

SANTÉ
ENVIRONNEMENT

MARS 2022

MÉTHODES

SANTÉ POST-INCENDIE 76 :
UNE ÉTUDE À L'ÉCOUTE DE VOTRE SANTÉ.
BILAN MÉTHODOLOGIQUE

Résumé

Santé post-incendie 76 : une étude à l'écoute de votre santé. Bilan méthodologique

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. Dans le cadre du dispositif d'étude 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 la nature 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 et la qualité de vie plusieurs mois après l'événement.

Au sein d'une zone exposée (Rouen et ses alentours) et d'une zone témoin (Le Havre et ses alentours), un échantillon de 10 777 personnes 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). La collecte du volet principal 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 ont répondu à ce volet, soit 95 % de l'objectif fixé de 5 000 répondants, correspondant à un taux de réponse de 47,6 %. Près des deux tiers des répondants ont répondu par Internet (64,3%). En outre, 1 306 questionnaires portant sur un enfant ont été recueillis sur les 1 400 souhaités. Ce volet principal a été complété par un volet spécifique visant à évaluer les conséquences sur différents troubles de la santé mentale. Il a été proposé dans un second temps aux personnes ayant accepté d'être recontactées à l'issue du volet principal. La collecte du volet santé mentale s'est déroulée du 12 novembre au 21 décembre 2020, et a été exclusivement réalisée par questionnaire auto-administré (Internet ou papier). Parmi les 4 779 répondants au volet principal, 2 885 (60,4%) ont consenti à être recontactés pour le volet santé mentale, parmi lesquels 1 968 (68,2%) ont répondu à l'enquête.

Le recours à Fidéli s'est révélé d'une grande importance : disponibilité des coordonnées géographiques des logements qui a permis de sélectionner les personnes selon la proximité de leur lieu de résidence à l'incendie ; richesse des informations de contact disponibles ; disponibilité de données socio-économiques. La représentativité de l'échantillon des répondants à l'enquête, au regard des critères statistiques utilisés, était satisfaisante.

Le protocole multimode retenu pour cette enquête a favorisé la participation en offrant aux personnes sélectionnées plusieurs opportunités de participer. Néanmoins, il est également associé à l'existence potentielle d'un effet mode. En effet, l'analyse spécifique menée dans cette étude suggère l'existence d'un effet mode associé à un biais de mesure (les répondants Internet diffèrent des répondants par téléphone car les réponses diffèrent selon le mode de recueil). Pour de futures enquêtes, l'identification et le traitement de ce biais nécessiteront un protocole adapté permettant de disposer d'une variable permettant d'expliquer la participation indépendamment de l'intérêt pour l'enquête.

MOTS-CLÉS : ENQUÊTE DE SANTÉ ; ACCIDENT INDUSTRIEL ; MÉTHODE D'ENQUÊTE ; MULTIMODE ; EFFET MODE

Abstract

Santé post-incendie 76: a study exploring health impacts of the Lubrizol fire on surrounding communities. Methodological review

On 26 September 2019, a major fire broke out at the industrial/commercial sites of Lubrizol and NL Logistique situated in the port area of Rouen, France, causing numerous environmental nuisances: plumes of black smoke, falling debris and soot deposits, widespread and odorous emissions of volatile organic compounds, flames and explosions. In addition to generating a feeling of insecurity and psychological stress, exposure to these phenomena – the nature and intensity of which varied depending on the distance from the fire and the plume – could carry health consequences for the surrounding populations. As part of a study aimed at assessing the short-, medium- and long-term health impact of the fire, a survey on perceived health was conducted to assess the nature and frequency of the population's exposure to the various fire-related nuisances, as well as any symptoms and health problems experienced during and after the event. The survey also made it possible to study people's physical and mental health status, along with their quality of life, several months after the event.

The Surveys Division of the French National Institute of Statistics and Economic Studies (INSEE) drew a random sample of 10,777 people from an exposed area (Rouen and its surroundings) and a control area (Le Havre and its surroundings) using the 2019 edition of the Housing and Individual Demographic Files (Fidéli) of fiscal origin. The main survey was administered from 4 September to 12 November 2020, using a multi-mode collection protocol (Internet or telephone). A total of 4,779 people responded, i.e., 95% of the target of 5,000 respondents, corresponding to a response rate of 47.6%. Nearly two-thirds of respondents replied by Internet (64.3%). Furthermore, 1,306 of the collected questionnaires concerned children (out of a target of 1,400). In addition to the main survey, a specific supplementary component aimed at assessing the consequences for different mental health disorders was proposed in a second phase to those who had given their consent during the main survey. The mental health component was conducted from 12 November to 21 December 2020 using exclusively self-administered questionnaires (Internet or paper). Of the 4,779 respondents to the main survey, 2,885 (60.4%) agreed to be contacted again for the mental health component, of which 1,968 (68.2%) completed the questionnaire.

The use of the Fidéli database proved highly valuable for several reasons: availability of geographical coordinates for homes (making it possible to select people according to the proximity of their residence to the fire), wealth of available contact information, and availability of socio-economic data. In terms of the statistical criteria used, the representativeness of the survey respondent sample was satisfactory.

The multi-mode protocol chosen for this survey encouraged participation by offering the selected individuals several opportunities to participate. However, it is also associated with a potential mode effect. Indeed, the specific analysis conducted in this study suggests the existence of a mode effect associated with measurement bias (Internet respondents differ from telephone respondents because their answers differ according to the mode of administration). For future surveys, the identification and processing of this bias will require an adapted protocol that includes a variable capable of explaining participation independently of interest in the survey.

KEY WORDS: HEALTH SURVEY; INDUSTRIAL ACCIDENT; SURVEY METHODOLOGY; MULTI-MODE; MODE EFFECT

Auteurs

Jean-Baptiste Richard, Anne-Laure Perrine, Abdelkrim Zeghnoun, Abdessatar Saoudi, Gaëlle Pédrone (Santé publique France)

Remerciements

- À l'ensemble de l'équipe du projet Santé Post Incendie 76 à Santé publique France : Pascal Empereur-Bissonnet, Franck Golliot, Imane Khireddine, Mélanie Martel, Yvon Mottref, Emmanuelle Le Lay, Anne-Claire Colleville, Pauline Morel, Myriam Blanchard, Maria El Haddad
- Aux équipes ayant assuré la réalisation de cette enquête : l'institut IPSOS Observer, l'ensemble des enquêteurs et enquêtrices ainsi que les équipes travaillant sur le plateau téléphonique de Plérin.
- À nos collègues de l'Insee et de la Drees pour leurs précieux conseils : F Beck, N Palliod, S Faivre, L Castell, P Sillard, A Schreiber, S Legleye, T Deroyon, A Leduc et P Raynaud.
- À Noémie Soullier pour sa relecture attentive.
- Aux membres du comité du label et du comité du secret statistique.
- 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.

Santé publique France remercie tous les habitants de Seine-Maritime qui ont participé à l'enquête et pris le temps de répondre au questionnaire.

Sommaire

Résumé.....	1
Abstract.....	2
Auteurs, remerciements	3
INTRODUCTION	6
CHAMP DE L'ENQUÊTE	7
Zone exposée	7
Les Gens du voyage résidant sur l'aire d'accueil du Petit Quevilly.....	7
Zone témoin	8
Unités enquêtées	8
DESCRIPTION DU SONDAGE	10
Taille d'échantillon et allocation par strate	10
Base de sondage et échantillonnage.....	10
VOLETS DE L'ENQUÊTE	12
Volet principal.....	12
Volet Santé mentale	12
TERRAIN DU VOLET PRINCIPAL DE L'ENQUÊTE	14
Étude pilote	14
Modes de contacts disponibles.....	14
Matériel d'enquête.....	15
Communication autour de l'enquête	17
Protocole de collecte des données du volet principal de l'enquête	17
Participation au volet principal de l'enquête.....	19
Participation au questionnaire enfants de l'enquête.....	23
Participation des Gens du voyage	23
Sollicitation de la hotline	23
TERRAIN DU VOLET SANTÉ MENTALE.....	25
Protocole de collecte du volet Santé mentale	25
Participation au volet Santé mentale	25
TRAITEMENT DE DONNÉES À CARACTÈRE PERSONNEL ET PROTECTION DU SECRET STATISTIQUE	28
APPARIEMENT AUX DONNÉES DU SNDS	29
TRAITEMENTS POST-COLLECTE.....	30
Géocodage des adresses.....	30
Recodage des questions ouvertes.....	30
Traitement de la non-réponse partielle	30
Traitement de la non-réponse totale	30
CARACTÉRISTIQUES DES RÉPONDANTS	32
Distances standardisées	32
Volet principal.....	32
Volet Santé mentale	35

ANALYSE COMPLÉMENTAIRE : UNE ÉVALUATION DU BIAIS DE SELECTION	37
Objectif et méthodes mises en œuvre	37
Lien entre variables d'intérêt et variables relatives aux délais de réponse	39
Score de propension (Internet vs téléphone), parmi les répondants	42
Résultats des régressions	44
Interprétation des résultats	46
CONCLUSION.....	49
Références bibliographiques	51
Annexe 1. Lettre d'information.....	53
Annexe 2. Correction de la non-réponse totale : informations complémentaires.....	55

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. Il a produit un important panache de fumées noires pendant 12 heures sur une trajectoire Nord-Est, traversant les régions Normandie et Hauts-de-France, associé à des retombées de suie constatées au sol. Des émissions diffuses et odorantes de composés organiques volatils ont eu lieu et des odeurs ont été senties plusieurs mois après l'événement, lors de la dépollution des sites incendiés. Les riverains les plus proches des sites industriels ont été exposés à la vision et aux bruits de l'incendie, ainsi qu'à un sentiment d'insécurité et un stress psychologique. A cela s'ajoute la dispersion aux alentours de débris de toiture en fibrociment contenant des fibres d'amiante.

Ces expositions, dont la nature et le niveau ont varié selon la distance à l'incendie et au panache, peuvent avoir des conséquences sur la santé pour les différentes populations exposées. En réponse à une saisine du ministère de la Santé, Santé publique France a mis en place un dispositif d'étude visant à évaluer l'impact sanitaire à court, moyen et long termes de l'accident industriel survenu le 26 septembre 2019 sur les sites des entreprises Lubrizol et NL Logistique à Rouen : le dispositif « Santé post incendie 76 » (1).

Ce dispositif est composé de quatre volets, complémentaires à la surveillance des effets sanitaires immédiats mise en place dans le mois suivant l'incendie (2). Un de ces volets a consisté à mettre en œuvre une enquête de santé perçue, intitulée 'Une étude à l'écoute de votre santé', avec pour objectif principal d'évaluer la nature et la fréquence de l'exposition de la population aux différentes nuisances consécutives à l'incendie, des symptômes et problèmes de santé ressentis pendant et après l'événement, ainsi que d'étudier l'état de santé physique et mental et la qualité de vie plusieurs mois après l'événement, au moment de la collecte des données (3). Santé publique France a inscrit son action dans une démarche participative, afin d'être au plus proche des réalités de terrain, en favorisant la concertation et la co-construction avec les représentants des personnes directement concernées par l'accident industriel. L'objectif a été d'améliorer la qualité et la pertinence de cette enquête, en les adaptant au contexte local et en répondant du mieux possible aux préoccupations et aux besoins exprimés par la population. Un Groupe santé a ainsi été mis en place à Rouen en janvier 2020, composé d'associations, de professionnels de santé (CHU de Rouen, Conseil de l'Ordre des médecins, Conseil de l'Ordre des infirmiers, union régionale des médecins libéraux), d'élus locaux, de la métropole Rouen Normandie, de l'Université de Rouen.

Le projet de cette enquête de santé a été présentée au Conseil national de l'information statistique (Cnis) lors de la séance du 28 novembre 2019 de la commission « Démographie et questions sociales ». Un avis d'opportunité favorable lui a été délivré le 9 décembre 2019. Le projet a également été présenté au Comité du Label de la statistique publique le 22 juin 2020, et a obtenu un avis d'examen favorable.

Le présent rapport vise à présenter la méthode d'enquête : définition de la zone d'étude, déroulé du terrain, bilan de la collecte, traitements post-collecte effectués afin de réduire le plus possible l'impact d'un éventuel biais de sélection sur les résultats de cette enquête.

CHAMP DE L'ENQUÊTE

Zone exposée

Cette zone géographique a été constituée dans l'objectif d'inclure les différents types d'exposition de la population et de concentrer l'observation épidémiologique sur la population jugée *a priori* la plus exposée aux divers facteurs de risque émis par l'accident industriel. Plusieurs sources d'informations ont été utilisées : la liste des communes de Seine-Maritime identifiées par un arrêté préfectoral daté du 14/10/2019 comme ayant été exposées aux fumées noires et à leurs retombées au sol (4) ; une extraction des coordonnées géographiques des signalements d'odeurs relevés par Atmo Normandie¹ pendant une durée de 2 mois après l'incendie ; l'existence d'une zone de mise à l'abri pour les personnes résidant à proximité de l'incendie, leur recommandant de ne pas quitter leur domicile, et plus globalement d'une exposition spécifique des personnes résidant à proximité de l'incendie (menace d'une atteinte vitale, d'une destruction de leurs biens, exposition aux flammes, aux explosions et au bruit générés par l'incendie).

Une stratification géographique a été utilisée dans le cadre de cette enquête, comprenant les quatre strates suivantes (Figure 1) :

- une strate de proximité du lieu de l'accident, définie par un critère purement géographique et divisée en deux cercles concentriques : 0-700 mètres et 700-1500 mètres du lieu de l'accident ;
- une strate « Sud-Ouest », qui regroupe 11 communes non exposées au panache de fumées noires mais ayant été le lieu d'au moins six signalements d'odeurs à Atmo Normandie au cours des deux mois qui ont suivi l'incendie ;
- une strate « Nord-Est proche », regroupant des zones exposées aux odeurs et situées sous le panache de fumées noires ;
- une strate « Nord-Est éloigné », comprenant les communes de Seine-Maritime ayant été exposées au panache et à ses retombées au sol, mais non concernées par une fréquence élevée de signalements d'odeurs.

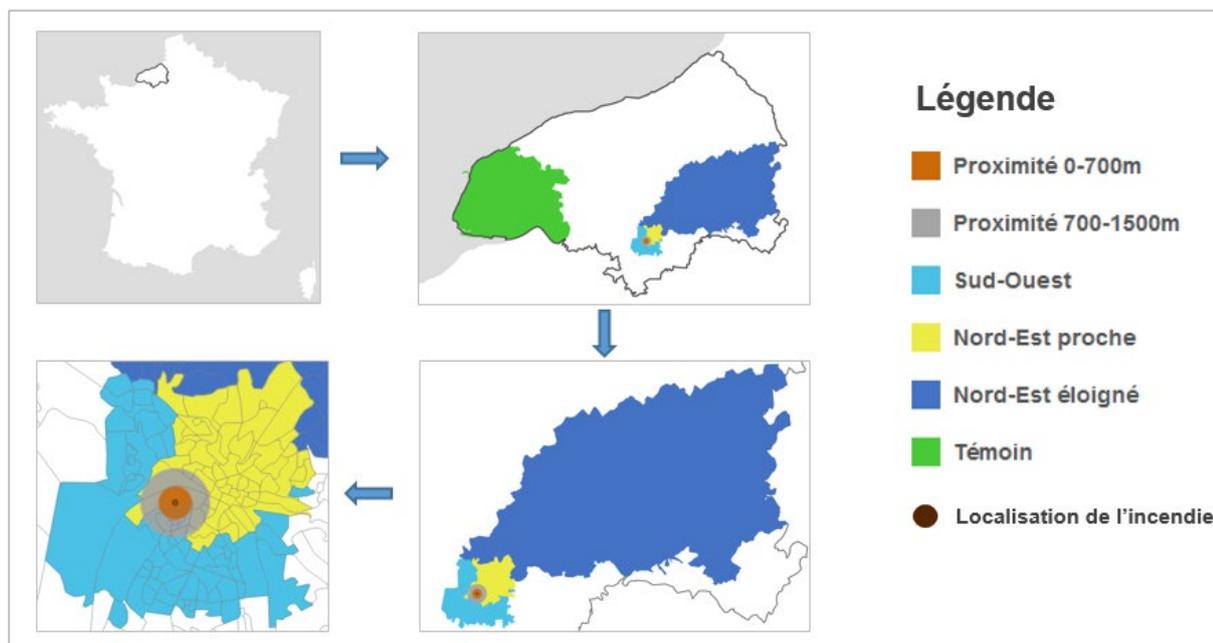
La zone exposée comprend 122 communes, toutes situées en Seine-Maritime, la plus éloignée étant distante d'environ 60 kilomètres du lieu de l'incendie.

Les Gens du voyage résidant sur l'aire d'accueil du Petit Quevilly

Les Gens du voyage qui résident sur l'aire d'accueil du Petit-Quevilly, à proximité immédiate de l'incendie, ne pouvaient pas être tirés au sort dans la base de sondage utilisée pour constituer l'échantillon d'enquête. Ils ne sont pas répertoriés dans la base de sondage. Mais grâce à la collaboration active de l'association Relais-Accueil des Gens du voyage, il a été possible de faire participer ce groupe de population à l'étude.

1. Association de surveillance de la qualité de l'air : <http://www.atmonormandie.fr/>

Figure 1. Stratification géographique de la zone enquêtée (France, département de Seine-Maritime)



Source : SPI76 : une étude à l'écoute de votre santé. Santé publique France 2021

Zone témoin

Une zone témoin a été choisie afin de disposer de résultats dans une population de référence non exposée à l'accident industriel. La ville du Havre et les communes situées dans les 29 kilomètres de sa périphérie nord-est a été choisie comme zone témoin, à partir des éléments suivants :

- une localisation non concernée par l'exposition aux nuisances liées à l'incendie ;
- une localisation géographique comprenant une exposition d'origine industrielle, de nature équivalente à celle de la zone exposée ;
- une localisation géographique suffisamment éloignée de Rouen pour limiter l'inclusion de personnes concernées par une exposition aux nuisances de l'incendie liée à leur activité professionnelle, mais située dans la même région que Rouen ;
- des caractéristiques sociodémographiques comparables à celle de la zone exposée, notamment du point de vue de la taille et de la densité de population (rurale/urbaine).

Unités enquêtées

Le champ de l'enquête a été défini par l'ensemble des adultes âgés de 18 ans et plus et des enfants âgés de moins de 18 ans, vivant dans un logement ordinaire² situé dans la zone d'étude au 26 septembre 2019 (jour de l'incendie des usines Lubrizol et NL Logistique), quel que soit le lieu de résidence actuel. La population effectivement interrogée est celle des adultes âgés de 18 ans et plus au moment de l'incendie. Les enfants de moins de 18 ans ont été interrogés indirectement à travers le questionnement de leurs parents si l'un d'entre eux était tiré au sort (un seul enfant par ménage était tiré au sort au sein du ménage dans lequel un adulte était tiré au sort ; un seul adulte par ménage pouvait être tiré au sort ; pour un enfant vivant en garde alternée, ses deux parents pouvaient être tirés au sort s'ils résidaient dans la zone d'étude lors de l'incendie).

2. Le logement ordinaire est un logement défini par opposition à un logement en résidence offrant des services spécifiques (résidences pour personnes âgées, pour étudiants, de tourisme, à vocation sociale, pour personnes handicapées...).

Plusieurs critères d'exclusion (certains au moment du tirage et d'autres au moment de la vérification de l'éligibilité) ont été définis :

- pour les habitants de la zone exposée : avoir emménagé dans la zone exposée après l'incendie ;
- pour les habitants de la zone exposée : avoir résidé dans la zone exposée à titre occasionnel le jour de l'incendie (pour les vacances, de façon intermittente, pour des raisons professionnelles, etc.) ;
- pour les habitants de la zone témoin : avoir été exposé, dans le cadre d'une activité professionnelle ou d'une autre activité, à l'incendie, que ce soit par ses odeurs, son panache de fumée, ses retombées de suie ou toute autre exposition ;
- pour les non-répondants par internet : ne pas parler français ou être en incapacité physique ou mentale de répondre au téléphone ;
- les mineurs tirés au sort dont l'adulte répondant dans le même foyer n'était pas le parent ou le ou le tuteur légal.

DESCRIPTION DU SONDAGE

Taille d'échantillon et allocation par strate

Le calcul du nombre de sujets nécessaire a reposé sur des hypothèses faites a priori en fonction des principaux objectifs de l'étude, en particulier : estimer la prévalence des symptômes ressentis avec une précision donnée ; comparer la prévalence des symptômes ressentis selon différentes zones d'expositions ; étudier la relation entre l'intensité de l'exposition et la fréquence des symptômes ressentis. Un objectif de 5 000 individus adultes et de 1 400 mineurs était visé sur l'ensemble de la zone enquêtée, permettant d'estimer une prévalence de 10% avec une précision relative proche de 10% pour les adultes et de 20 % pour les mineurs.

L'objectif était d'atteindre un effectif minimal de 1 000 répondants adultes dans toutes les strates géographiques. Cela a nécessité, compte tenu du faible nombre de personnes résidant dans la strate de proximité, d'interroger un adulte dans chacun des logements situés à moins de 700 mètres de l'incendie.

La taille de l'échantillon à tirer au sort a été déterminée à partir des hypothèses suivantes :

- un taux de mobilité, tenant compte des déménagements, décès et départs en institution ayant pu survenir entre la date de référence de la base de sondage (1^{er} janvier 2019) et la période de collecte (septembre 2020, soit 21 mois plus tard), estimé à 10% ;
- un taux de réponse estimé à 55%, et meilleur dans les strates de proximité (60%) que dans la zone témoin (40%).

En outre, un échantillon de réserve de 20% de la taille de chacune des strates de l'échantillon principal a été établi. Cet échantillon de réserve pouvait être ouvert avec des règles de déclenchement par strate, selon le taux de participation observé.

