

SANTÉ  
ENVIRONNEMENT

MAI 2024

MÉTHODES

APPORT DU SNDS DANS L'ÉTUDE  
DE LA NON-RÉPONSE À UNE ENQUÊTE  
DE SANTÉ : « SANTÉ POST-INCENDIE 76 :  
UNE ÉTUDE À L'ÉCOUTE DE VOTRE SANTÉ »

## Résumé

### Apport du SNDS dans l'étude de la non-réponse à une enquête de santé : « Santé Post-Incendie 76 : une étude à l'écoute de votre santé »

L'incendie survenu le 26 septembre 2019 sur les sites des entreprises Lubrizol et NL Logistique a été à l'origine de nombreuses nuisances environnementales. Dans le cadre du dispositif d'étude « Santé Post Incendie 76 », une enquête de santé a été mise en œuvre par Santé publique France afin d'estimer le type et la fréquence de l'exposition de la population aux différentes nuisances consécutives à l'incendie, les symptômes et problèmes de santé ressentis pendant et après l'événement. Au sein de la zone exposée (Rouen et ses alentours) et d'une zone non exposée (Le Havre et ses alentours), un échantillon de 10 777 personnes a été tiré au sort par la Division Sondages de l'Insee.

Il a été estimé à partir de cette enquête que 92 % de la population adulte étudiée a ressenti au moins une des nuisances ou pollutions générées par l'accident, des odeurs le plus souvent. Les deux-tiers de la population ont rapporté au moins un symptôme qu'ils ont attribués à l'accident : stress, anxiété, angoisse, irritation des yeux et de la gorge, toux, essoufflement, maux de tête, troubles du sommeil, etc. Un an après l'incendie, la perception des nuisances et pollutions générées par l'incendie a eu un effet négatif sur la qualité de vie liée à la santé des personnes exposées.

Ces premières analyses ont été réalisées à partir des données pondérées par les poids de sondage corrigés de la non-réponse totale, à l'aide la base de sondage Fidéli (Fichier démographique d'origine fiscale sur les logements et les individus) de l'Insee. Celle-ci a permis de disposer de variables pour les répondants comme pour les non-répondants, en particulier la distance de la résidence à l'incendie et de données socio-économiques. Les poids corrigés de la non-réponse ont été utilisés afin de produire le premier rapport de l'étude. Néanmoins, l'existence d'un biais de sélection endogène était suspectée, notamment si les personnes les plus affectées par les conséquences de l'incendie étaient plus enclines à répondre que les autres. L'absence de variables de santé pour les non-répondants ne permettait pas de vérifier l'hypothèse d'une participation à l'enquête indépendante de l'état de santé. Aussi, parallèlement à la réalisation de l'enquête, une demande d'accès aux données du système national des données de santé (SNDS) des personnes tirées au sort pour l'enquête avait été faite pour réaliser un appariement.

L'objectif était de mener une analyse de sensibilité du traitement de la non-réponse à partir de la construction d'indicateurs de recours aux soins et traduisant l'état de santé, disponibles à la fois chez les répondants et les non-répondants.

Ainsi, une association significative a été mise en évidence entre certains indicateurs de santé et la participation à l'enquête, notamment en zone exposée. Les personnes ayant consulté des professionnels de santé les 12 mois précédant l'enquête ont plus souvent participé à l'enquête que les autres. À l'inverse, les personnes dont les indicateurs indiquaient une santé dégradée ou une précarité économique ont moins souvent participé. Une association significative a été mise en évidence entre les indicateurs issus du SNDS (consultations chez le médecin, affection longue durée, hospitalisation etc.) et certaines variables clés mesurées par l'enquête, en particulier le score de santé physique, le score de santé mentale, ainsi que les symptômes de santé physique et psychique post-incendie en zone exposée.

L'identification de ces associations simultanées entre indicateurs issus du SNDS, participation à l'enquête et variables d'intérêt illustre l'existence d'une sélection endogène et en partie différentielle entre zone exposée et non exposée.

Toutefois, après inclusion de ces indicateurs de santé dans les modèles statistiques permettant de corriger la non-réponse à l'enquête, on constate une grande stabilité des estimations. Des modifications mineures de ces estimations ont été constatées en cohérence avec le biais de sélection précédemment mis en évidence, mais demeure d'une ampleur limitée. L'impact sur la santé des personnes les plus proches du lieu de l'incendie serait ainsi en réalité légèrement plus faible que ce qui a été montré avec les précédentes estimations produites sans recourir au SNDS.

En conclusion, la correction de la non-réponse est améliorée si on utilise les variables issues du SNDS dans le modèle statistique. Cependant la différence dans les estimations produites se révèle très faible, questionnant la pertinence de l'effort à fournir pour mettre en œuvre et traiter ces données appariées.

**MOTS CLÉS :** ENQUÊTE DE SANTÉ ; ACCIDENT INDUSTRIEL ; MÉTHODE D'ENQUÊTE ; APPARIEMENT SNDS ; INDICATEURS SANITAIRES SNDS ; BIAIS D'AUTO-SÉLECTION ; CORRECTION DE LA NON-RÉPONSE

**Citation suggérée :** Apport du SNDS dans l'étude de la non-réponse à une enquête de santé : « Santé Post-Incendie 76 : une étude à l'écoute de votre santé ». Saint-Maurice : Santé publique France, 2024. 46 p.  
[www.santepubliquefrance.fr](http://www.santepubliquefrance.fr)

ISSN : 2647-4816 / ISBN-NET : 979-10-289-0912-3 / RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE / DÉPÔT LÉGAL : MAI 2024

## Abstract

### Contribution of France's National Health Data System in studying non-response to the health survey launched following a large-scale industrial fire in Rouen

The fire that occurred on September 26, 2019 on the sites of the Lubrizol and NL Logistique companies was the cause of numerous environmental nuisances. As part of the "Santé Post Incendie 76" study system, a health survey was implemented by Santé publique France in order to estimate the type and frequency of exposure of the population to the various nuisances resulting from the fire, symptoms and health problems experienced during and after the event. Within the exposed zone (Rouen and its surroundings) and a non-exposed zone (Le Havre and its surroundings), a sample of 10,777 people was drawn at random by the INSEE Survey Division.

This survey showed that 92 % of the adult population studied felt at least one of the nuisances or pollution generated by the accident, most often odors. Two-thirds of the population reported at least one symptom that they attributed to the accident: stress, anxiety, distress, irritation of the eyes and throat, cough, shortness of breath, headaches, sleep problems, etc. One year after the fire, the perception of the nuisance and pollution generated by the fire had a negative effect on the quality of life linked to the health of those exposed.

These first analyzes were carried out using data weighted by the survey weights corrected for total non-response, using the Fidéli sampling frame (Demographic file of tax origin on housing and individuals) of the INSEE. This made it possible to have variables for both respondents and non-respondents, in particular the distance from the residence to the fire and socio-economic data. The weights corrected for non-response were used to produce the first report of the study. However, the existence of an endogenous selection bias was suspected, particularly if the people most affected by the consequences of the fire were more inclined to respond than the others. The absence of health variables for non-respondents did not make it possible to verify the hypothesis of participation in the survey independent of health status. In addition, at the same time as the survey was carried out, a request for access to data from the French health data system (SNDS) of the people drawn for the survey was made to carry out a matching.

The objective was to conduct a sensitivity analysis of the treatment of non-response based on the construction of indicators of healthcare use and reflecting health status, available for both respondents and non-respondents.

Thus, a significant association was highlighted between certain health indicators and participation in the survey, particularly in exposed areas. People who had consulted health professionals in the 12 months preceding the survey participated in the survey more often than others. Conversely, people whose indicators indicated poor health or economic insecurity participated less often. A significant association was shown between the indicators from the SNDS (consultations to the doctor, long-term illness, hospitalization etc.) and certain key variables measured by the survey, in particular the physical health score, the health score mental health, as well as physical and psychological health symptoms post-fire in exposed areas.

The identification of these simultaneous associations between indicators from the SNDS, participation in the survey and variables of interest illustrates the existence of an endogenous and partly differential selection between exposed and non-exposed areas.

However, after inclusion of these health indicators in the statistical models making it possible to correct for non-response to the survey, we note a great stability of the estimates. Minor

changes in these estimates were noted, consistent with the selection bias previously highlighted, but remain of limited magnitude. The impact on the health of people closest to the scene of the fire would therefore in reality be slightly lower than what was shown with previous estimates produced without using the SNDS.

In conclusion, the correction of non-response is improved if we use the variables from the SNDS in the statistical model. However, the difference in the estimates produced turns out to be very small, questioning the relevance of the effort required to implement and process this matched data.

**KEY WORDS :** HEALTH SURVEY; INDUSTRIAL ACCIDENT; SURVEY METHOD;  
DATA MATCHING; NATIONAL HEALTH INDICATORS; SELF-SELECTION BIAS;  
NON-RESPONSE CORRECTION

## Auteurs

Gaëlle Pédrone, Laetitia Bénézet, Anne-Laure Perrine, Maria El Haddad, Abdelkrim Zeghnoun, Abdessattar Saoudi, Jean-Baptiste Richard

Et le groupe de travail « Santé post incendie 76 » :

Myriam Blanchard, Pascal Empereur-Bissonnet, Franck Golliot, Emmanuelle Le-Lay, Yvon Motreff et Pauline Morel

## Remerciements

Laurence Manderau-Bruno (SpF), Marjorie Boussac (Cnam), Noémie Soullier (SpF), Mélanie Martel (SpF), Chloé Vigneron (SpF), Nahida Atiki (SpF), Christophe Bonaldi (SpF)

Aux membres du comité d'appui thématique, pour l'expertise qu'ils ont apportée tout au long de ce projet

Au Groupe Santé pour leur précieuse contribution à la définition du protocole et du questionnaire et au relai de communication

# Sommaire

Résumé.....	1
Abstract.....	3
Auteurs, remerciements .....	5
<b>INTRODUCTION .....</b>	<b>7</b>
<b>MÉTHODES .....</b>	<b>10</b>
Correction de la non-réponse .....	10
Étape 1 : Modélisation de la probabilité de réponse.....	10
Étape 2 : Formation des groupes de réponse homogènes (GRH).....	11
Étape 3 : Ajustement de la pondération initiale .....	11
Distances standardisées .....	11
Appariement.....	12
Données du SNDS mises à disposition .....	12
Indicateurs sanitaires du SNDS retenus .....	12
Analyses statistiques .....	13
Logiciel.....	14
Aspects éthiques .....	14
<b>RÉSULTATS .....</b>	<b>15</b>
Description de l'appariement .....	15
Description des indicateurs SNDS.....	16
Associations entre les indicateurs SNDS et la participation à l'enquête.....	17
Associations entre les indicateurs SNDS et les variables d'intérêt de l'enquête.....	19
Synthèse des résultats .....	24
Calcul des nouvelles pondérations .....	26
Comparaison des anciennes et nouvelles pondérations .....	29
Mise à jour des principaux résultats de l'enquête avec les nouveaux poids.....	34
Analyse de sensibilité pour les non consommateurs.....	40
<b>DISCUSSION .....</b>	<b>42</b>
Forces et limites .....	43
<b>CONCLUSION.....</b>	<b>44</b>
Références .....	45

# INTRODUCTION

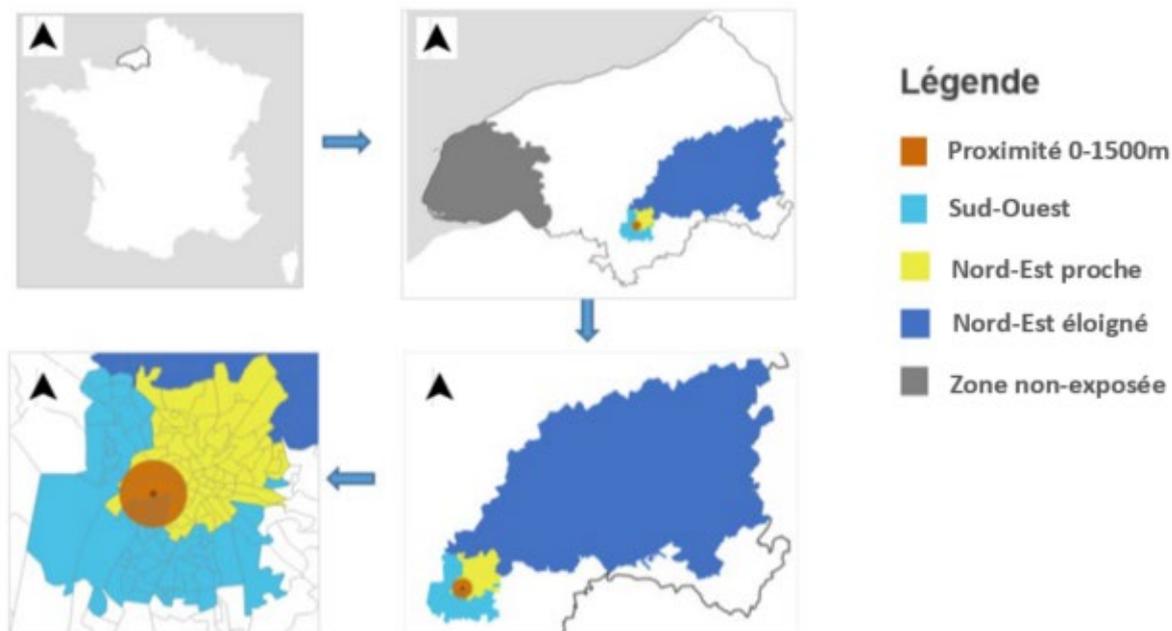
L'incendie survenu le 26 septembre 2019 sur les sites des entreprises Lubrizol et NL Logistique situées dans la zone portuaire de Rouen, a été à l'origine de nombreuses nuisances environnementales (panache de fumées noires, associé à des retombées de suie constatées au sol et des débris de toiture, émissions diffuses et odorantes de composés organiques volatils, flammes et explosions). Ces expositions, dont la nature et l'intensité ont varié selon la distance à l'incendie et au panache, en plus d'avoir pu générer un sentiment d'insécurité et un stress psychologique, peuvent avoir des conséquences sur la santé des différentes populations exposées (Direction régionale de l'environnement de l'aménagement et du logement de Normandie, 2019 ; Blanchard et al., 2021 ; Empereur-Bissonnet et al., 2021b ; Golliot et al., 2021). Dans le cadre du dispositif d'étude « Santé Post Incendie 76 » visant à évaluer l'impact sanitaire à court, moyen et long termes, une enquête de santé perçue a été mise en œuvre afin d'estimer le type et la fréquence de l'exposition de la population aux différentes nuisances consécutives à l'incendie, les symptômes et problèmes de santé ressentis pendant et après l'événement, et a également permis d'étudier l'état de santé physique et mentale perçues et la qualité de vie liée à la santé plusieurs mois après l'événement (Empereur-Bissonnet et al., 2021a).

Cette enquête par questionnaires a concerné la population résidente de la zone exposée (Rouen et ses alentours) et d'une zone non exposée (Le Havre et ses alentours), utilisée à titre de comparaison. Le questionnaire du volet principal de l'enquête s'est intéressé à la période de l'accident et à ses suites en termes d'exposition et de symptômes ressentis ainsi que la santé un an après l'incendie (incluant des informations sur la crise du covid-19). Ce volet a été complété par un volet spécifique visant à évaluer les conséquences sur différents troubles de la santé mentale, notamment l'anxiété, la dépression et le trouble stress post-traumatique. Il a été proposé dans un second temps aux personnes de l'ensemble de la zone d'étude (exposée et non exposée) ayant accepté d'être recontactées.

Un échantillon de 10 777 personnes résidant en zones exposée ou non exposée a été tiré au sort par la Division Sondages de l'Insee, dans le millésime 2019 du Fichier démographique d'origine fiscale sur les logements et les individus (Fidéli). Une stratification géographique (Figure 1) a été utilisée dans le cadre de cette enquête, comprenant les quatre strates suivantes : une strate de proximité (à moins de 1 500 mètres du lieu de l'accident), une strate Nord-Est proche (zone exposée aux odeurs et située sous le panache de fumées noires), une strate Nord-Est éloigné (zone exposée au panache et à ses retombées au sol mais peu exposée aux odeurs), une strate Sud-Ouest (non exposée au panache de fumées noires). L'objectif était d'interroger environ 1 000 personnes dans chacune des strates ainsi que dans la zone non exposée.

La collecte de l'enquête s'est déroulée du 4 septembre au 12 novembre 2020, selon un protocole de collecte multimode (Internet ou téléphone). Au final, 4 779 personnes de 18 ans et plus ont répondu à ce volet (dont 3 758 en zone exposée et 1 015 en zone non exposée), soit 95 % de l'objectif fixé de 5 000 répondants, correspondant à un taux de réponse de 47,6 %.

Figure 1. Situation du lieu de l'incendie, des zones exposée et non exposée et des différentes strates en France et en Seine-Maritime



Source : Santé Post Incendie 76 : une étude à l'écoute de votre santé. Santé publique France, 2021

Ces premières analyses ont été réalisées à partir des données, pondérées par les poids de sondage corrigés de la non-réponse totale. Le recours à la base de sondage Fidéli s'est révélé d'une grande importance pour cette étape de correction de la non-réponse, permettant de disposer de variables pour les répondants comme pour les non-répondants, en particulier la distance à l'incendie (via la disponibilité des coordonnées géographiques des logements) et de données socio-économiques comme la structure du ménage ou le niveau de vie des ménages (Richard et al., 2022). Ces poids corrigés de la non-réponse ont été utilisés afin de produire les estimations présentes dans le premier rapport de l'étude (Empereur-Bissonnet et al., 2021a).

Compte tenu du contexte de l'étude (post-accidentelle avec persistance des nuisances et inquiétude de la population), des modes de collectes retenus (internet, téléphone) et de l'interrogation d'individus dans une zone non exposée, moins concernés par l'enquête, l'existence d'un biais de sélection endogène était une des limites supposées de l'étude, c'est-à-dire que les personnes les plus concernées par les conséquences de l'incendie auraient été plus enclines à répondre que les autres. On a observé par exemple que le taux de participation était corrélé à la distance à l'incendie

(52,0 % dans la strate de proximité, 40,2 % en zone non exposée) de même que le fait de répondre par internet (67,4 % des répondants ont répondu par Internet en zone exposée, 53,1 % en zone non exposée) (Richard et al., 2022).

