

EpiCov : cinq questions sur une enquête inédite à propos de la crise sanitaire

Publié le 4 juin 2021 sur le blog de l'Insee : <https://blog.insee.fr/epicov-cinq-questions-sur-une-enquete-inedite-a-propos-de-la-crise-sanitaire/>

Temps de lecture : 7 minutes

François Beck, Patrick Sillard, Insee



Pour mesurer la dynamique de propagation de l'épidémie du Covid 19 et son impact sur les conditions de vie de la population, l'Inserm et la Drees, avec l'appui de l'Insee et de Santé publique France, ont créé de toutes pièces une enquête *ad hoc* intitulée EpiCov. Innovante à plusieurs titres, celle-ci a notamment combiné la réponse à un questionnaire et un auto-prélèvement sanguin en vue d'un test sérologique. Quel est l'apport de l'Insee à EpiCov ? Comment la collecte des données a-t-elle été menée ? Quels sont les premiers résultats ? François Beck et Patrick Sillard présentent les acquis fondamentaux de cette enquête d'un genre nouveau pour la statistique publique.

EpiCov, c'est quoi ?

EpiCov est une enquête élaborée dès la mise en place du confinement de mars 2020 dans le but d'éclairer la dynamique de propagation de l'épidémie et son impact sur les conditions de vie de l'ensemble de la population.

Ce qu'il faut d'abord retenir, c'est qu'elle s'appuie sur une cohorte, c'est-à-dire qu'il s'agit d'une enquête au cours de laquelle les mêmes personnes ont été interrogées plusieurs fois pour suivre au fil du temps l'évolution de leur état de santé. EpiCov, pour Épidémiologie et Conditions de vie, se singularise aussi par son ampleur. Cette enquête est menée auprès de 350 000 personnes de 15 ans et plus résidant en métropole (hors Ehpad) et 21 000 autres résidant à La Réunion, en Martinique et en Guadeloupe. Cette taille d'échantillon est exceptionnelle, les autres enquêtes nationales du service statistique public reposant sur des échantillons environ dix fois moins grands, à l'exception de l'enquête Emploi en continu. L'échantillon d'EpiCov a été construit de manière à surreprésenter les zones particulièrement touchées, les départements les moins peuplés et les personnes ayant les plus bas revenus, tout simplement pour mieux documenter la propagation de l'épidémie et ses conséquences sur les conditions de vie pour ces populations et ces zones en particulier.

L'autre singularité d'EpiCov est qu'elle repose sur un dispositif d'enquête inédit dans l'histoire de la statistique publique française : au questionnaire ont été adjoints des kits d'auto-prélèvement sanguin. Ces tests sérologiques ont été réalisés par 12 400 personnes lors de la première vague (mai 2020) puis 83 800 personnes lors de la seconde (octobre 2020). Ils ont permis d'estimer la proportion de personnes ayant été en contact avec le virus et, en croisant les résultats des tests avec les données issues des questionnaires, de connaître les profils sociodémographiques des personnes contaminées, d'analyser le lien entre le résultat du test et les symptômes déclarés ou encore les conditions de confinement des personnes. Bien entendu, les réponses à cette enquête, comme à toutes les enquêtes de la statistique publique, sont traitées de manière confidentielle dans le respect de la vie privée et des droits des personnes concernées.

Quel est l'apport de l'Insee à EpiCov ?

L'Institut national de la santé et de la recherche médicale (Inserm) et le service statistique du ministère de la Santé (Drees) sont les maîtres d'ouvrage d'EpiCov. Ils se sont associés à Santé publique France et à l'Insee afin de bénéficier de leur expérience en méthodologie d'enquête. L'institut a apporté son expertise en matière de méthodes de sondages en population nationale et de méthodologie de collecte par multimode. L'Insee a ainsi réalisé la sélection de ces 371 000 personnes à enquêter. Pour ce faire, il a mobilisé [la base Fidéli](#) (Fichier démographique d'origine fiscale sur les logements et les personnes), base exhaustive sur les logements et leurs occupants. Pour envoyer les questionnaires et les tests, les statisticiens en charge d'EpiCov ont pu tirer parti des informations de contact disponibles dans Fidéli, tous les logements étant associés à une adresse postale et 70 % d'entre eux étant associés à une adresse mail et à un téléphone. Au total, environ 135 000 personnes ont répondu à l'enquête. Ensuite, les méthodologues de l'Insee ont piloté les redressements des données pour faire en sorte que les résultats produits soient bien représentatifs de la population résidant en France.