Base de sondage et échantillonnage

La constitution de la base de sondage et l'échantillonnage ont été réalisés en collaboration avec la Division Sondages de l'Insee, dans le millésime 2019 du Fichier démographique d'origine fiscale sur les logements et les individus (Fidéli). Dans chacune des strates, la procédure a consisté à tirer un adulte maximum par logement, puis à tirer au sort un enfant parmi les enfants présents dans chaque logement où un individu adulte avait été échantillonné. Concernant la strate de proximité (0-700m), il s'agit d'une strate exhaustive dans laquelle tous les logements ont été sélectionnés avec le tirage d'un adulte par logement. Dans les autres strates, la méthode utilisée reposait sur un tirage systématique sur fichier trié sur les variables suivantes : le décile de niveau de vie, le type de ménage (en distinguant les couples avec enfants mineurs / majeurs et les familles monoparentales avec enfants mineurs / majeurs), l'identifiant de logement et l'identifiant de l'individu. La base de tirage pour les mineurs a été obtenue en appariant l'échantillon adulte avec la base de sondage, sur l'identifiant du logement. Un tirage stratifié par logement a ensuite été réalisé sur une base triée par l'identifiant de logement, l'âge et l'identifiant de l'individu. Au total, la base de sondage comprenait 687 805 individus, dont 539 249 adultes et 148 556 enfants, pour l'ensemble de la zone d'étude, comprenant la zone exposée et la zone témoin. Un échantillon de 10 777 adultes a été tiré au sort, 7 999 dans la zone exposée et 2 778 dans la zone témoin (Tableau 1), le taux de sondage variant de 57,8% dans la strate 0-700m à 1,9% dans la strate Nord-Est proche. Un échantillon de 3 234 enfants a également été tiré au sort : 2 348 dans la zone exposée et 886 dans la zone témoin. Notons par ailleurs que dans cette étude, nous n'avons pas eu recours à l'utilisation de l'échantillon

de réserve pour la zone exposée et témoin. L'objectif en termes d'inclusion a été atteint en utilisant uniquement l'échantillon principal pour chaque zone.

Tableau 1. Base de sondage et échantillon d'adultes et enfants tirés au sort, par strate

Strate	Base de sondage		Echantillon principal		Taux de sondage	
	Nombre d'adultes	Nombre d'enfants	Nombre d'adultes	Nombre d'enfants	Adultes	Enfants
Proximité 0-700m	1 061	238	613	134	57,8%	56,3%
Proximité 700-1500m	14 913	4 664	1 326	425	8,9%	9,1%
Sud-Ouest	93 436	25 908	2 020	600	2,2%	2,3%
Nord-Est proche	104 419	23 175	2 020	493	1,9%	2,1%
Nord-Est éloigné	55 693	16 521	2 020	696	3,6%	4,2%
Total zone exposée	269 522	70 506	7 999	2 348	3,0%	3,3%
Témoin	269 727	78 050	2 778	886	1,0%	1,1%
Total zone enquêtée	539 249	148 556	10 777	3 234	2,0%	2,2%

VOLETS DE L'ENQUÊTE

Volet principal

Le questionnaire du volet principal de l'enquête (accessible en ligne³) est constitué de deux parties qui se réfèrent à des temporalités différentes :

Une partie qui porte sur la période de l'accident et ses suites :

- Expositions perçues durant l'incendie ;
- Nuisances liées à l'incendie : odeurs, suies, débris ;
- Comportements adoptés suite à l'incendie ;
- Information et connaissance des recommandations ;
- Symptômes et problèmes de santé ressentis (adulte et enfant), recours aux soins ;
- Grossesse et allaitement ;

Une partie qui porte sur la situation au moment de l'interrogation, c'est-à-dire à moyen terme, environ 1 an après l'incendie, et composée des modules suivants :

- Santé globale : échelle internationale MOS SF-12 ;
- Indicateurs en lien avec la santé : taille, poids, recours aux soins, consommations de tabac et d'alcool ;
- Attitudes à l'égard de l'environnement ;
- Proximité et attitudes vis-à-vis de la Covid-19 ;
- Conditions de vie, profession ;
- Lien professionnel avec les entreprises Lubrizol et NL Logistique ;

En zone témoin, le questionnaire ne comporte pas les modules portant sur les expositions perçues durant l'incendie, les comportements adoptés suite à l'incendie, le ressenti et la confiance des personnes par rapport aux informations et la gestion de l'événement. Pour le reste, il est identique au questionnaire administré à l'échantillon exposé à l'accident industriel.

Volet Santé mentale

Le volet principal d'enquête a été complété par un volet spécifique visant à évaluer les conséquences sur différents troubles de la santé mentale. Il a été proposé dans un second temps aux personnes de l'ensemble de la zone d'étude (exposée et témoin) ayant accepté d'être recontactées. Le questionnaire du volet Santé mentale comprend trois instruments normalisés de mesure psychométrique de la santé mentale, traduits et validés en français :

- GAD-7 (*Generalized Anxiety Disorder*), pour estimer l'anxiété généralisée et en mesurer la sévérité au cours des deux semaines qui précèdent l'interrogation (5) ;
- PHQ-9 (*Patient Health Questionnaire*), pour mesurer la présence et l'intensité des symptômes dépressifs au cours des deux dernières semaines (6) ;
- PCL-5 (*PTSD Check List for DSM5*) pour recueillir les symptômes de stress post-traumatique et en évaluer l'intensité dans le dernier mois (7).

3. <https://www.santepubliquefrance.fr/regions/normandie/documents/questionnaire/2021/sante-post-incendie-76-une-etude-a-l-ecoute-de-votre-sante.-questionnaire>

Il comprend aussi certaines questions destinées à renseigner les facteurs d'ajustement spécifiques à la santé mentale : exposition au cours des 12 derniers mois à des évènements de vie marquants, antécédents de prise en charge psychologique (suivi par un thérapeute, prise de médicament psychotrope) et antécédents d'exposition à un événement potentiellement traumatique.

Pour les personnes qui auraient pu sentir mal après avoir répondu à ce questionnaire, plusieurs messages étaient affichés.

- Sur le questionnaire papier, un encadré « En cas de difficulté de répondre au questionnaire » présentait le message suivant : « Il est possible que vous soyez déstabilisé(e) par certaines questions ou en difficulté pendant ou à l'issue du remplissage de ce questionnaire. Dans ce cas, n'hésitez pas à faire une pause pour récupérer, à vous laisser du temps pour terminer de répondre à toutes les questions. Si vous ne vous sentez pas bien, il pourrait être nécessaire de consulter votre médecin généraliste afin de vous faire aider et, si nécessaire, réaliser un bilan de votre état émotionnel et psychologique. Vous trouverez des informations qui pourraient vous être utiles sur le site du Centre national de ressources et de résilience (CN2R) : <http://cn2r.fr/> ».
- Sur le questionnaire Internet un premier message était affiché au début du questionnaire : « Nous vous rappelons que, si vous en ressentez le besoin, vous pouvez consulter votre médecin (généraliste ou spécialiste) afin de faire avec lui un bilan approfondi de votre état émotionnel et psychologique ». Selon les réponses obtenues aux différentes échelles de questions (GAD-7, PHQ-9, PCL-5), le message suivant était également affiché en fin de questionnaire : « Vos réponses au questionnaire sont en faveur de symptômes d'anxiété élevés et/ou de symptômes dépressifs élevés et/ou de symptômes qui laissent penser que vous pourriez souffrir d'un état de stress post traumatique. Si vous ne l'avez pas déjà fait, nous vous conseillons de consulter votre médecin (généraliste ou spécialiste) afin de faire avec lui un bilan approfondi de votre état émotionnel et psychologique. »

TERRAIN DU VOLET PRINCIPAL DE L'ENQUÊTE

Étude pilote

La passation du questionnaire a fait l'objet de deux phases de test par téléphone, réalisées en juin et juillet 2020. Elles ont permis de valider l'acceptation de participer à l'enquête, la compréhension des questions et la fluidité de la passation. Ces tests avaient également pour buts de s'assurer que le délai de plusieurs mois entre l'événement et l'enquête entraînait peu de difficultés à répondre et d'évaluer la durée d'administration, avec pour objectif qu'elle ne dépasse pas 30 minutes par téléphone.

Modes de contacts disponibles

Les informations permettant de contacter les personnes tirées au sort provenaient de deux sources :

- les informations de contact dans Fidéli : nom, prénom, année de naissance (utilisée comme information de contact lorsque le nom/prénom était manquant), adresse postale, disponibles pour la totalité ou presque des individus, et numéro de téléphone fixe et/ou mobile et adresse mail, pour l'individu sélectionné ou pour la personne de référence du foyer fiscal ;
- un enrichissement téléphonique, réalisé par le prestataire de collecte, concernant les adresses pour lesquelles aucun numéro de téléphone n'était disponible dans Fidéli.

Au final, différents modes de contacts pouvaient être utilisés : l'adresse postale pour l'ensemble de l'échantillon, le mail pour 73 % des personnes, le téléphone mobile pour 51% et au moins un numéro de téléphone (fixe ou mobile) pour 80% d'entre elles (Tableau 2). À noter que pour une minorité d'adultes (1%), l'identité (nom et prénom) n'était pas disponible dans la base de sondage. Dans ces cas, l'année de naissance, issue de la base de sondage, était utilisée pour identifier la personne tirée au sort au sein d'un logement.

Tableau 2. Informations de contact disponibles pour la collecte, après enrichissement-échantillon adultes

Information de contact disponible (personne sélectionnée ou déclarant fiscal)	n	%
Adresse postale	10 777	100%
Identité connue (Nom/Prénom)	10 678	99%
Mail	7 844	73%
Numéro de téléphone (fixe ou mobile)	8 645	80%
<i>Dont numéro de téléphone mobile</i>	<i>5 507</i>	<i>51%</i>

Les informations de contact disponibles dans la base de sondage étaient liées à certaines caractéristiques sociodémographiques. Les personnes disposant d'un faible niveau de vie (1^{er} décile de niveau de vie) étaient en effet surreprésentées parmi les individus pour lesquels aucun mail n'était disponible (16,8% lorsqu'uniquement l'adresse postale était disponible ; 14,9% lorsque seul un téléphone était disponible en plus de l'adresse postale). Les jeunes adultes étaient surreprésentés au sein des personnes pour lesquelles seul un mail était disponible en plus de l'adresse postale (c'est-à-dire sans téléphone disponible ; 32% vs 19,8% dans l'ensemble de la base de sondage). A l'inverse, les habitants de zone rurale étaient sous-représentés au sein des personnes pour lesquelles seule l'adresse postale était disponible (11,8% vs 16,8% dans l'ensemble de la base) (Tableau 3).

Tableau 3. Caractéristiques sociodémographiques des individus selon les modes contacts disponibles dans la base de sondage

Contacts disponibles	% 1er niveau de décile	% 18-29 ans	% zone rurale
Uniquement adresse postale	16,8%	20,4%	11,8%
Adresse postale + mail uniquement	9,2%	32,0%	15,4%
Adresse postale + téléphone uniquement	14,9%	9,5%	17,9%
Adresse postale, mail et téléphone	7,2%	20,6%	17,7%
Global	9,7%	19,8%	16,8%

Matériel d'enquête

Pendant toute la période de la collecte, une ligne téléphonique (hotline, numéro vert et gratuit : 0800 942 590) ainsi qu'une adresse mail (contact@sante-post-incendie-76.fr) ont été mises en place afin de répondre aux interrogations et aider les personnes sélectionnées à répondre au questionnaire.

Des lettres annonces ainsi que des mails-annonces ont été envoyés à tous les adultes tirés au sort quelques jours avant le début de la collecte. Ces lettres et mails ont été adaptés en fonction de la zone enquêtée (zone exposée, zone témoin), et selon le fait que l'identité de la personne sélectionnée était connue ou non. Quatre modèles étaient ainsi déclinés :

- zone exposée avec nom/prénom (voir annexe 1) ;
- zone exposée sans nom/prénom ;
- zone témoin avec nom/prénom ;
- zone témoin sans nom/prénom.

Les lettres et mails annonces contenaient les informations suivantes :

- une présentation de l'enquête ;
- les identifiants personnels ainsi qu'un QR CODE pour se connecter au questionnaire Internet (*Computer Assisted Web Interview*) et au site Internet ;
- des explications pour répondre au questionnaire ;
- les démarches à suivre pour participer par téléphone ;
- les coordonnées de la hotline ;
- les mentions légales ;
- chaque courrier était, en outre, accompagné d'un dépliant⁴ de présentation de l'enquête, élaboré en prenant en compte les besoins de littératie, afin d'être compris par le plus grand nombre de participant. Son contenu était complémentaire à la lettre d'annonce.

4. www.santepubliquefrance.fr/regions/normandie/documents/depliant-flyer/2020/une-etude-a-l-ecoute-de-votre-sante-depliant-a5

Un site Internet dédié à l'étude⁵ était proposé afin de communiquer de plus amples informations sur l'étude, sur l'incendie et toutes autres informations sur le sujet, et afin de répondre en ligne au questionnaire (Figure 2). Dans un espace personnel sécurisé, la personne sélectionnée pouvait :

- cliquer sur un lien pour répondre à son questionnaire ;
- mettre à jour son profil ;
- prendre rendez-vous pour être recontactée par téléphone par un enquêteur.

Figure 2. Page d'accueil du site internet de l'étude « une étude à l'écoute de votre santé »

The screenshot shows the homepage of the website. At the top, there is a banner with a blue-tinted image of a city and a large plume of smoke. On the left of the banner is the logo for 'Santé publique France' and on the right is the French national emblem with the motto 'Liberté • Égalité • Fraternité' and 'RÉPUBLIQUE FRANÇAISE'. Below the banner is a navigation menu with the following items: ACCUEIL, MON PROFIL, RÉPONDRE À L'ÉTUDE, QUESTIONS/RÉPONSES, CONTACTS, TÉLÉCHARGEMENT.

The main content area is titled 'Santé Post Incendie 76 : volet complémentaire'. Below the title, it states: 'L'étude principale est terminée, un deuxième volet de l'enquête commence...'. To the left of this text is a map of France with a purple shaded region in the northwest, representing the area affected by the 2019 industrial fire in Rouen.

On the right side of the page, there are two sections. The first is 'POUR ACCÉDER À VOTRE QUESTIONNAIRE EN LIGNE', which includes a circular icon with a hand cursor. The second is 'POUR CONTACTER LA HOTLINE DÉDIÉE À L'ENQUÊTE', which lists contact options: 'Par téléphone' with the number '0 800 942 590' (Service à appel gratuit) and 'Par e-mail' with the address 'contact@sante-post-incendie-76.fr'. Below this is a section titled 'VOS COORDONNÉES' with the instruction: 'Pour gérer vos coordonnées, connectez-vous sur votre espace sécurisé «Mon profil»'.

Ce site proposait également un certain nombre de liens utiles renvoyant vers des documents en rapport avec cette enquête ou vers le site de Santé publique France :

- Comment répondre à l'enquête ;
- Nous contacter au sujet de l'enquête ;
- Le dispositif Santé Post Incendie 76 ;
- La réponse de Santé publique France à l'accident industriel du 26 septembre 2019 à Rouen ;
- L'incendie industriel de Rouen : que s'est-il passé ?
- Impact sanitaire immédiat : quelle surveillance a été mise en place ?
- Mentions légales - Santé Post Incendie 76.

5. www.sante-post-incendie-76.fr/ Site ouvert uniquement durant la collecte.

Communication autour de l'enquête

L'ensemble de la communication a été co-construite avec le Groupe santé (11 réunions 2020-2021) et discuté avec un Comité d'appui thématique. Différents outils de communication ont été créés pour faire connaître l'étude et pour mobiliser différents relais, à destination de différents publics (parties prenantes, acteurs relais, participants à l'étude et grand public) :

- un espace dédié⁶ pour le dispositif « Santé Post incendie 76 » sur le site Internet de Santé publique France : un effort d'écriture et de pédagogie a été réalisé pour améliorer l'accessibilité des contenus techniques : schémas, chronologie de l'enquête, questions-réponses, etc. ;
- une affiche au format 40x60 cm⁷, ayant pour objectif de promouvoir l'enquête en population générale. Elle a été transmise à différents relais pour être affichée dans les communes de la zone exposée (mairies, centres médico-sociaux, cabinets médicaux, pharmacies, PMI, associations de quartier, etc.) ;
- un dépliant de 4 pages format A4⁸ ayant pour objectif de présenter l'étude de manière détaillée et envoyé : aux professionnels de santé de la zone d'étude (pharmacies, centres de santé (dont dispensaires), maisons et pôles de santé pluridisciplinaires, centres de PMI, médecins généralistes) ; aux institutionnels (ARS de Normandie, Conseil départemental de Seine Maritime / conseillers départementaux, service communication de la Région Normandie, députés des circonscriptions pour la Normandie, Métropole de Rouen, Préfecture de Seine Maritime) ; aux mairies des communes concernées par l'étude. Cet envoi a été réalisé afin que les acteurs relais puissent bénéficier d'une information en amont du grand public et pouvoir répondre aux questions posées ;
- un dépliant de 4 pages format A5⁹ pour les participants à l'enquête, plus court, et donnant des éléments simples pour inviter à participer à l'étude ;
- des courriers, envoyés à l'intention des élus locaux (maires, députés, métropole, département), des professionnels de santé, des médecins généralistes et des pharmaciens de la zone exposée, ainsi qu'aux mairies de la zone témoin ;
- l'étude a été plusieurs fois présentée au comité de transparence et de dialogue (CTD) présidé par le préfet de Seine-Maritime et rassemblant l'ensemble des parties prenantes ;
- un communiqué de presse a été adressé au lancement de l'étude (juillet 2020).

Protocole de collecte des données du volet principal de l'enquête

La collecte des données du volet principal de l'enquête a été réalisée par questionnaire. Elle a été organisée selon un protocole multimode, en proposant d'abord une auto-administration du questionnaire sur Internet (CAWI) puis une administration par téléphone avec un enquêteur (CATI), la collecte par Internet pouvant se poursuivre jusqu'à la fin de la période du terrain de l'enquête. Seul l'individu sélectionné pouvait être interrogé pour répondre au questionnaire, le questionnaire ne pouvant être rempli qu'après validation de l'identité du répondant. Le protocole d'invitation et de relance a été adapté selon les modes de contacts disponibles (Tableau 4).

6. www.santepubliquefrance.fr/regions/normandie/articles/incendie-industriel-du-26-septembre-2019-a-rouen

7. www.santepubliquefrance.fr/regions/normandie/documents/affiche/2020/une-etude-a-l-ecoute-de-votre-sante

8. www.santepubliquefrance.fr/regions/normandie/documents/depliant-flyer/2020/une-etude-a-l-ecoute-de-votre-sante-depliant-a4

9. www.santepubliquefrance.fr/regions/normandie/documents/depliant-flyer/2020/une-etude-a-l-ecoute-de-votre-sante-depliant-a5

Tableau 4. Répartition des modes de contacts disponibles pour les individus tirés au sort, en plus de l'adresse postale

Groupe pour les contacts	n	%
Groupe 1 - Aucun	1 321	12%
Groupe 2 - Uniquement un mail	813	8%
Groupe 3 - Uniquement un téléphone	1 612	15%
Groupe 4 - Mail et téléphone	7 031	65%

Les principales étapes du protocole de collecte (Figure 3) ont été les suivantes :

- envoi du courrier postal d'invitation le vendredi 4 septembre 2020, suivi d'un mail d'invitation et d'un SMS d'invitation¹⁰, envoyés le lundi 7 septembre ;
- 4 relances par mail pour les personnes des groupes 2 et 4 et 4 relances par SMS¹¹ pour les personnes des groupes 3 et 4 avec un numéro de téléphone mobile ;
- 2 relances par courrier postal pour les personnes du Groupe 1 ;
- 1 relance par courrier postal pour les personnes du Groupe 3 avec uniquement un numéro de téléphone fixe.

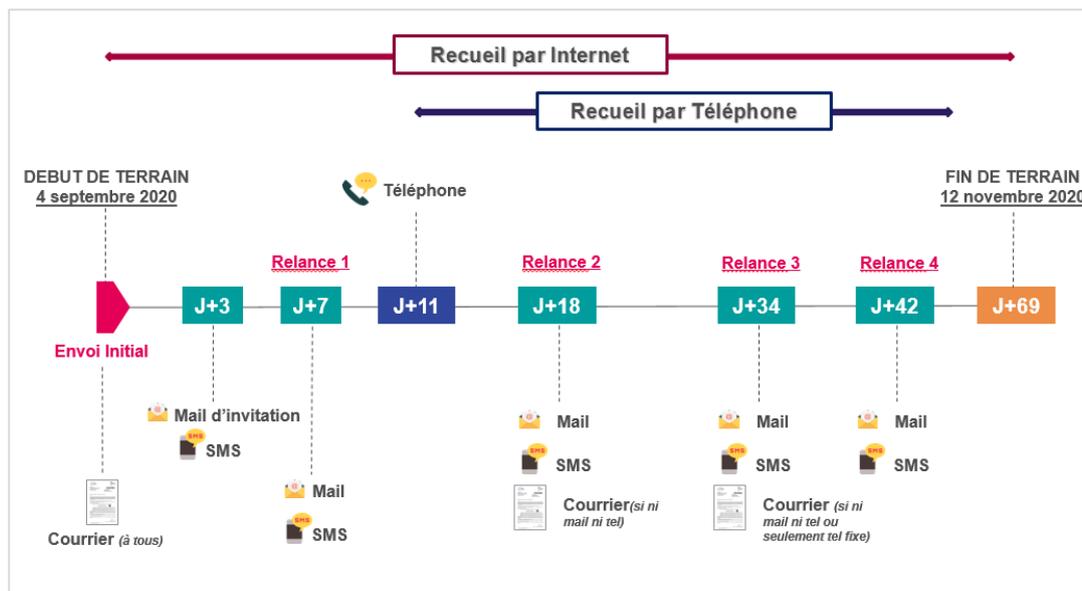
Une relance par téléphone avec passation de questionnaire a été faite entre le 15 septembre et le 17 octobre 2020. Le protocole d'appel était le suivant :

- chaque numéro de téléphone pouvait être appelé jusqu'à 20 fois, l'enquêteur laissant sonner jusqu'à 8 tonalités ;
- chaque numéro de téléphone pouvait être appelé au maximum 3 par jour ;
- dans le cas où plusieurs numéros sont disponibles pour joindre un individu : les appels sont priorisés sur mobile par rapport au fixe ; le nombre de tentatives est réparti en fonction du nombre de numéros disponibles (si 2 numéros sont disponibles, 10 tentatives sont réalisées sur le 1^{er} numéro puis 10 sur le 2nd) ;
- en cas de numéro occupé, un rappel est effectué 15 minutes plus tard ;
- un dépôt de message vocal est effectué après 2 appels sans réponse : « *Santé Publique France cherche à vous joindre pour participer à une étude de santé suite à l'accident industriel de Lubrizol et NL Logistique de septembre 2019. Merci du temps que vous pourrez nous accorder. Plus d'informations sur le site Internet www.sante-post-incendie-76.fr ou en appelant le 0800 942 590, numéro vert gratuit.* ».