L'absence de variables de santé pour les non-répondants ne permettait pas de vérifier l'hypothèse d'une participation à l'enquête indépendante de l'état de santé des individus, ni de prendre en compte un tel effet dans la correction de la non-réponse. Aussi, parallèlement à la réalisation de l'enquête, un protocole avait été déposé au Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé, afin de demander l'accès aux données du système national des données de santé (SNDS) correspondant aux personnes tirées au sort pour l'enquête. Un des objectifs de cet appariement était d'utiliser ces informations auxiliaires, directement liées aux recours aux soins et à l'état de santé, disponibles à la fois chez les répondants et les non-répondants, pour mener une analyse de sensibilité du traitement de la non-réponse mis en œuvre à partir des seules données disponibles dans Fidéli.

L'objectif de ce rapport est d'étudier l'apport de l'utilisation des indicateurs issus du SNDS dans la compréhension du mécanisme de non-réponse à cette enquête de santé post-incendie.

# MÉTHODES

## Correction de la non-réponse

Le principe de la correction de la non-réponse par repondération est de modifier le poids de sondage des individus répondants pour compenser le biais introduit par les non-répondants. Le calcul des pondérations a été effectué en deux phases. La première phase a consisté à établir des pondérations initiales relatives au plan de sondage. Elle a été effectuée par l'Insee au moment du tirage au sort. La seconde phase a consisté à ajuster les poids de sondage par rapport à la non-réponse totale observée au moment de la réalisation du terrain de l'enquête, la méthode retenue étant la méthode des scores (Haziza et Beaumont, 2007).

Les étapes de mise en œuvre du traitement de la non-réponse totale par la méthode des scores sont décrites ci-dessous.

Le plan de sondage étant stratifié, avec un tirage aléatoire simple au sein de chaque strate, la pondération initiale de chaque répondant est égale à l'inverse du taux de sondage dans la strate à laquelle il appartient. La probabilité d'inclusion est calculée en prenant comme dénominateur l'ensemble des adultes de 18 ans et plus. En notant  $h$  la strate de tirage,  $n_h^A$  le nombre d'adultes tirés au sort dans la strate  $h$  et  $N_h^A$  le nombre d'adultes de la strate  $h$  dans la base de sondage, le poids de sondage initial pour les adultes de 18 ans et plus tirés dans la strate  $h$  s'écrit :

$$W_h^A = \frac{N_h^A}{n_h^A}$$

### Étape 1 : Modélisation de la probabilité de réponse

La première étape de cette méthode consiste à modéliser la probabilité de réponse  $p_i$  pour chaque individu tiré au sort dans l'échantillon initial.

On note par  $a_i$  l'indicatrice de réponse qui vaut 1 si l'individu  $i$  a répondu à l'enquête et 0 sinon, tel que :

$$a_i = \begin{cases} 1 & \text{avec une probabilité } p_i \\ 0 & \text{avec une probabilité } 1 - p_i \end{cases}$$

On estime la probabilité de réponse  $p_i$  à l'aide d'un modèle de régression logistique de la forme

$\log\left(\frac{p_i}{1-p_i}\right) = \beta X_i$  où  $X_i$  est un vecteur de variables auxiliaires disponibles pour toutes les unités de l'échantillon initial et  $\beta$  un vecteur de paramètres inconnus.

La variable dépendante est binaire (1 pour répondant et 0 pour non-répondant). Les variables indépendantes sont des variables qualitatives et quantitatives ayant potentiellement une influence sur la probabilité de réponse à l'enquête (socioéconomique). La sélection de ces variables indépendantes a été faite a priori et/ou lors d'une étape de modélisation en utilisant un critère statistique.

On obtient avec le modèle la prédiction :  $\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \hat{\beta} X_i$ , le vecteur  $\hat{\beta}$  étant obtenu par la méthode du maximum de vraisemblance. On en déduit la probabilité  $\hat{p}_i$  qui est utilisée ensuite pour former les groupes de réponse homogènes.

La correction de la non-réponse totale à partir des variables sélectionnées dans la base Fidéli a été réalisée séparément en zone exposée et en zone non exposée et a mobilisé les variables suivantes :

âge de l'individu au moment de la collecte, mode de contact au niveau ménage, niveau de vie en décile (année n-1), sexe, lien familial de l'individu échantillonné avec le référent fiscal, statut d'occupation du logement, type de ménage, perception d'un salaire par l'individu cible, distance du domicile à l'usine (négative si arrière du panache, au sud-ouest de l'usine) et lieu de naissance de l'individu échantillonné. Une évaluation des différences de structure entre échantillons a été réalisée à l'aide de distances moyennes standardisées (voir ci-dessous), par rapport à la population source, c'est-à-dire l'échantillon initial tiré au sort par l'Insee, comprenant les répondants et les non-répondants, pondéré par les poids de sondage.

## Étape 2 : Formation des groupes de réponse homogènes (GRH)

En utilisant les probabilités de réponses estimées  $\hat{p}_i$ , on forme des GRH suivant la méthode des quantiles, en répartissant les unités en un certain nombre de groupes, définis a priori. Pour ce faire, on ordonne les valeurs de  $\hat{p}_i$  par ordre croissant, puis on divise l'échantillon en classes de tailles approximativement égales. L'idée est que si les probabilités de réponse  $p_i$  sont bien estimées, alors  $\hat{p}_i \approx p_i$ , et les GRH par rapport à  $\hat{p}_i$ , le seront également par rapport aux  $p_i$ , de sorte que le biais sera proche de 0 (Deroyon, 2017).

## Étape 3 : Ajustement de la pondération initiale

Une fois que les GRH sont constitués, le taux de réponse observé dans chaque groupe est obtenu en divisant la somme des poids de sondage chez les répondants du groupe par la somme des poids de sondage des répondants et non-répondants dans le même groupe. La pondération ajustée de la non-réponse totale pour chaque répondant est obtenue en divisant sa pondération initiale par le taux de réponse pondéré observé dans son GRH.

Afin de limiter l'impact des poids extrêmes, pour chaque échantillon et après traitement de la non-réponse, des troncatures à 95 % en zone exposée et à 98 % en zone non exposée (la dispersion des poids étant moindre dans cette zone car le taux de sondage était uniforme), ont été appliquées. En l'absence de marges disponibles pour la population source, aucun calage sur marge n'a été effectué.

## Distances standardisées

Les différences entre la base de sondage et les échantillons de répondants sont évaluées à l'aide de distances moyennes standardisées (SMD). Pour chaque modalité de variable catégorielle, nous calculons une distance selon la formule ci-dessous ( $p_A$  = % dans l'échantillon A,  $p_B$  = % dans l'échantillon B,  $q_A=1-p_A$ ,  $q_B=1-p_B$ ) :

$$d_{\text{modalité}} = (p_B - p_A) / \sqrt{(p_A q_A + p_B q_B) / 2}$$

La distance standardisée D pour une variable catégorielle est la moyenne des valeurs absolues des différences standardisées de chaque modalité de la variable catégorielle. Une valeur de la distance inférieure à 10 % est considérée comme une différence mineure (Austin et al.).

Les distances standardisées ont été calculées par rapport à la population source, c'est-à-dire l'échantillon initial tiré au sort par l'Insee, comprenant les répondants et les non-répondants (sans les hors champ), pondéré par les poids de sondage.

## Appariement

Après avoir exclu les personnes s'étant opposées à l'appariement (n = 26), les services de l'Insee ont procédé à une étape d'identification des personnes tirées au sort au Répertoire national d'identification des personnes physiques (95,6 % d'individus retrouvés), puis a déposé le fichier sur la plateforme SAFE de la Caisse nationale d'assurance maladie (CNAM). L'appariement avec les données du SNDS est donc un appariement direct sur la base du numéro d'inscription au répertoire des personnes physiques (NIR). Dans un second temps, la CNAM a extrait au sein du SNDS les données listées dans l'expression de besoin correspondant aux individus réellement identifiés et ayant eu des remboursements de soins de santé sur la période autorisée, puis a mis à disposition ces données dans un espace projet dédié. La base de données anonymisée de l'enquête a ensuite été chargée par la CNAM dans le portail de l'espace projet SNDS.

## Données du SNDS mises à disposition

L'extraction correspond au référentiel des bénéficiaires (principales informations caractérisant un bénéficiaire), aux données de consommations individuelles DCIR (remboursements de consultations médicales, biologie et pharmacie, indemnités journalières), aux données d'hospitalisation issues du Programme de Médicalisation des Systèmes d'Information (PMSI) (hospitalisation en médecine chirurgie, obstétrique, hospitalisation à domicile, hospitalisation en psychiatrie, hospitalisation en soins de suite et de réadaptation), aux données médico-administratives comme les affections longue durée (ALD), ainsi qu'à la cartographie des dépenses et des pathologies (ces tables annuelles ayant pour objectif d'analyser la répartition et la dynamique des dépenses d'assurance maladie rattachées aux prises en charge, en détaillant notamment les dépenses de soins de ville et les dépenses hospitalières, en fonction des pathologies).

Les données du DCIR étaient disponibles pour 2018-2022 (jusqu'en juin), les données d'hospitalisation du PMSI étaient disponibles pour 2018-2021 et les données de cartographie des pathologies pour les années 2018 et 2019 uniquement. Au total, 549 tables et 13 171 variables ont été mises à disposition par la CNAM en septembre 2022.

## Indicateurs sanitaires du SNDS retenus

**Temporalité** : pour évaluer le biais de sélection, les indicateurs ci-dessus ont été calculés pour la période de 1 an qui a précédé la passation du questionnaire (octobre 2019-septembre 2020).

Des indicateurs binaires ont été construits à partir des données du SNDS :

- Les soins de ville :
  - Recours à un médecin généraliste ;
  - Recours à un médecin spécialiste ;
  - Recours à un dentiste<sup>1</sup> ;
  - Recours à un laboratoire de biologie ;
  - Recours à un auxiliaire de santé (kinésithérapeute, infirmier) ;
- Les indemnités journalières (arrêts de travail) ;
- Les hospitalisations ;
- Les affection longue durée.

---

<sup>1</sup> Le recours à un dentiste est couramment utilisé dans les enquêtes de santé car il s'agit d'un bon proxy du fait d'être attentif et proactif dans son suivi médical.

Un indicateur spécifique de la santé mentale a également été construit, il s'agit de la délivrance d'au moins un psychotrope (anxiolytique, hypnotique, antidépresseur).

## Analyses statistiques

Pour étudier les associations entre participation à l'enquête et indicateurs sanitaires du SNDS dans la population, des régressions logistiques pondérées sur le poids de sondage ont été réalisées dans les deux zones. Les associations des différents indicateurs SNDS avec la participation ont été testées une à une dans un modèle ajusté sur les variables Fidéli retenues pour calculer le poids de correction de la non-réponse précédemment évoqué. Puis, pour choisir des indicateurs SNDS à retenir pour le calcul d'une nouvelle pondération, les indicateurs SNDS associés significativement à la participation ont été inclus simultanément dans le modèle multivarié en vérifiant l'absence de colinéarité dans les variables du modèle.

Pour étudier les associations entre les variables d'intérêt de l'enquête (scores de santé physique et mentale, symptômes physiques, symptômes psychiques, expositions à l'incendie, consultation d'un médecin) et les indicateurs SNDS, des régressions logistiques, linéaires et ordinales ont été réalisées (selon la nature de la variable d'intérêt : binaire, continue ou ordinale). Ces régressions étaient pondérées sur le poids de correction de la non-réponse évoqué plus haut.

Au final, le choix des indicateurs SNDS à inclure dans les modèles de non-réponse dans les zones exposée et non exposée s'est porté sur les indicateurs associés à la fois à la participation à l'enquête et aux principales variables d'intérêt de l'enquête au seuil de significativité de 5 %.

Après sélection des variables à inclure dans le modèle de non-réponse, la modélisation des probabilités de réponse a été réalisée à l'aide d'une régression logistique non pondérée.

Les personnes tirées au sort ayant pu être appariées avec le SNDS avec leur numéro d'inscription au répertoire (NIR), mais n'ayant aucune consommation de soins dans l'extraction du SNDS mise à disposition par la Cnam, ont été considérées comme non consommandes (tous les indicateurs SNDS valant 0). Afin de s'assurer de la robustesse des résultats avec cette hypothèse, une imputation par Hot Deck avec un donneur (Andridge et al. 2010) des indicateurs SNDS a été réalisée pour ces 431 personnes représentant 4,3 % des personnes appariées. Les variables Fidéli retenues pour réaliser cette imputation étaient le sexe, l'âge (en classes), le salaire (en classes) et le fait d'habiter dans un logement social.

Les étapes permettant de calculer les nouvelles pondérations tenant compte des indicateurs issus du SNDS sont résumées dans la Figure 2.

Figure 2. Étapes pour corriger la non-réponse à l'aide des indicateurs issus du SNDS



L'ensemble des analyses a été réalisé sous le portail SNDS avec le logiciel SAS® (SASGUIDE 8). La modélisation des probabilités de réponse a été réalisée sous SAS à l'aide de la procédure logistic non pondérée. Les procédures suivantes ont été utilisées : surveyFreq pour décrire les variables catégorielles, surveymeans pour décrire les variables continues, surveylogistic et surveyreg pour les modèles de régression et surveyimpute pour l'imputation par Hot Deck.

## Aspects éthiques

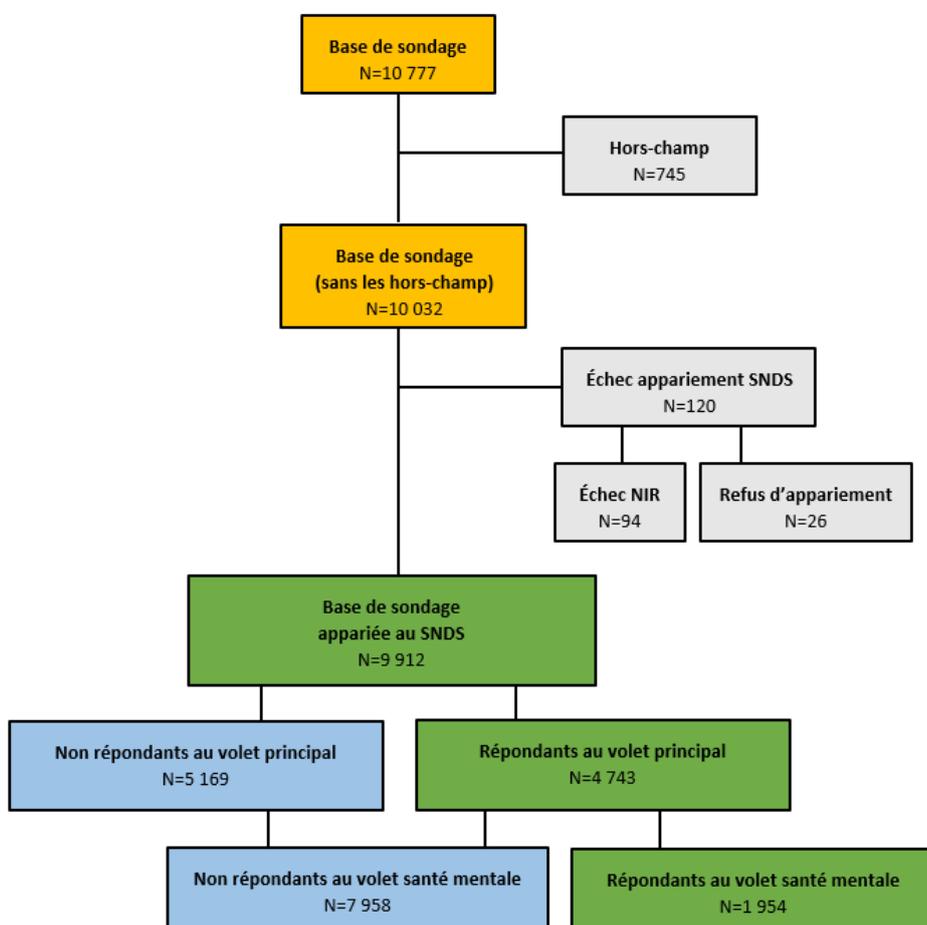
Le Comité du label de la statistique publique a donné son avis favorable à l'enquête le 11 juin 2020 et le Comité Éthique et Scientifique pour les Recherches, les Études et les Evaluations dans le domaine de la santé le 8 octobre 2020 (N° 2265921). Un avis favorable de la Commission nationale de l'informatique et des libertés (Cnil) pour l'appariement a été émis le 11 mars 2021 (N° MLD/AVL/AR212138).

# RÉSULTATS

## Description de l'appariement

Parmi les 10 032 personnes tirées au sort et considérées dans le champ de l'enquête, 9 912 personnes ont pu être appariées au SNDS (98,8 %). Seules 26 personnes se sont opposées à l'appariement avec le SNDS et pour 94 personnes, le NIR n'a pu être retrouvé par l'Insee (Figure 3). Parmi les personnes ayant pu être appariées, 95,7 % avaient des remboursements de soins pour la période 2018-2022 (les autres personnes sont considérées comme non consommandes), 47,9 % d'entre elles ont répondu au volet principal de l'enquête et 19,7 % au volet spécifique sur la santé mentale.

