



Conseil national
de l'information statistique

COMITÉ DU LABEL
DE LA STATISTIQUE PUBLIQUE

Montrouge, le 27 décembre 2023
N°2023_29578_DG75-L002

AVIS DE CONFORMITÉ

Enquête « Baromètre de Santé publique France »

Service producteur : Santé publique France

Opportunité : avis favorable émis le 9 juin 2022 par la Commission « Démographie et Questions Sociales »

Réunion du Comité du label du 8 novembre 2023 (commission « Ménages »)

Commission	Ménages
Type d'avis	Avis de conformité
Label d'intérêt général et de qualité statistique	Oui
Caractère obligatoire	Oui
Période de validité	2024
Publication JO	Oui
Périodicité	Ponctuelle

Descriptif de l'opération

L'enquête déclarative « Baromètre de Santé publique France » permet de suivre les principaux comportements, attitudes et perceptions liés aux prises de risque et à l'état de santé de la population résidant en France. Elle fournit de très nombreux indicateurs de santé, et une grande partie des informations obtenues est impossible à recueillir autrement : celles qui relèvent de comportements individuels ne donnant lieu à aucun enregistrement, mais surtout celles qui relèvent de la mesure des attitudes, des perceptions ou des opinions de la population. Le « Baromètre de Santé publique France », en explorant de tels comportements de santé et la perception des risques qui y sont associés, révèle les tendances sanitaires et sociales du moment et fournit ainsi des données indispensables à l'élaboration et l'évaluation de politiques de Santé publique. En plus d'être un outil d'évaluation, ce suivi particulier des populations fournit à l'ensemble des acteurs du champ sanitaire des données de cadrage qui vont aider à définir les actions à engager et à évaluer les actions en cours ou passées.

Cette enquête a pour objectifs de :

- Suivre les comportements de santé de la population française et leurs déterminants ;

- Renseigner les connaissances et les opinions de la population afin de documenter des enjeux de Santé publique ;
- Enrichir les connaissances sur l'état de santé de la population française.

Pour le pilote méthodologique prévu en 2023, deux échantillons indépendants seront générés : l'un par génération aléatoire de numéros de téléphone (fixe et mobile) ; l'autre par tirage au sort dans la base Fideli. À partir de 2024, tout l'échantillon sera tiré au sort dans la base Fideli. Les modalités de tirage au sort ainsi que la taille de l'échantillon nécessaire pour fournir des données régionales sont et seront discutés au préalable avec la division Sondages de l'Insee. Pour le pilote méthodologique de 2023, l'effectif prévu est de 5 000 répondants dans trois bras (génération aléatoire / Fideli téléphone / Fideli multimode). À partir de 2024, l'effectif sera de 20 000 répondants environ.

Le « Baromètre de Santé publique France », utilisé par les acteurs régionaux poursuit l'objectif de produire des estimations au niveau régional. Le recours à la base Fideli permet ainsi d'investiguer des territoires pour lesquels il est difficile d'avoir une précision suffisante au niveau régional avec la méthode de génération aléatoire utilisée jusqu'alors, la Corse par exemple. Des estimations sont également produites sur chacun des Drom hors Mayotte (à partir de 2024). Le cas de Mayotte pourrait faire l'objet d'une enquête spécifique comme ce fut le cas avec l'enquête Unono Wa Maore menée en 2018/2019. Cette éventuelle enquête devra à l'avenir être réfléchi et élaborée avec l'ensemble des partenaires institutionnels.

La comitologie autour de l'enquête repose sur :

- Un comité de conception interne à Santé publique France (constitué depuis 2021), composé des responsables scientifiques de l'enquête, de la Direction générale et des différentes Directions métiers. Il est chargé de concevoir la méthodologie et la programmation pluriannuelle des questionnaires de l'enquête, en adéquation avec les objectifs cibles de l'Agence à l'horizon 2030.
- Un comité de pilotage (en cours de constitution) auquel participeront les principaux partenaires et tutelles (SGMASS, DGS, Drees, etc.). Il est chargé de valider ou de formuler des recommandations sur la définition du périmètre, des objectifs et livrables attendus et des conditions de réalisation de cette enquête.
- Un comité d'exploitation et de valorisation, constitué des utilisateurs des bases de données de l'enquête. Il se réunit régulièrement pour échanger autour de l'exploitation et la valorisation des données de l'enquête.

La durée moyenne d'interview s'élève à 30 minutes environ.