10. « Santé publique France cherche à joindre Nom-Prénom pour une étude de santé suite à l'incendie de Lubrizol (2019). Répondez à cette enquête en cliquant sur « lien court ». Merci à vous ».

11. « Nous n'avons pas reçu votre réponse à l'étude de santé suite à l'incendie de Lubrizol (2019). Répondez à cette enquête en cliquant sur -lien court-. Merci à vous ».

Figure 3. Étapes du protocole de collecte des données du volet principal de l'enquête



Pour ce volet de l'enquête, 22 enquêteurs ont été formés, à partir d'un document de formation élaboré et présenté conjointement par Ipsos et Santé publique France. Cette formation abordait les points suivants : le contexte de l'étude, le dispositif global mis en place, le contenu détaillé du questionnaire, la phase contact et l'argumentation ainsi que des exercices de mise en situation.

En ce qui concerne les Gens du voyage, le questionnaire (volet principal et volet Santé mentale) a été administré en face-à-face les 2 et 3 décembre 2020 dans les locaux du relais d'accueil des Gens du voyage à Sotteville-lès-Rouen.

Participation au volet principal de l'enquête

La collecte s'est déroulée du 4 septembre au 12 novembre 2020. Parmi les 10 777 personnes tirées au sort, 755 se sont révélées hors-champ pour les raisons suivantes : elles étaient décédées, le dialogue était impossible, elles n'étaient pas présentes dans la zone exposée le jour de l'incendie ou elles avaient été exposées alors qu'elles résidaient en zone témoin (Tableau 5). Parmi les personnes dans le champ de l'enquête, 38,3% n'ont pu être contactées par téléphone, 14,1% ont refusé de participer. Au final, 4 773 personnes ont répondu au volet principal de l'enquête, correspondant à un taux de réponse de 47,6 %. Près des deux tiers des répondants ont répondu par CAWI (64,3%). La durée moyenne du questionnaire était de 24 minutes par téléphone (26 minutes en zone exposée, 19 minutes en zone témoin), et de 28 minutes par Internet.

Tableau 5. Bilan de participation à l'enquête principale

Echantillon initial - Adultes	10 777	
Hors-Champ	755	
HC - personne décédée	74	
HC- n'habitait pas dans la zone en septembre 2019	348	
HC - personne exposée de la zone témoin	100	
HC - personne contactée mais dialogue impossible	233	
Total dans le champ	10 022	100%
Non répondants (non répondant Internet, non répondant téléphone sans contact établi)	3 839	38,3%
Refus de participer, abandons	1 410	14,1%
Questionnaires complets	4 773	47,6%

La participation, de même que le mode de recueil, était très liée aux modes de contacts disponibles ainsi qu'à la strate géographique de résidence (Tableau 6). Les personnes dont on ne disposait que de l'adresse postale, et qui n'ont été informées de l'enquête que par courrier, ont participé à hauteur de 11,8%. C'est le cas d'un quart (25,5 %) des personnes pour lesquelles seul le mail était disponible en plus de l'adresse postale, 35,6% de celles pour lesquelles seul un numéro de téléphone était disponible en plus de l'adresse postale et 60,2 % de celles pour lesquelles l'ensemble de ces informations (adresse, mail et téléphone) étaient disponibles. Notons également que 11% des personnes tirées au sort n'ont reçu aucun des courriers postaux invitant à participer à l'enquête, ces derniers étant revenus en « n'habite pas à l'adresse indiquée » (NPAI), et qu'ainsi 187 (1,7 %) personnes tirées au sort n'ont pu être informées de l'enquête faute de moyens alternatifs pour les contacter.

Selon la zone géographique, la participation est apparue plus faible en zone témoin (40,2%) qu'en zone exposée (50,2%), et au sein de la zone exposée plus faible dans la strate Sud-Ouest (45,8%), située dans la direction opposée au panache de fumées.

Tableau 6. Participation selon le mode de contact (en plus de l'adresse postale) et selon la strate géographique

	Nombre de répondants	Participation (%)	% CAWI
Mode de contact			
Aucun	153	11,8%	82,4%
Uniquement un mail	202	25,5%	98,5%
Uniquement un téléphone	510	35,6%	21,8%
Mail et téléphone	3 908	60,2%	67,4%
Strate géographique			
Proximité 0-700m	301	52,0%	71,8%
Proximité 700-1500m	597	48,6%	67,8%
Sud-Ouest	864	45,8%	62,5%
Nord-Est proche	1 028	54,3%	74,6%
Nord-Est éloigné	968	50,9%	62,4%
Zone exposée	3 758	50,2%	67,4%
Zone témoin	1 015	40,2%	53,1%
Total zone enquêtée	4 773	47,7%	64,3%

La très grande majorité des questionnaires complets a été recueillie durant les 30 premiers jours de collecte (91% des questionnaires), 40% durant les 10 premiers jours, correspondant à la première phase de collecte dédiée exclusivement au recueil par Internet (Figure 4). Durant la collecte, certains pics ont été observés concernant le nombre de questionnaires recueillis quotidiennement, reflétant l'impact des différentes relances, notamment sur les réponses recueillies par Internet. Au sein des répondants par internet, alors que 7% ont répondu sans avoir reçu de mail (donc uniquement après le premier courrier d'invitation), 40% ont répondu après le mail d'invitation uniquement, 28% après la première relance par mail, 19% après la seconde relance par mail, tandis que seuls 6% d'entre eux ont répondu après les 3^{ème} et 4^{ème} relances (Figure 5). Par téléphone, 60% des répondants ont répondu lors des 4 premiers appels, 19% entre 5 et 8 appels, 11% entre 9 et 14 appels et 9 % ont répondu après plus de 15 appels (Figure 6). Le nombre d'appels moyen pour les répondants par téléphone s'élevait ainsi à 5,7 appels.

Figure 4. Répartition par jour de collecte et par mode du nombre de questionnaires recueillis et total de questionnaires recueillis

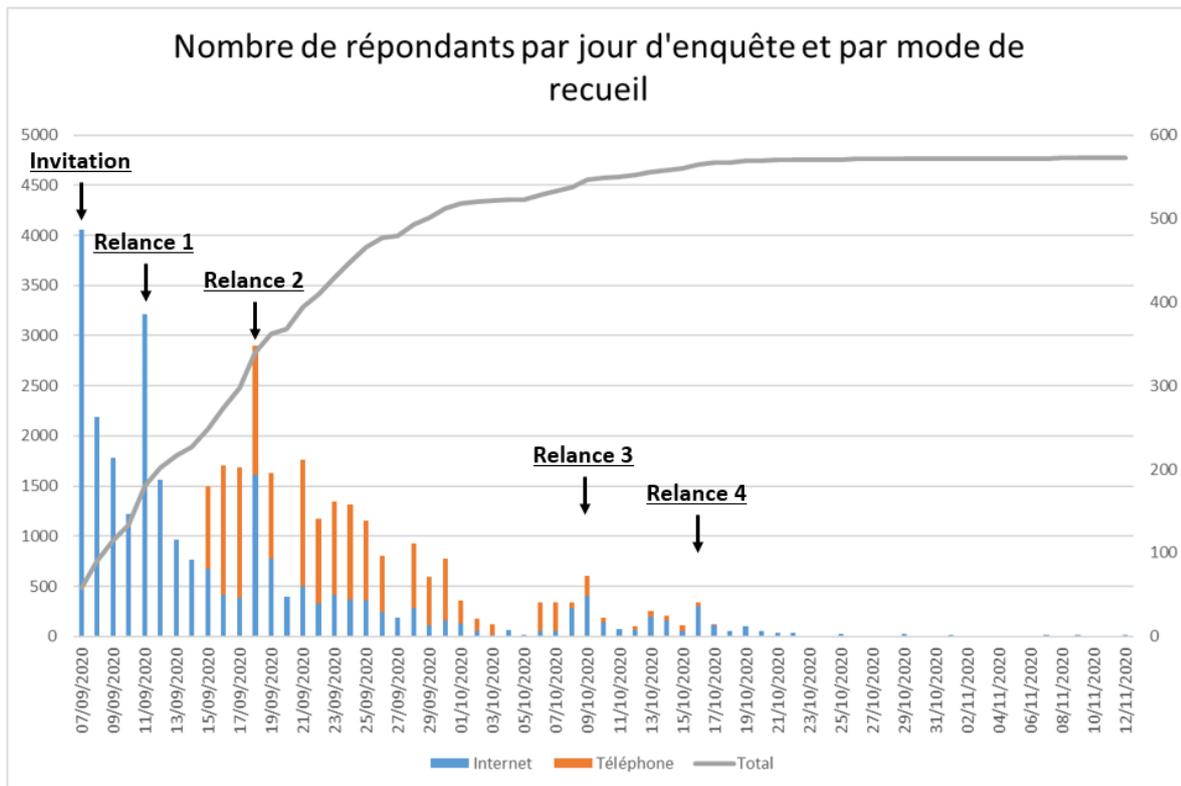


Figure 5. Répartition des répondants par Internet en fonction du nombre de mails envoyés

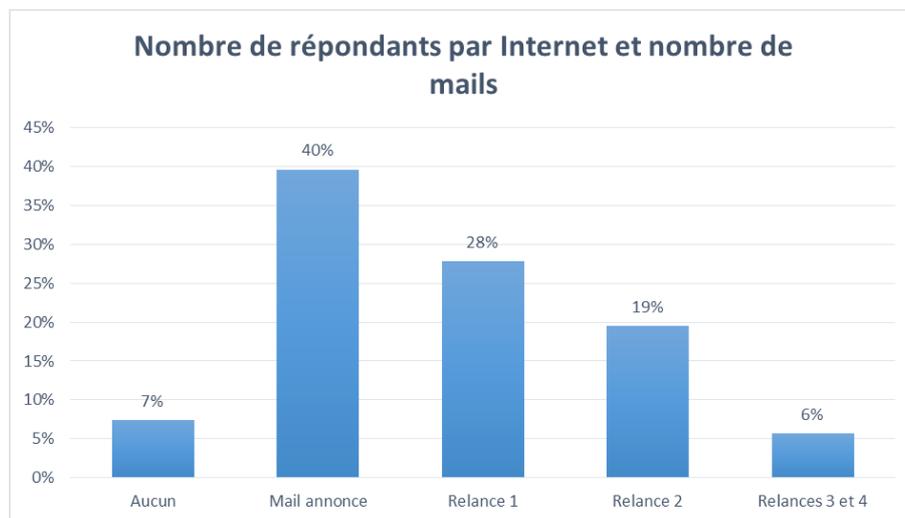
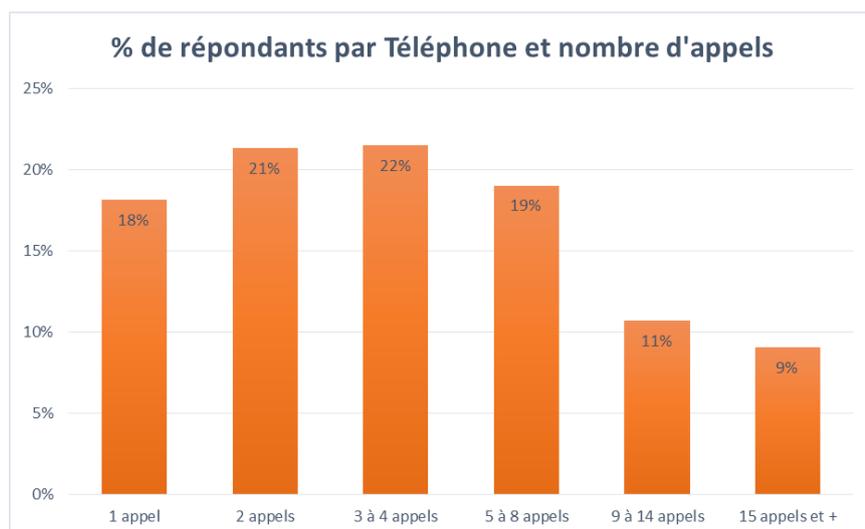


Figure 6. Répartition des répondants par téléphone en fonction du nombre d'appels



L'efficacité du protocole de collecte peut être appréciée par la distribution des répondants selon les relances effectuées et selon les informations de contact disponibles (Tableau 8). Ainsi, lorsque seule l'adresse postale était disponible, seuls 18% des répondants ont répondu après la première invitation par courrier, 45 % après une relance et 37 % après deux relances, indiquant que des relances postales supplémentaires auraient pu être mises en œuvre. Notons également la part importante de répondants par téléphone lorsqu'un numéro est disponible, de même que l'impact des appels téléphoniques pour inciter les personnes à répondre par Internet (un tiers des personnes pour lesquelles un mail et un téléphone sont disponibles).

Tableau 7. Distribution des répondants selon les informations de contact disponibles et les relances effectuées

	Aucun contact	Uniquement mail	Uniquement téléphone	Mail et téléphone
Nombre de répondants	153	202	510	3939
Effectifs dans la base de sondage	1320	812	1611	7024
Taux de réponse	11,6	24,9	31,7	55,6
Moment de réponse	Distribution des répondants (%)			
Après invitation initiale	18%			
Après 1ère relance courrier	45%			
Après 2ème relances courrier	37%			
Après invitation initiale		33%		
Après 1ère relance (mail)		22%		
Après 2ème relance (mail+courrier)		19%		
Après 3ème relances (mail+courrier)		15%		
Après 4ème relance (mail)		11%		
Après invitation initiale			24%	
Après appels téléphoniques et relance courrier			76%	
Après invitation initiale				35%
Après 1ère relance (mail)				12%
Après appels téléphoniques et relances (mails + courrier)				53%
<i>Dont réponse par Téléphone</i>				32%
<i>Dont réponse sur Internet</i>				21%

Participation au questionnaire enfants de l'enquête

Concernant le questionnaire enfants, parmi les 3 234 enfants tirés au sort, 580 étaient inéligibles car ils n'étaient pas l'enfant de la personne sélectionnée ou parce que le parent ou l'enfant lui-même ne résidait pas dans la zone exposée/témoin au moment de l'incendie (Tableau 9).

Dans près de la moitié des cas restants (49%), l'adulte tiré au sort n'a pas répondu à l'enquête principale. En revanche, la quasi-totalité des adultes dont un enfant était éligible a accepté de répondre aux questions concernant son enfant (1 306/1 353).

Tableau 8. Participation pour l'échantillon des enfants

	Zone exposée	Zone Témoin	Total zone enquêtée	
Echantillon initial	2348	886	3234	
Hors champ	394	186	580	
HC - pas enfant de la personne sélectionnée	295	72	408	
HC- Adulte ou enfant hors champ	100	112	172	
Total dans le champ	1954	700	2654	100%
Adulte non-répondant volet principal	892	409	1301	49,0%
Refus de l'adulte du module enfant	36	11	46	1,7%
Adulte répondant au module enfant	1026	280	1 306	49,20%

Participation des Gens du voyage

Pour les Gens du voyage, sur les 11 familles présentes sur l'aire d'accueil au moment de la collecte, 9 pouvaient être enquêtées car présentes au moment de l'incendie. Un adulte de chaque famille a été tiré au sort, ainsi qu'un enfant le cas échéant. Trois n'ont pas souhaité participer. Ce sont ainsi 6 adultes qui ont répondu à l'enquête, dont 3 étaient parents et ont répondu pour l'un de leurs enfants.

Sollicitation de la hotline

Durant toute la durée du terrain, la hotline mise en place a été sollicitée 889 fois, 84 % par téléphone et 16% par mail. Cela représente 8,3% de l'ensemble de la base des personnes tirées au sort, quoique d'autres personnes que celles tirées au sort pouvaient appeler ce numéro, diffusé par exemple dans les affiches grand public (Tableau 10). Notons également qu'une personne pouvait solliciter plusieurs fois ces services durant la période de l'enquête.

Les motifs d'appels portaient majoritairement sur une aide à la participation (60%) : prise de rendez-vous, demande d'informations, aide à la connexion, problèmes techniques. Un contact sur 6 (15%) visait à signaler une inéligibilité pour participer à l'enquête, tandis qu'une minorité des contacts étaient relatifs à l'exercice des droits d'oppositions.

Tableau 9. Motifs de contacts de la hotline

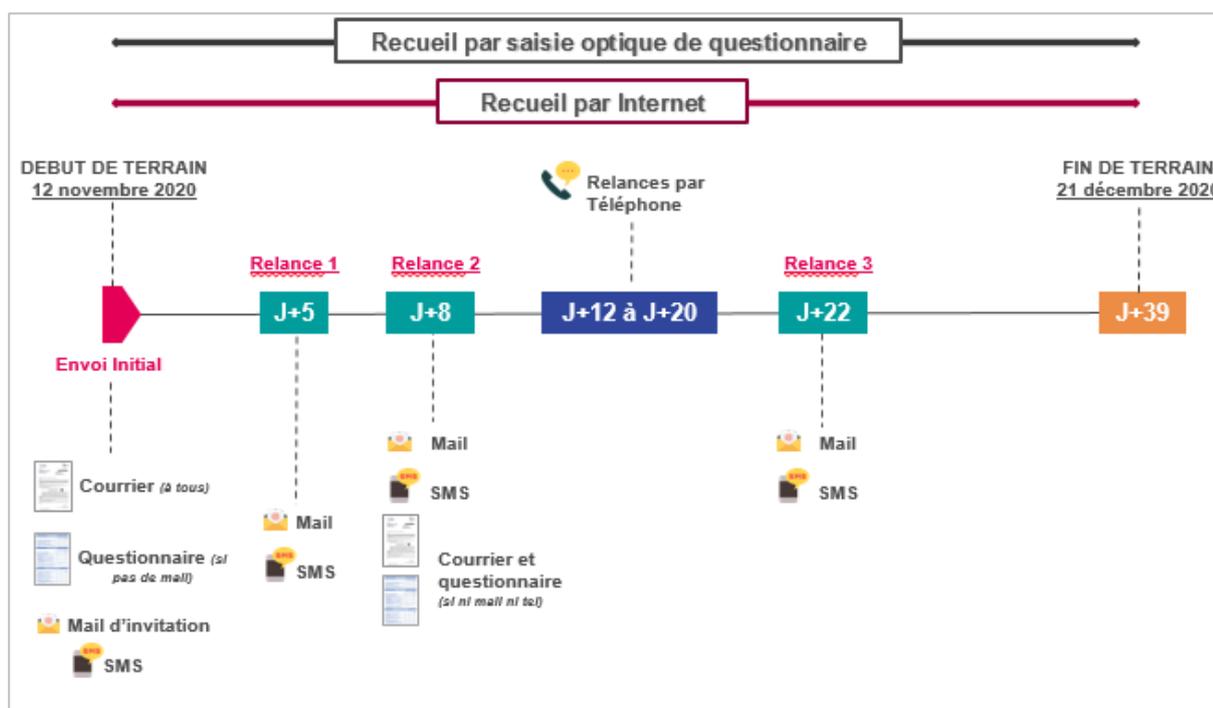
Hotline	n	%
Aide à la participation	532	60%
Prise de RDV	252	
Demande d'informations sur l'enquête	94	
Aide à la connexion, assistance questionnaire, remplissage, etc.	58	
Changement de coordonnées	51	
Souhait de répondre sur Internet	23	
Demande de codes de connexion	18	
Problème technique	16	
A déjà répondu au questionnaire	12	
N'a pas Internet	8	
Signaler une inéligibilité	132	15%
Personne sélectionnée hors de la région au moment de l'incendie	45	
Refus	43	
Personne sélectionnée en incapacité de répondre	25	
Personne sélectionnée décédée	19	
Exercice des droits	21	2%
Autres	204	23%
TOTAL	889	100%

TERRAIN DU VOLET SANTÉ MENTALE

Protocole de collecte du volet Santé mentale

Comme indiqué précédemment, le volet principal de l'enquête a été complété par un volet Santé mentale, proposé aux répondants de l'enquête principale ayant accepté d'être recontactés. La collecte s'est déroulée à la suite du volet principal, du 12 novembre au 21 décembre 2020. Le recueil des données de ce volet était exclusivement réalisé par questionnaire auto-administré (Internet ou papier). L'invitation à participer était envoyée par courrier postal, mail et SMS. Une relance par courrier et 3 relances par mail et SMS ont été réalisées. Des incitations supplémentaires par téléphone ont été mises en place auprès des non-répondants par des enquêteurs du prestataire de collecte, durant une période de 8 jours. Les étapes principales du protocole de collecte de ce volet d'enquête sont présentées sur la figure 7.

Figure 7. Protocole de collecte du volet Santé mentale



Participation au volet Santé mentale

Parmi les 4 779 répondants au volet principal, 2 885 (60,4%) ont consenti à être recontactés pour le volet Santé mentale (Tableau 11). Au final, parmi ces 2 885 personnes, 1 968 (68,2%) ont répondu à l'enquête, 890 personnes n'ont pas répondu, 27 personnes ont répondu de manière trop incomplète pour que leurs réponses puissent être prises en compte et 25 personnes ont répondu de manière incomplète mais permettant de tenir compte des réponses. La grande majorité des répondants (90,6%) a répondu par CAWI, les autres ayant répondu par questionnaire papier. Sur les 1 968 répondants, 1 627 vivaient dans la zone exposée et 341 dans la zone témoin. Le taux de réponse était plus important dans la zone exposée (69,2%) que dans la zone témoin (63,7%). Dans la zone exposée, le taux de réponse était le plus élevé dans la strate Nord-Est-proche (74,0%) et le plus faible dans la strate Sud-Ouest (63,0%).

Durant la collecte, plusieurs pics ont été observés concernant le nombre de questionnaires recueillis quotidiennement, reflétant l'impact des différentes relances, notamment sur les réponses recueillies

par Internet ou lorsque des incitations à participer étaient faites par téléphone (Figure 8). Au sein des répondants CAWI, alors que 26% ont répondu après le premier mail d'invitation, 25 % ont répondu après la première relance par mail, 30% après la seconde relance par mail et les appels téléphoniques, et 18 % après une troisième relance par mail (Figure 9). Il est également intéressant de noter l'impact de l'incitation à participer par téléphone qui, menée durant 8 jours et après 2 relances par mail, a permis de recueillir 489 questionnaires CAWI, soit 28% de l'ensemble des questionnaires CAWI (l'impact sur le nombre de questionnaires papiers reçus étant plus difficile à estimer).