**Figure 3. Description des personnes tirées au sort selon le statut d'appariement au SNDS et de réponse aux deux volets de l'enquête**



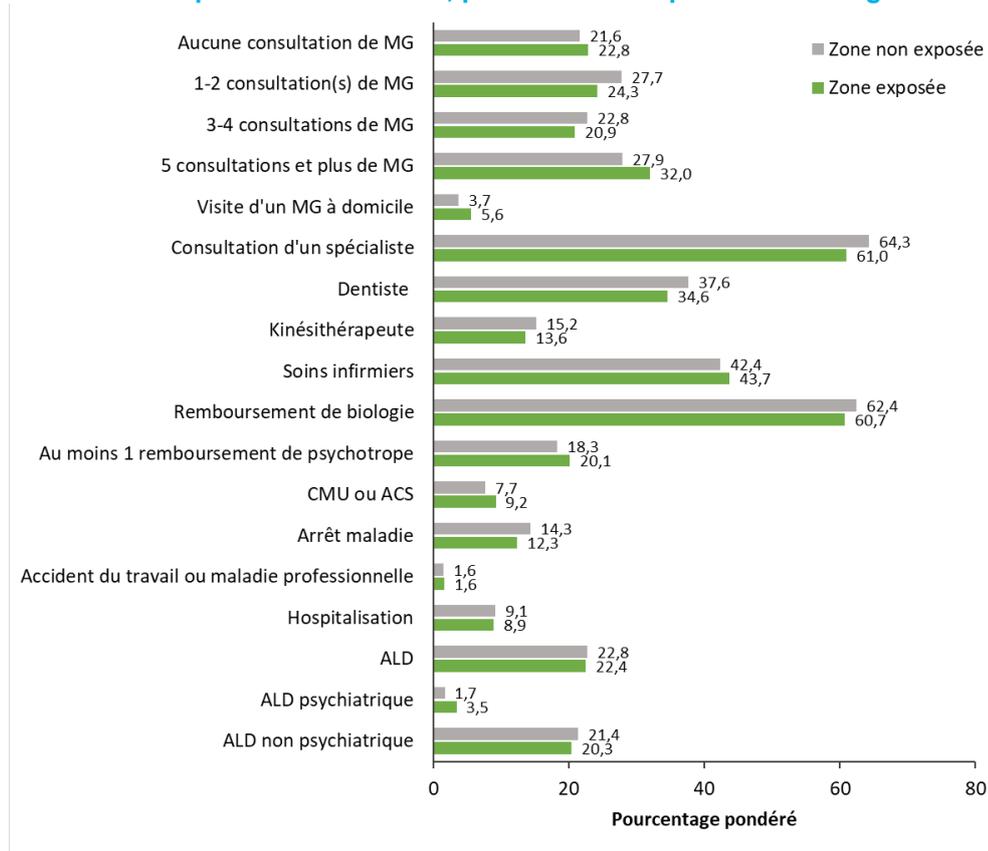
## Description des indicateurs SNDS

À partir des données extraites du SNDS, 15 indicateurs ont été calculés pour chacune des 9 912 personnes tirées au sort qui ont pu être appariées au SNDS. Tous les indicateurs qui suivent ont été calculés pour les 12 mois ayant précédé la passation du questionnaire (1<sup>er</sup> octobre 2019-30 septembre 2020) :

- Le nombre de consultations chez un médecin généraliste ;
- Avoir reçu au moins une visite à domicile d'un médecin généraliste ;
- Avoir consulté au moins une fois un médecin spécialiste ;
- Avoir consulté au moins une fois un dentiste ;
- Avoir consulté au moins une fois un kinésithérapeute ;
- Avoir reçu au moins une fois des soins infirmiers ;
- Avoir eu au moins un remboursement de biologie (laboratoire d'analyse) ;
- Avoir été remboursé au moins une fois d'une délivrance de psychotrope ;
- Être bénéficiaire de la couverture maladie universelle (CMU) ou de l'aide au paiement d'une complémentaire santé (ACS) ;
- Avoir perçu des indemnités journalières pour un arrêt maladie ;
- Avoir perçu des indemnités journalières pour un arrêt de travail ou une maladie professionnelle ;
- Avoir été hospitalisé ;
- Être en ALD ;
- Être en ALD pour pathologie psychiatrique ;
- Être en ALD pour pathologie non psychiatrique.

En zone exposée, l'année qui a précédé l'enquête, 77,2 % de la population adulte avaient consulté un médecin généraliste, 5,6 % avaient reçu une visite à domicile d'un médecin généraliste, 61,0 % avaient consulté un médecin spécialiste, 34,6 % un dentiste, 13,6 % un kinésithérapeute, 43,7 % avaient reçu des soins infirmiers, 60,7 % avaient été remboursés pour un acte de biologie, 20,1 % pour une prescription de psychotrope, 9,2 % étaient bénéficiaires de la CMU/ACS, 12,3 % avaient eu un arrêt de travail pour maladie et 1,6 % un arrêt pour accident du travail ou maladie professionnelle, 8,9 % avaient été hospitalisés, 22,4 % étaient en ALD et 3,5 % en ALD psychiatrique (Figure 4). Ces pourcentages étaient comparables pour la zone non exposée.

**Figure 4. Description des indicateurs SNDS les 12 mois qui ont précédé l'enquête, selon la zone, chez les 9 912 personnes étudiées, pondérés sur le poids de sondage**



Les indicateurs SNDS associés à la participation à l'enquête sont plus nombreux en zone exposée qu'en zone non exposée (Figure 5). Ceci est probablement lié à un manque de puissance (larges intervalles de confiance) dans la zone non exposée où il y avait 4 fois moins de personnes tirées au sort que dans la zone exposée. Les indicateurs SNDS en faveur d'une participation, en particulier dans la zone exposée, sont le fait d'avoir consulté des professionnels de santé : médecin généraliste, spécialiste, dentiste, avoir reçu des soins infirmiers ou un remboursement pour un acte de biologie. Ces indicateurs reflètent une bonne prise en charge de sa santé et des comportements de prévention.

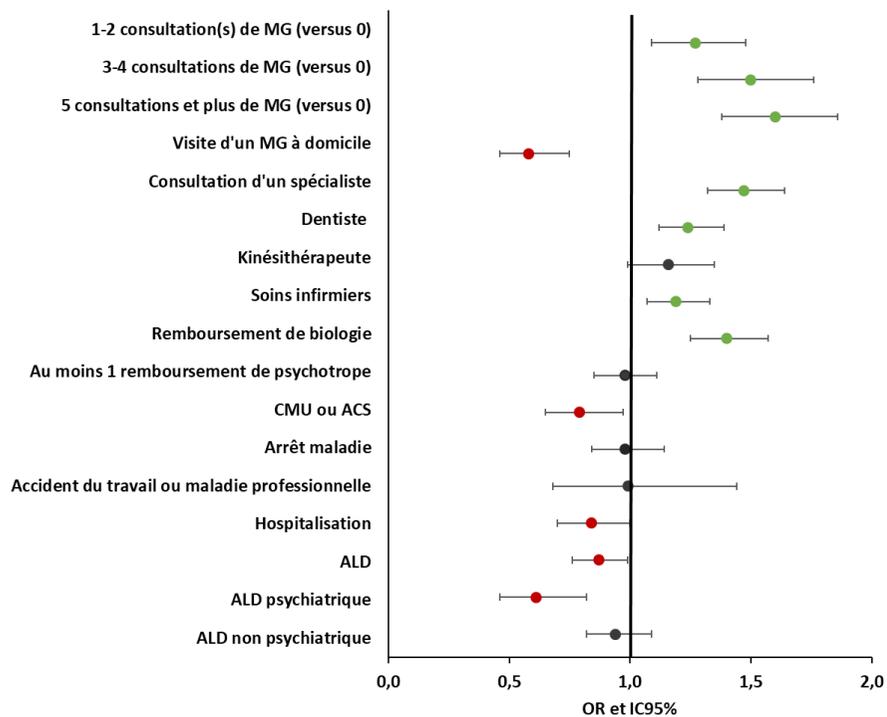
À l'inverse, les indicateurs associés à une moindre participation relèvent d'une part d'une santé dégradée (avoir eu une visite à domicile du médecin généraliste, avoir été hospitalisé, être bénéficiaire d'une ALD), d'autre part d'une précarité économique comme en témoigne l'indicateur de CMU/ACS.

La délivrance de psychotrope et les arrêts pour maladie ou accident du travail ne sont pas associés à la participation dans les deux zones.

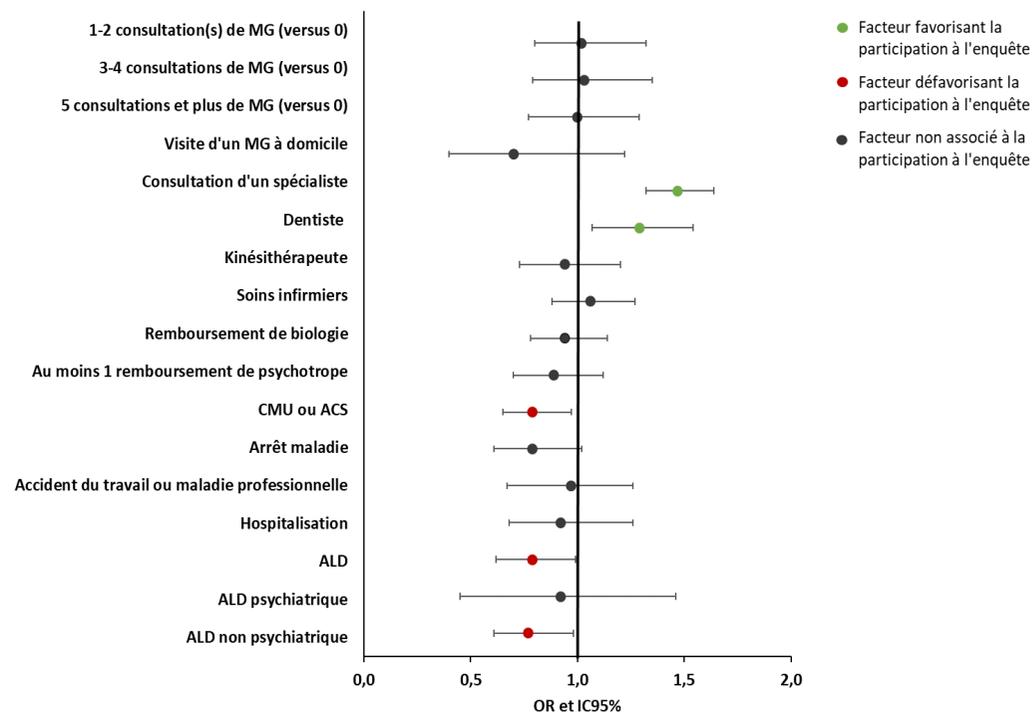
Pour tester un effet différentiel des différents indicateurs SNDS sur la participation à l'enquête, entre les deux zones, des modèles testant l'interaction entre la zone et les indicateurs ont été réalisés. Ces interactions étaient significatives pour le nombre de consultations chez un médecin généraliste ( $p = 0,015$ ) et le remboursement d'un acte de biologie ( $p = 0,001$ ). Aucune différence d'effet des autres indicateurs sur la participation entre les deux zones n'a été mise en évidence.

Figure 5. Associations entre les indicateurs SNDS et la participation à l'enquête dans les deux zones

Zone exposée



Zone non exposée



## Associations entre les indicateurs SNDS et les variables d'intérêt de l'enquête

Pour étudier les liens entre les indicateurs SNDS et les principales variables d'intérêt de l'enquête, des modèles de régression ont été réalisés : régression linéaire pour les scores de santé physique et mentale, régressions logistiques pour les variables binaires et régressions ordinales pour la variable dénombrant le nombre d'expositions. Ces modèles sont ajustés sur les variables Fidéli incluses dans les modèles de non réponse, les indicateurs SNDS étant testés un par un dans ces modèles ajustés (Figure 6 à Figure 11).

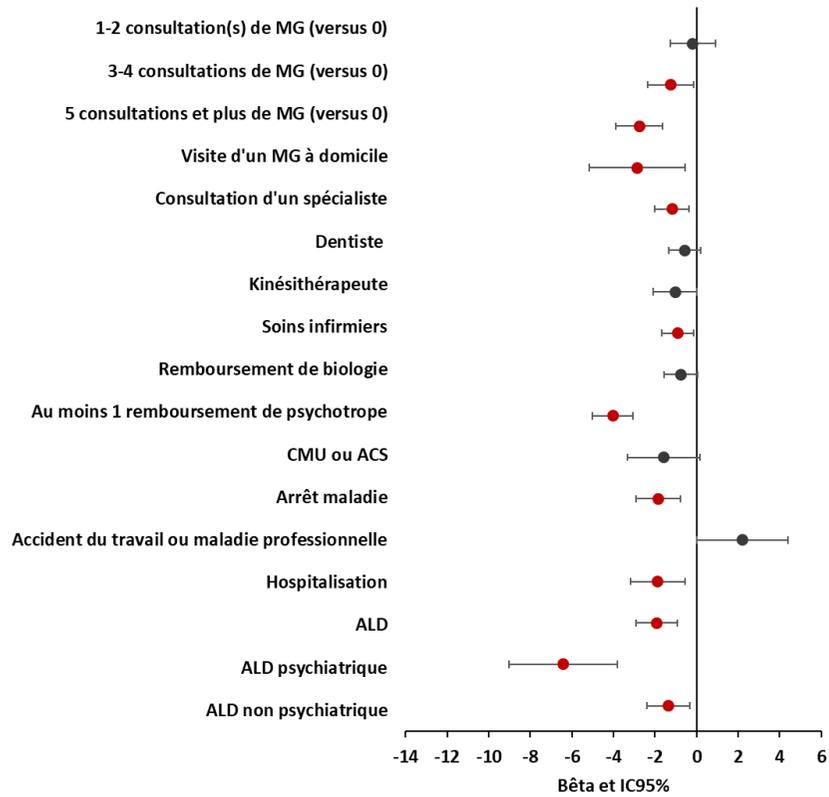
Le score de santé mentale (MCS) a été calculé avec la deuxième version de la forme courte du Medical Outcomes Study Short Form Health Survey (SF-12v2) (Ware 1996, Gandek 1998). Plus le score est élevé et meilleure est la santé mentale déclarée par le répondant. Les indicateurs SNDS associés à un moins bon score de santé mentale sont le fait d'avoir consulté un médecin (généraliste ou spécialiste), les soins infirmiers, les remboursements de psychotropes, les arrêts maladies, l'hospitalisation et le fait d'être en ALD durant les 12 mois précédant l'enquête (Figure 6).

Le score de santé physique (PCS) a également été calculé avec le SF-12v2. Plus le score est élevé et meilleure est la santé physique déclarée. Les indicateurs SNDS associés à un moins bon score de santé physique sont le fait d'avoir consulter un médecin (généraliste ou spécialiste), les consultations chez un kinésithérapeute, les soins infirmiers, les remboursements de biologie, les remboursements de psychotropes, être bénéficiaire de la CMU/ACS, les arrêts maladies, l'hospitalisation et le fait d'être en ALD (Figure 7).

En zone exposée, les symptômes psychiques post-incendie étaient associés à un nombre important de consultations chez le médecin généraliste, à un remboursement de biologie, un remboursement de psychotrope et à un accident du travail ou une maladie professionnelle durant les 12 mois précédant l'enquête (Figure 8). Les symptômes physiques post-incendie étaient associés à un nombre important de consultations chez le médecin généraliste, une consultation chez un kinésithérapeute et le remboursement de psychotrope (Figure 9). Le nombre d'expositions post-incendie déclarées n'était pas associé aux indicateurs sanitaires, à l'exception des arrêts maladie (moins d'expositions déclarées chez les personnes ayant eu un arrêt maladie, Figure 10). Enfin, le fait d'avoir consulté un médecin suite aux symptômes post-incendie était associé aux consultations multiples chez un médecin généraliste, une consultation d'un spécialiste, d'un kinésithérapeute, aux soins infirmiers, aux remboursements de biologie ou de psychotropes, au fait d'être bénéficiaire de la CMU/ACS et être en ALD non psychiatrique (Figure 11).

**Figure 6. Associations entre les indicateurs SNDS et le score de santé mentale (MCS) de la population dans les deux zones (N = 3 758 en zone exposée et N = 1 015 en zone non exposée), Bêta étant le coefficient de régression associé à chaque variable**

