartenant à son réseau dans le sous-groupe « infirmières/infirmiers ». Selon Rolland *et al.*⁽⁸⁾, la stigmatisation liée à la COVID-19 réduit les erreurs d'approximation liées au biais de mémoire. Ces trois biais sont généralement à l'origine d'une sous-estimation du nombre de vrais cas d'infection dans la population cible. Il est à noter que l'ampleur de ces biais peut être variable en fonction du niveau de circulation du virus dans la population générale.

Au-delà de ces biais classiquement énumérés par les auteurs de la méthode APR, il est probable qu'un *biais chronologique* émerge. En effet, les répondants peuvent situer de manière erronée la date exacte à laquelle une personne de leur réseau a été infectée. Toutefois, ce biais est probablement faible à cause de la courte période de rappel. D'un autre côté, il n'est pas exclu qu'un *biais de désirabilité* ait une influence sur les résultats rendus par la méthode APR, car les personnes atteintes de la COVID-19 sont relativement stigmatisées⁽¹²⁾. Enfin, il est à noter que la stigmatisation liée à la COVID-19 peut être associée au niveau de circulation du virus et de la virulence des sous-lignées ou variants qui sont en émergence.

La méthode APR repose principalement sur des questions adressées aux répondants pour déterminer leur réseau de connaissances. Ces questions portent sur des sous-populations spécifiques (par exemple, « *combien connaissez-vous de personnes, adultes et mineures, dont le nom de famille est Tremblay?* »). Les résultats obtenus par la méthode APR sont grandement influencés par les sous-populations incluses dans l'enquête dont le choix et la sélection restent tributaires des scientifiques responsables de mener l'enquête.

Par ailleurs, il est important de souligner qu'il est possible d'avoir un chevauchement des réseaux de connaissances des différents répondants. Toutefois, le caractère aléatoire de l'échantillon des répondants de la population générale tend à réduire cette limite⁽⁸⁾. De plus, il a été démontré que le chevauchement entre les réseaux de connaissances des répondants n'affectait pas à la baisse les estimés⁽¹³⁾.

Il est important de souligner que la technique APR ne permet pas d'avoir accès aux caractéristiques du réseau de connaissances des répondants. Donc, elle ne permet pas de faire des analyses par variables sociodémographiques (région, âge, sexe, etc.)

L'évaluation de la qualité des données qui proviennent des répondants est difficile. Toutefois, puisque la technique repose essentiellement sur le réseau de connaissances des répondants, les observations extrêmes

ont été ramenées au 99^e percentile afin de réduire l'impact qu'auraient des valeurs extrêmes sur les méthodes d'estimation indirectes.

Enfin, notre expérience personnelle des premières semaines de collecte de données a montré que certains participants étaient réticents à participer à l'étude, car ils ne comprenaient pas le but des questions posées à propos des sous-populations. Pour y remédier, nous avons mis en place une équipe dédiée à répondre aux questions des participants par retour de courriel.