Tableau 10. Bilan de participation au volet Santé mentale

Répondants au volet principal	4 779	
Refus d'être recontacté	1 894	39,6%
Échantillon initial – répondants au volet principal ayant consenti à être recontacté	2 885	60,4%
Non répondants total	917	31,8%
Répondants au questionnaire papier	185	6,4%
Répondants CAWI	1 783	61,8%
Questionnaires complets	1 968	68,2%
Zone exposée (Rouen)	1 627	69,2%
Zone Témoin (Le Havre)	341	63,7%
Proximité 0-700m	155	69,5%
Proximité 700-1500m	257	67,6%
Sud-Ouest	315	63,0%
Nord-Est proche	474	74,0%
Nord-Est éloigné	426	70,2%

Figure 8. Répartition par jour de collecte et par mode du nombre de questionnaires Santé mentale recueillis

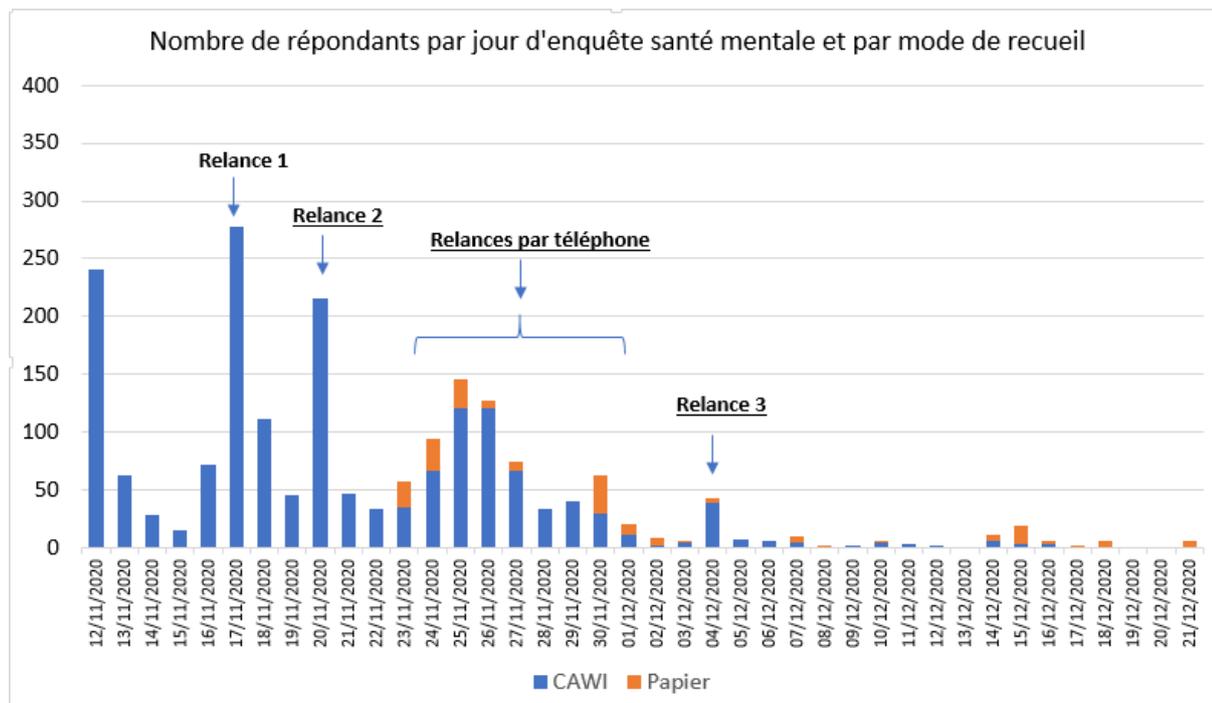
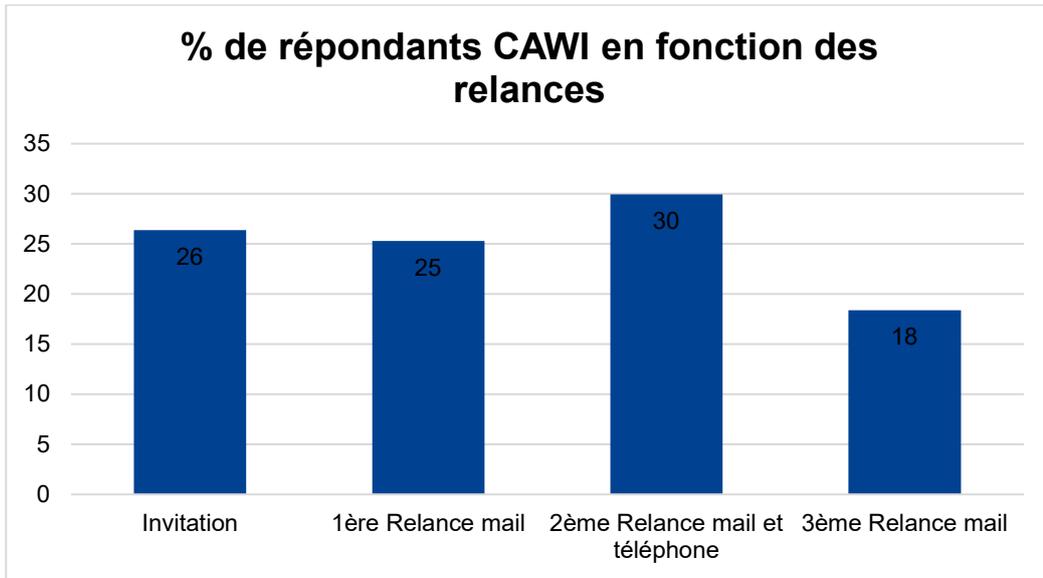


Figure 9. Répartition des répondants CAWI au questionnaire Santé mentale en fonction des relances



TRAITEMENT DE DONNÉES À CARACTÈRE PERSONNEL ET PROTECTION DU SECRET STATISTIQUE

Le recueil des données de santé pour l'enquête « une étude à l'écoute de votre santé » s'est fondé sur le consentement des personnes concernées. Le traitement de données à caractère personnel, mis en œuvre pour l'exécution d'une mission d'intérêt public, a fait l'objet d'une analyse d'impact sur la protection des données et a été consigné dans le registre des activités de traitement de Santé publique France. Santé publique France est responsable de ce traitement et fait appel à Ipsos comme sous-traitant pour réaliser la collecte. Cette sous-traitance a obtenu l'avis favorable du Comité du secret statistique.

Les personnes sélectionnées pour l'étude bénéficiaient d'un droit d'accès, d'opposition, de rectification, d'effacement ou de limitation de traitement pour les données les concernant ou concernant leur enfant pendant la période de conservation des données d'identification, c'est-à-dire du début de la collecte à trois mois après la fin de la collecte. Les personnes étaient informées de leur liberté de participer et de leurs droits quant à l'accès et à la rectification des données recueillies. Les informations utiles pour exprimer un refus ont été mentionnées dans la lettre-avis, le courriel-avis, la plaquette d'information, le site Internet de l'étude, et rappelées lors de la passation du questionnaire par l'enquêteur. Les refus pouvaient être exprimés de manière explicite par téléphone (numéro vert), par mail, par courrier ou sur le site Internet de l'étude. Les demandes d'exercice des droits pouvaient être adressées par courrier électronique ou par courrier postal au délégué à la protection des données (DPO) de Santé publique France ou auprès du prestataire de collecte Ipsos (enquêteurs ou service de hotline).

APPARIEMENT AUX DONNÉES DU SNDS

Les données concernant l'échantillon tiré au sort vont être enrichies (sous réserve de non-opposition de la part des personnes) des informations sur les remboursements de soins et le statut vital des personnes interrogées (soins en ville et à l'hôpital, remboursements de médicaments, causes médicales de décès) issues du Système national des données de santé (SNDS). Cet appariement avec les données du SNDS a fait l'objet d'une autorisation du CESREES (8 octobre 2020) et de la CNIL (11 mars 2021). L'accès à ces données auxiliaires permettra de répondre à trois objectifs : documenter l'état de santé de la population d'étude au moment de l'enquête et de manière prospective ; utiliser ces informations, disponibles à la fois chez les répondants et les non-répondants, afin d'évaluer le mécanisme de participation à l'enquête en fonction de variables liées à l'état de santé ; enrichir les réponses au questionnaire par des données de santé à des fins d'analyses statistiques croisant les deux sources d'information.

TRAITEMENTS POST-COLLECTE

Géocodage des adresses

Les coordonnées géographiques du logement de chaque personne tirée au sort, disponibles dans la base de sondage Fidéli, ont été fournies par l'Insee. Parmi les répondants à l'enquête, 320 personnes ont toutefois déclaré une adresse de résidence au moment de l'incendie (septembre 2019) différente de celle figurant dans Fidéli. Ces adresses ont été mises à jour lors du recueil par questionnaire, puis géocodées¹². Ces informations ont été utilisées dans une base de données spécifique pour la cartographie des expositions aux nuisances liées à l'incendie.

Recodage des questions ouvertes

Certaines questions du questionnaire proposaient une réponse entièrement en texte libre : sujets de préoccupation, nature de l'exposition pour les personnes exposées en zone témoin, diagnostic fait par le médecin, problème psychologique vécu par l'enfant. L'information contenue par ces variables a alors été synthétisée en plusieurs variables se déclinant parfois en plusieurs modalités.

D'autres questions comportaient une modalité « autre, préciser » : symptômes ou problèmes de santé, personne ayant nettoyé les suies, type d'odeur, diplôme, catégorie socio-professionnelle. Dans ce cas, lorsque cela était possible, la variable initiale a été corrigée en réaffectant l'information dans la modalité appropriée, ne laissant dans la catégorie « autre » que les éléments ne pouvant être reclassés, soit en un seul groupe, soit déclinés en modalités si l'effectif le permettait.

Traitement de la non-réponse partielle

Les données manquantes des variables issues du questionnaire ont été imputées par la méthode d'imputation par équations chaînées (8, 9). Cette méthode très flexible permet à la fois d'imputer des variables quantitatives et qualitatives. Elle a été implémentée avec le package ICE de Stata. Compte tenu du taux très faible de données manquantes (moins de 1% de valeurs manquantes pour la majorité des variables¹³), une imputation simple a été réalisée.

Traitement de la non-réponse totale

Le principe de la correction de la non-réponse est de modifier le poids de sondage des individus répondants pour compenser le biais de sélection introduit par les non-répondants. Les individus classés comme hors-champ (personnes décédées, pas présentes dans la zone exposée le jour de l'incendie ou exposées alors qu'elles résidaient en zone témoin, dialogue impossible) ont été exclus des analyses de non-réponse.

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 (10). Cette méthode se déroule en trois étapes essentielles, décrites en Annexe 2.

12. <https://adresse.data.gouv.fr/csv>

13. 11,3% de valeurs manquantes pour le poids corporel, et entre 4 et 7% pour les variables taille, consommation d'alcool, ressenti sur les finances du ménage, diplôme et catégorie professionnelle et sociale.

La correction de la non-réponse totale dans la présente étude a été réalisée séparément (cf. Tableau 11 pour les effectifs) en zone exposée et en zone témoin, pour les trois échantillons suivants :

- volet principal : répondants adultes ;
- volet principal : répondants enfants (questionnaires portant sur l'enfant de l'adulte répondant) ;
- volet Santé mentale : répondants adultes. Dans ce cas la correction de la non-réponse a été mise en œuvre à partir des poids corrigés de la non-réponse des individus ayant participé au volet principal.

Tableau 11. Effectifs des répondants et non-répondants

	Répondants		Non-Répondants		Total
	n	%	n	%	
Zone exposée - adultes	3 764	50%	3 736	50%	7 500
Zone exposée - adultes, volet Santé mentale	1 627	43%	2 137	57%	3 764
Zone témoin - adultes	1 015	40%	1 513	60%	2 528
Zone témoin - adultes, volet Santé mentale	341	34%	674	66%	1 015
Zone exposée - enfants	1 026	53%	928	47%	1 954
Zone témoin - enfants	280	40%	420	60%	700

Afin de limiter l'impact des poids extrêmes, pour chaque échantillon et après traitement de la non-réponse, une troncature a été appliquée. Ce sont ces jeux de poids tronqués qui ont été utilisés pour l'ensemble des analyses statistiques. En l'absence de marges disponibles pour la population source, aucun calage sur marge n'a été effectué.

Les informations relatives à ces traitements sont disponibles en Annexe 2 : pour chaque jeu de poids sont présentées les variables utilisées, la constitution des groupes de répondants, et les statistiques descriptives du facteur de repondération, de la pondération initiale, de la pondération corrigée de la non-réponse et de la pondération tronquée.

CARACTÉRISTIQUES DES RÉPONDANTS

Distances standardisées

Les différences entre échantillons 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, $q_A=1-p_A$, $q_B=1-p_B$) :

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

Et pour une variable continue (x la moyenne et s l'écart-type) :

$$d = \frac{100 * (x_a - x_b)}{\sqrt{\frac{s_a^2 + s_b^2}{2}}}$$

La distance standardisée D pour une variable catégorielle est la moyenne des valeurs absolues des différences standardisées de chaque catégorie. Une valeur de la SMD inférieure à 10% en valeur absolue est considérée comme le signe d'une différence mineure (11).

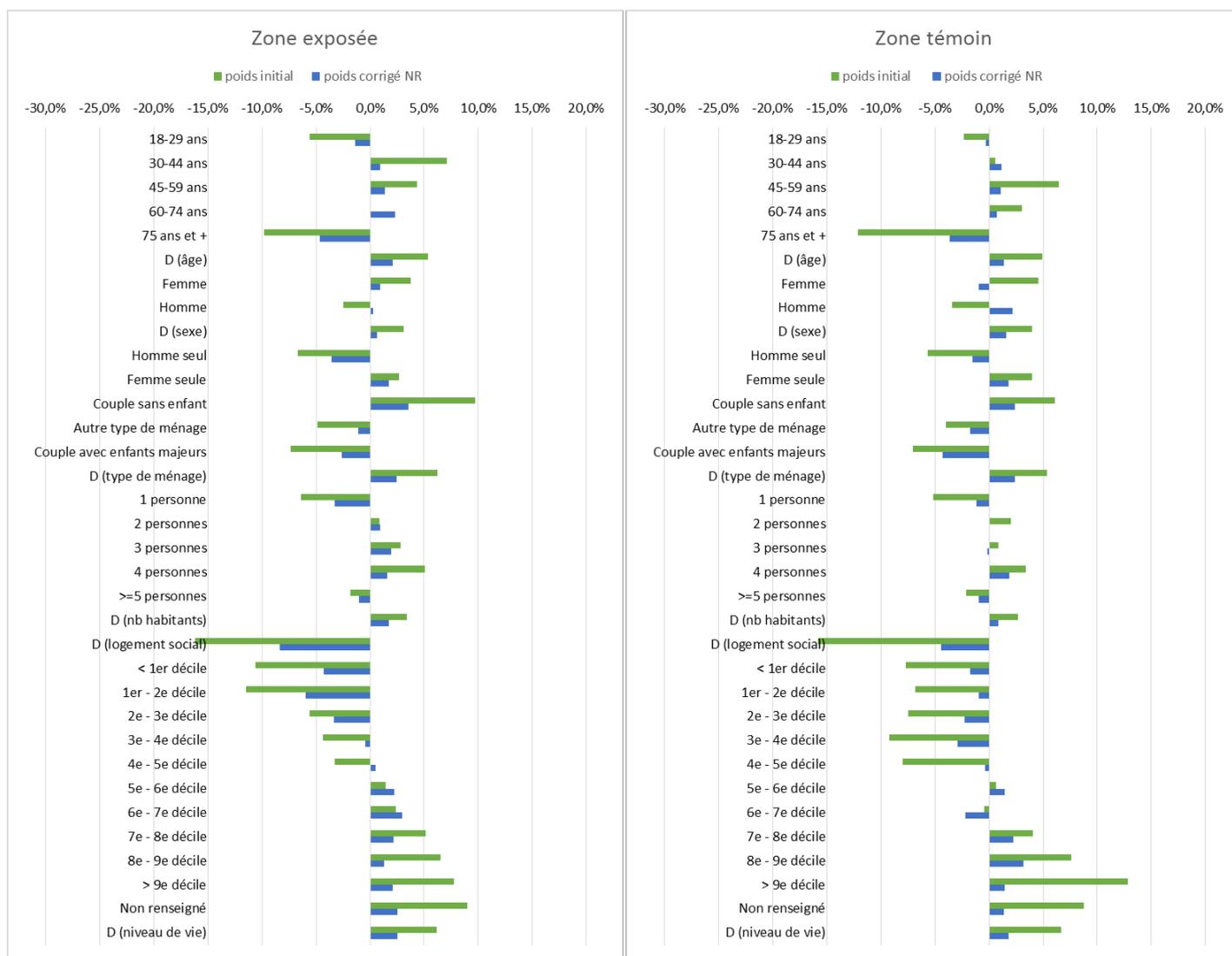
Les distances standardisées ont été calculées par rapport à la population source, c'est-à-dire l'échantillon initial tiré au sort par l'Insee, comprenant les répondants et les non-répondants (sans les hors champ), pondéré par le poids de sondage.

Volet principal

Dans la Figure 10, sont présentées les caractéristiques des répondants pondérées par le poids initial et le poids final (corrigé de la non-réponse totale et tronqué).

On observe d'une part que la correction de la non-réponse permet de diminuer, pour l'ensemble des variables utilisées, la SMD par rapport à la distribution de l'échantillon tiré au sort pondéré ; d'autre part qu'après correction de la non-réponse, les SMD sont toutes inférieures à 10%, ce qui n'était pas le cas pour certaines modalités avant traitement de la non-réponse totale : personnes résidant en logement social et les deux premiers déciles de niveau de vie.

Figure 10. Distances standardisées, par rapport à la population source, de l'échantillon des répondants au volet principal avec le poids de sondage et le poids corrigé de la non-réponse, en zone exposée et en zone témoin



En zone exposée comme en zone témoin (Tableau 12), la comparaison des distributions de quelques variables sociodémographiques entre l'échantillon tiré au sort (pourcentages pondérées par le poids de sondage) et l'échantillon des répondants (pourcentages pondérés par le poids corrigé de la non-réponse) montre des résultats très proches correspondant à des SMD < 10%. Pour le sexe, les pourcentages sont quasi-identiques dans la zone exposée, ils sont légèrement différents pour la zone témoin. Pour l'âge, on observe un écart autour de 1 point pour les 75 ans et plus. Concernant la taille du ménage et la présence d'enfants de moins de 17 ans, les pourcentages sont également proches. Concernant le niveau de vie des ménages, l'écart le plus important entre les pourcentages pondérés de la population exposée ou témoin et la population de référence se situe autour de 1 point.

Tableau 12. Distribution des caractéristiques de la population source et de l'échantillon des répondants pondéré, en zone exposée et en zone témoin

Variables	Zone exposée		Zone témoin	
	Population source* (% pondérés)	Répondants ** (% pondérés - [IC95])	Population source* (% pondérés)	Répondants ** (% pondérés - [IC95])
Sexe				
Homme	46,7	46,8 [45 - 48,7]	47,7	48,2 [44,9 - 51,5]
Femme	53,3	53,2 [51,3 - 55]	52,3	51,8 [48,5 - 55,1]
Âge en classes				
18-29 ans	18,8	18,2 [16,7 - 19,7]	16,9	16,8 [14,3 - 19,3]
30-44 ans	25,1	25,6 [24,1 - 27,2]	24,3	24,7 [21,8 - 27,5]
45-59 ans	24,2	24,8 [23,2 - 26,3]	24,6	25,2 [22,4 - 28,0]
60-74 ans	20,7	21,7 [20,2 - 23,2]	23,1	23,4 [20,7 - 26,2]
75 ans et plus	11,2	9,8 [8,6 - 11,0]	11,0	9,9 [7,7 - 12,1]
Niveau de vie en décile				
< 1 ^{er} décile	10,1	8,8 [7,7 - 10]	8,2	7,7 [5,8 - 9,6]
Entre le 1 ^{er} et le 2e décile	9,0	7,3 [6,3 - 8,4]	9,5	9,2 [7,1 - 11,3]
Entre le 2e et le 3e décile	8,9	8,8 [7,7 - 9,9]	9,4	8,6 [6,6 - 10,6]
Entre le 3e et le 4e décile	8,8	9,0 [7,9 - 10,1]	9,6	9,5 [7,4 - 11,6]
Entre le 4e et le 5e décile	8,7	9,3 [8,2 - 10,4]	9,6	10,0 [8,0 - 12,0]
Entre le 5e et le 6e décile	8,7	9,5 [8,4 - 10,6]	9,9	9,2 [7,4 - 11,1]
Entre le 6e et le 7e décile	8,5	9,1 [8,1 - 10,1]	9,9	10,6 [8,6 - 12,6]
Entre le 7e et le 8e décile	8,5	8,9 [7,9 - 9,9]	9,7	10,6 [8,7 - 12,5]
Entre le 8e et le 9e décile	8,9	9,5 [8,5 - 10,5]	9,5	9,9 [8,2 - 11,6]
> 9e décile	9,9	10,6 [9,6 - 11,7]	8,3	8,7 [7,1 - 10,4]
Manquant	10,0	9,0 [7,9 - 10,1]	6,4	5,9 [4,2 - 7,6]
Nombre de personnes dans le logement				
Une personne	26,2	24,8 [23,1 - 26,4]	20,4	19,9 [17,2 - 22,7]
Deux personnes	32,9	33,4 [31,6 - 35,1]	34,4	34,5 [31,3 - 37,6]
Trois personnes	17,1	17,9 [16,5 - 19,3]	17,8	17,7 [15,2 - 20,3]
Quatre personnes	14,1	14,7 [13,4 - 15,9]	18,8	19,5 [16,9 - 22,1]
>=5 personnes	9,6	9,3 [8,3 - 10,4]	8,7	8,4 [6,6 - 10,2]
Présence d'enfants <17 ans				
Non	70,2	69,1 [67,5 - 70,8]	67,1	65,8 [62,7 - 68,9]
Oui	29,8	30,9 [29,2 - 32,5]	32,9	34,2 [31,1 - 37,3]

* Échantillon initial des répondants et non-répondants ; ** échantillon des répondants après traitement de la non-réponse et troncature

Volet Santé mentale

Concernant l'échantillon des répondants du volet Santé mentale, la correction de la non-réponse totale a également permis de diminuer les SMD relatives à l'échantillon tiré au sort pondéré pour la majorité des variables, mais les écarts observés après traitement de la non-réponse apparaissent légèrement plus importants, relativement aux répondants du volet principal (Figure 11). En zone exposée comme en zone témoin, la SMD entre les pourcentages obtenus avec les pondérations finales et les pourcentages pondérés de l'échantillon initial est plus importante que la SMD observée pour le volet principal, mais elle demeure globalement limitée : les personnes résidant en logement social, disposant des plus faibles revenus (2 premiers déciles), et dans une moindre mesure les personnes les plus jeunes de la zone exposée (18-29 ans) ainsi que les plus âgées (75 ans et +) sont légèrement sous-représentées (Tableau 13). Les ménages d'une personne en zone exposée et de trois personnes en zone témoin sont également sous-représentés.