Zone exposée



Zone non exposée

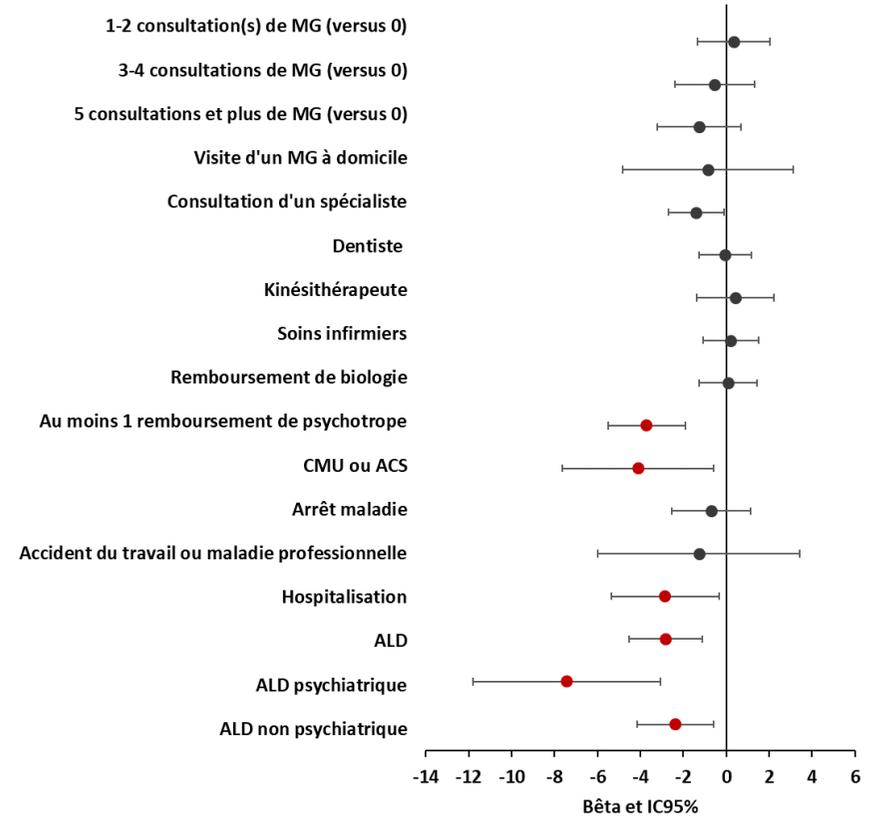
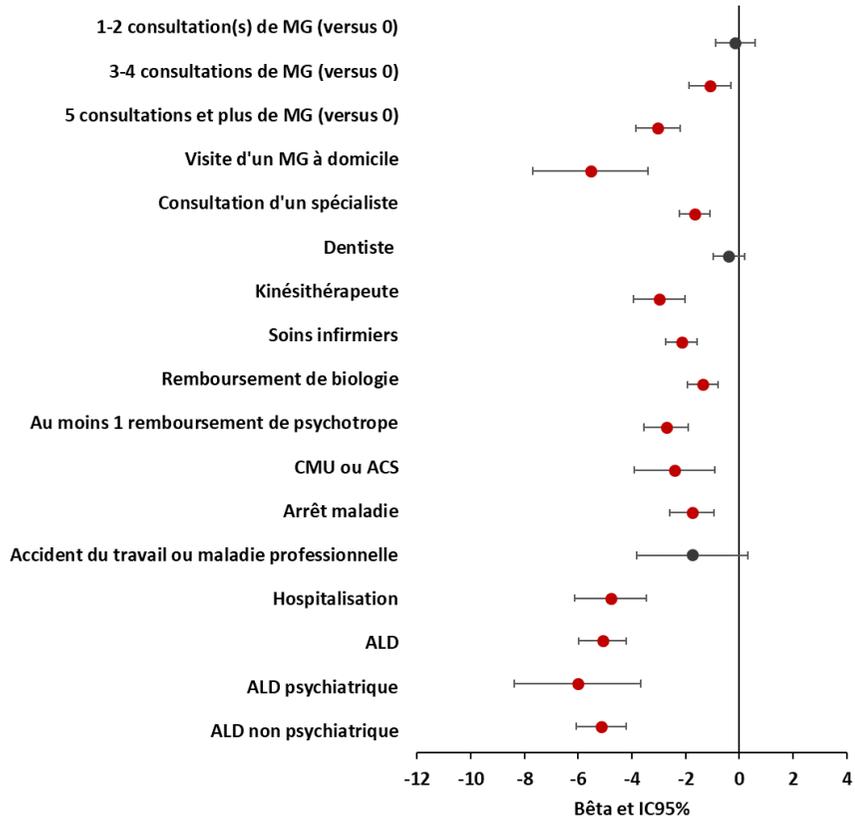
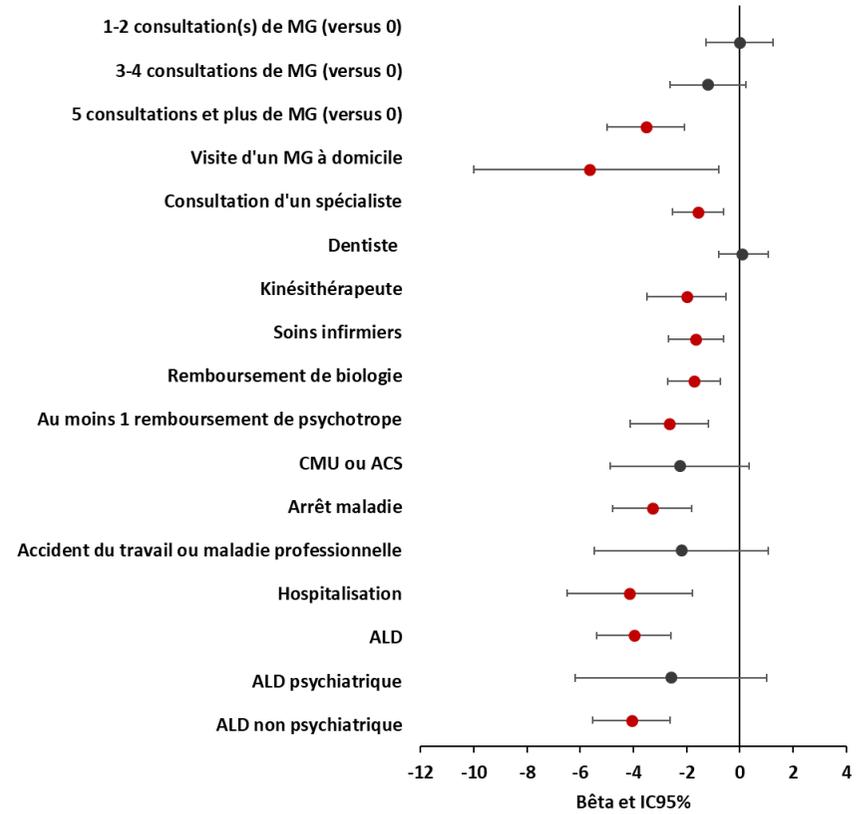


Figure 7. Associations entre les indicateurs SNDS et le score de santé physique (PCS) de la population dans les deux zones (N = 3 758 en zone exposée et N = 1 015 en zone non exposée), Bêta étant le coefficient de régression associé à chaque variable

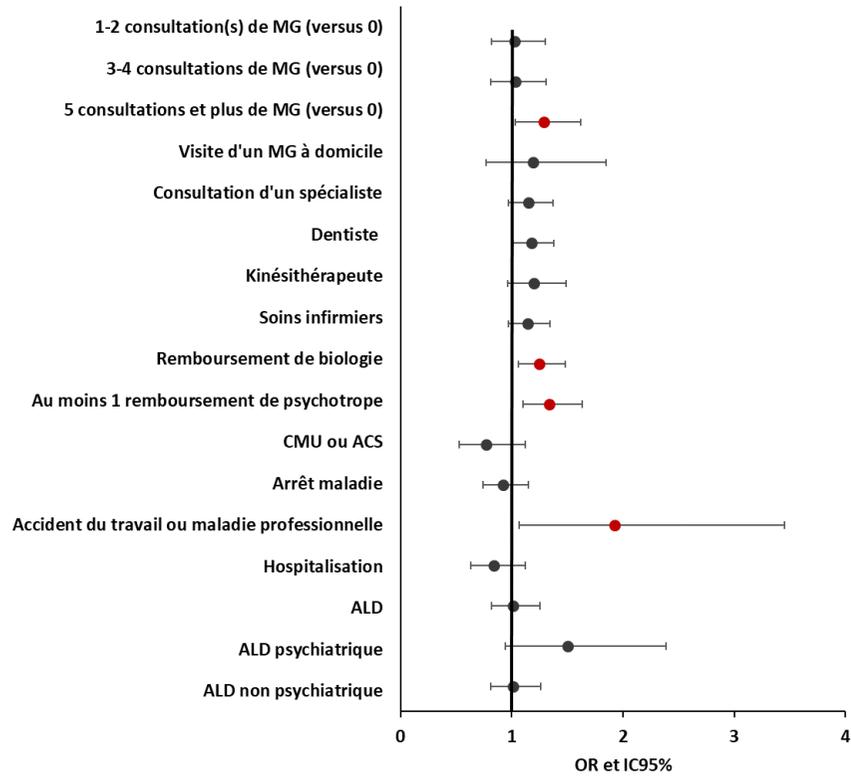
Zone exposée



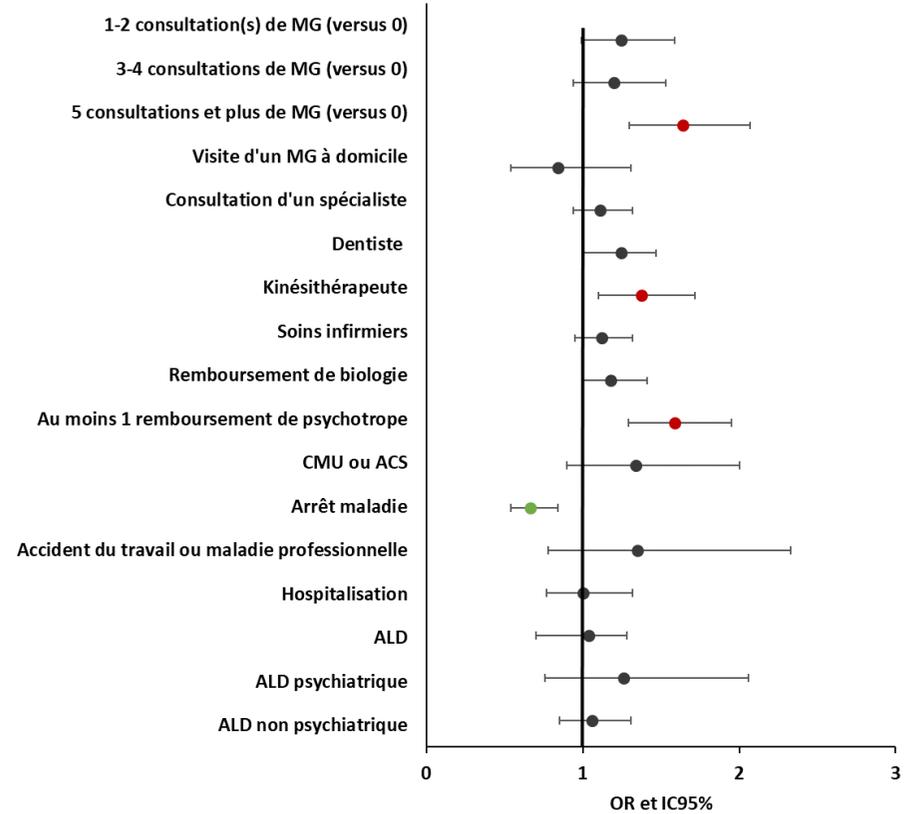
Zone non exposée



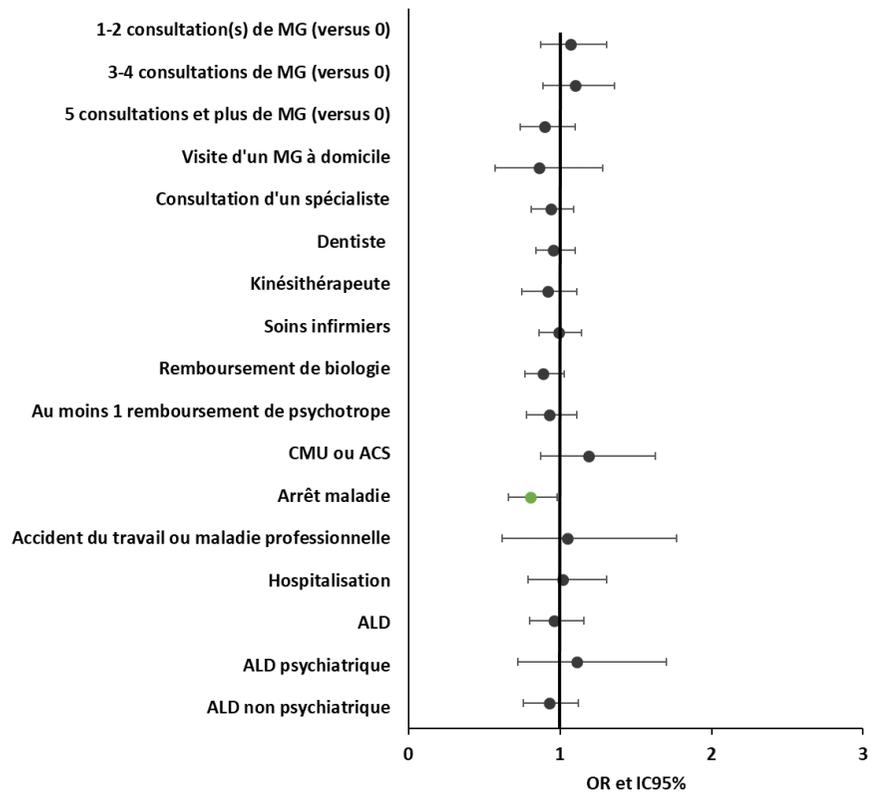
**Figure 8. Associations entre les indicateurs SNDS et l'existence d'au moins un symptôme psychique post-incendie de la population de la zone exposée (N = 3 758)**



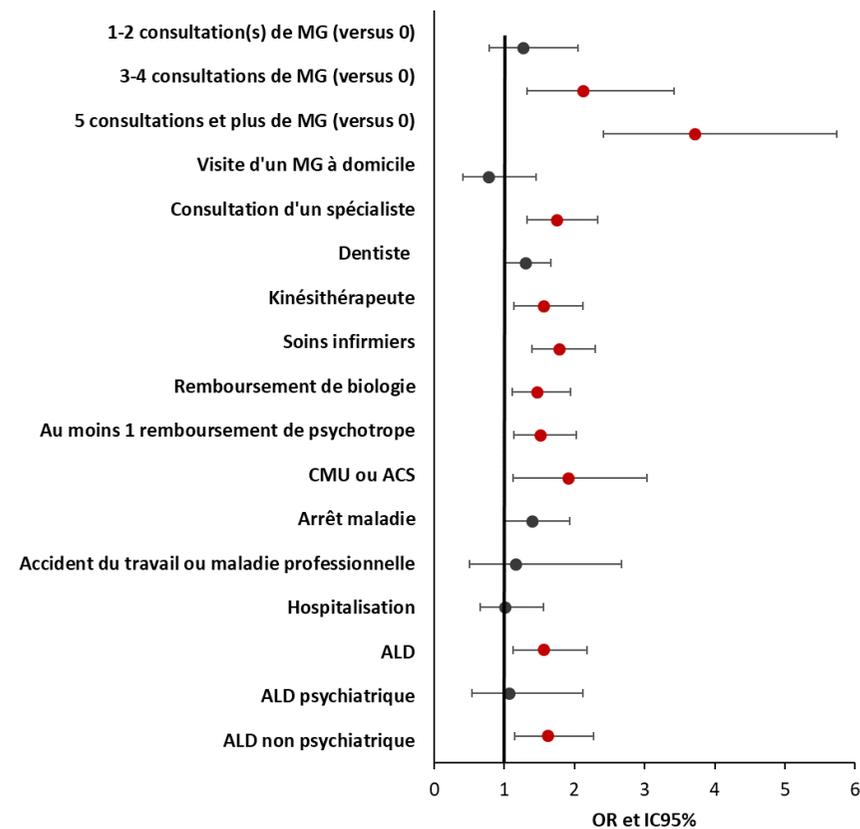
**Figure 9. Associations entre les indicateurs SNDS et l'existence d'au moins un symptôme physique post-incendie de la population de la zone exposée (N = 3 758)**



**Figure 2. Associations entre les indicateurs SNDS et le nombre d'expositions à l'incendie déclarées par la population de la zone exposée (N = 3 758)**



**Figure 1. Associations entre les indicateurs SNDS et le fait d'avoir consulté un médecin suite aux symptômes post-incendie dans la population de la zone exposée (N = 3 758)**



## Synthèse des résultats

**Tableau 1. Récapitulatif des liens entre les indicateurs SNDS, la participation et les variables d'intérêt de l'enquête en zone exposée (effet positif (+), effet négatif (-) ou effet non significatif), modèles multivariés et pondérés**

Zone exposée Indicateurs SNDS	Participation	Variables d'intérêt de l'enquête					
		Score de santé mentale	Score de santé physique	Symptômes psychiques	Symptômes physiques	Nombre d'expositions	Consultation d'un médecin suite au(x) symptôme(s)
<b>Nombre de consultations chez un médecin généraliste</b>							
1-2 ( <i>versus 0</i> )							
3-4 ( <i>versus 0</i> )	+						+
5 et plus ( <i>versus 0</i> )	+	-	-				+
<b>Visite d'un médecin généraliste à domicile</b>	-	-	-		+		
<b>A vu un médecin spécialiste</b>	+						
<b>A vu un dentiste</b>	+				+		
<b>A vu un kinésithérapeute</b>			-				
<b>A eu des soins infirmiers</b>			-				+
<b>Au moins un remboursement de biologie</b>	+		-	+			-
<b>Au moins un remboursement de psychotrope</b>		-	-	+	+		
<b>Au moins un remboursement en CMU ou ACS</b>	-		-				
<b>Au moins un arrêt de travail pour maladie</b>							
<b>Au moins un arrêt de travail pour accident du travail ou maladie professionnelle</b>				+		+	
<b>Au moins une hospitalisation</b>	-		-				
<b>Affection longue durée</b>	-	-	-				+
<b>Affection longue durée psychiatrique</b>	-	-	-				
<b>Affection longue durée non psychiatrique</b>			-				

**Tableau 2. Récapitulatif des liens entre les indicateurs SNDS, la participation et les variables d'intérêt de l'enquête en zone non exposée (effet positif (+), effet négatif (-) ou effet non significatif), modèles multivariés et pondérés**

Zone non exposée Indicateurs SNDS	Participation	Variables d'intérêt de l'enquête	
		Score de santé mentale	Score de santé physique
<b>Nombre de consultations chez un médecin généraliste</b>			
1-2 (versus 0)			
3-4 (versus 0)			
5 et plus (versus 0)			-
<b>Visite d'un médecin généraliste</b>			-
<b>A vu un médecin spécialiste</b>	+		
<b>A vu un dentiste</b>	+		
<b>A vu un kinésithérapeute</b>			
<b>A eu des soins infirmiers</b>			
<b>A un au moins un remboursement de biologie</b>			
<b>A eu au moins un remboursement de psychotrope</b>		-	
<b>Au moins un remboursement en CMU ou ACS</b>	-		
<b>Au moins un arrêt de travail pour maladie</b>			-
<b>Au moins un arrêt de travail pour accident du travail ou maladie professionnelle</b>			
<b>Au moins un hospitalisation</b>			-
<b>Affection longue durée</b>	-	-	-
<b>Affection longue durée psychiatrique</b>		-	
<b>Affection longue durée non psychiatrique</b>	-		-

Au final, le choix des indicateurs SNDS à inclure dans les modèles de non-réponse dans les zones exposée et non exposée s'est porté sur les indicateurs associés à la fois à la participation à l'enquête et au moins une des principales variables d'intérêt de l'enquête dans les modèles multivariés. Ainsi, les indicateurs SNDS retenus pour la construction des modèles de non-réponse sont :

- Le nombre de consultations chez un médecin généraliste ;
- La visite d'un médecin généraliste à domicile ;
- La consultation d'un dentiste ;
- Le remboursement d'au moins un acte de biologie ;
- Bénéficiaire d'au moins un remboursement de la CMU/ACS ;
- Et être en ALD.

L'indicateur « au moins une hospitalisation » n'a pas été retenu car il était associé, au seuil de 5 %, à la participation en zone exposée et non associé à la participation en zone non exposée dans les modèles multivariés.

## Calcul des nouvelles pondérations

Les variables incluses dans les modèles, permettant de calculer les nouvelles pondérations tenant compte des variables Fidéli et des indicateurs sanitaires SNDS retenus à l'étape précédente, sont décrites dans le Tableau 3.

