



Conseil national
de l'information statistique

COMITÉ DU LABEL
DE LA STATISTIQUE PUBLIQUE

Montrouge, le 12 mai 2020
N°2020_10492_DG75-L002

AVIS D'EXAMEN

Enquête EPI Idémiologie et Conditions de Vie liées au COVID 19

EpiCov - vague T1

Service producteur : Co-Maîtrise d'ouvrage Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) - Inserm (Centre d'épidémiologie et de santé des populations (CESP), Inserm U 1018, 94800 Villejuif).

Opportunité : Décision d'opportunité émise le 14 avril 2020, dans le cadre de la procédure d'urgence prévue en application de l'article 2 du décret n° 2009-318 du 20 mars 2009 relatif au Conseil national de l'information statistique, au comité du secret statistique et au comité du label de la statistique publique – commission « Démographie et questions sociales ».

Consultation télématique du Comité du label du 10 au 20 avril 2020 – commission « Ménages ».

Commission	Ménages
Type d'avis	Avis d'examen
Label d'intérêt général et de qualité statistique	Non
Caractère obligatoire	Non
Période de validité	Avril-mai 2020
Visa pour parution JO	Oui (vague T1)

Partie descriptive

Porté par l'Inserm et la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) du ministère des Solidarités et de la Santé, en lien avec leurs partenaires (Insee, Santé publique France, CNRS, Ined, Université Paris-Saclay), le projet EpiCov est une large étude épidémiologique, adossée à une grande enquête statistique, proposant de fournir une cartographie du statut immunitaire de la population et de sa dynamique, sur l'ensemble du territoire, via la collecte d'échantillons biologiques couplée à des questionnaires.

L'objectif du projet EpiCov est de fournir d'une part, une cartographie précise du statut immunitaire de la population sur des paramètres de santé, de conditions de vie et d'inégalités sociales, et d'autre part, un suivi de la dynamique épidémique à court, moyen et long terme. Le déploiement dans des délais exceptionnellement rapides d'une cohorte de surveillance épidémique à grande échelle et statistiquement représentative à l'échelon départemental permettra en particulier de nourrir les modélisations de l'épidémie.

Le projet repose sur une grande enquête nationale auprès d'un échantillon représentatif, sélectionné aléatoirement par l'Insee, d'environ 350 000 personnes de 15 ans ou plus (pour obtenir 200 000 répondants), résidant sur le territoire (France métropolitaine, Martinique, Guadeloupe et La Réunion). Les personnes sont invitées à répondre à un questionnaire (en ligne ou téléphonique) d'une durée de 20 à 30 minutes et, en parallèle, pour 100 000 d'entre elles qui l'accepteront, à réaliser à leur domicile un prélèvement de quelques gouttes de leur sang, qui sera renvoyé par la Poste pour déterminer s'il y a eu contact avec le virus. Les réponses seront traitées de manière confidentielle par les équipes de recherche, dans le respect de la réglementation en vigueur (Secret statistique, CNIL, RGPD).

Ce projet permettra d'éclairer les dimensions spatiales, temporelles, socio-démographiques et familiales de l'épidémie et des mesures de confinement. Il ambitionne de décrire la fréquence des symptômes du Covid-19, de nourrir les modélisations de l'évolution de l'immunité en population, de fournir des estimations de la fréquence de l'exposition au virus en incluant les formes asymptomatiques pourvoyeuses de transmission, et d'évaluer l'évolution de la santé, du bien-être et des comportements des personnes vivant en France. Il inclut des sous-groupes définis géographiquement ou socio-démographiquement. L'ensemble des départements français (à l'exception de la Guyane et de Mayotte) sont inclus dans l'enquête.

La collecte des données est prévue du 30 avril au 24 mai pour la première vague.

Des premiers résultats à partir des réponses aux questionnaires devraient être disponibles à l'échelle nationale à la fin du mois de mai.

Les premiers résultats sur les prélèvements biologiques seront donnés au plus tôt, en fonction de la disponibilité de tests sérologiques qualifiés, actuellement en cours de développement, et de la capacité des plates-formes de test, à partir de fin mai.

Un comité de liaison collectivités-chercheurs sera mis en place pour présenter l'étude, discuter les premiers résultats et leurs conséquences, et diffuser l'information concernant des études complémentaires lancées sur les territoires.