Le contexte sanitaire a-t-il influé sur la collecte des données de l'enquête ?

Les conditions particulières du confinement ont compliqué la collecte sur plusieurs aspects. Tout d'abord, la mobilisation d'enquêteurs pour mener des entretiens en face-à-face a été contrainte par les possibilités de déplacement. Comme pour les autres enquêtes du service statistique public, la collecte a été adaptée. L'équipe pluridisciplinaire d'EpiCov a fait le choix d'un protocole dit « multimode », c'est-à-dire combinant plusieurs supports de réponse, en s'appuyant très massivement sur internet. Ce protocole a également offert la possibilité de répondre par téléphone à un échantillon aléatoire des individus n'ayant pas répondu sur Internet.

L'autre difficulté a été d'ordre méthodologique. Certaines personnes sélectionnées pour l'enquête ont été particulièrement motivées pour répondre, car elles souhaitaient mieux comprendre la situation sanitaire ou simplement avoir accès à un test de dépistage à une période durant laquelle les tests étaient très peu disponibles. Cela a généré un biais dit « d'auto-sélection », en d'autres termes le fait que les personnes se sentant particulièrement concernées par le thème de l'enquête ont été plus enclines que les autres à participer. Il a ainsi fallu construire une modélisation statistique pour neutraliser ce biais.

Quels sont les principaux résultats de l'enquête EpiCov ?

L'enquête Epicov a d'abord établi que, en mai 2020, 4,5 % de la population de France métropolitaine avait déjà été touchée par le virus. La proportion de personnes testées positives variait de 1,5 % en région Bourgogne-Franche-Comté à 6,7 % dans le Grand Est et 9,2 % en Île-de-France. Lors de cette première vague, le virus a touché en premier lieu les 30-49 ans (6,5 %) tandis que seuls 1,3 % des personnes âgées de plus de 65 ans (hors Ehpad) avaient été infectées. Vivre dans une commune à forte densité urbaine, exercer une profession dans le domaine du soin ou vivre avec un nombre élevé de personnes dans le même logement ont été des facteurs associés à un risque plus élevé d'avoir un test positif. Ce premier volet de l'enquête Epicov a également permis de constater un effet cumulatif des inégalités sociales : les personnes qui habitent dans des communes de forte densité, dans des logements surpeuplés et qui ont travaillé à l'extérieur lors du premier confinement, qui constituent les groupes sociaux les plus concernés par le risque d'exposition, sont aussi celles qui ont été les plus contaminées par le virus.

Quels sont les enseignements à tirer de cette enquête pour l'Insee ?

EpiCov a constitué un puissant catalyseur d'innovations méthodologiques. Très rapidement, nos équipes ont mis en évidence un biais d'auto-sélection combiné à un effet de mode concernant les symptômes déclarés par les enquêtés : les personnes répondant sur internet ont déclaré plus de symptômes que celles répondant par téléphone, à caractéristiques sociodémographiques égales. Neutraliser ce biais constituait un défi car cette correction était inédite à l'Insee et très peu documentée dans la littérature spécialisée. En s'appuyant sur les différences observées entre les individus ayant répondu sur Internet et ceux ayant répondu par téléphone, l'Insee a développé [une méthode permettant de corriger efficacement l'effet de mode comme le biais d'auto-sélection.](#)

De façon générale, ce protocole d'enquête, ayant fait intervenir un échantillon massif et un questionnement multimode à la fois auto-administré (internet) et intermédié (téléphone), sera source d'enseignements opérationnels importants pour le développement des enquêtes multimodes au sein de la statistique publique. Il a, par exemple, permis de confirmer l'intérêt de l'approche séquentielle dans les enquêtes multimodes, en d'autres termes le fait de proposer des modes différents de collecte au fur et à mesure de l'avancée de l'enquête pour limiter la non-réponse. Cette approche nous a conduit à optimiser le recours à internet (donc à minimiser les coûts de l'enquête) sans pour autant dégrader le taux de réponse global par rapport au multimode concurrentiel, lequel consiste à offrir dès le début de la collecte le monde de son choix à l'enquêté.