**Tableau 3. Variables incluses dans les modèles de non-réponse, en zones exposée et non exposée**

	Zone exposée	Zone non exposée
<b>Variables Fidéli</b>	Âge de l'adulte au moment du terrain Mode de contact au niveau ménage Niveau de vie en décile (année n-1) Sexe Lien familial de l'individu échantillonné avec le référent fiscal Statut d'occupation du logement (propriétaire/locataire) Type de ménage Perception d'un salaire par l'individu-cible (année n) Distance domicile-usine (en m), négative si arrière du panache (sud-ouest de l'usine) Lieu de naissance de l'individu échantillonné	Âge de l'adulte au moment du terrain Mode de contact au niveau ménage Niveau de vie en décile (année n-1) Sexe Lien familial de l'individu échantillonné avec le référent fiscal Indicatrice de logement social O/N Type de logement Perception du chômage de l'individu-cible (année n)
<b>Indicateurs SNDS</b>	Nombre de consultations chez un médecin généraliste Visite d'un médecin généraliste à domicile Consultation d'un dentiste Remboursement d'au moins un acte de biologie Bénéficiaire d'au moins un remboursement de la CMU/ACS Être en ALD	Nombre de consultations chez un médecin généraliste Visite d'un médecin généraliste à domicile Consultation d'un dentiste Remboursement d'au moins un acte de biologie Bénéficiaire d'au moins un remboursement de la CMU/ACS Être en ALD

Après modélisation, les 18 groupes homogènes de répondants de la zone exposée sont décrits selon leur taux de réponse (Tableau 4) ainsi que les 10 groupes homogènes de répondants de la zone non exposée (Tableau 5).

Les rapports des poids de sondage corrigés de la non-réponse tronqués (max/min, P75/P25, P95/P5) sont comparables aux rapports des poids des pondérations initiales en zone exposée (Tableau 6) et non exposée (Tableau 7).

**Tableau 4. Répartition des répondants et non-répondants dans les GRH en zone exposée**

GRH	Refus/abandons/Non contactés		Répondants		Taux de réponse pondéré
	n	%	n	%	
1	342	92%	28	8%	0,08
2	322	87%	48	13%	0,13
3	302	81%	69	19%	0,19
4	292	79%	78	21%	0,21
5	266	72%	104	28%	0,26
6	247	67%	124	33%	0,33
7	240	65%	130	35%	0,36
8	204	55%	166	45%	0,44
9	196	53%	175	47%	0,47
10	173	47%	197	53%	0,53
11	149	40%	221	60%	0,59
12	145	39%	226	61%	0,62
13	130	35%	240	65%	0,64
14	361	32%	750	68%	0,67
15	101	27%	269	73%	0,73
16	89	24%	282	76%	0,74
17	66	18%	304	82%	0,82
18	48	13%	323	87%	0,88

**Tableau 5. Répartition des répondants et non-répondants dans les GRH en zone non exposée**

GRH	Refus/abandons/Non contactés		Répondants		Taux de réponse pondéré
	n	%	n	%	
1	239	96%	11	4%	0,04
2	203	81%	48	19%	0,19
3	188	75%	62	25%	0,25
4	169	67%	82	33%	0,33
5	146	58%	104	42%	0,42
6	142	57%	109	43%	0,43
7	127	51%	123	49%	0,49
8	115	46%	136	54%	0,54
9	90	36%	160	64%	0,64
10	77	31%	174	69%	0,69

**Tableau 6. Statistiques descriptives des différentes pondérations en zone exposée**

Paramètres	Pondérations initiales (poids de sondage)	Pondérations corrigées de la non-réponse Fidéli tronquées	Pondérations corrigées de la non-réponse Fidéli + SNDS non tronquées	Pondérations corrigées de la non-réponse Fidéli + SNDS tronquées
Somme	250 783	250 783	250 783	250 783
Ecart-type	17	40	64	40
Variance	296	1 638	4 154	1 582
Ecart interquartile	24	49	44	49
Moyenne	34	67	67	67
Mode	28	87	77	86
Coeff Variation (%)	51	60	97	60
Max 100%	97	159	655	158
95%	52	150	159	147
90%	52	134	111	125
75% Q3	52	87	78	87
50 % Médiane	46	69	59	67
25 % Q1	28	38	34	38
10%	11	16	14	15
5%	2	6	3	5
0% Min	1	3	1	3
Rapport P75/P25	2	2	2	2
Rapport P95/P5	26	25	53	29
Rapport Max/Min	97	53	575	53

**Tableau 7. Statistiques descriptives des différentes pondérations en zone non exposée**

Paramètres	Pondérations initiales (poids de sondage)	Pondérations corrigées de la non-réponse Fidéli tronquées	Pondérations corrigées de la non-réponse Fidéli + SNDS non tronquées	Pondérations corrigées de la non-réponse Fidéli + SNDS tronquées
Somme	243 175	243 175	243 175	243 175
Ecart-type	0,9	101	226	97
Variance	0,8	10 106	51 026	9 515
Ecart interquartile	0	106	82	86
Moyenne	97	242	241	241
Mode	97	145	140	154
Coeff Variation (%)	0,9	42	94	40
Max 100%	97	501	2 206	507
95%	97	501	507	507
90%	97	467	392	429
75% Q3	97	276	233	256
50 % Médiane	97	209	198	217
25 % Q1	97	170	152	166
10%	97	145	140	154
5%	97	145	140	154
0% Min	97	145	105	154
Rapport P75/P25	1	2	2	2
Rapport P95/P5	1	4	4	3
Rapport Max/Min	1	4	21	3

## Comparaison des anciennes et nouvelles pondérations

Dans la Figure 12 et la Figure 13 sont présentées les distances standardisées des indicateurs sanitaires SNDS et des variables Fidéli selon les pondérations (poids de sondage, pondération avec les variables Fidéli tronquée, pondération avec variables Fidéli + SNDS tronquée et pondération avec les variables Fidéli + SNDS non tronquée).

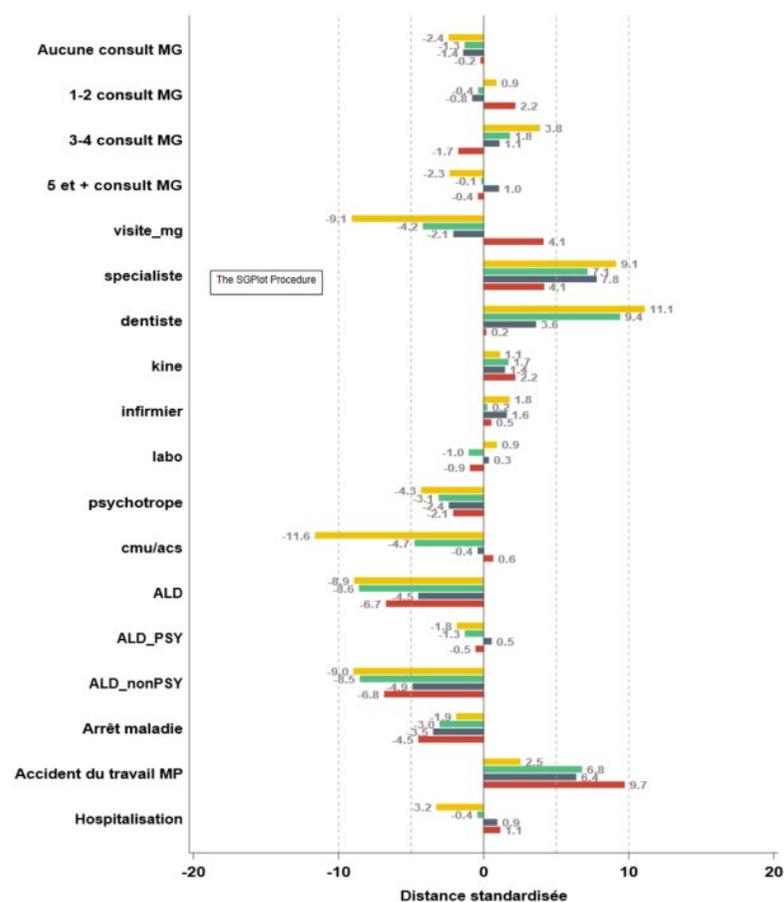
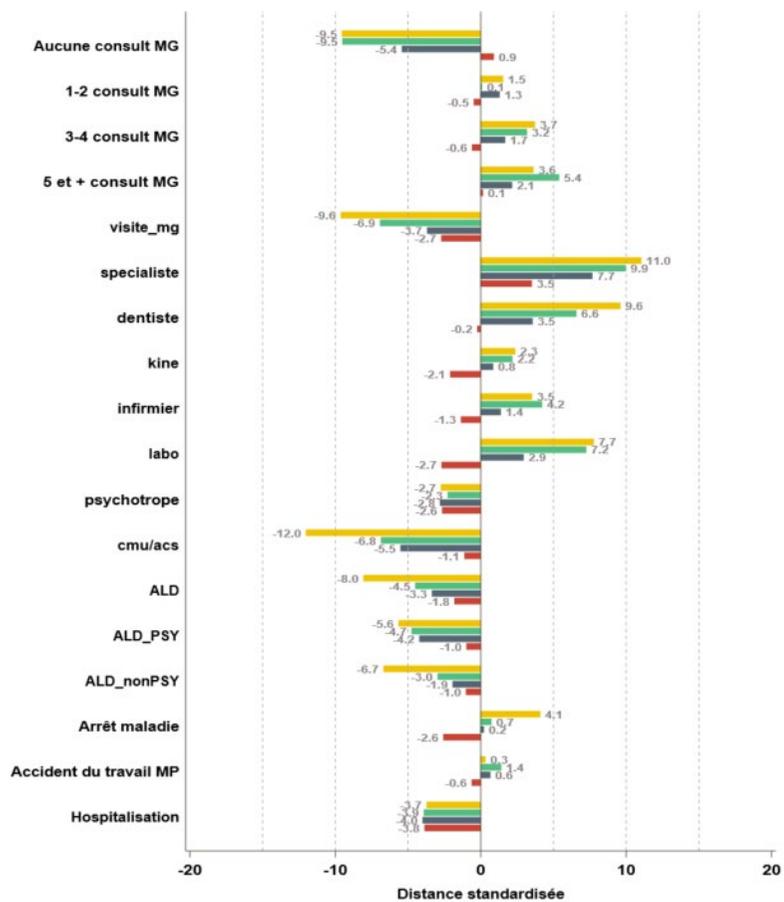
On observe d'une part que la correction de la non-réponse permet de diminuer, pour l'ensemble des variables utilisées (Fidéli et SNDS), la distance standardisée par rapport à la distribution de l'échantillon tiré au sort pondéré sur le poids de sondage ; d'autre part qu'après correction de la non-réponse, les distances standardisées sont toutes inférieures à 10 %, ce qui n'était pas le cas pour quelques modalités avant traitement de la non-réponse totale.

En plus des distances standardisées, les pourcentages pondérés de ces différentes variables sont décrits dans la Figure 14 et la Figure 15. On constate que les pourcentages varient peu selon les pondérations retenues.

Figure 12. Distances standardisées des indicateurs SNDS, par rapport à la population source, de l'échantillon des répondants au volet principal avec respectivement le poids de sondage, le poids corrigé de la non-réponse avec les variables Fidéli, le poids corrigé de la non-réponse avec les variables Fidéli et SNDS tronqué et le poids corrigé de la non-réponse avec les variables Fidéli et SNDS non tronqué, en zone exposée et en zone non exposée

Zone exposée

Zone non exposée



■ Poids de CNR + SNDS non tronqué    ■ Poids de CNR + SNDS tronqué  
■ poids de CNR    ■ poids de sondage

Figure 13. Distances standardisées des variables Fidéli, par rapport à la population source, de l'échantillon des répondants au volet principal avec respectivement le poids de sondage, le poids corrigé de la non-réponse avec les variables Fidéli, le poids corrigé de la non-réponse avec les variables Fidéli et SNDS tronqué et le poids corrigé de la non-réponse avec les variables Fidéli et SNDS non tronqué, en zone exposée et en zone non exposée

Zone exposée

Zone non exposée

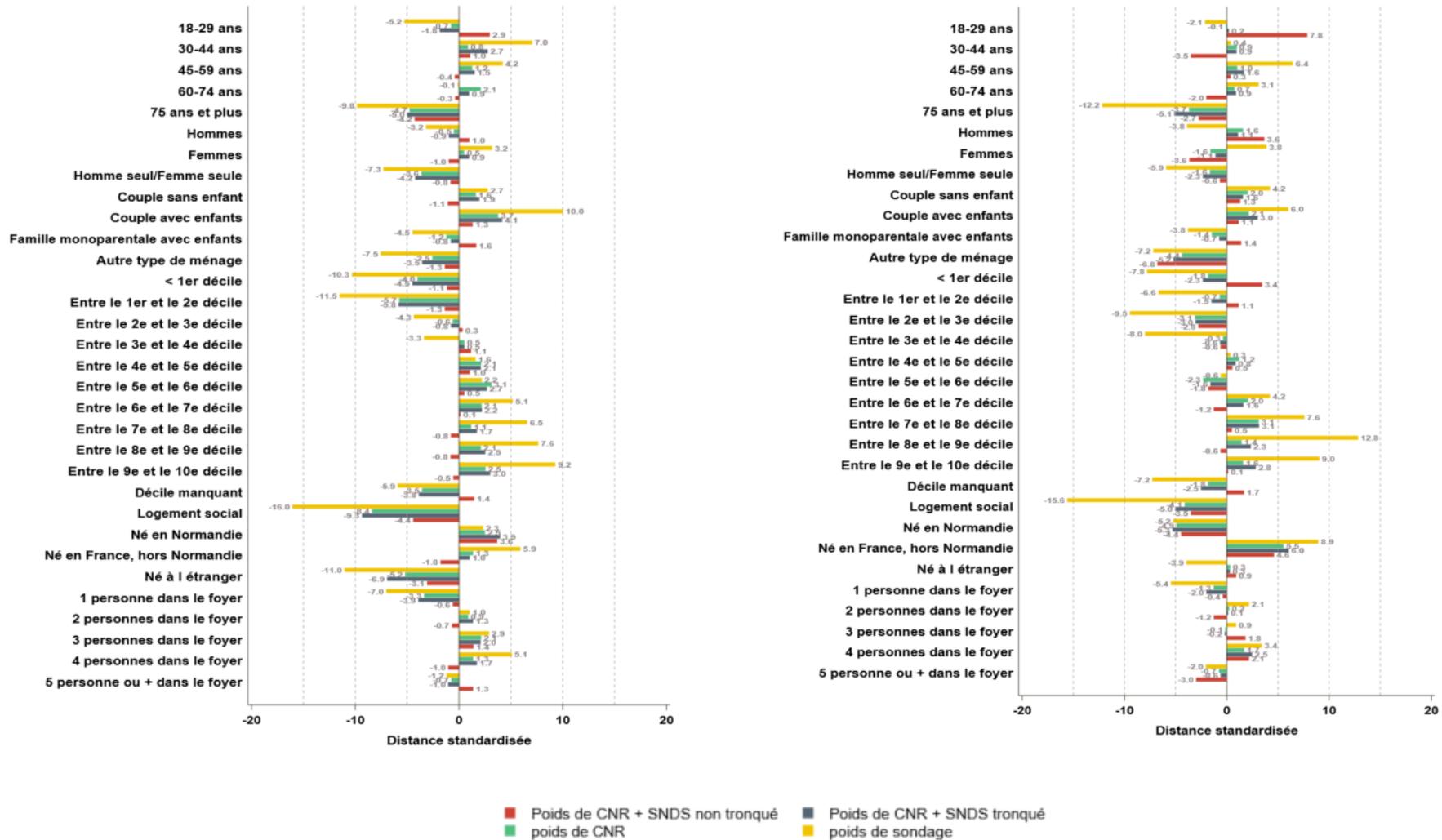
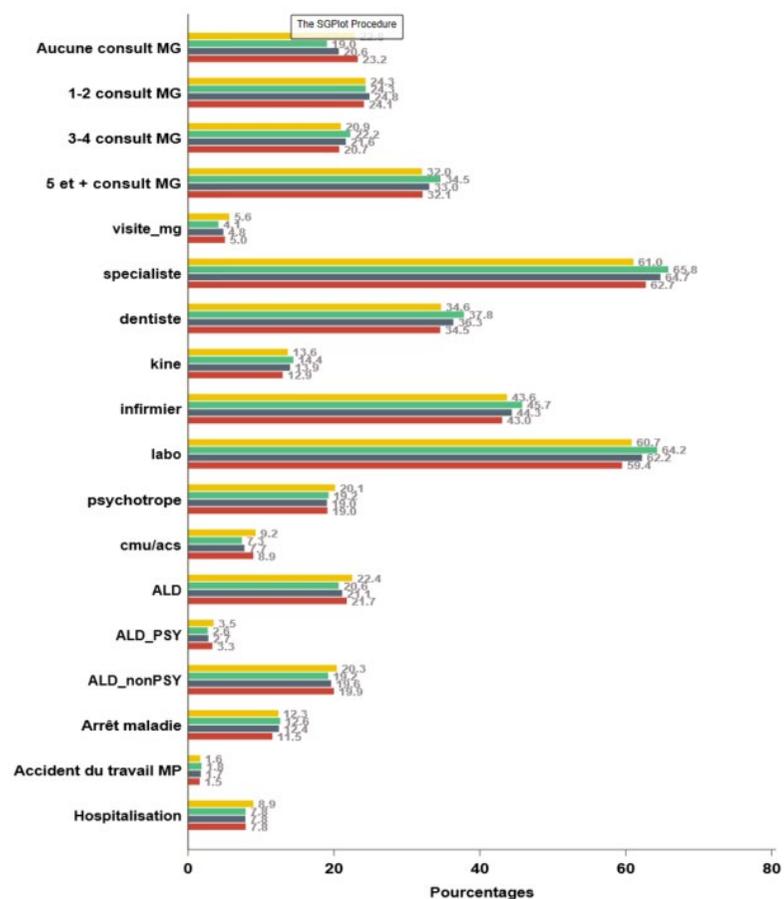


Figure 14. Pourcentages pondérés des indicateurs SNDS, pour l'ensemble des personnes tirées au sort avec le poids de sondage et les répondants avec respectivement le poids corrigé de la non-réponse avec les variables Fidéli, le poids corrigé de la non-réponse avec les variables Fidéli et SNDS tronqué et le poids corrigé de la non-réponse avec les variables Fidéli et SNDS non tronqué, en zone exposée et en zone non exposée