Une deuxième vague se déroulera en juin. Cette opération pourra être répétée régulièrement pour suivre la dynamique de l'épidémie et l'évolution des conditions sanitaires et sociales dans le pays, dans tous les territoires, toutes les classes d'âge, tous les grands groupes sociaux.

~~~~~

***Le Comité du label de la statistique publique émet les recommandations ou observations suivantes :***

#### **Remarques générales**

- Le Comité du label a instruit le dossier EpiCov dans un contexte très particulier en termes de délais et de procédure dématérialisée, sur la base de documents évolutifs au fur et à mesure de l'instruction par les services. Il a décliné les principes généraux de son examen en aménageant ses méthodes d'instruction au contexte de l'urgence de l'enquête. Cette méthode et l'absence d'interaction approfondie avec les services ne suffisent cependant pas à garantir un examen aussi détaillé que dans le cadre d'une procédure standard, ce qui ne permet pas la délivrance d'un avis de conformité, en l'état du dossier présenté au Comité.
- Le Comité du label demande à pouvoir procéder à un examen des opérations de collecte de la vague T2 (et suivantes), selon ses procédures habituelles.

#### **Concertation**

- Le Comité du label a pris note de la concertation menée pour la définition des objectifs de l'enquête et des thématiques retenues. Cette concertation a associé nombre de chercheurs : ont ainsi été associées des équipes de l'Inserm sur l'épidémiologie des maladies infectieuses, la santé mentale et les inégalités de santé, les équipes de l'Ined sur les questions relatives à

la situation professionnelle, la vie quotidienne et affective et les inégalités sociales de santé, notamment celles liées à l'origine, les équipes du CNRS sur les problématiques de confiance dans les discours publics et de suivi scolaire des enfants et de l'Université Paris-Dauphine pour le recours et le renoncement aux soins. Hors sphère recherche, la concertation auprès d'autres acteurs comme les partenaires sociaux ou d'autres acteurs institutionnels a cependant été très limitée, faute de temps.

S'il comprend la nécessité d'un tel fonctionnement pour la vague T1, le Comité du label souhaite que les services élargissent la concertation pour la vague T2, tout en gardant un objectif prioritaire de comparaison entre les deux vagues pour les variables principales de l'enquête, qu'il conviendra d'apprécier en évolution, donc avec un questionnaire stabilisé autant que faire se peut. Il pourra cependant y avoir des évolutions justifiées par certains défauts constatés lors de la 1<sup>ère</sup> vague, voire compte tenu de l'évolution très rapide de la situation sanitaire, sociale ou économique, qui pourraient justifier de faire évoluer le dispositif dès la 2<sup>ème</sup> vague. Le sujet des vagues ultérieures, après la vague T2, est à cet égard encore plus ouvert et doit être anticipé par les services pour l'ouvrir à la concertation.

## Méthodologie

- Le Comité du label souligne la clarté des documents relatifs à l'échantillonnage de l'enquête et prend acte de l'apport de la mobilisation très active de la division Sondages de l'Insee. Sous certaines hypothèses de taux de réponse, les intervalles de confiance ont été estimés pour calibrer la taille de l'échantillon de façon à satisfaire les objectifs fixés en termes d'estimation de la prévalence du Covid-19, à différents niveaux géographiques. La stratégie de tirage de 20 sous-échantillons de type « modèle réduit » lui apparaît comme une bonne façon de gérer les incertitudes à venir et de pouvoir s'y adapter.
- La stratégie d'échantillonnage de la partie « tests sérologiques », dont le volume est contraint par la capacité de tests, est par contre moins claire à ce stade et devra être précisée (ce point doit aussi être mieux traité dans le protocole de l'enquête, lors du recueil de l'accord éventuel des personnes).
- Le Comité du label note que les personnes sont interrogées à leur adresse au moment de l'enquête (qui peut différer de celle au moment de l'échantillonnage), la représentativité étant traitée par calage sur la population à la date de l'enquête. L'échantillon tiré, représentatif de début 2018 (date de la base de sondage Fidéli), sera recalé sur la population 2020, ce qui conduit à faire l'hypothèse a priori raisonnable qu'il n'y a pas de spécificité des entrants entre 2018 et 2020.
- Le Comité du label alerte sur la possibilité que certaines personnes soient interrogées au titre de Sapris et d'EpiCov et demande aux services de s'assurer de la façon dont ce problème sera géré le cas échéant.