Figure 11. Distances standardisées, par rapport à la population source, de l'échantillon des répondants au volet Santé mentale avec le poids de sondage et le poids corrigé de la non-réponse, en zone exposée et en zone témoin

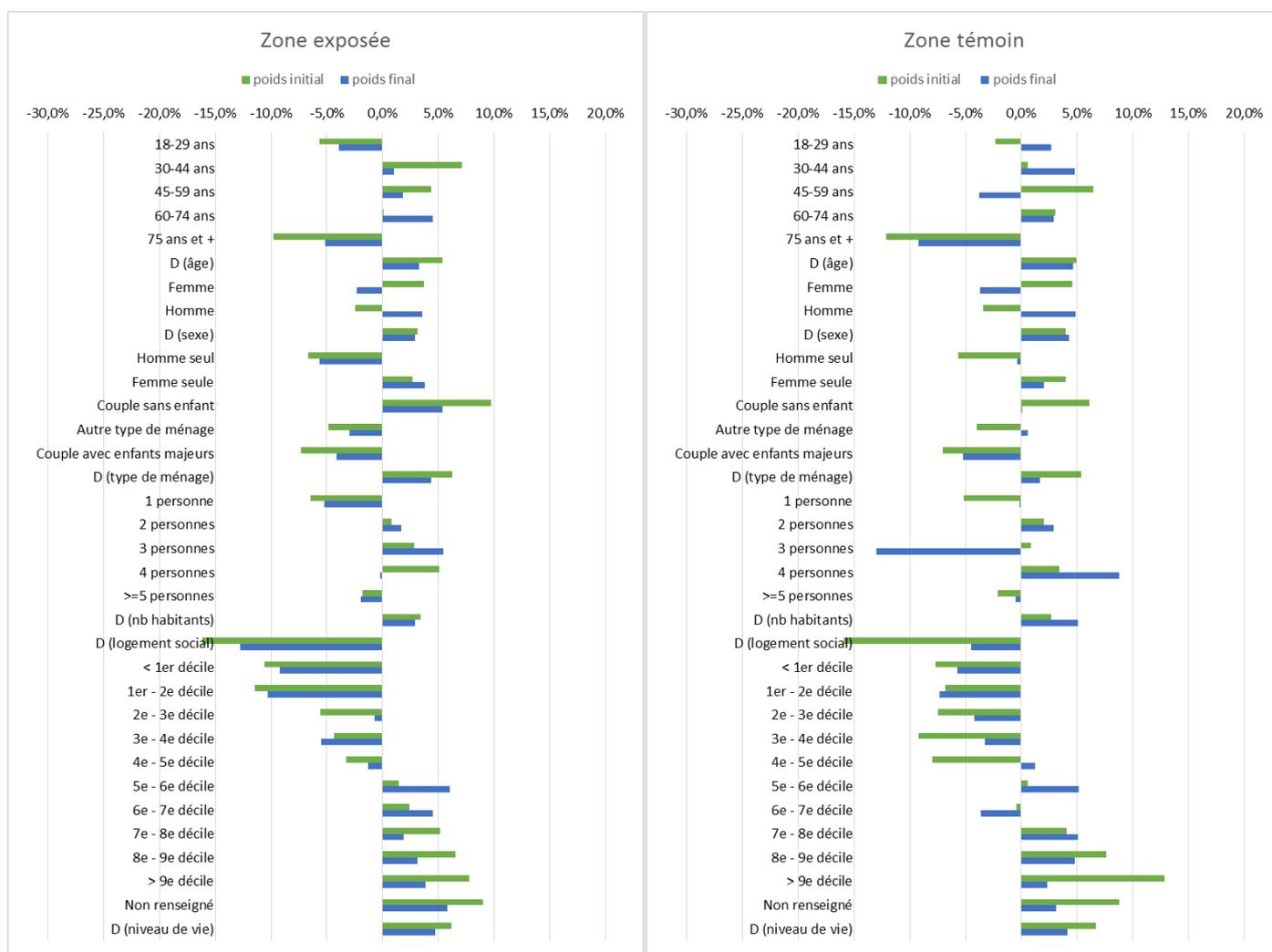


Tableau 13. Pourcentages pondérés de certaines caractéristiques de la population source et de l'échantillon final du volet Santé mentale, en zone exposée et en zone témoin

Variables	Zone exposée		Zone témoin	
	Population source* (% pondérés)	Répondants ** (% pondérés - [IC95])	Population source* (% pondérés)	Répondants ** (% pondérés - [IC95])
Sexe				
Un homme	46,7	48,4[45,5 - 51,4]	47,7	49,5[43,3 - 55,8]
Une femme	53,3	51,6[48,6 - 54,5]	52,3	50,5[44,2 - 56,7]
Âge en classes				
18-29 ans	18,8	17,3[15 - 19,6]	16,9	18[12,6 - 23,3]
30-44 ans	25,1	25,6[23,1 - 28]	24,3	26,4[20,8 - 31,9]
45-59 ans	24,2	25[22,5 - 27,5]	24,6	23,0[18,1 - 27,9]
60-74 ans	20,7	22,6[20,1 - 25]	23,1	24,4[19,3 - 29,4]
75 ans et plus	11,2	9,6[7,6 - 11,6]	11,0	8,3[4,8 - 11,8]
Niveau de vie en décile				
< 1 ^{er} décile	10,1	7,5[5,8 - 9,1]	8,2	6,7[2,9 - 10,4]
Entre le 1 ^{er} et le 2 ^e décile)	9,0	6,3[4,7 - 7,8]	9,5	7,5[3,7 - 11,2]
Entre le 2 ^e et le 3 ^e décile	8,9	7,4[5,7 - 9,1]	9,4	8,5[4,7 - 12,3]
Entre le 3 ^e et le 4 ^e décile	8,8	8,5[6,7 - 10,3]	9,6	10,0[6,1 - 13,9]
Entre le 4 ^e et le 5 ^e décile	8,7	10,5[8,6 - 12,4]	9,6	11,1[6,9 - 15,4]
Entre le 5 ^e et le 6 ^e décile	8,7	10,0[8,2 - 11,7]	9,9	8,8[5,3 - 12,4]
Entre le 6 ^e et le 7 ^e décile	8,5	9,1[7,5 - 10,7]	9,9	11,5[7,7 - 15,2]
Entre le 7 ^e et le 8 ^e décile	8,5	9,4[7,8 - 11,0]	9,7	11,1[7,8 - 14,4]
Entre le 8 ^e et le 9 ^e décile	8,9	10,0[8,4 - 11,6]	9,5	10,2[7,1 - 13,2]
> 9 ^e décile	9,9	11,7[10,0 - 13,3]	8,3	9,2[6,5 - 12,0]
Manquant	10,0	9,8[7,9 - 11,6]	6,4	5,5[2,3 - 8,6]
Nombre de personnes dans le logement				
Une personne	26,2	24,0[21,4 - 26,5]	20,4	20,3[15,2 - 25,4]
Deux personnes	32,9	33,7[30,9 - 36,6]	34,4	35,8[30 - 41,6]
Trois personnes	17,1	19,2[16,8 - 21,5]	17,8	13,1[9,1 - 17,1]
Quatre personnes	14,1	14,1[12,2 - 15,9]	18,8	22,3[16,9 - 27,7]
>=5 personnes	9,6	9,1[7,4 - 10,8]	8,7	8,5[4,9 - 12,2]
Présence d'enfants <17 ans				
Non	70,2	69,0[66,3 - 71,7]	67,1	65,6[59,7 - 71,6]
Oui	29,8	31,0[28,3 - 33,7]	32,9	34,4[28,4 - 40,3]

* Échantillon des répondants et non-répondants ; ** échantillon des répondants après traitement de la non-réponse

ANALYSE COMPLÉMENTAIRE : UNE ÉVALUATION DU BIAIS DE SÉLECTION

Objectif et méthodes mises en œuvre

L'objectif de cette analyse est de fournir des éléments permettant d'évaluer le biais de non-réponse à l'enquête, en essayant en particulier de mesurer si le mécanisme de non-réponse est ignorable ou non-ignorable. Le protocole de collecte mis en œuvre dans cette enquête, reposant en premier lieu sur un mode de réponse auto-administré, pourrait en effet être à l'origine d'une éventuelle sélection endogène si l'intérêt pour le thème de l'enquête est associé à la participation et aux variables d'intérêt (12). La question soulevée est donc la suivante : y a-t-il eu une auto-sélection des répondants due à leur intérêt pour le sujet de l'enquête ? Nous proposons d'évaluer l'existence d'une sélection endogène par des analyses portant sur les différences entre modes de collecte : internet (CAWI) / téléphone (CATI), dans le volet principal de l'enquête. L'hypothèse sous-jacente est qu'en cas d'existence d'une sélection endogène, celle-ci devrait être d'autant plus marquée parmi les répondants sur Internet car 1) ce mode de recueil était le mode proposé aux participants en première intention ; 2) il repose, davantage que pour le CATI, sur une action volontaire de la part du répondant ; 3) il fait l'objet d'une insistance moindre (4 mails de relance vs plus de 15 relances téléphoniques sur CATI).

Cette analyse vise à évaluer l'existence d'un effet lié au mode de collecte et, le cas échéant, à différencier la part attribuable à un biais de sélection différentielle (en contrôlant les observables, les répondants CAWI diffèrent des répondants CATI car la composition de l'échantillon diffère) de celle attribuable à un biais de mesure (les répondants CAWI diffèrent des répondants CATI car les réponses diffèrent selon le mode de recueil).

Les variables mobilisées dans cette analyse sont présentées Tableau 14 et définies comme suit :

- **variables d'intérêt de l'enquête** : expositions aux nuisances, symptômes ressentis suite à l'incendie, état de santé au moment de la collecte ;
- **variables liées à la participation et supposées non affectées d'un biais de mesure lié au mode** : disponibles dans la base de sondage ou recueillies par questionnaire ;
- **variables relatives aux délais de réponse, supposées liées à l'intérêt pour l'enquête et à une sélection différentielle selon le mode de collecte** : répondants précoces vs répondants tardifs (ces derniers représentant les plus réfractaires à répondre à l'enquête/les plus difficiles à joindre), répondants avant vs après relance (mail ou plusieurs appels téléphoniques). Deux autres variables supposées être liées à un intérêt pour l'enquête ont également été testées : l'exposition aux différentes nuisances liées à l'incendie en zone exposée, l'inquiétude vis-à-vis de l'environnement en zone témoin.

Tableau 14. Variables d'intérêt et variables retenues pour le score de propension et dans les régressions, en zone exposée et en zone témoin

Variables d'intérêt	Zone exposée	Zone témoin
Post-accidentel		
Expositions aux odeurs, fumées, suie, débris, visuel (O/N)	x	
Au moins 1 symptôme lié à l'incendie + type de symptômes (O/N)	x	
Au moins 1 symptôme physique lié à l'incendie (O/N)	x	
Au moins 1 symptôme psychique lié à l'incendie (O/N)	x	
Avoir consulté suite aux symptômes liés à l'incendie (O/N)	x	
Santé actuelle		
État de santé général (excellent/très bon/ bon/médiocre/mauvais)	x	x
SF12- score de santé physique (PCS) (Continu 0-100)	x	x
SF12 - score de santé mentale (MCS) (Continu 0-100)	x	x
Maladie chronique (O/N)	x	x
Avoir consulté un médecin généraliste dans les 12 derniers mois (O/N)	x	x
Variables utilisées pour le score de propension		
Variables issues de la base de sondage		
Quartier prioritaire de la ville	x	
Logement social	x	
Niveau de vie en décile	x	x
A perçu un salaire en année -1	x	x
A perçu une pension en année n-1	x	
Type de logement	x	x
Téléphone fixe / téléphone mobile / mail disponible	x	
Lieu de naissance		x
Variables du questionnaire (supposées non affectées d'un biais de mesure lié au mode)		
Situation financière perçue	x	x
Niveau de diplôme le plus élevé	x	x
Taille du foyer	x	x
Usage du téléphone fixe / téléphone mobile / Internet	x	x
Âge (splines)	x	x
Sexe	x	x
Distance à l'usine (négative si sens opposé au panache), en splines	x	x
Ancienneté de résidence dans la région		x
Situation professionnelle actuelle	x	x
Variables additionnelles utilisées dans les régressions		
Expositions aux odeurs, fumées, suie, débris, visuel*	x	
Répondants les plus tardifs (dernier quintile de répondants au sein de chaque mode) **	x	x
Répondants après relance***	x	x
Attitudes à l'égard de l'environnement (inquiétude de sa qualité ou vis-à-vis d'un risque industriel)		x

* L'exposition aux nuisances est étudiée comme variable d'intérêt et comme variable explicative pour l'étude des symptômes

** Cette variable distingue les répondants les plus réfractaires à répondre à l'enquête/les plus difficiles à joindre

*** Au moins 1 relance par mail / au moins 1 relance par courrier / au moins 3 appels téléphoniques avant d'avoir participé

Cette analyse a été menée en plusieurs étapes :

- comparaison du lien, selon le mode de collecte, entre les variables relatives aux délais de réponse et les variables d'intérêt (supposé plus important parmi les répondants Internet) ;
- constitution d'un score de propension par modélisation de la réponse sur Internet plutôt que sur téléphone (variables présentées dans le Tableau 14) et vérification des propriétés équilibrantes du score en zone exposée et en zone témoin ;
- estimation de l'effet lié au mode de collecte par différentes régressions des variables d'intérêt (outcome) :
 - régression non pondérée : ajustée sur les facteurs de confusion¹⁴ et sur le mode de collecte ;
 - régression pondérée (inverse du score de propension) : ajustée sur les facteurs de confusion et sur le mode de collecte ;
 - régression pondérée (inverse du score de propension) : ajustée sur les facteurs de confusion, sur le mode de collecte et sur les variables additionnelles (Tableau 13).

Si le modèle est correctement spécifié, les deux premières régressions visent à mesurer le biais de mesure lié au mode de collecte en contrôlant tous les facteurs de confusion (rendre comparables les échantillons de répondants Internet et téléphone), sous l'hypothèse d'une absence de sélection différentielle entre les modes de collecte. La troisième régression est une analyse de sensibilité des résultats des deux premières régressions, par l'ajout dans les modèles de variables supposées liées à cette sélection différentielle. Les modèles de régression utilisés sont des régressions de Poisson à variance robuste, permettant de fournir des estimations de risques relatifs (RR), plus facilement interprétables que les odds-ratios (13).

Lien entre variables d'intérêt et variables relatives aux délais de réponse

Le Tableau 15 présente les estimations des différentes variables d'intérêt parmi les répondants ayant répondu avant ou après une première relance (courrier, mail ou au moins 3 appels téléphoniques). Les caractéristiques des répondants les plus volontaires pour répondre à l'enquête, c'est-à-dire ayant répondu avant relance, diffèrent selon le mode de recueil. Sur CAWI, alors qu'aucune différence ne s'observe concernant les indicateurs relatifs à la santé actuelle, les premiers répondants déclarent plus que les autres des symptômes en lien avec l'incendie (la différence étant plus importante parmi les symptômes de nature psychique), et un peu plus souvent avoir été exposés aux différentes pollutions liées à l'incendie, à l'exception des retombées de débris. Sur CATI, les premiers répondants déclarent à l'inverse moins d'expositions et de symptômes que les répondants après relance. Ils présentent en revanche des indicateurs de santé actuelle plus dégradés. Ces différences sont de même nature en zone témoin s'agissant des répondants CATI.

L'analyse menée auprès des répondants ayant répondu le plus tardivement (dernier quintile de répondants au sein de chaque mode) indique que ces derniers ont déclaré moins de symptômes liés à l'incendie, quel que soit le mode de collecte, alors qu'aucune différence n'apparaît concernant les expositions. Les répondants les plus tardifs sur CATI déclarent également moins souvent un état de santé général médiocre ou mauvais que les premiers répondants, en zone exposée comme en zone témoin (Tableau 16).

Un lien semble donc exister entre l'empressement à répondre et les variables d'intérêt de l'enquête, de nature différente selon le mode de collecte. Cependant, ces premiers résultats n'excluent pas

14. Les facteurs de confusion, liés aux variables explicatives et à expliquer, correspondent aux mêmes variables que celles utilisées pour établir le score de propension : variables de la base de sondage et variables du questionnaire non affectées d'un biais de mesure.

que les écarts observés soient liés à des caractéristiques sociodémographiques différentes des répondants (selon le mode et/ou selon le délai de réponse) : les personnes les plus joignables au téléphone peuvent par exemple être plus souvent des retraités, en moins bonne santé générale que les actifs, et habitant plus souvent des zones rurales, qui sont aussi plus éloignées de l'usine.

Tableau 15. Estimations des différentes variables d'intérêt selon la zone (exposée/témoin), le mode de recueil (CATI/CAWI) et le fait d'avoir répondu avant ou après relance (% pondérés par le poids corrigé de la non-réponse)

Zone exposée	CATI		CAWI		CATI+CAWI	
	avant relance	après relance	avant relance	après relance	avant relance	après relance
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	88,1%	83,7%	82,6%	82,0%	83,4%	82,7%
Avoir une maladie chronique	59,1%	50,2%	39,0%	36,2%	41,9%	41,6%
Score moyen de santé mentale (MCS)	47,1	48,1	46,9	47,4	46,9	47,7
Score moyen de santé physique (PCS)	46,8	50,4	53,2	52,9	52,3	51,9
Santé générale médiocre ou mauvaise	28,4%	14,5%	9,1%	8,1%	11,9%	10,6%
Post-accidentel						
Au moins 1 symptôme lié à l'incendie	53,5%	66,3%	79,4%	73,7%	75,7%	70,8%
Au moins 1 symptôme psychique lié à l'incendie	28,9%	32,8%	51,0%	43,5%	47,9%	39,4%
Au moins 1 symptôme physique lié à l'incendie	36,3%	51,5%	65,9%	60,1%	61,7%	56,8%
Avoir consulté suite aux symptômes liés à l'incendie	7,5%	12,5%	13,0%	10,8%	12,2%	11,5%
Avoir entendu du bruit ou été réveillé	28,4%	28,5%	32,7%	30,4%	32,1%	29,6%
Avoir eu des retombées de suies	39,6%	41,7%	43,9%	40,3%	43,2%	40,9%
Avoir eu des retombées de débris	3,9%	5,2%	4,8%	5,3%	4,6%	5,3%
Avoir été sous le panache	21,6%	24,9%	30,9%	26,7%	29,6%	26,0%
Avoir perçu des odeurs	48,4%	60,9%	75,6%	73,0%	71,7%	68,3%
Zone témoin	CATI		CAWI		CATI+CAWI	
	avant relance	après relance	avant relance	après relance	avant relance	après relance
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	91,3%	81,3%	77,2%	77,4%	80,6%	79,5%
Avoir une maladie chronique	50,7%	53,1%	68,6%	64,5%	64,3%	58,2%
Score moyen de santé mentale (MCS)	51,5	48,7	48,2	49,5	49,0	49,1
Score moyen de santé physique (PCS)	46,7	51,1	53,5	53,2	51,9	52,0
Santé général médiocre ou mauvaise	21,1%	11,1%	5,2%	6,3%	9,0%	8,9%

Ci-dessus en gras, différence standardisée entre avant et après relance > 10

Tableau 16. Estimations des différentes variables d'intérêt selon la zone (exposée/témoin), le mode de recueil (CATI/CAWI) et le fait d'avoir répondu tardivement ou non (% pondérés par le poids corrigé de la non-réponse)

Zone exposée	CATI		CAWI		CATI+CAWI	
	4 premiers quintiles	Dernier quintile	4 premiers quintiles	Dernier quintile	4 premiers quintiles	Dernier quintile
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	85,4%	80,4%	82,6%	81,2%	83,6%	81,0%
Avoir une maladie chronique	51,7%	50,1%	38,4%	34,5%	42,8%	38,7%
Score moyen de santé mentale (MCS)	47,7	48,7	47,1	47,6	47,3	47,9
Score moyen de santé physique (PCS)	49,7	50,8	52,9	53,4	51,8	52,7
Santé générale médiocre ou mauvaise	17,9%	10,9%	8,8%	7,8%	11,8%	8,6%
Post-accidentel						
Au moins 1 symptôme lié à l'incendie	64,7%	64,8%	76,5%	73,7%	72,6%	71,3%
Au moins 1 symptôme psychique lié à l'incendie	33,9%	27,1%	48,0%	41,8%	43,3%	37,8%
Au moins 1 symptôme physique lié à l'incendie	50,2%	47,3%	63,6%	58,7%	59,1%	55,7%
Avoir consulté suite aux symptômes liés à l'incendie	64,7%	64,8%	76,5%	73,7%	72,6%	71,3%
Avoir entendu du bruit ou été réveillé	28,0%	30,0%	31,5%	30,6%	30,3%	30,4%
Avoir eu des retombées de suies	41,5%	41,4%	41,3%	42,4%	41,3%	42,1%
Avoir eu des retombées de débris	5,2%	4,6%	5,3%	4,7%	5,3%	4,7%
Avoir été sous le panache	25,5%	21,0%	28,2%	28,3%	27,3%	26,3%
Avoir perçu des odeurs	59,5%	58,7%	74,6%	72,1%	69,6%	68,5%
Zone témoin						
	CATI		CAWI		CATI+CAWI	
	4 premiers quintiles	Dernier quintile	4 premiers quintiles	Dernier quintile	4 premiers quintiles	Dernier quintile
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	81,9%	83,5%	77,5%	76,6%	79,7%	80,0%
Avoir une maladie chronique	52,8%	52,8%	66,1%	64,5%	59,7%	58,7%
Score moyen de santé mentale (MCS)	49,2	48,2	48,9	49,8	49,1	49,0
Score moyen de santé physique (PCS)	50,4	51,7	53,3	53,4	51,9	52,5
Santé générale médiocre ou mauvaise	13,3%	8,4%	6,5%	4,2%	9,8%	6,3%

Ci-dessus en gras, différence standardisée entre avant et après relance > 10

Score de propension (Internet vs téléphone), parmi les répondants

Deux modèles ont été réalisés indépendamment en zone exposée et zone témoin. Sur la plage commune du score de propension, une repondération du poids de sondage par l'inverse du score de propension a été réalisée pour rendre les populations de répondants CATI et CAWI les plus comparables possibles (neutralisation du biais de composition des échantillons de répondants au sein de chaque mode). La propriété équilibrante du score de propension a été vérifiée en calculant les SMD entre répondants CAWI/CATI, et en comparant les SMD utilisant les données pondérées par le poids corrigé de la non-réponse à celles obtenues par le poids inverse du score de propension :

- pour les variables du modèle : variables d'équilibrage ;
- pour les autres variables socio-économiques non utilisées dans le modèle : variables de contrôle (Tableau 17) ;
- Pour les variables d'intérêt (Tableau 18).