Zone exposée



Zone non exposée

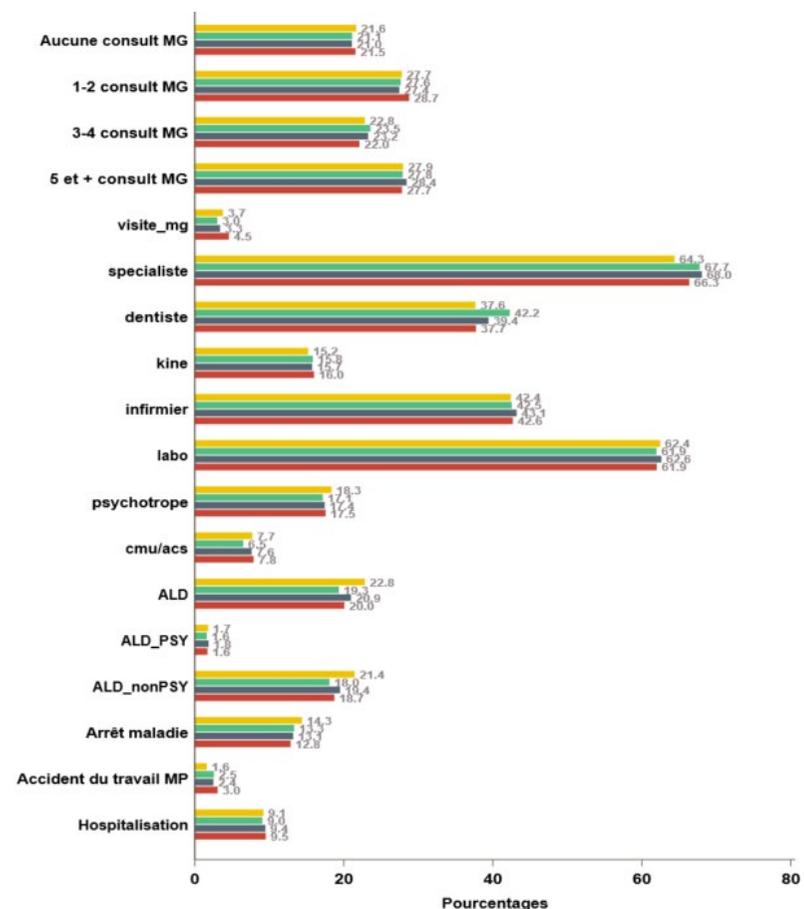
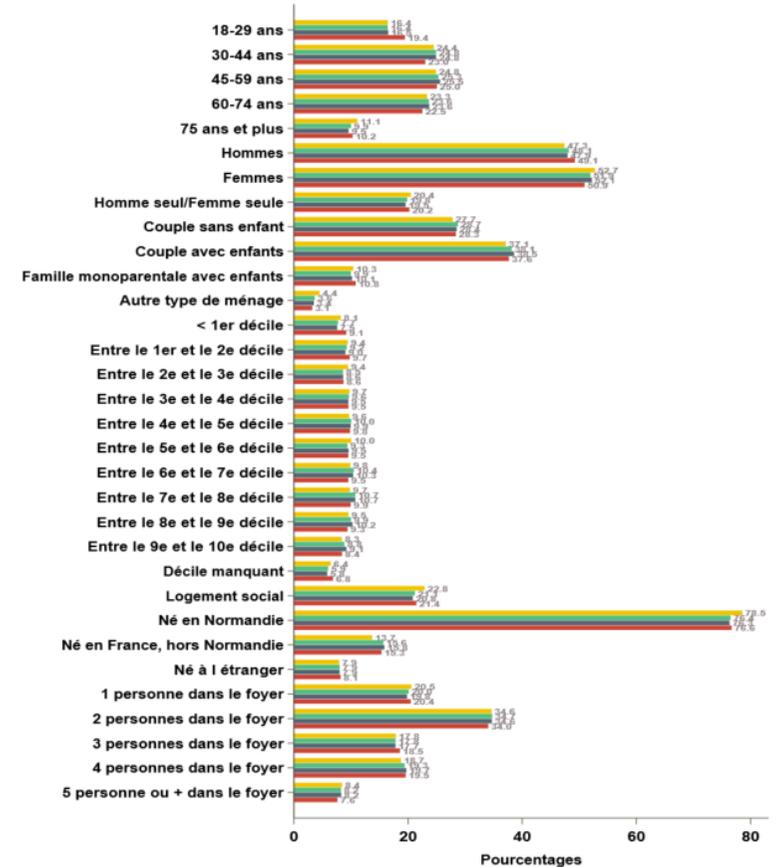
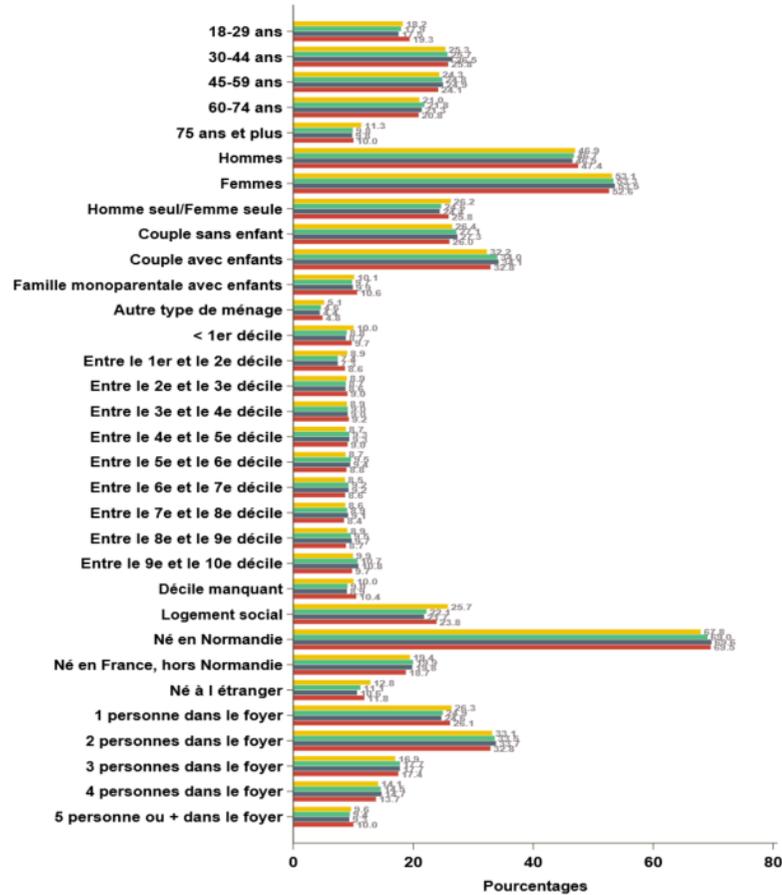


Figure 15. Pourcentages pondérés des variables Fidéli, pour l'ensemble des personnes tirées au sort avec le poids de sondage et les répondants avec respectivement le poids de sondage, le poids corrigé de la non-réponse avec les variables Fidéli, le poids corrigé de la non-réponse avec les variables Fidéli et SNDS tronqué et le poids corrigé de la non-réponse avec les variables Fidéli et SNDS non tronqué, en zone exposée et en zone non exposée

Zone exposée

Zone non exposée



■ Poids de CNR + SNDS non tronqué   ■ Poids de CNR + SNDS tronqué  
 ■ poids de CNR   ■ poids de sondage

## Mise à jour des principaux résultats de l'enquête avec les nouveaux poids

En comparant les résultats principaux de l'enquête obtenus avec l'ancienne et la nouvelle pondération de correction de la non-réponse, on constate une grande stabilité des estimations. Le score de santé mentale varie au maximum de deux dixièmes de points (Figure 16), même chose pour le score de santé physique (Figure 17). Concernant les variables de santé déclarée, les pourcentages diffèrent d'au maximum un point (Figure 18 à Figure 23). Concernant les expositions déclarées, les pourcentages varient d'1,2 points au maximum, avec des écarts les plus importants (tout en restant modestes) dans la strate de proximité (Figure 24 à Figure 26).

**Figure 16. Score moyen de santé mentale (MCS) selon la strate des répondants à l'enquête, pondérés sur les différents poids**

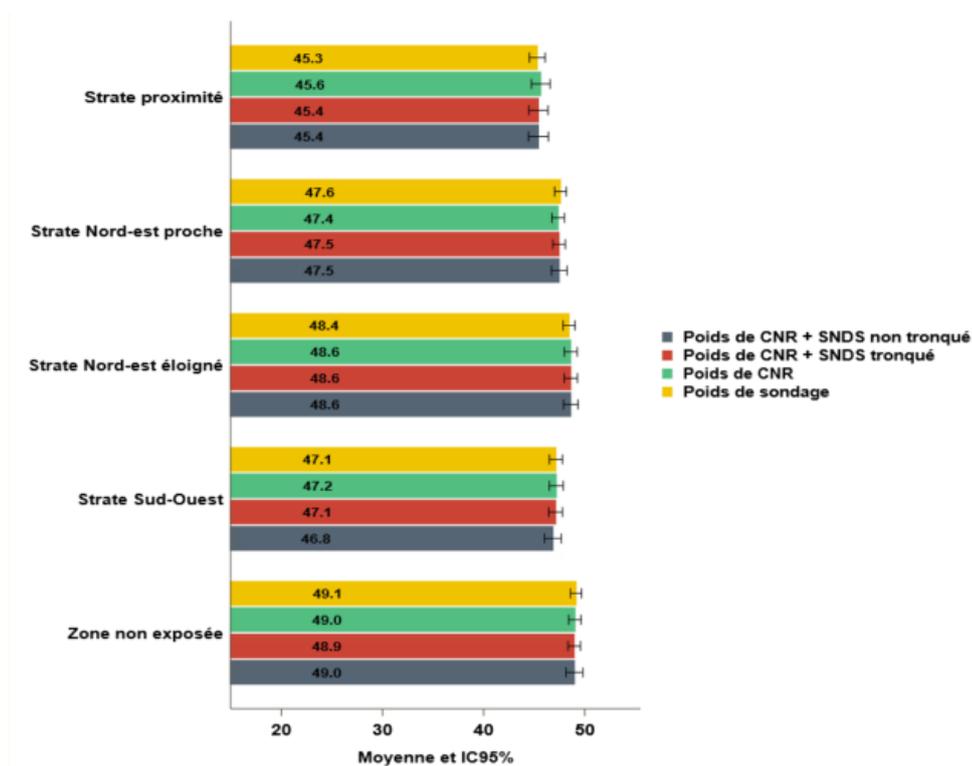


Figure 17. Score moyen de santé physique (PCS) selon la strate des répondants à l'enquête, pondérés sur les différents poids

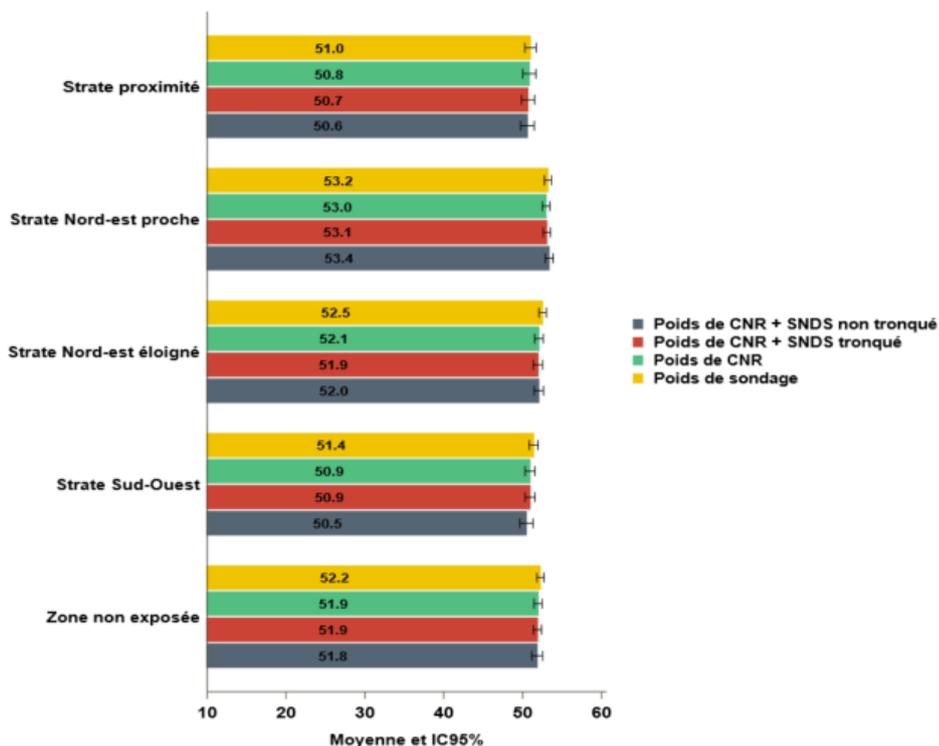


Figure 18. Pourcentages de répondants ayant déclaré une santé médiocre ou mauvaise selon la strate, pondérés sur les différents poids

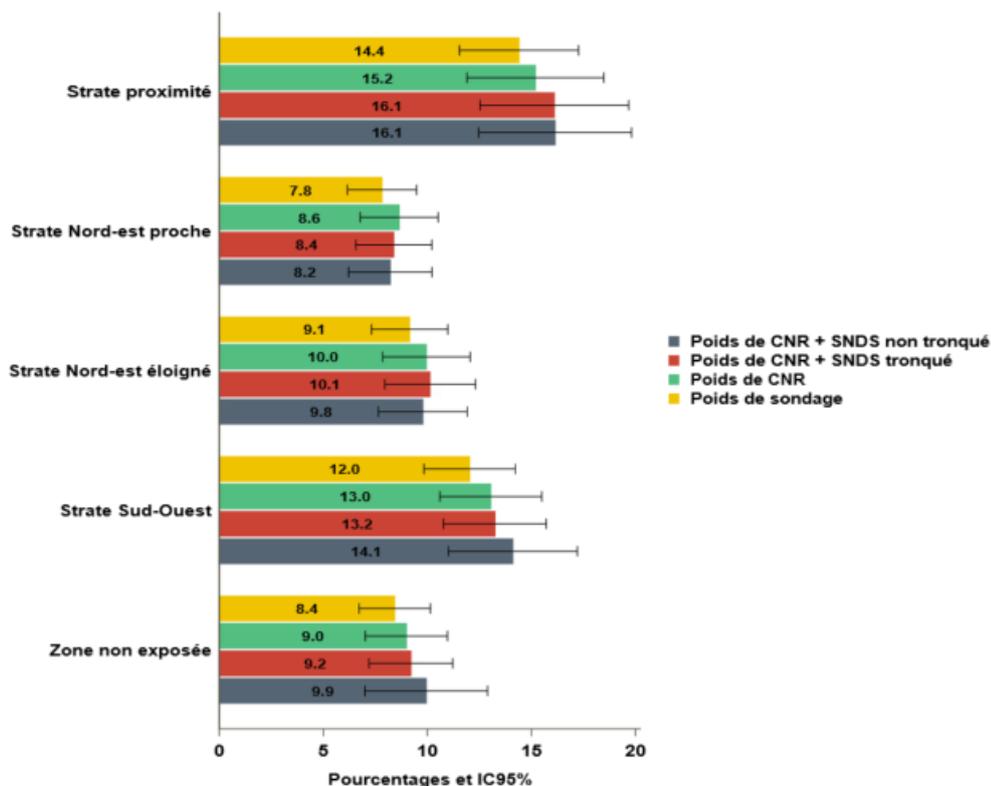


Figure 19. Pourcentages de répondants ayant déclaré être beaucoup limité pour les efforts physiques selon la strate, pondérés sur les différents poids

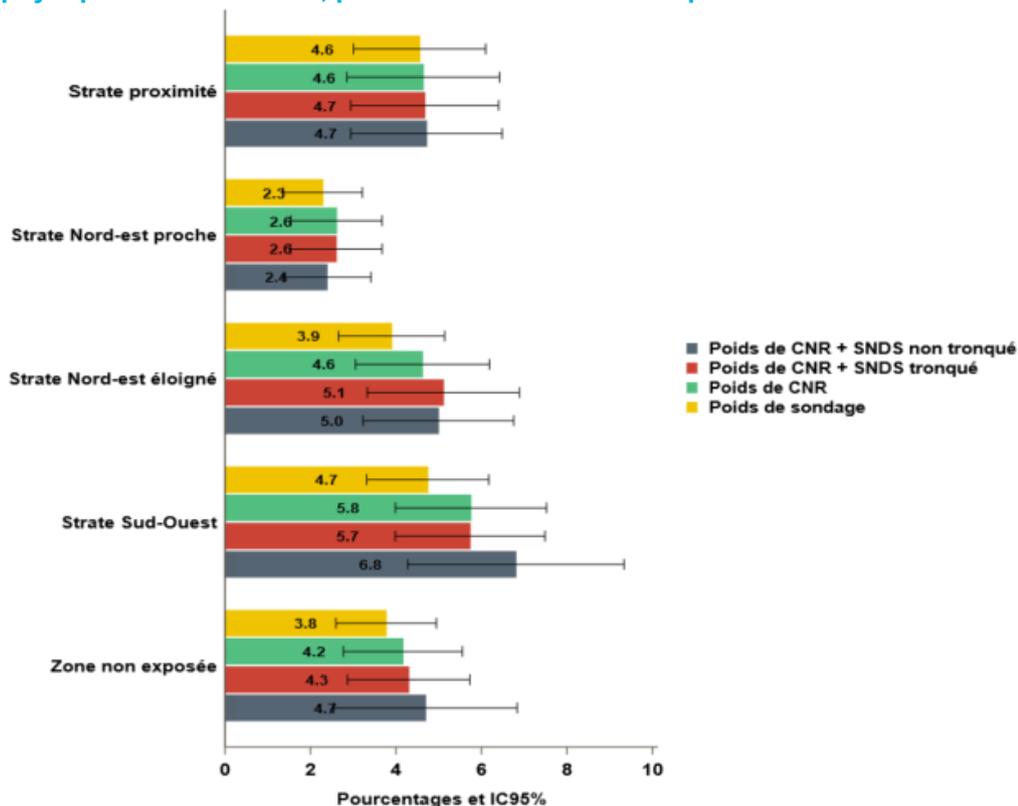


Figure 20. Pourcentages de répondants ayant déclaré être beaucoup limité pour monter les étages d'escalier selon la strate, pondérés sur les différents poids