- Le Comité du label note que différentes pistes de correction de la non-réponse sont envisagées et que les informations disponibles (base de sondage Fidéli avec des informations de niveau de vie notamment, informations sur la probabilité d'être contacté, facteurs contextuels, etc.) permettront de mettre en œuvre ces méthodes. Il encourage les services à dégager les moyens adéquats sur ce point, compte tenu notamment des échéances très volontaristes en termes de diffusion des premiers résultats. Il souhaite être destinataire de la documentation de la méthode qui sera mise en œuvre, l'enjeu étant d'être représentatif en population générale. Celle-ci devra être précisément définie dans les publications (hors Ehpad, etc.).
- Le Comité du label identifie deux biais de sélection potentiels. D'une part, il y a les personnes pour lesquelles on dispose de coordonnées de contact téléphonique ou mél, mais qui ne répondent pas spontanément au questionnaire auto-administré ; d'autre part, celles qui ne sont joignables que par courrier, faute d'autre coordonnée de contact.
  - S'agissant du premier risque, le comité note que les relances ne seront effectuées que sur une partie de l'échantillon (10 à 15 %), en raison de contraintes de moyens enquêteurs chez le prestataire sur les 15 jours de collecte, ce qui peut permettre de redresser le biais de non-réponse spontanée. Les pondérations de ces individus devraient cependant être ramenées à l'échantillon total, ce qui accroît sensiblement la variance des estimateurs.

- S'agissant du second risque, les personnes ne peuvent être contactées que par courrier. Or les délais postaux étant notablement rallongés pendant la crise sanitaire, ces personnes peuvent répondre en retard à l'enquête. Se pose la question de la date de fin de collecte pour ces personnes qui sont potentiellement atypiques. Le Comité du label suggère de procéder en deux temps pour permettre la prise en compte de ces éventuels retardataires. Cette intégration des répondants tardifs en vague 1 nécessitera de produire deux jeux de pondérations (pondérations provisoires, puis définitives).
- Enfin, le Comité du label note qu'il est prévu des appariements ultérieurs des réponses à l'enquête avec le SNDS (second semestre 2020). Il souligne l'intérêt de cette démarche. Il suggère de mobiliser ces possibilités d'appariements sur une durée plus longue, de façon à apprécier des conséquences de plus long terme de l'infection par le virus.

## Protocole

- Le Comité du label n'a pas été destinataire du cahier des charges du prestataire, notamment des consignes fixées et du recueil éventuel de parodontées mesurant l'intensité de l'effort de collecte des individus ne répondant pas en mode auto-administré et qui pourraient être utilisées pour des redressements.
- S'agissant des tests sérologiques, le Comité du label note que les personnes sont informées de cette possibilité et ont donc l'opportunité de donner leur accord ou non. Pour être loyal en termes de collecte, il serait utile de préciser aux personnes :
  - si elles seront destinataires des résultats individuels les concernant ;
  - que ce test est susceptible de leur être proposé s'ils donnent leur accord, mais que ce n'est pas certain, si le nombre de volontaires excède les capacités de test. Dans ce cas un courrier les informant de cette situation serait utile.

## Diffusion

- Le Comité du label a pris note des publications grand public prévues, telles que mentionnées dans la fiche du Cnis.
- Il a également pris note de l'existence d'un groupe d'exploitation, rassemblant à ce jour l'ensemble des concepteurs de l'enquête. Il est indiqué que ce groupe doit être élargi à d'autres chercheurs intéressés, sous réserve d'obtenir les autorisations nécessaires pour l'accès à ces données sensibles.  
Le Comité du label appuie cette motion d'ouverture du groupe d'exploitation, dans le respect des procédures prévues pour l'accès aux données de santé. Il appuie la nécessité que cette ouverture des bases de données se fasse dans un premier temps dans le cadre coordonné du groupe d'exploitation, sur la base d'une documentation qui peut être construite et partagée par les membres du groupe. S'agissant de l'intégration des nouveaux chercheurs au groupe d'exploitation, le Comité du label souhaite disposer d'informations plus précises sur le calendrier de cet élargissement, sa publicité et son fonctionnement.

## Questionnaire

- Globalement, et compte tenu des délais, le Comité du label approuve le choix d'avoir utilisé le questionnaire Sapis comme support du questionnaire EpiCov (il aurait cependant souhaité disposer d'une synthèse des écarts entre les deux questionnaires). Ce choix lui paraît être la seule option possible pour pouvoir démarrer une telle enquête dans des délais aussi contraints. En conséquence, le Comité du label a *de facto* limité ses remarques à celles qui lui paraissent les plus structurantes ou qui portent sur l'interaction avec les enquêtés dans le contexte d'EpiCov.
- Le fait de distinguer un questionnaire court et un questionnaire long (environ 10 % de l'échantillon total, soit environ 20 000 répondants) semble adapté aux objectifs, selon le degré de précision attendu pour les différents types d'indicateurs.

- Sur le fond :