La propriété équilibrante du score de propension est vérifiée en zone exposée : toutes les SMD calculées pour les variables d'équilibrage comme pour les variables de contrôle sont inférieures à 10% (exception faite de la situation professionnelle actuelle : 10,1%). Concernant les variables d'intérêt (Tableau 17), les SMD entre répondants CAWI et CATI apparaissent très atténuées lorsque le poids inverse du score de propension est utilisé. En zone exposée, une SMD <10% est observée pour 10 des 14 variables étudiées. Une SMD comprise entre 10% et 15% est observée pour 3 variables : avoir été sous le panache, avoir entendu du bruit ou été réveillé, score de santé mentale (MCS). Enfin, une SMD de 15,8% est observée pour le fait d'avoir déclaré au moins un symptôme de nature psychique en lien avec l'incendie (39,2% sur CATI vs 47,0% sur CAWI).

En zone témoin, la propriété équilibrante du score est vérifiée pour les variables d'équilibrage. Concernant les variables de contrôle, 5 variables présentent une SMD comprise entre 10% et 15%, et 2 variables une SMD >15%. Notons que pour ces deux variables, la distance standardisée calculée avec le poids corrigé de la non-réponse était très élevée (55% pour la situation professionnelle actuelle, 42% concernant la présence d'un mail dans la base de sondage). Concernant les variables d'intérêt, les SMD diminuent après application du poids inverse du score de propension (relativement au poids corrigé de la non-réponse), mais restent supérieures à 10% pour 3 variables : avoir consulté un médecin généraliste au cours des 12 derniers mois (83,8% sur CATI vs 77,7% sur CAWI), avoir une maladie chronique (47,7% sur CATI vs 39,6% sur CAWI), ou déclarer un état de santé général médiocre ou mauvais (11,1% sur CATI vs 7,7% sur CAWI). En contrôlant sur les variables introduites dans les modèles du score de propension (voir Tableau 14), les répondants CATI apparaissent ainsi en moins bon état de santé que les répondants CAWI.

Tableau 17. Distances standardisées CATI/CAWI pour les variables d'équilibrage et les variables de contrôle, avec les poids corrigés de la non-réponse (NR) et l'inverse du score de propension, en zone exposée et en zone témoin

Variables d'équilibrage	Zone exposée		Zone témoin	
	Poids NR	Poids score de propension	Poids NR	Poids score de propension
	SMD	SMD	SMD	SMD
Sexe	6,5%	1,0%	13,7%	1,2%
Âge	62,1%	1,2%	54,4%	2,8%
Type de logement : Maison	11,9%	3,2%	16,7%	1,4%
Niveau de vie en décile	37,3%	8,4%	22,5%	6,9%
Ressenti des ménages sur leur situation financière actuelle	10,8%	3,4%	16,8%	4,2%
Diplôme le plus élevé	56,2%	6,6%	53,7%	3,8%
Utilisation de téléphone fixe au sein du foyer	55,4%	3,7%	40,3%	2,9%
Utilisation de téléphone portable au sein du foyer	44,7%	2,2%	22,6%	3,2%
Utilisation d'Internet	70,2%	7,2%	66,6%	3,2%
Lieu de naissance de l'individu échantillonné	-	-	15,5%	5,6%
Présence d'un mail ménage	32,8%	0,6%	-	-
Présence d'un tél portable ménage	6,3%	4,0%	-	-
Présence d'un tél fixe ménage	50,4%	0,1%	-	-
QPV	2,6%	0,8%	-	-
Logement social	20,8%	1,8%	-	-
Variables de contrôle	Poids NR	Poids score de propension	Poids NR	Poids score de propension
	SMD	SMD	SMD	SMD
Ménage pauvre à 60 % année n-1	11,4%	1,3%	1,5%	4,9%
Perception du chômage	4,9%	0,5%	6,2%	2,4%
Perception d'une pension alimentaire	5,1%	3,2%	9,9%	0,6%
Perception d'APL	7,2%	0,1%	11,0%	5,7%
Lien familial de l'individu échantillonné avec le référent fiscal	14,4%	2,4%	32,1%	10,3%
Statut d'occupation du logement	9,5%	7,8%	7,5%	2,0%
Nombre d'enfants âgé de entre de 0 ans et 17 ans	22,2%	6,7%	17,3%	13,4%
Type de ménage	21,3%	8,3%	32,3%	11,1%
Lieu de naissance de l'individu échantillonné	11,4%	2,7%	-	-
Nombre d'année de résidence dans la région	32,1%	6,7%	-	-
Situation professionnelle actuelle	59,4%	10,1%	55,6%	15,2%
Perception de minima sociaux	1,3%	6,4%	4,1%	4,1%
Présence téléphone portable	-	-	14,6%	14,0%
Présence d'un mail ménage	-	-	42,3%	24,9%
Présence d'un téléphone fixe ménage	-	-	47,2%	12,1%
QPV	-	-	0,9%	1,5%
Logement social	-	-	21,8%	3,6%

10%<SMD≤15%

SMD>15%

Tableau 18. Distances standardisées CATI/CAWI et estimations (%) des principales variables d'intérêt parmi les répondants CATI et CAWI, en zone exposée et en zone témoin, à partir des poids corrigés de la non-réponse et du poids inverse du score de propension

Zone exposée	Poids NR			Poids score de propension		
	SMD	CATI (%)	CAWI (%)	SMD	CATI (%)	CAWI (%)
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	5,6%	84,3	82,2	6,1%	81,4	83,7
Avoir une maladie chronique	28,5%	51,3	37,2	4,3%	42,0	39,9
Score moyen de santé mentale (MCS)	7,6%	48,0	47,2	12,6%	48,4	47,2
Score moyen de santé physique (PCS)	37,5%	50,0	53,0	7,1%	52,0	52,5
Santé générale médiocre ou mauvaise	23,8%	16,3	8,5	4,3%	10,8	9,5
Post-accidentel						
Au moins 1 symptôme lié à l'incendie	24,2%	64,7	75,7	4,2%	73,1	74,9
Au moins 1 symptôme psychique lié à l'incendie	28,8%	32,3	46,2	15,8%	39,2	47,0
Au moins 1 symptôme physique lié à l'incendie	25,5%	49,6	62,2	0,5%	60,3	60,6
Avoir consulté suite aux symptômes liés à l'incendie	0,1%	11,5	11,6	6,6%	13,1	11,0
Avoir entendu du bruit ou été réveillé	8,3%	35,1	39,2	12,0%	43,6	37,8
Avoir eu des retombées de suies	0,3%	41,4	41,6	4,7%	44,4	42,0
Avoir eu des retombées de débris	0,3%	5,0	5,1	4,0%	4,8	5,7
Avoir été sous le panache	2,7%	30,5	31,7	10,7%	35,4	30,4
Avoir perçu des odeurs	31,4%	59,2	73,9	9,6%	68,2	72,6
Zone témoin	Poids NR			Poids score de propension		
	SMD	CATI (%)	CAWI (%)	SMD	CATI (%)	CAWI (%)
Santé actuelle						
Avoir consulté un médecin généraliste dans les 12 derniers mois	18,6%	83,2	75,7	15,4%	83,8	77,7
Avoir une maladie chronique	25,6%	45,6	33,1	16,5%	47,7	39,6
Score moyen de santé mentale (MCS)	35,5%	50,5	53,2	8,0%	51,8	52,4
Score moyen de santé physique (PCS)	4,5%	48,9	49,3	3,4%	49,1	49,4
Santé générale médiocre ou mauvaise	15,5%	11,8	7,3	11,8%	11,1	7,7

10% < SMD ≤ 15%

SMD > 15%

Résultats des régressions

L'effet du mode de collecte sur les variables d'intérêt a été évalué par la mise en œuvre des trois types de modélisations (Tableau 19). En cohérence avec les différences observées précédemment, l'effet lié au mode de collecte est significatif :

- en **zone exposée** concernant le score de santé mentale, la déclaration d'un symptôme psychique lié à l'incendie, une exposition au bruit ou au panache de fumée : les répondants CAWI déclarent plus de symptômes psychiques que les répondants CATI, et présentent un score de santé mentale plus faible. Inversement, les répondants CATI semblent avoir été plus exposés que les répondants CAWI au bruit et aux retombées de suies. Notons que l'ajout des variables additionnelles (modèle 3), n'a que peu d'impact sur les RR présentés,

de même que l'ajout des variables relatives aux délais de réponse, pourtant susceptible de capter une sélection différentielle selon le mode de collecte (Tableau 20) ;

- en **zone témoin** concernant l'état de santé général et la déclaration d'une maladie chronique : les répondants CAWI apparaissent globalement en meilleure santé que les répondants CATI. Par ailleurs, les différences entre CATI et CAWI ont tendance à diminuer lorsque les variables additionnelles sont ajoutées dans les différents modèles, suggérant l'existence d'une différence entre les échantillons, non prise en compte dans le score de propension. Les répondants après relance (Tableau 20) ont significativement plus de maladies chroniques que les premiers répondants.

Tableau 19. Association entre les différentes variables d'intérêt et le mode de recueil (CATI en référence, RR et coefficients de régressions linéaires pour les scores continus MCS et PCS), selon 3 modèles de régression, en zone exposée et en zone témoin

Zone exposée	Modèle 1*	Modèle 2**	Modèle 3***
	RR CAWI (réf=CATI)	RR CAWI (réf=CATI)	RR CAWI (réf=CATI)
Santé actuelle			
Avoir consulté un médecin généraliste dans les 12 derniers mois (réf=non)	1.03 [1.00;1.07]	1.03 [0.99;1.08]	1.03 [0.98;1.07]
Avoir une maladie chronique (réf=non)	0.93 [0.86;1.01]	0.96 [0.87;1.06]	0.93 [0.84;1.04]
Score moyen de santé mentale (MCS)	-1.5 [-2.1;-0.79]	-1.27 [-2.0;-0.51]	-1.3 [-2.1;-0.44]
Score moyen de santé physique (PCS)	0.51 [-0.03;1.0]	0.49 [-0.11;1.1]	0.37 [-0.26;1.0]
Santé générale médiocre ou mauvaise (réf=bonne santé)	0.86 [0.70;1.04]	0.85 [0.67;1.09]	0.89 [0.69;1.15]
Post-accidentel			
Au moins 1 symptôme lié à l'incendie	1.04 [1.00;1.09]	1.03 [0.98;1.08]	1.03 [0.98;1.09]
Au moins 1 symptôme psychique lié à l'incendie	1.19 [1.09;1.29]	1.20 [1.08;1.33]	1.21 [1.09;1.35]
Au moins 1 symptôme physique lié à l'incendie	1.03 [0.97;1.09]	1.01 [0.94;1.08]	1.02 [0.95;1.09]
Avoir consulté suite aux symptômes liés à l'incendie	0.90 [0.75;1.09]	0.82 [0.65;1.03]	0.82 [0.63;1.06]
Avoir entendu du bruit ou été réveillé	0.92 [0.86;1.00]	0.87 [0.79;0.96]	0.86 [0.78;0.96]
Avoir eu des retombées de suies	0.97 [0.90;1.05]	0.94 [0.86;1.04]	0.92 [0.83;1.02]
Avoir eu des retombées de débris	1.50 [1.07;2.10]	1.19 [0.82;1.72]	1.13 [0.75;1.72]
Avoir été sous le panache	0.80 [0.72;0.89]	0.84 [0.74;0.94]	0.82 [0.72;0.94]
Avoir perçu des odeurs	1.08 [1.03;1.13]	1.07 [1.01;1.13]	1.06 [1.00;1.13]
Zone témoin	Modèle 1*	Modèle 2**	Modèle 3***
	RR CAWI (réf=CATI)	RR CAWI (réf=CATI)	RR CAWI (réf=CATI)
Santé actuelle			
Avoir consulté un médecin généraliste dans les 12 derniers mois (réf=non)	0,95 [0,89 ; 1,01]	0,93 [0,87 ; 1] *	0,93 [0,87 ; 1,01]
Avoir une maladie chronique (réf=non)	0,80 [0,69 ; 0,93]**	0,84 [0,72 ; 0,98]*	0,90 [0,77 ; 1,06]*
Score moyen de santé mentale (MCS)	0,54 [-0,57 ; 1,64]	0,35 [-0,81 ; 1,51]	0,28 [-0,95 ; 1,52]
Score moyen de santé physique (PCS)	0,85 [-0,07 ; 1,76]	0,59 [-0,42 ; 1,6]	0,38 [-0,76 ; 1,52]
Santé générale médiocre ou mauvaise (réf=bonne santé)	0,60 [0,38 ; 0,93]*	0,66 [0,44 ; 1]	0,70 [0,45 ; 1,09]

* Régression non pondérée avec facteurs de confusion et mode de collecte

** Régression pondérée (poids score de propension) avec facteurs de confusion et mode de collecte

*** Régression pondérée (poids score de propension) avec facteurs de confusion, mode de collecte et variables additionnelles

Tableau 20. Association entre les différentes variables d'intérêt et les variables relatives aux délais de réponse (RR et coefficients de régressions linéaires pour les scores continus MCS et PCS), pour le modèle 3 de régression, en zone exposée et en zone témoin

Zone exposée	Modèle 3***	
	RR Dernier quintile de répondants (réf=4 premiers quintiles)	RR répondants après relance (réf=avant relance)
Santé actuelle		
Avoir consulté un médecin généraliste dans les 12 derniers mois	0.99 [0.93;1.05]	0.98 [0.94;1.03]
Avoir une maladie chronique	0.91 [0.80;1.04]	0.95 [0.85;1.06]
Score moyen de santé mentale (MCS)	0.57 [-0.47;1.6]	-0.10 [-1.0;0.76]
Score moyen de santé physique (PCS)	0.55 [-0.23;1.3]	-0.27 [-0.94;0.40]
Santé générale médiocre ou mauvaise	0.82 [0.58;1.16]	1.07 [0.81;1.42]
Post-accidentel		
Au moins 1 symptôme lié à l'incendie	0.97 [0.91;1.04]	1.02 [0.97;1.08]
Au moins 1 symptôme psychique lié à l'incendie	0.87 [0.75;1.00]	1.07 [0.97;1.18]
Au moins 1 symptôme physique lié à l'incendie	0.91 [0.83;1.00]	1.04 [0.97;1.12]
Avoir consulté suite aux symptômes liés à l'incendie	1.02 [0.75;1.38]	0.94 [0.71;1.24]
Avoir entendu du bruit ou été réveillé	1.04 [0.92;1.19]	0.99 [0.88;1.11]
Avoir eu des retombées de suies	1.10 [0.96;1.25]	0.91 [0.82;1.01]
Avoir eu des retombées de débris	1.12 [0.68;1.84]	0.89 [0.58;1.35]
Avoir été sous le panache	0.89 [0.75;1.07]	0.98 [0.85;1.13]
Avoir perçu des odeurs	0.99 [0.92;1.06]	0.99 [0.94;1.06]
Zone témoin	Modèle 3***	
	RR Dernier quintile de répondants (réf=4 premiers quintiles)	RR répondants après relance (réf=avant relance)
Santé actuelle		
Avoir consulté un médecin généraliste dans les 12 derniers mois	1,04 [0,96 ; 1,12]	0,99 [0,9 ; 1,09]
Avoir une maladie chronique	1,01 [0,85 ; 1,21]	1,34 [1,07 ; 1,67]*
Score moyen de santé mentale (MCS)	-0,7 [-2,02 ; 0,62]	0,35 [-1,23 ; 1,94]
Score moyen de santé physique (PCS)	0,43 [-0,68 ; 1,54]	-0,56 [-2,01 ; 0,88]
Santé générale médiocre ou mauvaise	0,88 [0,5 ; 1,54]	1,14 [0,65 ; 2,02]

*** Régression pondérée (poids score de propension) avec variables auxiliaires + variables additionnelles

Interprétation des résultats

La lecture de ces différents résultats suggère que des effets de mode (biais de sélection différentiel et/ou biais de mesure) : 1) existent et 2) diffèrent entre zone exposée et zone témoin, ce qui, compte-tenu du sujet d'une étude post-incendie, était attendu.

En zone exposée, pour la plupart des variables d'intérêt, les différences observées selon l'empressement à répondre et selon le mode de collecte apparaissent d'ampleur modérée. Le score de propension équilibre bien les échantillons de répondants (Internet et téléphone), sur l'ensemble des variables du modèle ainsi que sur des variables de contrôle. Parmi les différences qui demeurent, le fait d'avoir été exposé au bruit et au panache apparaît plus fréquent au sein des répondants CATI, avec des SMD comprises entre 10% et 15% en utilisant l'inverse du score de

propension et les RR associés au mode de collecte de l'ordre d'environ 0,8. Par ailleurs, les répondants CAWI déclarent plus souvent que les répondants CATI un symptôme de nature psychique lié à l'incendie, et un score de santé mentale inférieur à celui des répondants CATI (RR ~ 1,2). Ces derniers apparaissent donc plus affectés ou préoccupés par la catastrophe elle-même ou ses conséquences.

En l'absence d'un instrument (variable expliquant la participation mais indépendante de la variable d'intérêt) qui aurait permis de modéliser simultanément la participation et la variable d'intérêt, nous ne pouvons pas, avec les données dont nous disposons, établir la part de cette différence attribuable à un biais de sélection différentiel entre les modes de celle attribuable à un effet de mesure (12). La sélection différentielle reflèterait par exemple le fait que les personnes les plus impactées psychologiquement par l'incendie auraient plus volontiers participé que les autres, et que cette sélection endogène serait ainsi mise en évidence car les deux modes de collecte reposent sur différents leviers de participation. Si cette hypothèse ne peut être écartée, les éléments suivants sont en faveur d'un biais de mesure, qui reflèterait cette fois-ci non pas un biais de composition d'échantillons mais un biais lié au mode de recueil des données, en l'occurrence avec ou sans enquêteur :

- la littérature scientifique s'accorde à observer que, dans le cadre d'échelles de santé perçue, les scores des différentes dimensions sont plus péjoratifs lorsque le questionnaire est complété directement par la personne interrogée (Internet dans notre cas) que lorsqu'il est administré par téléphone. Ce biais dit de « désirabilité sociale » est étant notamment observé pour l'évaluation de la santé mentale (14-19) ;
- le fait que l'ajout des variables additionnelles, supposées liées à une sélection différentielle selon l'intérêt pour l'enquête, ne modifient pas les associations avec le mode de collecte ;
- le fait que parmi les différentes variables d'intérêt, seules celles relatives à la santé mentale, supposées plus sensibles, présentent un effet de mode résiduel.

En zone témoin, un effet de mode résiduel est également observé, en faveur d'une participation plus importante des personnes en meilleur état de santé sur CAWI et il est possible que l'enquête surestime l'état de santé de la population résidant au Havre, avec une participation sans doute moindre des personnes ayant consulté lors des douze derniers mois ou ayant une maladie chronique. La diminution des RR avec l'introduction des variables additionnelles renforce cette hypothèse d'une différence entre les échantillons, non prise en compte dans le score de propension. En outre, et de manière plus globale, les effectifs plus réduits dans cette zone (~1 000 répondants) n'ont pas permis d'estimer un score de propension équilibrant l'ensemble des variables de contrôle. Le biais engendré par les différences de caractéristiques observables est moins bien corrigé, et l'analyse des différences entre modes, liée à la qualité des propriétés équilibrantes du score de propension, de moindre qualité.

Dans la zone exposée, comme dans la zone témoin, un effet de mode résiduel semble donc persister, pouvant être le résultat d'une sélection endogène et/ou d'un effet de mesure lié au mode de collecte. L'attribution de nos écarts à un biais de mesure uniquement reposerait sur l'absence de sélection endogène, et sur le fait que les variables mobilisées dans le score de propension expliquent intégralement le fait de répondre sur un mode plutôt qu'un autre. Cette hypothèse ne peut être vérifiée.

Cet effet de mode semble plus important en zone témoin qu'en zone exposée. Ceci pourrait s'expliquer par une participation plus faible (-10 points en zone témoin) et faisant appel à des motivations de participer différentes de celle de la zone exposée, directement impactée par l'incendie de septembre 2019.

Notons également que la distance à l'incendie, variable disponible dans la base de sondage, est une variable mobilisée dans la correction de la non-réponse et très corrélée aux variables d'intérêt de la zone exposée, notamment l'exposition aux différentes nuisances de l'incendie, permettant de

diminuer en partie un éventuel biais d'auto-sélection, ce qui n'est pas le cas en zone témoin dans laquelle l'éloignement à l'incendie n'est pas lié aux variables d'intérêt.

Etant donné que peu de variables sont concernées par cet effet de mode, notamment en zone exposée, et ne pouvant identifier rigoureusement la part attribuable de cet effet, aucun traitement correctif n'a été mis en œuvre (un tel traitement aurait pu être envisagé par exemple si l'effet de mode avait pu être attribué à un biais de mesure uniquement (20)). Des précautions d'interprétation concernant l'analyse de ces variables devront être prises.

Compte tenu de la littérature et des résultats présentés, sous l'hypothèse que l'effet de mode constaté soit principalement issu d'un biais de mesure entre modes de recueil et que les répondants par téléphone ont tendance à sous-déclarer leurs problèmes psychologiques, il apparaît possible que :

- d'une part, l'état de santé mentale des habitants de la zone exposée soit plus dégradé que celui estimé par l'étude (du fait que près du tiers de ces personnes a répondu par téléphone) ;
- d'autre part, la différence mesurée entre les estimations de l'état de santé mentale dans la zone exposée et dans la zone témoin soit légèrement surévaluée (du fait d'une plus grande proportion de répondants par téléphone en zone témoin qu'en zone exposée).

Par ailleurs, l'état de santé de la population résidant au Havre est probablement surestimé dans cette enquête, ce dont il faudrait tenir compte dans les comparaisons entre zones (exposée et témoin) qui ne concernaient toutefois pas les variables principales d'intérêt (symptômes et autres variables de santé post-incendie) directement liées à l'incendie, mais uniquement l'état de santé actuel, tel qu'évalué par le SF-12.

