

APPRENDRE DES PARADONNÉES POUR AMÉLIORER LES PROTOCOLES DE COLLECTE : L'EXEMPLE D'EPICOV

Marine Bondon (), Géraldine Charrance (*), Paul Cochet (*), Aude Leduc (**),*

Thomas Merly-Alpa ()*

() Ined, Service des Enquêtes et Sondages*

*(**) Insee, Division Logement*

marine.bondon@ined.fr, geraldine.charrance@ined.fr, paul.cochet@ined.fr,
aude.leduc@insee.fr, thomas.merly-alpa@ined.fr

Mots-clés : Paradoxonnées, collecte, multimode, Fidéli, Covid-19

Domaine concerné : Paradoxonnées, mesures et impact de la pandémie de Covid

Résumé

La collecte de l'enquête EpiCov (vague 1) s'est déroulée entre le 2 mai et le 1^{er} juin 2020. Elle a reposé sur un échantillon de 350 000 personnes en métropole issu de Fidéli. Deux modes de collecte ont été employés : internet complété par le téléphone pour un cinquième des lots aléatoires composant l'échantillon. Environ 130 000 personnes ont répondu en métropole dont 30 000 par téléphone. La collecte a été l'occasion de recueillir des paradoxonnées sur les deux modes de collecte : nombre de connexions au site internet de collecte, matériel utilisé, jour et heure de remplissage du questionnaire, type de numéros de téléphone disponibles, historique des appels, protocole de relance, etc.

Cet article rend compte de l'exploitation des paradoxonnées de EpiCov avec un double objectif : i) documenter les potentialités de Fidéli en tant que base de sondage pour une collecte multimode ; ii) aboutir à des préconisations pour les protocoles de collecte d'enquêtes futures.

Pour répondre à l'objectif d'évaluation de Fidéli comme base de sondage, une analyse de la disponibilité et de la validité des coordonnées de contact présentes dans le fichier est menée et déclinée en fonction des caractéristiques sociales disponibles dans Fidéli (sexe, âge, décile de revenu, quartier prioritaire de la politique de la ville, etc.).

L'apport de l'enrichissement annuaire est également étudié, globalement et en fonction des profils socio-démographiques. Par la suite, l'article s'intéresse à la rentabilité des différents lots du protocole de collecte. Il s'attache à mettre en lien les taux de réponse selon les protocoles (monomode versus multimode), les caractéristiques sociales des enquêtés, les modes de collecte et la richesse des coordonnées de contact à disposition (adresse mél ou non, numéro de téléphone ou non, numéro de mobile ou non, etc.). Grâce à la multiplicité des coordonnées disponibles, le protocole de relance a été particulièrement étoffé dans l'enquête EpiCov (mails, SMS, messages vocaux, etc.). L'efficacité des différents moyens de relance est analysée et questionnée.

Le nombre de connexions et, pour les lots concernés, d'appels nécessaires au remplissage du questionnaire ainsi que le mode de réponse finalement utilisé sont également analysés et déclinés en fonction des protocoles et des profils sociodémographiques. Une analyse des abandons est également prévue, notamment en lien avec le type de matériel utilisé pour le remplissage du questionnaire en ligne.

En dernier lieu, un focus est fait sur les lots multimode, divisés en deux groupes. Pour certains lots, la collecte s'est déroulée selon un multimode concurrentiel où les collectes internet et téléphone ont été ouvertes dès le début. Pour un autre lot, elle s'est déroulée selon un multimode séquentiel, la collecte téléphone ayant été ouverte dans un second temps, après la collecte internet. Une comparaison est menée pour en tirer des conclusions sur la manière d'exploiter au mieux la complémentarité des modes de collecte. La taille importante de l'échantillon EpiCov permet de nuancer les résultats selon différents types de sous-populations d'étude.

Étant donné que la collecte de la vague 1 d'EpiCov s'est déroulée pendant une période très spécifique (la crise sanitaire et en particulier, en partie, pendant le premier confinement), les résultats concernant la collecte téléphonique seront, dans la mesure du possible, mis en perspective avec ceux de la partie téléphonique de la collecte de l'enquête EHIS (European Health Interview Survey), qui a eu lieu en 2019 et pour laquelle les données ont également été exploitées.

Bibliographie

[1] Warszawski, J. et al., A national mixed-mode seroprevalence random population-based cohort on SARS-Cov-2 epidemic in France : the socio-epidemiological EpiCov study, 2021, preprint MedRxiv.