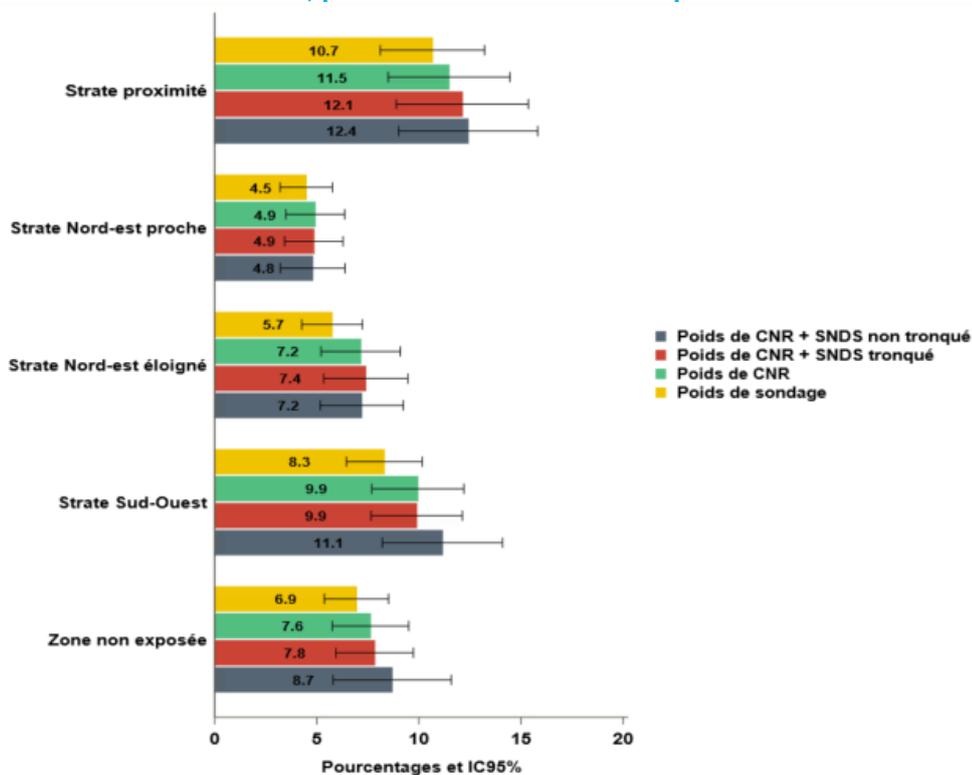


Figure 21. Pourcentages de répondants ayant déclaré au moins un symptôme psychique post-incendie en zone exposée, selon la strate, pondérés sur les différents poids

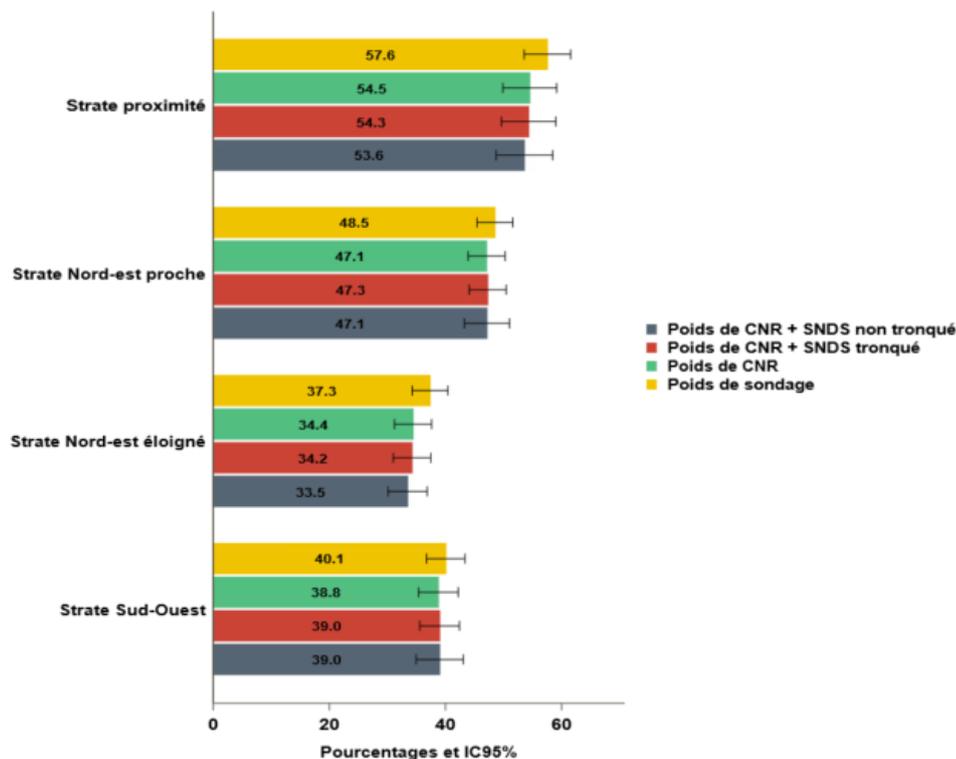


Figure 22. Pourcentages de répondants ayant déclaré au moins un symptôme physique post-incendie en zone exposée, selon la strate, pondérés sur les différents poids

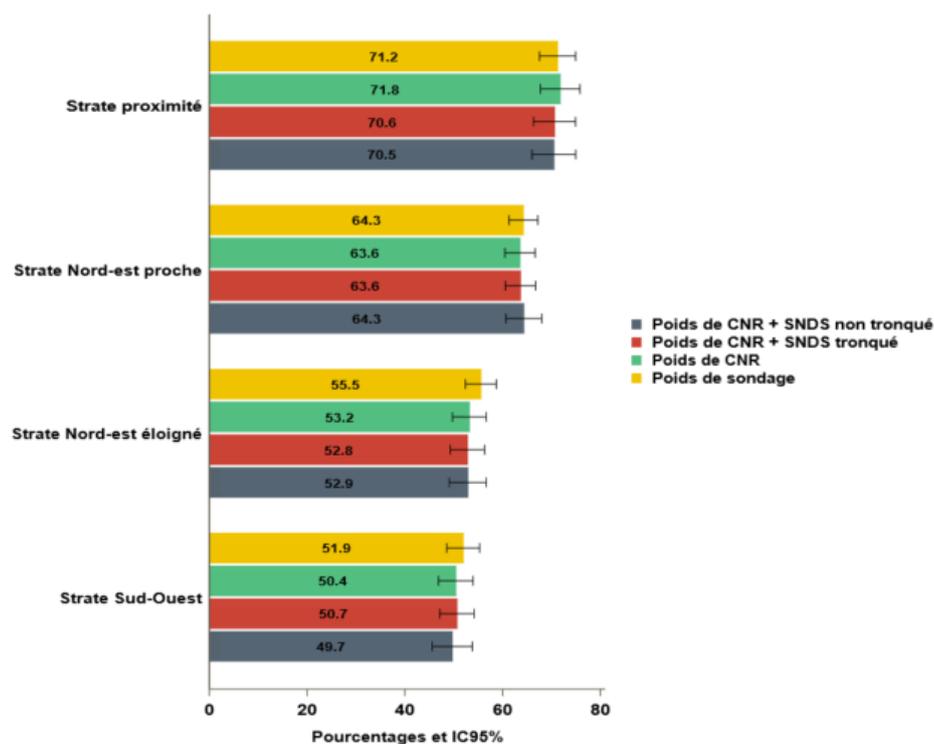


Figure 23. Pourcentages de répondants ayant consulté au moins une fois un médecin suite aux symptômes post-incendie, en zone exposée, selon la strate, pondérés sur les différents poids

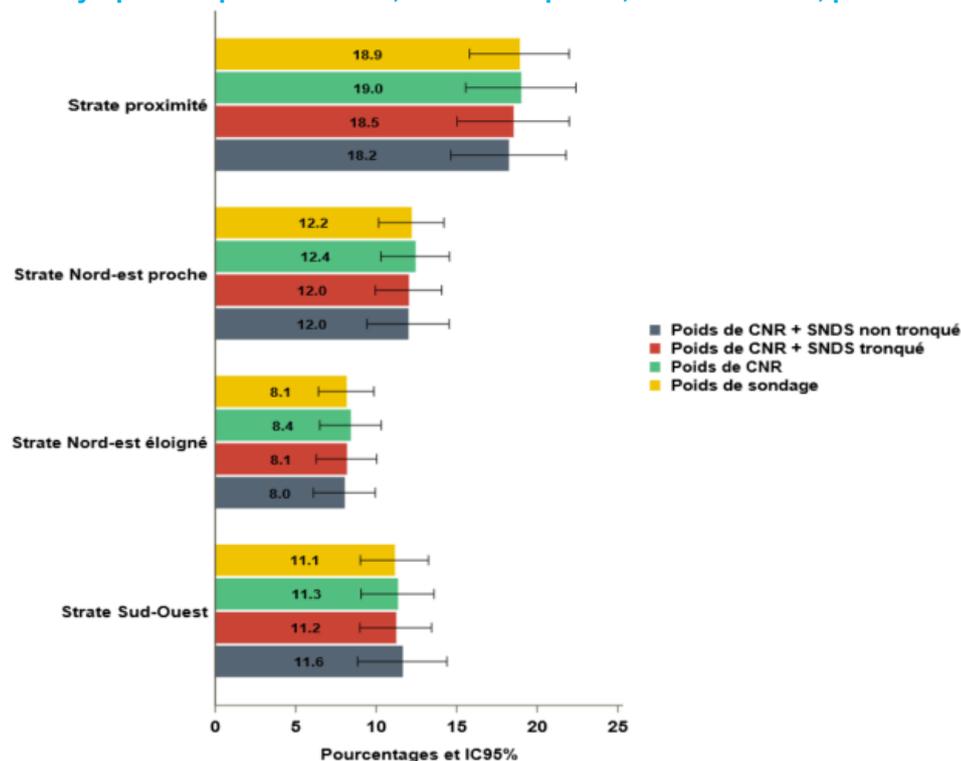


Figure 24. Pourcentages de répondants ayant déclaré avoir perçu des odeurs, en zone exposée, selon la strate, pondérés sur les différents poids

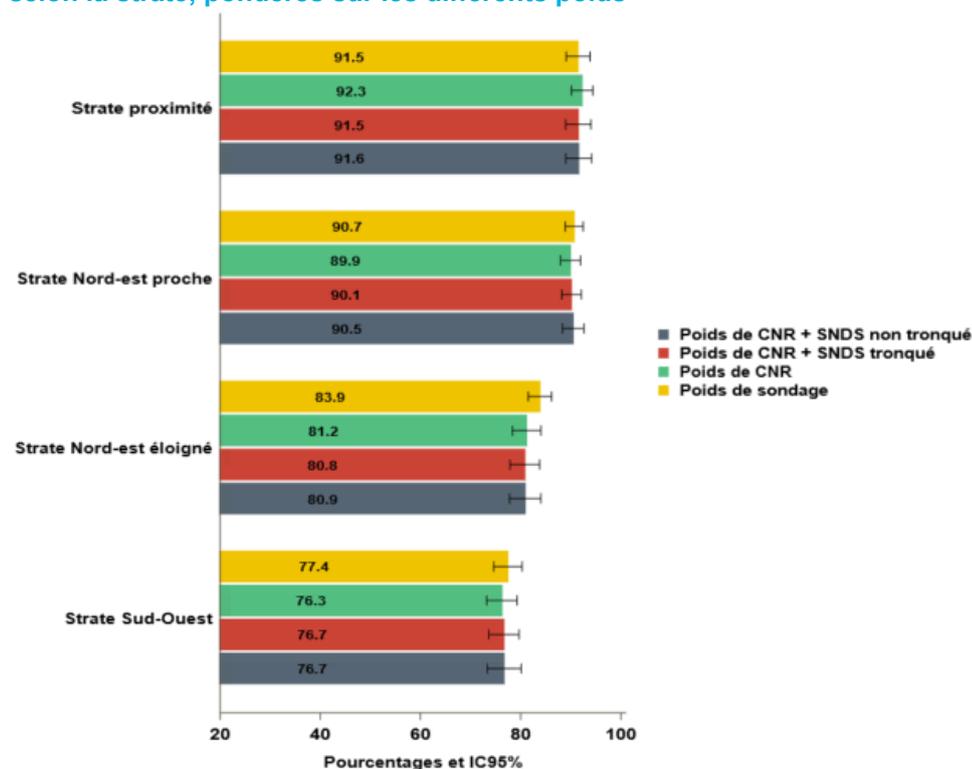


Figure 25. Pourcentages de répondants ayant déclaré avoir été exposé aux fumées, en zone exposée, selon la strate, pondérés sur les différents poids

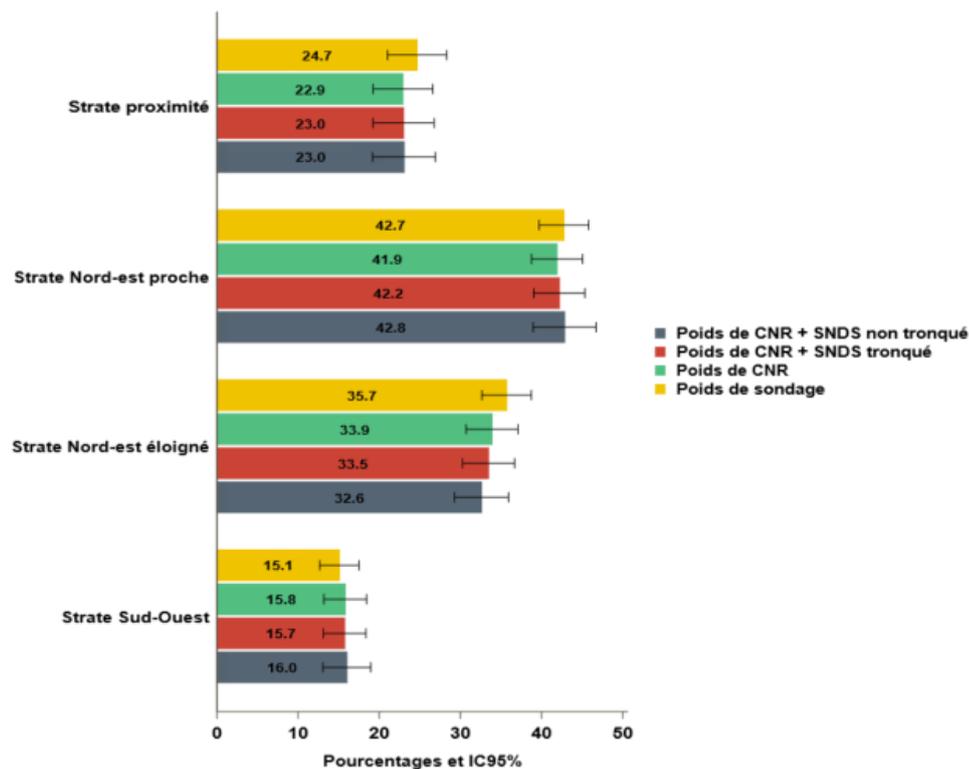
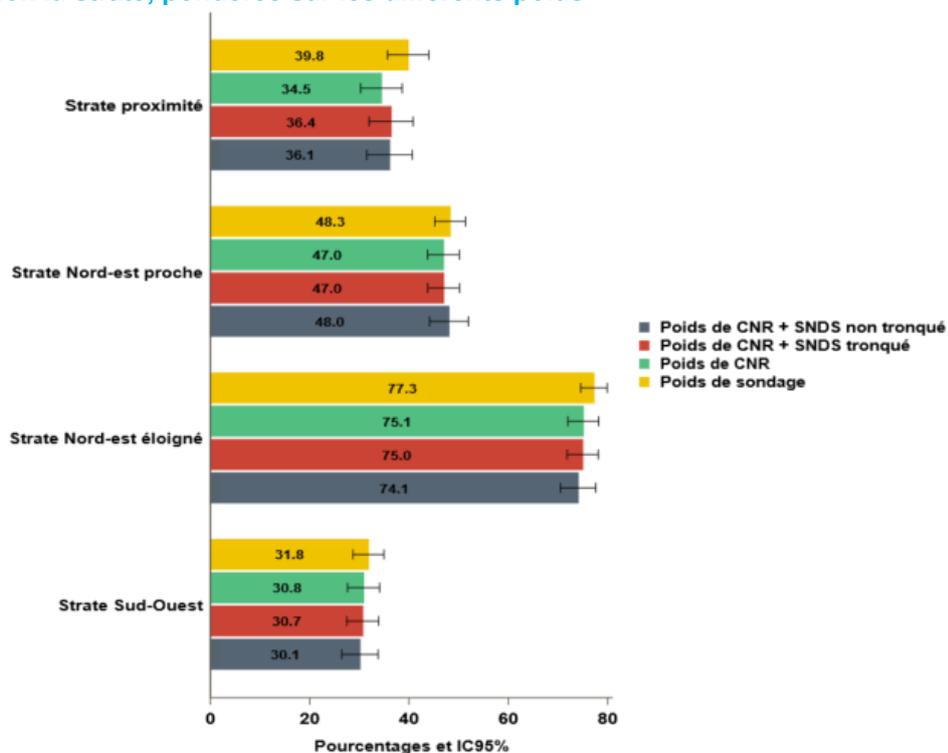


Figure 26. Pourcentages de répondants ayant déclaré avoir été exposé aux suies, en zone exposée, selon la strate, pondérés sur les différents poids



## Analyse de sensibilité pour les non consommateurs

Suite à l'appariement des individus tirés au sort avec le SNDS par l'Insee, il s'est avéré que 431 individus n'avaient aucun remboursement connu dans le SNDS (personnes absentes du référentiel des bénéficiaires extrait par la Cnam entre 2017 et juin 2022), ce qui représente 4,3 % des personnes appariées. Pour ce travail, ces personnes ont été considérées comme non consommatrices mais les études menées sur le SNDS montrent qu'il y a en général entre 1 et 2 % de personnes non consommatrices, c'est-à-dire n'ayant aucun recours au système de santé lors d'une période donnée. Afin de tester l'effet de cette hypothèse sur les résultats, un travail d'imputation a été mené pour attribuer des valeurs aux indicateurs sanitaires de ces 431 personnes.

En calculant les pourcentages pondérés sur le poids de sondage des indicateurs avant et après imputation, on constate que les pourcentages augmentent au maximum de 3 points en zone exposée (Figure 27). Plus particulièrement pour les indicateurs sanitaires retenus dans le modèle de correction de la non-réponse : nombre de consultations chez un médecin généraliste (maximum +1,6 points), visite d'un médecin généraliste à domicile (+0,2 point), consultation d'un dentiste (+1,9 points), remboursement d'un acte de biologie (+3,1 points), bénéficiaire de la CMU/ACS (+0,8 point), ALD (+1,5 points).

En zone non exposée, on constate que les pourcentages augmentent au maximum de 2,6 points (Figure 28). Plus particulièrement pour les indicateurs sanitaires retenus dans le modèle de correction de la non-réponse : nombre de consultations chez un médecin généraliste (maximum +1,2 points), visite d'un médecin généraliste à domicile (+0,1 point), consultation d'un dentiste (+1,4 points), remboursement d'un acte de biologie (+2,6 points), bénéficiaire de la CMU/ACS (+0,2 point), ALD (+0,8 point).