4. Références

- (1) MINISTÈRE DE LA SANTÉ ET DES SERVICES SOCIAUX « Pandémie de la COVID-19 – Modification des priorités de dépistage et de gestion des cas et des contacts - Salle de presse - MSSS », <<https://www.msss.gouv.qc.ca/ministere/salle-de-presse/communiqué-3371/>> (consulté le 25 mai 2022).
- (2) BOISCLAIR, D., R. BORGÈS DA SILVA, V. BOUCHER, N. DE MARCELLIS-WARIN, P.-C. MICHAUD et I. PEIGNIER (janvier 2022). *Combien de personnes ont développé des symptômes ou contracté la covid-19 au Québec ? Une étude exploratoire* [en ligne], CIRANO, <<https://doi.org/10.54932/KWYT2364>> (consulté le 11 mars 2022).
- (3) BANESHI, M. R., A. RASTEGARI et A. A. HAGHDOOST (2021). « Review of Size Estimation Methods », dans *Methods in Epidemiology*, Springer, p. 1-15.
- (4) STATISTIQUE CANADA (2019). *Guide de l'utilisateur, Fichiers de microdonnées de 2018 et 2017-2018 de l'Enquête sur la santé dans les collectivités canadiennes (ESCC) - Composante annuelle*.
- (5) KILLWORTH, P. D., C. MCCARTY, H. R. BERNARD, G. A. SHELLEY et E. C. JOHNSEN (1998). « Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach », *Evaluation review*, vol. 22, n° 2, p. 289-308.
- (6) HABECKER, P., K. DOMBROWSKI et B. KHAN (2015). « Improving the network scale-up estimator: Incorporating means of sums, recursive back estimation, and sampling weights », *PloS one*, vol. 10, n° 12, p. e0143406.
- (7) STATISTIQUE CANADA (2 septembre 2021). « Échantillonnage non probabiliste », <<https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch13/nonprob/5214898-fra.htm>> (consulté le 27 mai 2022).
- (8) ROLLAND-GUILLARD, L., G. CHARRANCE et E. MORAND (2020). *Quelles méthodes pour estimer la taille d'une population difficile à enquêter?*, [en ligne], Institut National d'Études Démographiques, <<https://hal.archives-ouvertes.fr/hal-02928767/document>> (consulté le 24 mai 2022).
- (9) FEEHAN, D. M., et M. J. SALGANIK (août 2016). « Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations », *Sociological methodology*, vol. 46, n° 1, p. 153-186.
- (10) OCAGLI, H., D. AZZOLINA, G. LORENZONI, S. GALLIPOLI, M. MARTINATO, A. S. ACAR, P. BERCHIALLA, D. GREGORI, et INCIDENT STUDY GROUP (2021). « Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol », *International journal of environmental research and public health*, vol. 18, n° 11, p. 5713.
- (11) BLAZEK, S., et D. FEEHAN « Network Scale-Up Method », dans *Methods of prevalence estimation in modern slavery. An introductory overview*, USA, Global fund to END MODERN SLAVERY.
- (12) MICONI, D., Z. Y. LI, R. L. FROUNFELKER, V. VENKATESH et C. ROUSSEAU (1 mars 2021). « Socio-cultural correlates of self-reported experiences of discrimination related to COVID-19 in a culturally diverse sample of Canadian adults », *International Journal of Intercultural Relations*, vol. 81, p. 176-192.
- (13) SNIDERO, S., R. CORRADETTI et D. GREGORI (2004). « The network scale-up method: A simulation study in case of overlapping sub-populations », *Advances in Methodology and Statistics*, vol. 1, n° 2, p. 395-405.

Annexe 1 : Questionnaire

Estimation du nombre de cas de COVID-19 au Québec

Q0. Êtes-vous cette personne ?

1. Oui **Consigne ==> poursuivre le questionnaire**
2. Non ==> faites suivre le courriel à la personne concernée ou demander à la personne concernée d'y répondre en cliquant de nouveau sur le lien inclus dans votre message. Quitter cette page. Merci

Q1. Quel âge avez-vous?

1. De 18 à 24 ans
2. De 25 à 29 ans
3. De 30 à 34 ans
4. De 35 à 39 ans
5. De 40 à 44 ans
6. De 45 à 49 ans
7. De 50 à 54 ans
8. De 55 à 59 ans
9. De 60 à 64 ans
10. De 65 à 69 ans
11. De 70 à 74 ans
12. 75 ans ou plus

Q2. Au total, combien de personnes habitent chez vous en vous incluant ? Compter aussi les enfants qui habitent chez vous, que ce soit en permanence ou de temps en temps (en garde partagée)

- ___ personne(s)
1. Une seule (moi-même)

Q3. Parmi ces personnes, combien d'enfants des groupes d'âge suivants habitent avec vous (la moitié du temps ou plus)?

Filtre : Q2.1 = > 1

1. 0 à 4 ans : (numérique entier [0,+)
2. 5 à 12 ans : (numérique entier [0,+)
3. 13 à 17 ans : (numérique entier [0,+)

Q4. Actuellement, quel est votre statut en ce qui concerne la vaccination contre la COVID-19?

1. J'ai reçu 4 doses de vaccins contre la COVID-19
2. J'ai reçu 3 doses de vaccins contre la COVID-19
3. J'ai reçu 2 doses de vaccins contre la COVID-19
4. J'ai reçu 1 seule dose de vaccin contre la COVID-19
5. Je n'ai reçu aucune dose de vaccin contre la COVID-19
99. Je préfère ne pas répondre

Q5. Dans les 7 derniers jours, avez-vous ressenti des symptômes que l'on pourrait associer à la COVID-19? En vous basant sur la liste suivante présentée par le Gouvernement du Québec, veuillez sélectionner chacun des groupes de symptômes ressentis.

Considérez-les 7 derniers jours, en comptant aujourd'hui.

A. Fièvre (38,1 °C (100,6 °F) et plus (température buccale)),

1. Oui
2. Non

B. Symptômes généraux (*perte soudaine d'odorat sans congestion nasale, perte de goût, grande fatigue, perte d'appétit importante, douleurs musculaires généralisées non liées à un effort physique, mal de tête*)

1. Oui
2. Non

C. Symptômes respiratoires (*toux nouvelle ou aggravée, essoufflement, difficulté à respirer, mal de gorge, nez qui coule ou congestion nasale de cause inconnue*)

1. Oui
2. Non

D. Symptômes gastro-intestinaux (*nausées, vomissements, diarrhée, maux de ventre*)

1. Oui
2. Non

Q6. Avez-vous été en isolement dû à la COVID-19 dans les 7 derniers jours, incluant aujourd'hui?