L'appariement des données de cette enquête au SNDS offrira la perspective de pouvoir confirmer ou non certaines de ces hypothèses en introduisant dans la modélisation de la non-réponse des indicateurs de recours aux soins mesurés et disponibles pour répondants comme non-répondants.

CONCLUSION

L'étude « Santé Post-incendie 76 : une étude à l'écoute de votre santé » a été mise en œuvre dans un contexte singulier, avec d'une part l'importance d'évaluer rapidement les nuisances et symptômes de santé ressentis par les personnes ayant été exposées à l'incendie industriel, d'autre part la production d'une enquête de qualité statistique représentative de cette population, dans le contexte de la crise sanitaire liée à la COVID-19 ayant débuté début 2020.

La préparation de cette enquête a nécessité la mise en place d'une équipe projet pluridisciplinaire au sein de l'agence (épidémiologie, communication, dialogue avec la société, statistique, méthodologie d'enquête, sociologie etc.), la réalisation d'une pré-enquête qualitative, ainsi que de nombreux échanges avec différents partenaires (Drees, Insee, représentants des habitants de la zone exposée) qui ont permis de définir le protocole (définition de la zone d'étude, de la zone témoin, élaboration du questionnaire, élaboration du plan d'échantillonnage et du protocole de collecte) et de recueillir les données en septembre et octobre 2020, soit 1 an après l'incendie. La construction de l'enquête (protocole, questionnaire, zone d'étude, documents d'information) s'est faite dans le cadre d'une démarche participative avec des personnes concernées par l'accident, organisé par Santé publique France et pilotée par le professeur Marcel Calvez. Chaque étape a fait l'objet d'une présentation aux CTD, rassemblant l'ensemble des parties prenantes.

Le protocole multimode mis en place a tenu compte des conséquences de la crise sanitaire liée à la Covid-19 sur l'activité des instituts de sondage, qui a réduit fortement les capacités d'enquête téléphonique de même qu'en face-à-face. Il a permis d'atteindre 95% des objectifs fixés, 4 773 répondants adultes ayant répondu à l'enquête sur les 5 000 souhaités initialement, 1 306 mineurs sur les 1 400 souhaités. Le taux de réponse à l'enquête a atteint 48% (50% pour la zone exposée), avec une participation importante des personnes par Internet (64 % dans l'ensemble). Le recours à la base de sondage Fidéli s'est révélé d'une grande importance pour cette enquête. La disponibilité des coordonnées géographiques (X,Y) des logements a permis de sélectionner les participants à l'enquête selon la proximité de leur lieu de résidence à l'incendie. La richesse des informations de contact disponibles a permis de favoriser la participation en contactant nominativement la très grande majorité des personnes sélectionnées et en diversifiant les modes d'information et de relances : courrier postal, mail, envois de SMS, appels téléphoniques. Enfin, certaines données socio-économiques, comme la structure du ménage ou le niveau de vie des ménages, ont été précieuses tant pour alléger le questionnaire du recueil de ce type de données que pour modéliser et traiter en partie la non-réponse totale. Au regard des critères statistiques utilisés concernant les variables socio-économiques disponibles, la représentativité de l'échantillon des répondants à l'enquête est satisfaisante.

Néanmoins, ce protocole présente plusieurs limites. Tout d'abord, le protocole séquentiel retenu repose sur deux modes de recueil différents, Internet et téléphone. S'il favorise la participation en offrant plusieurs opportunités de participer aux personnes sélectionnées (21), il est également associé à l'existence potentielle d'un effet mode. En effet, par téléphone, la présence d'un enquêteur permet de motiver la participation, notamment des personnes qui se sentent le moins concernées, mais peut augmenter le biais de désirabilité sociale sur des sujets sensibles ou sur certaines opinions, tandis que la participation en mode auto-administré (Internet dans notre cas) limite ce dernier biais (22), mais augmente celui de *satisficing* (réponses sans effort, systématiques, voire abandon en cours de questionnaire). Par ailleurs, la participation par Internet demande une motivation plus importante, et le biais d'auto-sélection lié à la participation de personnes plus intéressées par le sujet de l'enquête devient ainsi plus important. Notons que dans notre enquête, la proportion de répondants par Internet est ainsi plus importante dans la zone exposée (64,7%) que dans la zone témoin (53,1%), et que l'analyse spécifique menée dans cette étude suggère l'existence d'un effet mode, que nous supposons davantage associé à un biais de mesure qu'à une sélection endogène. L'identification et le traitement de ce biais, recourant par exemple au modèle de Heckman, aurait nécessité un protocole adapté, permettant de disposer d'une variable permettant

d'expliquer la participation indépendamment de l'intérêt pour l'enquête : affectation aléatoire de sous-échantillons à différentes combinaisons de modes de recueil et/ou à des efforts différentiels pour favoriser la participation (nombre de relances, incitations financières) (12).

Par ailleurs, compte tenu du protocole de collecte mis en place, la participation à l'enquête s'est avérée très liée aux informations de contacts disponibles dans la base de sondage (12% lorsque seule l'adresse postale était disponible contre 60 % lorsqu'un mail et un téléphone étaient en plus disponibles), l'accès à ces informations pouvant refléter elles-mêmes un certain degré de confiance vis-à-vis des pouvoirs publics. Pour des raisons liées aux délais de collecte, au coût de l'enquête ainsi qu'au contexte sanitaire, aucun recueil en face-à-face ni par questionnaire papier n'étaient proposés (en dehors du cas particulier des Gens du voyage), alors qu'il aurait certainement permis de favoriser la participation des personnes moins joignables ou plus réfractaires à participer.

Notons enfin que l'appariement, pour les personnes sélectionnées et n'ayant pas exercé leur droit d'opposition, à leurs consommations de soins dans le SNDS permettra de mener d'utiles analyses de sensibilité, notamment concernant la participation à l'enquête en lien avec certains indicateurs d'état de santé.

Références bibliographiques

1. 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
2. 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)
3. 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)
4. 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
5. Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*. 2006;166(10):1092-7
6. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*. 2001;16(9):606-13
7. Weathers FW, Litz BT, Keane TM, et al. The ptsd checklist for dsm-5 (pcl-5). *Scale available from the National Center for PTSD at www.ptsd.va.gov*. 2013;10
8. Royston P. ICE: Stata module for multiple imputation of missing values. 2014
9. Royston P, White IR. Multiple imputation by chained equations (MICE): implementation in Stata. *J Stat Softw*. 2011;45(4):1-20
10. Haziza D, Beaumont JF. On the construction of imputation classes in surveys. *International Statistical Review*. 2007;75(1):25-43
11. 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
12. Castell L, Sillard P. Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman. 2021
13. Zou G. A modified poisson regression approach to prospective studies with binary data. *American journal of epidemiology*. 2004;159(7):702-6
14. McHorney CA, Kosinski M, Ware Jr JE. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Medical care*. 1994;551-67
15. Perkins JJ, Sanson-Fisher RW. An examination of self-and telephone-administered modes of administration for the Australian SF-36. *Journal of clinical epidemiology*. 1998;51(11):969-73

16. Bowling A, Bond M, Jenkinson C, et al. Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *Journal of Public Health*. 1999;21(3):255-70
17. Powers JR, Mishra G, Young AF. Differences in mail and telephone responses to self-rated health: use of multiple imputation in correcting for response bias. *Australian and New Zealand journal of public health*. 2005;29(2):149-54
18. Lyons RA, Wareham K, Lucas M, et al. SF-36 scores vary by method of administration: implications for study design. *Journal of Public Health*. 1999;21(1):41-5
19. Epstein JF, Barker PR, Kroutil LA. Mode effects in self-reported mental health data. *Public Opinion Quarterly*. 2001;65(4):529-49
20. Legleye S, Razafindranovona T, de Peretti G. Agréger les échantillons d'une enquête multimode en limitant l'effet de mesure: une proposition d'imputation raisonnable et pragmatique. 2018
21. De Leeuw ED, Hox JJ, Dillman DA. Mixed-mode surveys: when and why. *International handbook of survey methodology*. 2008:299-316
22. Schlenger WE, Silver RC. Web-based methods in terrorism and disaster research. *Journal of traumatic stress: official publication of the international society for traumatic stress studies*. 2006;19(2):185-93

Annexe 1. Lettre d'information



Vos contacts

Tél. : 0800 942 590 (Numéro vert et gratuit)
Courriel : contact@sante-post-incendie-76.fr
Site internet : www.sante-post-incendie-76.fr
Votre identifiant : PLU10003
Votre mot de passe : XVPSITA369



Objet : Santé Post Incendie 76 : une étude à l'écoute de votre santé

Monsieur,

L'agence Santé publique France, missionnée par le ministère chargé de la Santé, réalise une étude de santé auprès des habitants des communes touchées par l'incendie industriel de Lubrizol et NL Logistique, survenu le 26 septembre 2019. Cette enquête s'intéresse à votre perception des pollutions générées par l'incendie, aux symptômes et problèmes de santé que vous avez ressentis, à votre accès aux informations au cours de l'événement, ainsi qu'à votre santé actuelle et votre qualité de vie. Vous avez été tiré au sort pour participer à l'étude, et vous avez reçu ou allez peut-être recevoir un mail ou un SMS à ce sujet.

Cette enquête est reconnue d'intérêt général par le Conseil national de l'information statistique (Cnis).

Votre participation est essentielle afin de permettre la prise en compte de la diversité des caractéristiques de la population interrogée et d'assurer ainsi la qualité des résultats de l'étude.

Vous pouvez répondre au questionnaire :

- En ligne, en vous connectant à l'adresse sécurisée ou en scannant le QR code ci-dessous.

<https://etude.sante-post-incendie-76.fr>

Mot de passe : XVPSITA369



Cette adresse est à inscrire dans la barre d'adresse de votre navigateur. Puis vous saisissez le mot de passe indiqué ci-dessus (merci de faire attention à bien respecter les caractères : minuscules / majuscules)

- Par téléphone, un enquêteur de l'institut Ipsos prendra prochainement contact avec vous pour vous interroger, avec ce numéro : 02 36 80 10 52. Vous pouvez également convenir d'un rendez-vous en appelant le 0800 942 590 ou sur le site www.sante-post-incendie-76.fr, en saisissant votre mot de passe.

À la suite de l'enquête, que vous ayez ou non répondu au questionnaire, sous réserve de votre non-opposition, des données sur vos consommations de soins qui sont disponibles auprès de la Caisse nationale d'assurance maladie (Cnam) pourront être recueillies. Comme vos réponses à l'enquête, elles seront traitées de manière confidentielle, en respectant votre vie privée.

Vous trouverez au verso de cette lettre, des informations sur le traitement de données mis en œuvre dans le cadre de cette étude ainsi que les modalités d'exercice de vos droits.

Je vous donne l'assurance que, comme la loi en fait la plus stricte obligation, vos réponses resteront confidentielles et serviront uniquement à l'établissement de statistiques.

En vous remerciant par avance de votre participation, je vous prie d'agréer, Monsieur, l'expression de ma considération distinguée.



Comment vos coordonnées ont été obtenues ?

Vos coordonnées ont été obtenues à partir des bases de logements de l'Insee qui sont constituées à partir de sources administratives.



Que deviennent les informations que vous donnerez ?

Les résultats de l'enquête seront rendus publics à partir de fin 2020. Les données obtenues seront très précieuses pour évaluer l'impact de cet incendie industriel, prendre d'éventuelles mesures de santé publique et améliorer la gestion sanitaire d'autres accidents industriels.

Vu l'avis favorable du Conseil national de l'information statistique, l'enquête Santé post incendie 76 : une étude à l'écoute de votre santé a obtenu le visa n°2020X712SA du Ministre de l'économie et des finances, valable pour l'année 2020 – Arrêté en date du 03/07/2020.

Confidentialité et traitement de vos données à caractère personnel

Cette enquête met en œuvre un traitement de données à caractère personnel fondé sur la réalisation d'une mission d'intérêt public et dont Santé publique France a la responsabilité. Le règlement général 2016/679 du 27 avril 2016 sur la protection des données (RGPD), la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, ainsi que la loi n°51-711 du 7 juin 1951 relative à l'obligation, la coordination et le secret en matière de statistiques, s'appliquent à la présente enquête.

À cette fin, les données suivantes vous concernant ou concernant votre enfant seront recueillies dans la mesure où ces données sont nécessaires aux objectifs visés : données démographiques et sociales, données de santé (incluant symptômes, recours aux soins), situations professionnelles et financières actuelles, consommation de tabac et d'alcool. Ces données ne feront pas apparaître vos noms et prénoms et seront associées à un identifiant anonyme.

Les réponses à ce questionnaire sont protégées par le secret statistique et destinées à Santé publique France.

Le prestataire Ipsos est seul destinataire des données d'identification (nom et coordonnées), habilité par le Comité du secret de la statistique publique, et les Archives de France, conformément aux dispositions de l'article L. 212-4 du code du patrimoine. Elles seront conservées par Ipsos pendant 3 mois à compter de la fin de la collecte, et par Santé publique France pendant 2 ans si vous acceptez de les communiquer pour être recontacté à des fins d'enquêtes ultérieures.

Cette enquête n'est pas obligatoire.

Sous réserve de votre non-opposition, seront collectées auprès de l'Assurance maladie des données concernant le suivi de votre santé (médicaments, consultations, hospitalisations, examens complémentaires). Ces données ainsi que vos réponses au questionnaire seront conservées pendant 10 ans à compter de la fin de la collecte pour les besoins de l'enquête par Santé publique France. Elles seront archivées au-delà de cette durée. À tout moment, leur usage et leur accès seront strictement contrôlés et limités à l'élaboration de statistiques ou à des travaux de recherche scientifique.

L'analyse de ces données sera effectuée par Santé publique France ou des organismes de recherche, à partir de données ne permettant pas de vous identifier directement (exclusion du nom, prénom, adresses).

Exercice de vos droits d'opposition, d'accès et de rectification

Conformément à la loi, vous pouvez exercer un droit d'accès, d'opposition, de rectification, d'effacement ou de limitation de traitement pour les données vous concernant ou concernant votre enfant pendant la période de conservation des données d'identification. Ces droits sont applicables à la collecte de données, dans le cadre de l'enquête, à l'appariement ultérieur des données recueillies avec des données de l'Assurance Maladie et aux utilisations ultérieures de ces données.

Ces droits peuvent être exercés auprès de l'enquêteur qui vous contactera, ou en vous adressant par courriel à : dpo@santepubliquefrance.fr ou par courrier à : Santé publique France - DPO - 12 rue du Val d'Osne, 94415 Saint-Maurice Cedex. Pour toute question relative au traitement de vos données, vous pouvez contacter le délégué à la protection des données de Santé publique France à l'adresse dpo@santepubliquefrance.fr. Vous pouvez, si vous l'estimez nécessaire, adresser une réclamation auprès de la CNIL.

Agence nationale de santé publique
12, rue du Val d'Osne 94415 Saint-Maurice Cedex France
Standard +33(0)1 41 79 67 00

www.santepubliquefrance.fr

Annexe 2. Correction de la non-réponse totale : informations complémentaires

Méthode mise en œuvre pour la correction de la non-réponse totale

Étapes de mises œuvres du traitement de la non-réponse totale par la méthode des scores :

Étape 1 : Modélisation du score

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 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 β un vecteur de paramètres inconnus.

On obtient avec ce 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 le score \hat{p}_i qui servira de critère d'homogénéité des classes. Notons que lors de cette étape est également sélectionné l'ensemble des variables auxiliaires qui expliquent le comportement de réponse. La sélection de ces variables peut se faire de façon a priori ou lors d'une étape de modélisation en utilisant un critère statistique.

Étape 2 : Formation des classes homogènes

En utilisant les probabilités de réponses estimées \hat{p}_i , on forme les classes homogènes suivant la méthode des quantiles égaux, en répartissant les unités en un certain nombre de groupes, définis a priori. Pour ce faire, on ordonne les valeurs de \hat{p}_i en 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 classes homogènes par rapport à \hat{p}_i , le seront également par rapport aux p_i , de sorte que le biais sera proche de 0.

Étape 3 : Ajustement de la pondération initiale

Une fois que les groupes de réponse homogènes (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.

Volet principal-échantillon des adultes

Concernant les adultes, 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 sa strate de sondage. 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 :

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

La modélisation des probabilités de réponse a été réalisée sous SAS à l'aide de la procédure logistic non pondérée. 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. Pour la zone exposée, la variable distance au lieu de l'incendie a été forcée dans le modèle de la correction de la non-réponse. Cette variable est très corrélée à certains indicateurs d'intérêt mais aussi à la participation à l'enquête. La liste des variables retenues dans le modèle utilisé pour prédire les probabilités de réponse pour les deux zones (exposée et témoin) est présentée dans le Tableau A2.1. Les statistiques descriptives des pondérations sont présentées Tableau A2.2, Tableau A2.3 et Figure A2.1 pour la zone exposée, et Tableau A2.4, Tableau A2.5 et Figure A2.2 pour la zone témoin.

Tableau A2.1. Liste des variables des modèles de non-réponse pour l'échantillon adultes, en zone exposée et en zone témoin

Zone exposée	Zone témoin
Âge de l'adulte au moment du terrain	Âge de l'adulte au moment du terrain
Mode de contact au niveau ménage	Mode de contact au niveau ménage
Niveau de vie en décile (année n-1)	Niveau de vie en décile (année n-1)
Sexe	Sexe
Lien familial de l'individu échantillonné avec le référent fiscal	Lien familial de l'individu échantillonné avec le référent fiscal
Statut d'occupation du logement (propriétaire / locataire)	Indicatrice de logement social O/N
Type de ménage	Type de logement
Perception d'un salaire par l'individu-cible (année n)	Perception du chômage de 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	

Tableau A2.2. 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	341	91%	33	9%	0,09
2	330	88%	45	12%	0,13
3	318	85%	57	15%	0,14
4	277	74%	97	26%	0,27
5	258	69%	117	31%	0,31
6	254	68%	121	32%	0,32
7	239	64%	135	36%	0,36
8	414	55%	336	45%	0,43
9	172	46%	203	54%	0,54
10	311	42%	438	58%	0,59
11	508	34%	991	66%	0,66
12	192	26%	557	74%	0,74
13	71	19%	304	81%	0,82
14	51	14%	324	86%	0,88

Les rapports des poids (max/min, P75/P25, P95/P5) de la pondération corrigée de la non-réponse tronquée sont comparables aux rapports des poids de la pondération initiale.

Tableau A2.3. Statistiques descriptives des différentes pondérations en zone exposée

Paramètres	Facteur repondération	Pondérations initiales (poids de sondage)	Pondérations corrigées de la non-réponse tronquées
Somme	-	252 586	252 586
Ecart-type	0	17	41
Variance	0	293	1 646
Ecart interquartile	0	24	50
Moyenne	1	34	67
Mode	1	28	87
Coeff Variation (%)	31	51	60
100Max 100%	1	52	159
95%	1	52	150
90%	1	52	134
75% Q3	1	52	87
50% Médiane	1	46	69
25% Q1	0	28	38
10%	0	11	15
5%	0	2	6
0% Min	0	1	3
Rapport P75/P25	2	2	2
Rapport P95/P5	3	26	27
Rapport Max/Min	10	52	52

Figure A2.1. Répartition des poids avant et après correction de la non-réponse totale en zone exposée

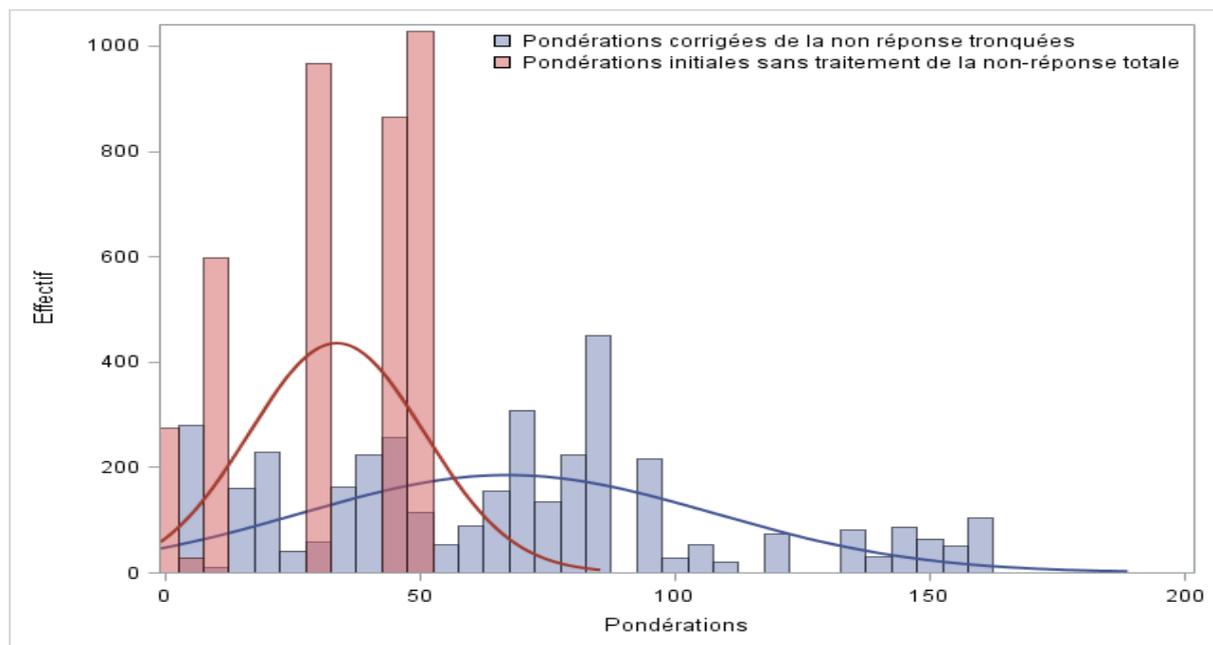


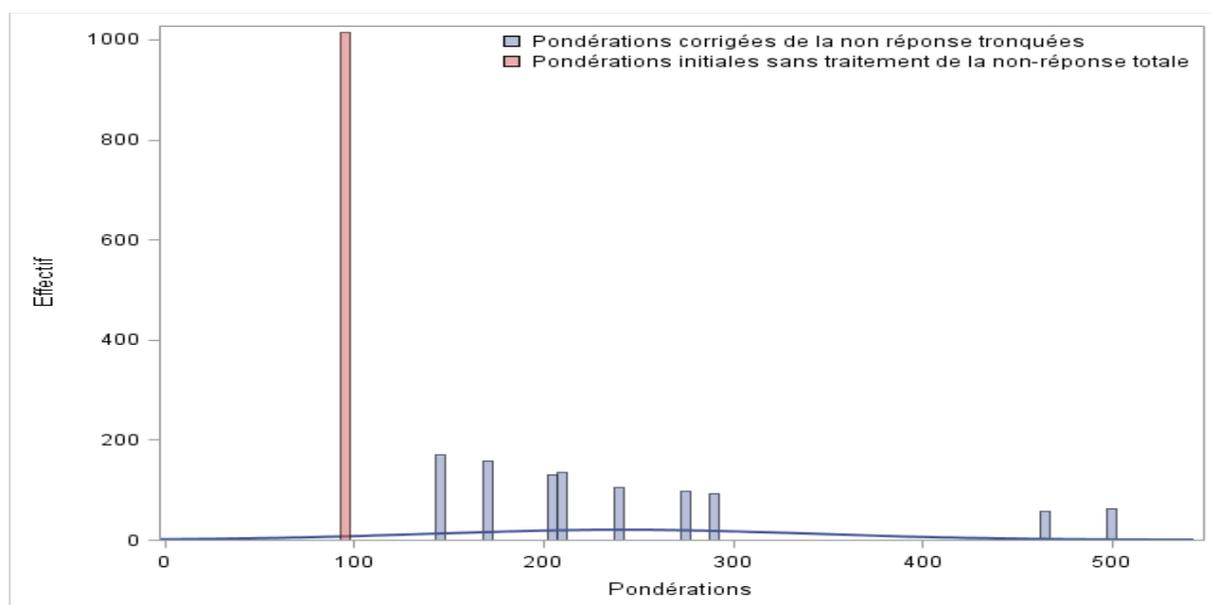
Tableau A2.4. Répartition des répondants et non-répondants dans les GRH en zone témoin

GRH	Refus/abandons/Non contactés		Répondants		Taux de réponse pondéré
	n	%	n	%	
1	238	94%	14	6%	0,06
2	204	81%	49	19%	0,19
3	195	77%	58	23%	0,23
4	160	63%	93	37%	0,37
5	155	61%	98	39%	0,39
6	132	56%	105	44%	0,44
7	131	49%	137	51%	0,51
8	121	48%	132	52%	0,52
9	94	37%	159	63%	0,63
10	83	33%	170	67%	0,67

Tableau A2.5. Statistiques descriptives des différentes pondérations en zone Témoin

Paramètres	Facteur d'ajustement	Pondérations initiales	Pondérations corrigées de la non-réponse tronquées
Somme	-	245 454	245 454
Ecart-type	0	0	101
Variance	0	0	10 120
Ecart interquartile	0	0	106
Moyenne	1	97	242
Mode	1	97	145
Coeff Variation	31	0	42
100Max 100%	1	97	501
95%	1	97	501
90%	1	97	467
75% Q3	1	97	276
50% Médiane	1	97	209
25% Q1	0	97	170
10%	0	97	145
5%	0	97	145
0% Min	0	97	145
Rapport P75/P25	2	1	2
Rapport P95/P5	4	1	4
Rapport Max/Min	12	1	4

Figure A2.2. Répartition des poids avant et après correction de la non-réponse totale – zone témoin



Volet principal-échantillon des enfants

Les enfants étaient tirés au sort conditionnellement au tirage de l'adulte. Une fois l'adulte tiré au sort, une personne mineure était sélectionnée parmi les mineurs vivant dans le même ménage.