Outre les équipes de Santé publique France, il existe de nombreux autres utilisateurs des données produites, parmi lesquels :

- Le ministère de la Santé et de la prévention et les autres partenaires institutionnels ;
- Les équipes de recherche universitaire (CNRS, Inserm) qui sont nombreuses à valoriser les données de ces enquêtes : les bases de données existantes sont en effet mises à disposition gratuitement pour des analyses complémentaires.
- La presse et le grand public notamment par l'alimentation de ces données dans Géodes, outil cartographique en open data permettant de visualiser sous forme de cartes dynamiques, de tableaux ou de graphiques l'ensemble des indicateurs de Santé publique France.

A l'avenir, les données seront mises à disposition sur des plateformes de diffusion (Quêtelet, Centre d'accès sécurisé aux données ou *Health Data Hub*).

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

Remarques générales

- Le Comité demande au service d'élargir la concertation, notamment au monde de la recherche, et de la mettre en place de façon plus précoce pour l'enquête 2026. Il rappelle que la concertation ne vise pas à prendre systématiquement en compte toutes les demandes formulées, mais à les entendre et à justifier, le cas échéant, pourquoi elles sont ou ne sont pas retenues.
- Concernant le champ de l'enquête, le Comité s'interroge sur l'opportunité :
 - d'inclure des personnes plus jeunes ;
 - d'inclure les personnes résidant dans des communautés. Elles pourraient constituer des populations aux comportements de santé spécifiques. Le Comité invite le service à s'informer sur les réflexions en cours dans la statistique publique en la matière et à se rapprocher de la Drees pour échanger sur ses pratiques.
- Le Comité alerte le service sur le taux de réponse anticipé (39 %), qui reste très en deçà des standards de la statistique publique. Pour l'examen de l'enquête 2026, il précise qu'il sera très attentif au niveau effectif de ce taux et aux mesures que le service proposera, le cas échéant, pour l'améliorer.

Méthodologie, questionnaire, protocole

* Traitements aval :

- Le Comité souligne très positivement les contacts pris par le service avec le département des Méthodes statistiques de l'Insee afin d'évaluer les effets de mode et de sélection. Il demande de poursuivre ces bonnes pratiques en les étendant aux questions de correction de la non-réponse partielle ou de la non-réponse endogène par des stratégies du type variabilité des probabilités d'inclusion. Concernant la correction de la non-réponse endogène, le Comité note que, d'après les travaux sur l'enquête Lubrizol, les données de la base de sondage ont été jugées suffisantes et que l'apport complémentaire du SNDS serait faible en comparaison du coût de mise en œuvre de l'appariement. Le Comité souligne néanmoins que la sélection endogène pourrait être liée à des variables inobservables et qu'il pourrait être intéressant de tester une stratégie du type de celle décrite par Castell et Sillard 2021¹.
- Le Comité constate des différences sensibles entre les questionnaires papier et Internet. Les conséquences de ces différences vont au-delà du traitement des effets de mode, puisque certains modules ne sont pas repris dans la version papier par exemple. Le Comité demande au service de documenter les méthodes retenues pour traiter ces différences et de les décrire dans le prochain dossier. Le service pourra également présenter à cette occasion les stratégies d'intégration des données des sous-échantillons méthodologiques dans les résultats finaux.

* Protocole :

1 Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman : <https://www.insee.fr/fr/statistiques/5237610>

- Le Comité note que les termes du cahier des charges avec le prestataire sont suffisamment « souples » pour une adaptation efficiente aux contraintes du multimode ou pour autoriser une réaction rapide à d'éventuelles dérives mises en évidence par le suivi hebdomadaire de la collecte.
- Concernant la mise en conformité avec la nouvelle réglementation Arcep, le Comité note que le protocole proposé est de commencer l'enquête téléphonique avec des numéros polyvalents validés (NPV). Ces derniers pouvant être identifiés par les enquêtés et réduire la réussite des contacts, une seconde phase s'appuiera sur un mode manuel. Le surcoût de cette dernière reste à évaluer.

* Questionnaire :

- Le Comité considère qu'une question liée au Covid aurait trouvé sa place dans le questionnaire, à l'instar d'autres épidémies évoquées dans certaines questions. Il formule en annexe quelques remarques, que le service est invité à étudier dans ses instances de concertation.
- Le Comité retient que la possibilité d'exploiter le recueil de la PCS à deux chiffres permettra d'enrichir l'étude du thème des conditions de travail dans une future édition. Les traitements de l'édition 2024 n'utiliseront, quant à eux, que la PCS à un chiffre.