1. Oui
2. Non

Q7. Dans les 7 derniers jours incluant aujourd'hui, avez-vous eu un résultat de test positif à la COVID-19?

1. Oui
2. Non
3. Pas de résultat de test, mais je crois que je l'ai contractée (autodiagnostic en fonction de mes symptômes des 7 derniers jours)

Le concept de « **connaître quelqu'un** » fait référence à des personnes qui vivent au Québec que vous connaissez et qui vous connaissent et avec qui vous avez eu des contacts dans les deux dernières années (en face à face, par téléphone, sur les réseaux sociaux ou par courrier/courriel.)

Q8A. Selon votre meilleure estimation, combien connaissez-vous de gens, adultes et mineurs, ayant contracté la COVID-19 (avec ou sans confirmation par un test positif) au cours des 7 derniers jours (incluant aujourd'hui)?

- (nombre 0 à xxx)

Les prochaines questions sont destinées à estimer la taille de votre réseau de connaissances au Québec selon une nouvelle approche.

Nous souhaitons savoir le nombre total de personnes que vous connaissez parmi les différentes catégories ci-dessous, **peu importe si elles ont eu la COVID-19 ou non**.

Il est tout à fait possible que vous ne connaissiez **aucune personne** correspondant aux questions ci-dessous, si tel est le cas, **inscrivez "0"** pour pouvoir poursuivre le sondage.

Q8B – I. Selon votre meilleure estimation, combien vous connaissez

		Nombre (0 à XXX)
B	de personnes, adultes et mineures, dont le nom de famille est Tremblay?	
C	de personnes, adultes et mineures, dont le nom de famille est Gagnon?	
D	de personnes qui exercent le métier d’infirmières/infirmiers (incluant les auxiliaires)?	
E	de personnes qui exercent le métier d’ingénieurs?	
F	de personnes qui exercent le métier d’avocat?	
G	de femmes ayant accouché dans les 12 derniers mois?	
H	de personnes qui résident dans une résidence pour personne âgée (RPA), en ressource intermédiaire ou familiale (RI-RTF) ou en CHSLD?	
I	de personnes, adultes et mineures, dont la langue maternelle (première langue apprise et encore comprise) n’est ni le français ni l’anglais?	

Q9. Quel est votre plus haut niveau de scolarité atteint?

1. Aucun certificat ou diplôme
2. Diplôme d’études secondaires (DES) ou attestation d’équivalence
3. Certificat ou diplôme d’apprenti ou d’une école de métiers
4. Certificat ou diplôme d’un collège, d’un cégep ou d’un autre établissement non universitaire
5. Certificat ou diplôme universitaire inférieur au baccalauréat
6. Diplôme universitaire (Baccalauréat, Certificat ou diplôme universitaire supérieur au baccalauréat, Grade en médecine, en dentisterie, en médecine vétérinaire ou en optométrie, Maîtrise et Doctorat acquis)

Q10. Êtes-vous né(e) au Canada?

1. Oui **CONSIGNE: aller à Q12**
2. Non

Q11. Depuis combien d'années vivez-vous au Canada?

1. Moins de 5 ans
2. 5 à 10 ans
3. Plus de 10 ans
8. Ne sait pas

Q12. Quelle langue parlez-vous le plus souvent à la maison?

1. Français
2. Anglais
3. Autre

Méthodes d'estimation du nombre de cas de COVID-19 dans la population générale au Québec

AUTEURS

Denis Hamel, statisticien
Jérémy Sylvain-Morneau, statisticien
Bureau d'information et d'études en santé des populations, INSPQ

Radhouene Doggui, épidémiologiste
Direction des risques biologiques, INSPQ

SOUS LA COORDINATION DE

Marie-Claude Gariépy, chef d'unité scientifique
Direction des risques biologiques, INSPQ

REMERCIEMENTS

Pour la mise en œuvre du questionnaire et de la collecte en ligne :

Miguel Breault-Mallette, technicien de recherche
Bureau d'information et d'études en santé des populations, INSPQ

Josiane Rivard, technicienne de recherche
Centre de recherche du CHU de Québec — Université Laval

Pour du soutien logistique :

Maude Dionne, conseillère scientifique
Direction des risques biologiques, INSPQ

RÉVISEUSE

Céline Plante, conseillère scientifique
Bureau d'information et d'études en santé des populations, INSPQ

MISE EN PAGE

Sylvie Lafond, agente administrative
Direction des risques biologiques, INSPQ