En notant :

- h la strate de tirage ;
- W_h^A le poids de sondage de l'adulte tiré au sort ;
- f_E le nombre d'enfants du ménage ;

Le poids initial de l'enfant tiré au sort est donné par : $W_h^E = f_E * W_h^A$

Lorsque les enfants vivaient en garde alternée chez deux parents vivant dans la même zone (exposée/témoin), leurs poids ont été ajustés en utilisant la méthode de partage des poids. Des questions spécifiques figurent dans le questionnaire afin de déterminer l'éligibilité des parents en cas de garde alternée.

Le traitement de la non-réponse pour l'échantillon des enfants a été réalisé séparément en zone exposée et témoin et avec la même méthode que celle utilisée pour l'échantillon des adultes. La liste des variables incluses dans le modèle de la correction de la non-réponse pour les zones exposée et témoin est présentée dans le Tableau A2.6. Les statistiques descriptives des pondérations sont présentées Tableau A2.7 et Figure A2.3 pour la zone exposée, et Tableau A2.8 et Figure A2.4 pour la zone témoin.

Tableau A2.6. Liste des variables du modèle de la correction de la non-réponse totale concernant les enfants

Zone Exposée	Zone Témoin
Type de ménage	Type de ménage
Âge de l'adulte échantillonné au moment du terrain	Âge de l'adulte échantillonné au moment du terrain
Sexe de l'adulte	Sexe de l'adulte
Niveau de vie en décile (année n-1)	Niveau de vie en décile (année n-1)
Mode de contact au niveau ménage	Mode de contact au niveau ménage
Âge de l'enfant échantillonné au moment du terrain	Âge de l'enfant échantillonné au moment du terrain
Logement social	Logement social
Perception du chômage de l'adulte (année n)	Nombre d'années écoulées depuis l'année d'emménagement
Présence d'enfants de 0 à 17 ans au 1er janvier 2019	
Distance domicile-usine (en m), négative si arrière du panache (sud-ouest de l'usine)	
Perception d'un salaire par l'adulte (année n)	
Nombre de personnes du logement	

Tableau A2.7. Statistiques descriptives des différentes pondérations – enfants, zone exposée

Paramètres	Facteur d'ajustement	Pondérations initiales	Pondérations corrigées de la non-réponse tronquées
Somme	-	67 780	67 780
Ecart-type	0	28	50
Variance	0	800	2 511
Ecart interquartile	0	32	51
Moyenne	1	35	66
Mode	0	28	70
Coeff Variation	33	82	76
100Max 100%	1	278	189
95%	1	93	175
90%	1	69	156
75% Q3	1	46	82
50% Médiane	1	28	51
25% Q1	0	14	31
10%	0	6	16
5%	0	3	7
0% Min	0	1	5
Rapport P75/P25	2	3	3
Rapport P95/P5	7	31	25
Rapport Max/Min	7	555	40

Figure A2.3. Répartition des poids avant et après correction de la non-réponse totale – enfants, zone exposée

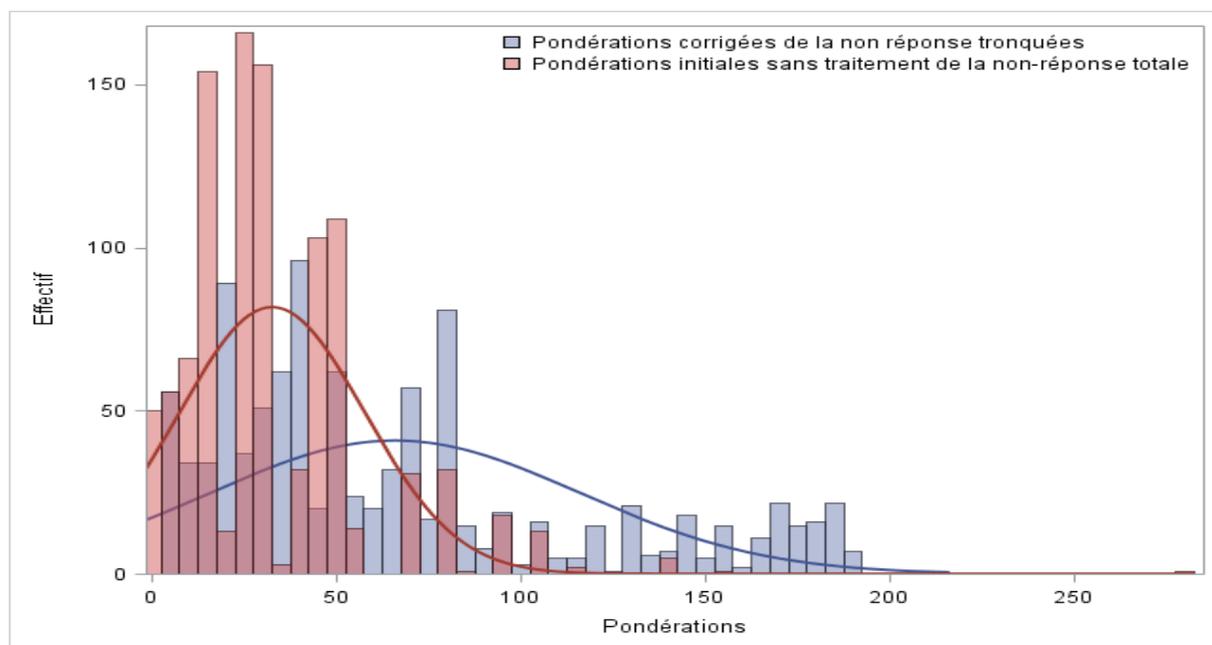
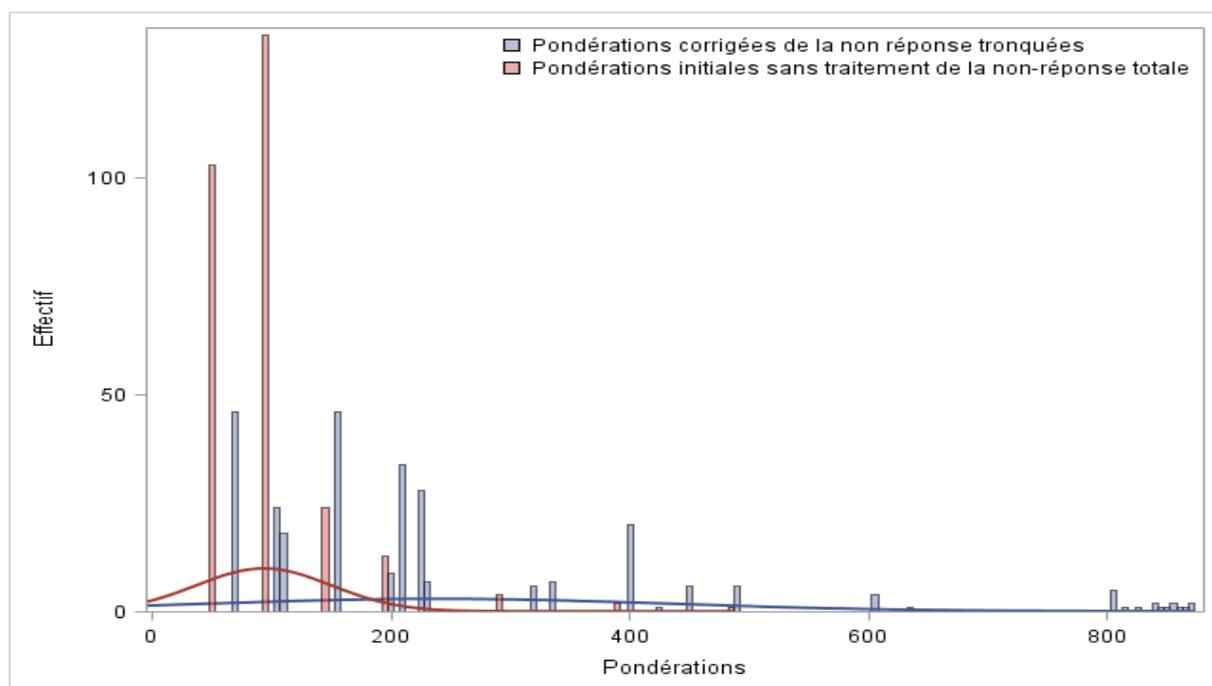


Tableau A2.8. Statistiques descriptives des différentes pondérations – enfants zone Témoin

Paramètres	Facteur d'ajustement	Pondérations initiales	Pondérations corrigées de la non-réponse tronquées
Somme		69 033	67 346
Ecart-type	0	57	194
Variance	0	3 193	37 600
Ecart interquartile	0	49	121
Moyenne	1	94	241
Mode	1	97	68
Coeff Variation	33	60	81
100Max 100%	1	485	871
95%	1	194	803
90%	1	146	471
75% Q3	1	97	230
50% Médiane	1	97	201
25% Q1	0	49	109
10%	0	49	68
5%	0	49	68
0% Min	0	49	68
Rapport P75/P25	2	2	2
Rapport P95/P5	6	4	12
Rapport Max/Min	6	10	13

Figure A2.4. Répartition des poids avant et après correction de la non-réponse totale – enfants, zone témoin



Volet Santé mentale

Un jeu de poids spécifique a été calculé afin de tenir compte de la non-réponse totale au volet Santé mentale. Comme pour le volet principal de l'enquête, ce traitement de la non-réponse a été réalisé séparément en zone exposée et en zone témoin et en utilisant la même méthode, mais en repartant des poids corrigés de la non-réponse des répondants au volet principal. La liste des variables incluses dans le modèle de correction de la non-réponse pour les zones exposée et témoin est présentée dans le Tableau A2.9. En plus des informations disponibles dans la base de sondage sur les répondants et non-répondants, des informations collectées dans le questionnaire principal ont été utilisées pour le traitement de la non-réponse au volet Santé mentale. Les statistiques descriptives des pondérations sont présentées Tableau A2.10, Tableau A2.11 et Figure A2.5 pour la zone exposée, et Tableau A2.12, Tableau A2.13 et Figure A2.6 pour la zone témoin.

Tableau A2.9. Liste des variables du modèle de la correction de la non-réponse totale pour le volet Santé mentale

Zone Etude	Zone Témoin
Âge de l'adulte échantillonné au moment du terrain	Âge de l'adulte échantillonné au moment du terrain
Distance domicile-usine (en m), négative si arrière du panache (sud-ouest de l'usine)	Distance domicile-usine (en m), négative si arrière du panache (sud-ouest de l'usine)
Niveau de vie en décile (année n-1)	Niveau de vie en décile (année n-1)
Sexe	Sexe
Perception de minima sociaux (année n-1)	Perception de minima sociaux (année n-1)
Lien familial de l'individu échantillonné avec le référent fiscal	Lien familial de l'individu échantillonné avec le référent fiscal
A eu au moins une exposition	A eu au moins une exposition
Diplôme le plus élevé de l'adulte participant	Diplôme le plus élevé de l'adulte participant
Utilisation d'Internet	Utilisation d'Internet
État de santé général	État de santé général
Situation professionnelle actuelle de l'adulte participant	Disponibilité d'une adresse mail
Maladie ou problème de santé chronique ou durable	
Consultation d'un médecin	
Consultation d'un psychologue, psychiatre ou psychothérapeute	
A eu un symptôme tel que stress ou panique	
MCS : Score agrégé de santé mentale	
A eu au moins un symptôme physique lié à l'incendie	
Consommation de boissons alcoolisées	
Logement: Quartiers de la politique de la ville(QPV)	
Type de ménage	

Tableau A2.10. Répartition des répondants et non-répondants dans les GRH, volet Santé mentale en zone exposée

GRH	Refus/abandons/Non contactés		Répondants		Taux de réponse pondéré
	n	%	n	%	
1	160	85%	18	15%	0,15
2	152	81%	36	19%	0,17
3	138	73%	50	27%	0,27
4	131	70%	57	30%	0,28
5	136	72%	53	28%	0,29
6	131	70%	57	30%	0,31
7	233	62%	143	38%	0,38
8	114	61%	74	39%	0,38
9	222	59%	155	41%	0,39
10	109	58%	79	42%	0,41
11	94	50%	94	50%	0,5
12	94	50%	94	50%	0,52
13	242	43%	323	57%	0,56
14	127	34%	249	66%	0,64
15	54	29%	135	71%	0,7

Tableau A2.11. Statistiques descriptives des différentes pondérations, volet Santé mentale en zone exposée

Paramètres	Facteur d'ajustement	Pondérations initiales	Pondérations corrigées de la non-réponse tronquées
Somme	-	252 586	252 603
Ecart-type	0	41	108
Variance	0	1 650	11 688
Ecart interquartile	0	50	147
Moyenne	0	67	155
Mode	1	87	147
Coeff Variation	30	61	70
100Max 100%	1	159	411
95%	1	150	380
90%	1	134	345
75% Q3	1	87	221
50% Médiane	1	69	132
25% Q1	0	37	74
10%	0	15	28
5%	0	6	15
0% Min	0	1	11
Rapport P75/P25	2	2	3
Rapport P95/P5	3	27	26
Rapport Max/Min	5	159	39

Figure A2.5. Répartition des poids avant et après correction de la non-réponse totale, volet Santé mentale en zone exposée

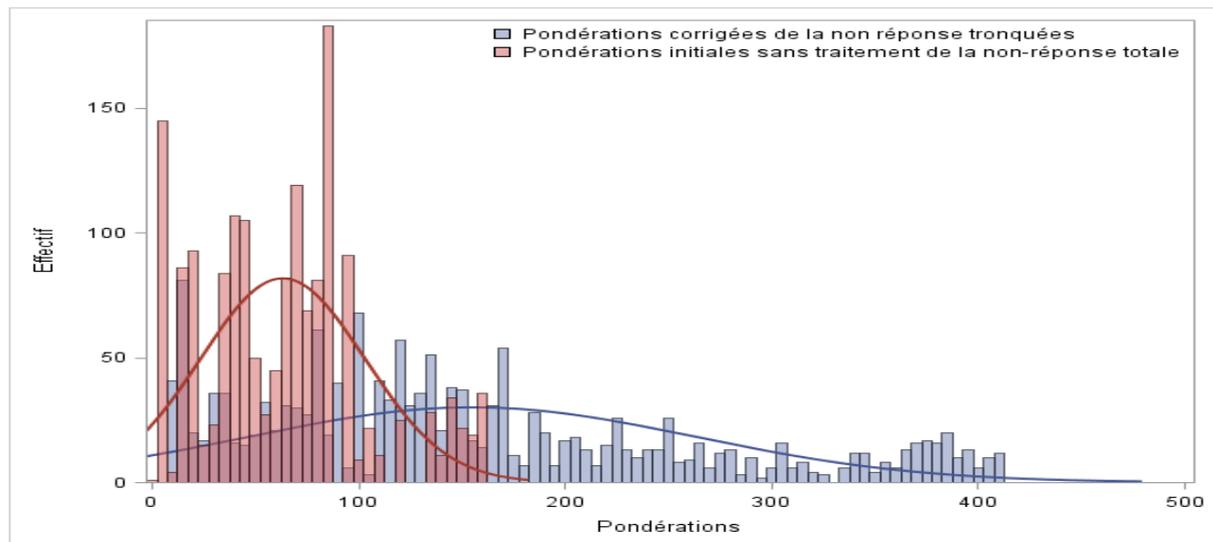


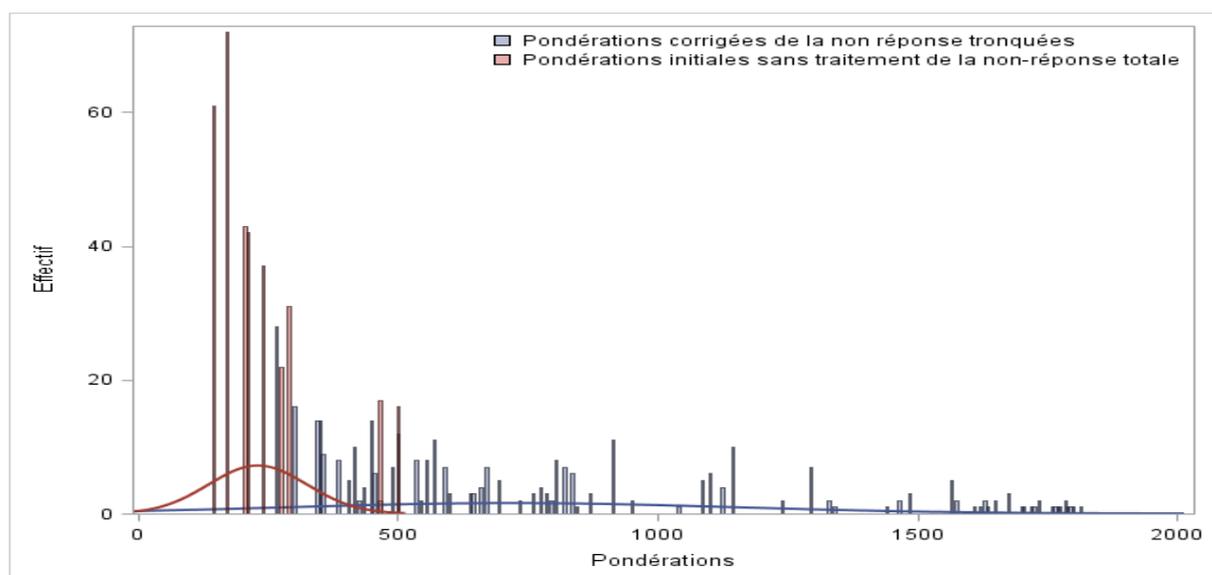
Tableau A2.12. Répartition des répondants et non-répondants dans les GRH, volet Santé mentale en zone témoin

GRH	Refus/abandons/Non contactés		Répondants		Taux de réponse pondéré
	n	%	n	%	
1	90	89%	11	11%	0,11
2	162	80%	41	20%	0,2
3	149	73%	54	27%	0,28
4	67	66%	35	34%	0,34
5	63	62%	38	34%	0,38
6	59	58%	43	42%	0,41
7	47	47%	54	53%	0,52
8	37	36%	65	64%	0,64

Tableau A2.13. Statistiques descriptives des différentes pondérations, volet Santé mentale en zone témoin

Paramètres	Facteur d'ajustement	Pondérations initiales	Pondérations corrigées de la non-réponse tronquées
Somme	-	245 454	245 453
Ecart-type	0	101	431
Variance	0	10 120	186 164
Ecart interquartile	0	106	530
Moyenne	0	242	720
Mode	1	145	265
Coeff Variation	38	42	60
100Max 100%	1	501	1 814
95%	1	501	1 676
90%	1	467	1 484
75% Q3	1	276	915
50% Médiane	0	209	568
25% Q1	0	170	384
10%	0	145	299
5%	0	145	265
0% Min	0	145	265
Rapport P75/P25	2	2	2
Rapport P95/P5	3	4	6
Rapport Max/Min	6	4	7

Figure A2.6. Répartition des poids avant et après correction de la non-réponse totale, volet Santé mentale en zone témoin



Citation suggérée : 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

ISSN : 2647-4816 / ISBN-NET 979-10-289-0767-9 / RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION,
SANTÉ PUBLIQUE FRANCE / DÉPÔT LÉGAL : MARS 2022