Diffusion

- Le Comité invite le service à mettre en place une stratégie de communication afin d'anticiper l'accompagnement de résultats potentiellement affectés d'éventuelles ruptures de séries.
- Le Comité demande au service de communiquer et d'organiser un partage d'expérience sur l'ensemble de sa méthodologie et en particulier sur les conclusions provenant de l'exploitation des tests méthodologiques.
- Le Comité note positivement les discussions en cours avec l'équipe Progedo Quetelet pour une mise à disposition sur le site. Il encourage la mise à disposition au CASD, également envisagée.

Environnement juridique

- Le Comité note que le service est en contact étroit avec les équipes de la Cnil ; il attire néanmoins son attention sur les points suivants :
 - Le dossier ne précise pas quelles seront les modalités de conservation des données pour la réalisation de l'enquête.
 - Les données directement identifiantes doivent être traitées et transmises de façon séparée des données de santé, et enregistrées dans une base de données distincte.
 - Les courriers électroniques et SMS adressés aux participants ne doivent révéler aucune information sur l'état de santé réel ou supposé du participant auquel il est proposé de participer à l'étude.
 - Pour se prévaloir de l'anonymat d'un jeu de données, le responsable de traitement doit réaliser une analyse permettant de démontrer que ses processus d'anonymisation respectent les trois critères définis par l'avis n°05/2014 sur les techniques d'anonymisation adoptés par le groupe de l'Article 29 (G29) le 10 avril 2014. À défaut, si ces trois critères ne peuvent être réunis, une étude des risques de ré-identification doit être menée. Par conséquent, il sera nécessaire, le cas échéant, de remplacer dans le

dossier le terme « anonymisé » qui est inapproprié par les termes données « codées » ou « pseudonymisées », en précisant que les noms et prénoms des personnes ne seront pas recueillis.

- Le caractère obligatoire ou non de l'enquête doit être mentionné dès le premier contact avec les personnes afin qu'elles puissent ou non exercer leurs droits.
- Enfin, la Cnil n'étant pas partenaire de l'étude, celle-ci ne devra pas être mentionnée comme telle sur le site web dédié à cette enquête.

Le Comité du label émet un avis de conformité à l'enquête « Baromètre de Santé publique France », et par délégation du Président du Cnis, lui attribue le label d'intérêt général et de qualité statistique, avec proposition d'octroi du caractère obligatoire et visa ministériel.

Cet avis est valide pour l'année 2024.

La Présidente du Comité du label de la
statistique publique

Signé : Dominique BONNANS

* * *

ANNEXE

Remarques sur le questionnaire :

1. Sur la situation de couple (DF2, réponse 2 (page 10)) : le Comité suggère de rappeler que les réponses sont totalement confidentielles afin d'éviter une sous-déclaration des situations de couple (par infobulle pour la version web, ou par consigne enquêteur).
2. Sur le nombre et l'âge des enfants (c'est-à-dire des personnes à charge avec responsabilité forte) : considérant que sur certains modules (sommeil, santé mentale) il est important de connaître l'étendue des responsabilités familiales des interrogés, qui n'est pas la même selon l'âge et le nombre des enfants, le Comité recommande, si la question DF2=1 (page 10) de recueillir de l'information sur :
 - le nombre d'enfants ;
 - les années de naissance de ces enfants.

Ou de recueillir une information pour savoir si l'interrogé(e) a des enfants de 0-2 ans à charge (ex : problématique de santé mentale post partum)

3. Sur le module sommeil : sur SOM9, le Comité propose de préciser dans l'intitulé si sont incluses (ou non) les causes exogènes de réveil (réveil nocturne d'enfant, environnement bruyant).
4. Sur les informations permettant d'approcher le niveau de vie :
 - le Comité estime important de connaître la situation à l'emploi de la personne avec qui l'interrogé vit en couple (Si DF2=2) ;
 - le Comité approuve l'inclusion d'une question sur le niveau de vie subjectif (RS1 page 107 du questionnaire) ;
 - le Comité recommande par ailleurs de conserver à toutes les étapes d'élaboration de la base de données le décile de niveau de vie utilisé pour le tirage de l'échantillon dans FIDELI afin de mieux approcher, au-delà de la seule CSP, ce niveau de vie.