Au regard des faibles variations de pourcentages après imputation dans les deux zones, il est raisonnable de penser que le choix de considérer comme non consommateurs les individus non retrouvés dans le référentiel des bénéficiaires n'a eu qu'un très faible impact sur les résultats présentés dans ce rapport.

Figure 27. Pourcentages pondérés des indicateurs SNDS avant et après imputation en zone exposée

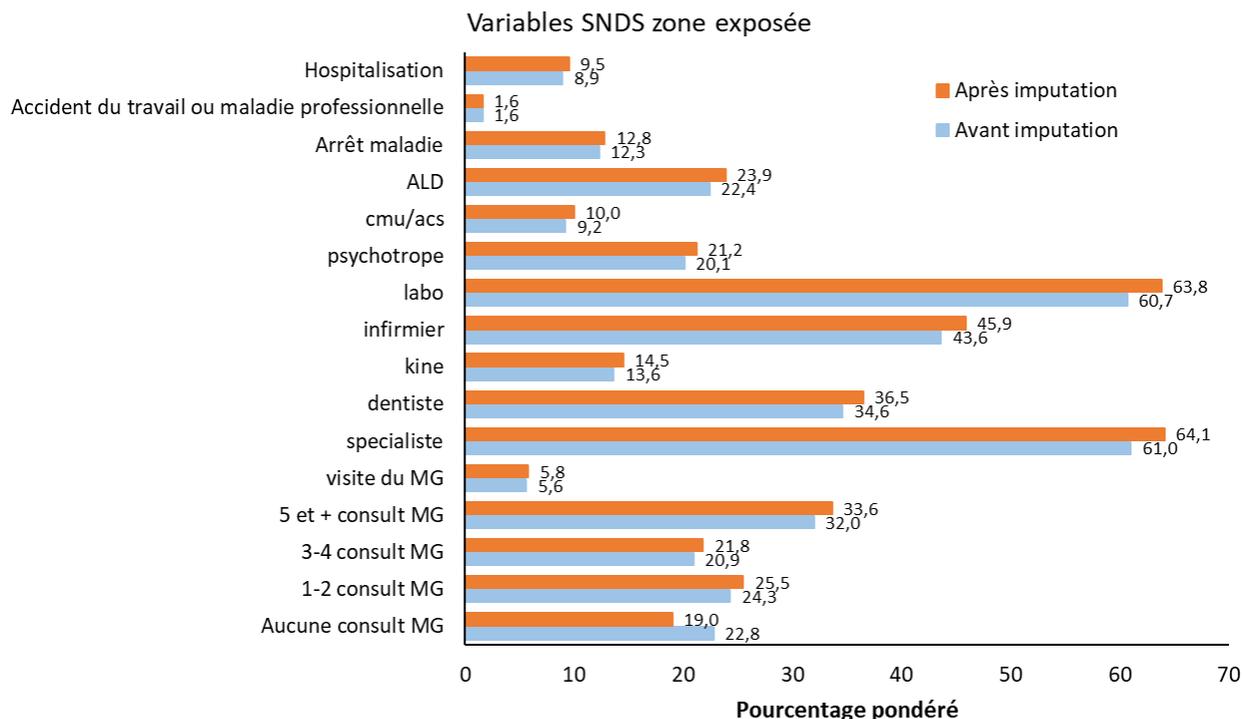
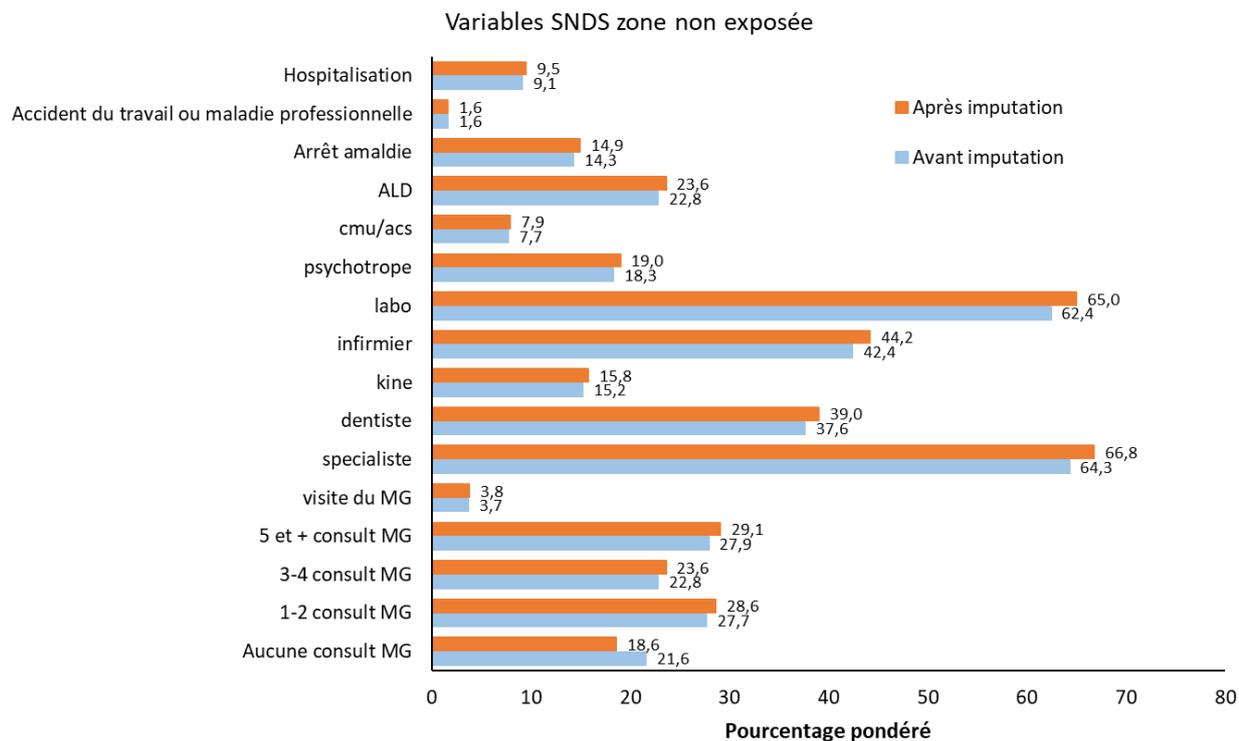


Figure 28. Pourcentages pondérés des indicateurs SNDS avant et après imputation en zone non exposée



# DISCUSSION

Cette analyse permet d'évaluer l'apport des indicateurs sanitaires disponibles dans le SNDS dans la compréhension du mécanisme de participation à une enquête de santé post-incendie.

Elle repose sur un appariement direct des personnes tirées au sort, sur la base du numéro d'inscription au répertoire des personnes physiques, aux données disponibles dans le SNDS. Cet appariement, mis en œuvre par l'Insee et la CNAM, a été possible pour 98,8 % de la base des personnes tirées au sort pour participer à l'enquête.

Elle complète le travail préalable de description de la participation à cette enquête, mené à partir des données de la base Fidéli uniquement [Richard 2 022], et apporte des réponses à la question de l'existence d'une sélection endogène à ce type d'enquête, qui se caractériserait par une participation plus importante des personnes concernées par le sujet de l'enquête, ainsi que de ses conséquences sur les estimations produites.

Tout d'abord, une association est mise en évidence entre certains indicateurs de santé et la participation à l'enquête, notamment en zone exposée. Les personnes ayant consulté des professionnels de santé (médecin généraliste, spécialiste, dentiste, avoir reçu des soins infirmiers ou un remboursement pour un acte de biologie) durant les 12 mois précédant l'enquête ont plus participé que les autres. A l'inverse, les personnes dont les indicateurs indiquaient une santé dégradée (avoir eu une visite à domicile du médecin, avoir été hospitalisé, être bénéficiaire d'une affection longue durée) ou une précarité économique (bénéficiaire de la CMU/ACS) ont moins participé. Une sélection différentielle entre zone exposée et non exposée est particulièrement significative : les personnes ayant consulté plusieurs fois un médecin généraliste durant les 12 mois précédant l'enquête ont plus souvent participé que les autres en zone exposée (OR = 1,60 [1,38-1,86]), ce qui n'est pas le cas en zone non exposée (OR = 1,00 [0,77-1,29]).

En outre, une association significative est mise en évidence entre les indicateurs SNDS et certaines variables clés mesurées par l'enquête, en particulier le score de santé physique, le score de santé mentale, ainsi qu'en zone exposée les symptômes de santé physique et psychique post-incendie. Un plus grand nombre de consultations chez un médecin généraliste durant les 12 mois précédant l'enquête était en particulier lié aux déclarations de tels symptômes et à de moins bons scores de santé (mentale ou physique).

L'identification de ces associations simultanées entre indicateurs SNDS, participation à l'enquête et variables d'intérêt illustre l'existence d'une sélection endogène et en partie différentielle entre zone exposée et non exposée. En zone exposée, les personnes ayant consulté plusieurs fois un médecin généraliste ont été plus enclines à participer et ont plus souvent déclaré des symptômes physique ou psychique post-incendie, laissant supposer une surestimation de la part de personnes ayant déclaré de tels symptômes dans les analyses préalablement publiées sans mobilisation des indicateurs issus du SNDS.

Toutefois, après inclusion des indicateurs de santé dans les modèles permettant de corriger la non-réponse à l'enquête, on constate une grande stabilité des estimations. Les scores de santé mentale ou physique ne varient au maximum que de deux dixièmes de points. Concernant les symptômes post-incendie, l'écart maximal est constaté dans la strate de proximité pour les symptômes psychiques (53,6 % vs 54,5 % avant prise en compte des indicateurs SNDS) et physiques (70,5 % vs 71,8 % avant prise en compte des indicateurs SNDS). Cette modification des estimations est donc cohérente avec le biais de sélection précédemment mis en évidence, mais demeure d'une ampleur limitée. L'impact sur la santé des personnes les plus proches du lieu de l'incendie serait ainsi en réalité légèrement plus faible que ce qui a été montré avec les précédentes estimations produites sans l'appariement avec le SNDS.

## Forces et limites

Lorsqu'il s'agit de construire des modèles de non-réponse, il y a des arguments en faveur d'analyses non pondérées des données (par exemple : ne pas atténuer des effets dans des sous-populations ayant un faible poids de sondage) et d'autres arguments en faveur d'une pondération sur le poids de sondage (par exemple : meilleure prise en compte du schéma de stratification). Dans ce travail, les deux méthodes ont été testées et conduisaient à des résultats comparables. Nous avons donc choisi de ne présenter que les résultats pondérés.

Cette analyse mobilise les sources de données les plus complètes possibles pour ce type d'étude, par un appariement des données issues de Fidéli (données liées au logement et données fiscales de l'individu) et du SNDS (données de remboursements de soins), et sont disponibles pour la quasi-totalité de l'échantillon tiré au sort, répondants et non-répondants. Elle fait ainsi partie des rares études permettant de décrire la participation à une enquête de santé en fonction d'indicateurs de santé, et les résultats obtenus apparaissent cohérents avec certains constats publiés précédemment, notamment celui concernant une participation associée positivement au recours à un praticien de ville (généraliste, dentiste, spécialiste), négativement au fait d'avoir un état de santé dégradé (hospitalisé, ALD, maladie grave) (Santin 2014, Soullier 2018).

Ces résultats soulignent également que les estimations corrigées du biais de non-réponse mesurable par les indicateurs du SNDS demeurent proches des estimations produites avec une correction de la non-réponse reposant exclusivement sur les données issues de Fidéli. Les analyses de sensibilité menées pour évaluer l'impact de la troncature des poids ainsi que du classement des personnes qui n'avaient aucune consommation connue dans le SNDS soulignent la robustesse de ces résultats.

Certaines limites peuvent néanmoins être évoquées. D'une part, les indicateurs retenus et issus du SNDS ne sont peut-être pas suffisamment spécifiques à l'événement survenu, et le biais de sélection n'a peut-être que partiellement été appréhendé et corrigé. Si les données avaient été disponibles dans le SNDS, il aurait par exemple été intéressant de disposer d'un indicateur relatif aux consultations d'un psychologue ou d'autres professions paramédicales. Ainsi, d'autres variables inobservables, et donc non prises en compte dans cette analyse complémentaire, pourraient être à l'origine d'une sélection endogène. Enfin, les résultats de cette analyse se situent dans un contexte singulier d'une enquête menée suite à important accident industriel, et auprès de personnes qui ont été exposées à des intensités diverses et dans un contexte de crainte générale sur l'impact à moyen et long-terme de cette exposition, augmentant ainsi substantiellement le risque de sélection endogène. Ils peuvent difficilement être extrapolables au cas d'enquêtes de santé menées en population générale, et les analyses sur données appariées au SNDS qui seront produites à partir des enquêtes EpiCOV ou EHIS 2019 (Warszawski 2020, Leduc 2021) apporteront un éclairage intéressant sur ce sujet. Soulignons enfin que la mise en œuvre de tels traitements statistiques nécessite des délais d'accès aux données et des ressources à anticiper : acteurs impliqués pour l'appariement, démarches auprès de la CNAM pour obtenir les données, habilitation, formations, développements d'algorithmes et travail sur le portail de la CNAM.

Une perspective pour compléter ce travail et pour de futures enquêtes serait d'étudier plus précisément les corrélations et les liens qui existent entre les variables Fidéli et les indicateurs SNDS à l'aide d'analyses factorielles des correspondances multiples.

# CONCLUSION

Grâce à l'appariement aux données du SNDS, nous avons pu mettre en évidence que les personnes tirées au sort pour participer à l'enquête « Santé Post Incendie 76 : une étude à l'écoute de votre santé » qui avaient consulté des professionnels de santé au cours des 12 mois précédant l'enquête ont été plus enclines à participer, à l'inverse des personnes dont les indicateurs indiquaient une santé dégradée ou une précarité économique. Ces relations étaient plus fortes pour les personnes dans la zone exposée que la zone non exposée. Ces indicateurs SNDS se révèlent également associés aux variables d'intérêt de l'enquête. Ainsi, la correction de la non-réponse est améliorée si on utilise les variables SNDS dans le modèle. Cependant la différence dans les estimations produites se révèle très faible, questionnant la pertinence de l'effort à fournir pour mettre en œuvre et traiter ces données appariées.

## Références

Golliot F, Blanchard M, Empereur-Bissonnet P, et al. Évaluation de l'impact sanitaire de l'incendie industriel survenu à Rouen. Dispositif d'études « Santé post-incendie 76 ». Environnement, Risques & Santé. 2021 ; 20(2):164-70

Blanchard M, Leduc G, Sinno-Tellier S, et al. Surveillance de l'impact sanitaire immédiat de l'incendie du 26 septembre 2019, Rouen. Environnement, Risques & Santé. 2021 ; 1(1)

Empereur-Bissonnet P, Richard J-B, Motreff Y, et al. Impact sanitaire à court et moyen termes de l'incendie industriel de Rouen en 2019 : matériels et méthodes d'enquête. Environnement, Risques & Santé. 2021 ;1(1)

Direction régionale de l'environnement de l'aménagement et du logement de Normandie. Arrêté du 14 octobre 2019, imposant à la société LUBRIZOL France des prescriptions de mesures d'urgence pour son site situé dans les communes de Rouen et de Petit-Quevilly. DREAL Normandie. 2019

Golliot F., Blanchard M., Empereur-Bissonnet P., Le Lay E., Richard J.B., Moisan F., et al. Evaluation de l'impact sanitaire de l'incendie industriel survenu à Rouen. Dispositif d'études « Santé post-incendie 76 ». Environ Risque Sante. 2021 ; 20 :164-70

Empereur-Bissonnet P, Perrine A-L, Pédrone G, El Haddad M, Zeghnoun A, Richard J-B, et al. Santé Post Incendie 76 - Une étude à l'écoute de votre santé. Étude épidémiologique par questionnaire sur l'incendie industriel du 26 septembre 2019 à Rouen (France). Rapport principal. Saint-Maurice : Santé publique France ; 2021. 81 p.

Morel P, El Haddad M, Perrine A-L, Empereur-Bissonnet P, Richard J-B, Golliot F, et al. Santé Post Incendie 76. Une étude à l'écoute de votre santé. Étude épidémiologique par questionnaire suite à l'incendie industriel du 26 septembre 2019 à Rouen (France). Résultats du volet complémentaire sur la santé mentale. Saint-Maurice : Santé publique France ; 2021. 24 p.

Richard JB, Perrine AL, Zeghnoun A, Saoudi A, Pédrone G. Santé post-incendie 76 : une étude à l'écoute de votre santé. Bilan méthodologique. Saint-Maurice : Santé publique France, 2022. 67 p. Disponible à partir de l'URL : [www.santepubliquefrance.fr](http://www.santepubliquefrance.fr)

Haziza D, Beaumont JF. On the construction of imputation classes in surveys. International Statistical Review. 2007 ;75(1):25-43

Deroyon T, La correction de la non-réponse par repondération. Insee. Département des méthodes statistiques. Version n° 1, diffusée le 10 octobre 2017

Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine. 2015 ;34(28):3661-79

Andridge, R. R., and Little, R. J. A. (2010). "A Review of Hot Deck Imputation for Survey Non-response." International Statistical Review 78 :40–64.

Santin G, Geoffroy B, Bénézet L, Délézire P, Chatelot J et al. In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse. Journal of Clinical Epidemiology. 2014. 67(6) :722-730

Soullier R, Geoffroy-Pérez B, Guégen A, Bénézet L, Chatelot J et al. Correction de la non-réponse et estimation de prévalences : résultats issus de trois cohortes épidémiologiques ciblant les risques professionnels. 10e colloque sur les sondages, 24 au 26 octobre 2018

Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey : construction of scales and preliminary tests of reliability and validity. Med Care. 1996 ;34(3):220-33

Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries : results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998 ;51(11):1171-8

Warszawski et al., 2020, “En mai 2020, 4,5 % de la population en France métropolitaine a développé des anticorps contre le SARS-CoV-2 – Premiers résultats de l’enquête nationale EpiCoV”, Études et Résultats, n°1167, DREES, octobre

Leduc A, Deroyon T, Rochereau T, Renaud A. Premiers résultats de l’enquête santé européenne (EHIS) 2019 - Les dossiers de la DREES n° 78 - avril 2021