  - Le Comité du label a noté dans la version initialement transmise qu'une partie importante des questions de la partie B portait sur les symptômes ressentis au cours des 15 derniers jours, c'est-à-dire environ 4 semaines après le début du confinement (démarrage de la collecte envisagé au plus tôt le 27 avril). Ces questions ont ensuite été modifiées pour porter sur l'ensemble de la période depuis le début du confinement.
  - Le Comité du label avait compris initialement que la dimension « ménage » sur un sous-échantillon spécifique serait mise en œuvre dès la vague T1. En réalité, ce sous-échantillon sera mis en œuvre uniquement en vague T2. Ceci pose la question du rajout de certaines questions sur les symptômes ressentis par les personnes cohabitant au sein du ménage. Cette suggestion de rajout n'a cependant pas été retenue pour des contraintes de durée du questionnaire. Le Comité prend note de cette décision pour la vague T1 et souligne la nécessité que la dimension « ménage » soit bien prise en compte dans la vague T2 de l'enquête, ainsi que la thématique articulation de la vie familiale et de la vie professionnelle.
  - Le Comité du label a regretté qu'il n'y ait pas davantage d'informations sur le degré d'anxiété par rapport à la situation économique du ménage, qui peut expliquer une partie des comportements étudiés. Le Comité a pris note de la décision des services de ne garder que l'échelle de dépression disponible dans d'autres enquêtes en population générale eu égard à la priorité donnée aux questions pour lesquelles une référence avant le confinement était disponible.
  - Le dossier initial présenté à la commission Ménages ne faisait état d'aucune justification concernant la collecte des données relatives à la vie sexuelle, à la consommation d'alcool, de tabac ou de drogues. La Cnil demande de justifier ces questions au regard de la finalité de l'enquête, ce qui n'est pas le cas en l'état du dossier ou de la présentation en opportunité au Cnis. Des précisions ont été apportées par le service sur l'intérêt de ces questions. Le Comité du label demande à ce que ces explications soient portées à la connaissance des enquêtés, par exemple par le biais du site de collecte.
  - Par ailleurs, il est demandé aux enquêtés d'indiquer le motif de leur dernière sortie, y compris lorsque ce dernier n'est pas visé dans l'attestation de sortie dérogatoire. La Cnil conseille de ne pas recueillir le motif précis de la sortie et de ne proposer, en plus des motifs visés dans les attestations, que la réponse « autre motif ».
  - Le Comité du label a pris note du rajout de questions permettant de cerner l'origine migratoire des personnes enquêtées, rajout effectué tardivement à l'initiative des services. L'objectif est de pouvoir caractériser les personnes issues de l'immigration, qui représentent les groupes sociaux potentiellement les plus touchés par l'épidémie de Covid-19. Le Comité du label a cependant conseillé aux services d'expliquer aux enquêtés pourquoi les personnes sont interrogées sur le lieu de naissance de leurs parents, ce qui ne va pas de soi au vu du thème de l'enquête. Il rappelle aussi que ce facteur devra être analysé toutes choses égales d'ailleurs.
- Autres remarques :
  - La possibilité pour les mineurs de s'opposer au droit d'accès de leurs parents relève des termes de l'article 70 de la loi Informatique et Libertés, lorsque le mineur participant à l'étude est âgé de quinze ans ou plus ce qui sera le cas en l'espèce.
  - Sur la forme, le projet de questionnaire démarre par une introduction qui est orientée recherche plutôt que statistique publique (« *Les chercheurs de l'Institut National de la Santé et de la Recherche Médicale lance une grande enquête nationale sur le coronavirus en collaboration avec le ministère des solidarités et de la santé* »). Le Comité du label suggère de revoir cette introduction qui devrait présenter les deux maîtres d'ouvrage de façon plus symétrique. De même, le terme « chercheurs » est utilisé à divers endroits du questionnaire sans mention de l'autre donneur d'ordre. Des modifications ont été apportées sur ces points.

## Lettres-avis

- Des propositions rédactionnelles ont été faites par le secrétariat du Comité du label sur la version destinée aux personnes majeures. Le Comité du label rappelle cependant que la responsabilité du respect des règles de conformité au RGPD incombe au délégué à la protection des données du ou des services.

**Le Comité du label de la statistique publique émet un avis d'examen favorable permettant la délivrance par l'Insee d'un échantillon issu de Fidéli pour réaliser l'enquête EpiCov de la Drees et de l'Inserm. Il propose d'inscrire la vague T1 de l'enquête EpiCov au programme des enquêtes de la statistique publique, avec visa ministériel.**

**Cet avis est valable pour avril-mai 2020.**

La présidente du comité du label de la  
statistique publique

A handwritten signature in black ink, appearing to read 'Nicole Roth', written in a cursive style.

Nicole ROTH