Par ailleurs, l'expérience d'EpiCov met en évidence la nécessité de développer des sites de collecte des données *responsive*, autrement dit adaptés à tous les supports de consultation. En effet, parmi les enquêtés ayant répondu sur internet, 20 % l'ont fait sur un smartphone et 7 % sur une tablette, pratiques qui, de surcroît, sont avant tout le fait des catégories d'individus globalement moins enclins à participer à nos enquêtes (les jeunes en général, les jeunes actifs en particulier).

Enfin, avec EpiCov, nous avons enrichi nos protocoles de relance, ces procédures qui nous permettent de solliciter à nouveau les enquêtés qui n'auraient pas répondu. Nous avons eu recours aux SMS, pratique encore peu répandue à l'Insee mais qui se développe notamment grâce aux données de contact présentes dans la base de données Fidéli. Nous avons également eu recours à des messages vocaux préenregistrés qui n'ont pas posé de problème particulier et qui présentent l'avantage d'être très peu coûteux.

L'ESSENTIEL :

- L'enquête EpiCov, pour Épidémiologie et Conditions de vie, a été menée auprès d'environ 135 000 personnes de plus de 15 ans vivant en France métropolitaine, à la Réunion, en Martinique et en Guadeloupe.
- Cette enquête a combiné l'envoi d'un questionnaire et un auto-prélèvement sanguin en vue d'un test sérologique pour rendre compte de la propagation de l'épidémie de la Covid-19 et de ses conséquences sur les conditions de vie des personnes.
- Comportant de nombreuses innovations dans le recueil comme le traitement des données, EpiCov constitue une étape importante dans le développement du multimode, c'est-à-dire la mise à disposition de plusieurs canaux de réponse, pour les enquêtes auprès des ménages de l'Insee en particulier, et des enquêtes de la statistique publique en général.

Pour en savoir plus :

- [Plateforme de collecte dédiée à l'enquête EpiCov](#)
- [Enquête EPIdémiologie et Conditions de Vie liées au COVID 19 – EpiCov – vague T2 – 2020X711SA](#), Cnis
- Warszawski J, Bajos N et al., Meyer L., de Lamballerie X., Seng R., Beaumont A., Slama R., Rahib D., Lydié N., Hisbergues M., Legendre B., Barlet M., Rey S., Raynaud P., Leduc A., Costemalle V., Beck F., Legleye S., Castell L., Favre-Martinoz C., Paliod N., Sillard P. [« En mai 2020, 4,5 % de la population en France métropolitaine a développé des anticorps contre le SARS-CoV-2 »](#), DREES – INSERM, *Études et résultats* n° 1167, octobre 2020

- Bajos N, Warszawski J et al., « [Les inégalités sociales au temps du COVID-19. Questions de santé publique](#) », octobre 2020
- Pauline Givord, Julien Silhol (Insee). Avec la contribution de François Beck, Laura Castell, Cyril Favre-Martinoz, Stéphane Legleye, Nicolas Paliod, Patrick Sillard (Insee), Muriel Barlet, Aude Leduc, Philippe Raynaud (Drees), « [Confinement : des conséquences économiques inégales selon les ménages](#) », *Insee Première* n° 1822, octobre 2020
- François Beck, « [La plateforme Coleman remplit plusieurs fonctions permettant de collecter des données sur le Web dans le cadre des enquêtes ménages.](#) », Insee, rapport d'activité 2019
- Jean-Luc Tavernier, « [La statistique publique à l'épreuve de la crise sanitaire](#) », Blog de l'Insee, 6 mai 2020
- Patrick Sillard, Sébastien Faivre, Nicolas Paliod, Ludovic Vincent, « [Pour les enquêtes auprès des ménages, l'Insee rénove ses échantillons](#) », *Courrier des statistiques* n° 4, juin 2020
- Laura Castel et Patrick Sillard, « [Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman](#) », Documents de travail n° M2021/02, mars 2021

Source : Insee, Blog de l'Insee <https://blog.insee.fr/epicov-cinq-questions-sur-une-enquete-inedite-a-propos-de-la-crise-sanitaire/>