

PUBLIC VOICE, LE PANEL PROBABILISTE

Yves Fradier (), Nicolas Sauger (**)*

() Kantar Public, Directeur Grandes Enquêtes et Méthodes*

*(**) Sciences Po, Directeur du CDSP, Professeur*

Yves.fradier@kantar.com

Mots-clés : panel, aléatoire, collecte, multimode

Domaine : 4. Collecte de données d'enquêtes

Résumé

Cet article présente les étapes successives de la constitution du premier lot de 5 000 panélistes Public Voice en France.

Public Voice couvre 19 pays d'Europe, mais cette présentation se concentre sur la France uniquement.

Il a été décidé de lancer ce panel probabiliste, dédié uniquement aux enquêtes de la Statistique Publique, pour trois raisons :

- Contribuer à pallier la baisse des taux de participation aux enquêtes,
- Ainsi que la baisse des budgets consacrés aux études,
- L'évolution de la crise sanitaire rendant toujours incertaine la pérennité du face à face, par ailleurs.

Pour la France, nous avons produit deux phases pilotes, recourant au CATI et au face à face. Après analyse des résultats, nous avons finalement retenu le CATI pour la phase en grandeur réelle, pour des questions de couverture de la population (meilleure avec le CATI selon cette expérience), et de coût et délais (deux fois inférieurs).

Nous avons augmenté la part des numéros de fixes par rapport au pilote pour obtenir un effectif plus important de personnes non connectées à Internet. Cette part de la population a la même importance pour nous que la part majoritaire qui, elle, est connectée.

Nous présentons dans cet article la conception et la mise en œuvre des différentes phases de terrain : le pilote CATI, le pilote face à face, et la phase CATI en grandeur réelle.

A chaque étape sont présentés des éléments de comparaison entre la structure de notre panel et celle de la population.

Une dimension supplémentaire inédite est apportée : la comparaison entre la structure des personnes acceptant d'entrer dans le panel comparée à celle des personnes qui refusent. C'est-à-dire que sans autre donnée en entrée qu'un numéro de téléphone, nous avons recueilli des données

sociodémographiques en amont de la question de consentement à Public Voice, ce qui nous permet cette comparaison.

Nous présentons également des comparaisons entre les données recueillies dans le questionnaire de « profiling » placé à la fin du recrutement, vs d'autres sources de données reconnues pour la rigueur méthodologique de leur provenance.

Au final, nous considérons que ce panel peut constituer une réponse à la nécessité de continuer à produire des données d'enquêtes avec des taux de participation, des durées de terrain, et des coûts acceptables.

Abstract en Anglais

This article presents the successive stages in the constitution of the first batch of 5 000 Public Voice panellists in France.

Public Voice covers 19 European countries, but this presentation focuses on France only.

It was decided to launch this probabilistic panel, dedicated solely to Public Statistics surveys, for two reasons: to help offset the decline in participation rates in surveys, and the decline in budgets devoted to studies.

The evolution of the health crisis still makes the sustainability of face-to-face surveys uncertain.

For France, we produced two pilot phases, using CATI and face-to-face. After analysing the results, we finally chose CATI for the full-scale phase, for reasons of population coverage (better with CATI according to this experience), and cost and time (twice as low).

We increased the share of landline numbers compared to the pilot to obtain a larger number of people not connected to the Internet. This part of the population has the same importance for us as the majority part, which is connected.

In this article we present the design and implementation of the different field phases: the CATI pilot, the face-to-face pilot, and the full-scale CATI phase.

At each stage, elements of comparison between the structure of our panel and that of the population are presented.

An additional dimension is added: the comparison between the structure of people who accept to enter the panel compared to those who refuse. That is, without any input data other than a telephone number, we have collected socio-demographic data before the Public Voice consent question, which allows us to make this comparison.

We also present comparisons between the data collected in the profiling questionnaire placed at the end of the recruitment process, vs. other data sources known for the methodological rigour of their provenance.

In the end, we consider that this panel can be a response to the need to continue to produce survey data with acceptable participation rates, field times and costs.

Introduction

Kantar Public lance Public Voice, qui est un panel probabiliste, au niveau français et européen. La création *ex nihilo* d'un nouveau panel est un long processus semé de péripéties, et très coûteux.

Pourtant elle s'impose pour plusieurs raisons :

- Tout d'abord la baisse des taux de réponse aux enquêtes en général¹, en particulier auprès de certaines cibles (les jeunes, les PCS-). Cette baisse des taux de réponse a pour conséquence très directe, la hausse des coûts de production : il faut mettre en œuvre plus de dispositifs, et augmenter le taux d'insistance. Finalement, on obtient moins de données pour plus cher.
- Ensuite la baisse des budgets de recherche. Les commanditaires des enquêtes de la Statistique Publique sont plus que jamais aujourd'hui, on l'a vu lors de la crise COVID, en attente de données d'enquêtes. Mais ces enquêtes sont de moins en moins dotées des budgets nécessaires.
- Par ailleurs, les deux dernières années ont eu un impact négatif fort sur les enquêtes en face à face. Dans certains pays, les terrains face à face ont été arrêtés, dans tous, les taux de participation ont baissé par la hausse des refus par crainte de la contagion.
- Cette situation nous pousse tous à trouver des solutions qui satisfassent à la fois les ambitions méthodologiques, les contraintes budgétaires et le contexte sanitaire.

Notre réponse est la création du panel Public Voice. Notre présentation porte sur la constitution de la partie française du panel.

1. Les Pilotes

Avant de lancer la production en grandeur réelle, nous avons décidé de piloter le recrutement, en ayant recours à deux modes de collecte distincts : le face à face et le CATI. Ces deux pilotes ont été conduits en juillet 2021.

1.1 Résultat du pilote face à face

Nous avons tiré au sort 25 communes sur l'ensemble du territoire métropolitain. Puis un ou deux points de chute par commune, pour un total de 31. Ce sont les adresses de départ fournies aux enquêteurs. Chacun d'eux doit ensuite observer des règles de *random route* pour les adresses suivantes.

160 recrutements ont été réalisées, à partir de 1 062 adresses.

Le résultat en termes de structure des panélistes recrutés vs la population est en ligne, sauf sur trois points essentiellement : les 16-24 ans (7 % de recrutés vs 14 % dans la population), la taille d'agglomération (44 % de recrutés parmi les communes de moins de 2 000 habitants vs 23 % dans la population) et la taille du ménage spécifiquement pour les personnes vivant seules : 29 % vs 19 % dans la population).

¹ Par exemple, le taux de réponse a baissé pour les enquêtes en face-à-face Cadre de vie et sécurité (de 72 % à 66 % entre 2012 et 2021) et SRCV (de 85 % à 80 % entre 2010 et 2019).

1.2 Résultat du pilote CATI

Nous avons procédé à une génération aléatoire de numéros de téléphones fixes (à l'exclusion des 09) et mobiles (06 et 07), dans une proportion de 80 % de mobiles et 20% de fixes.

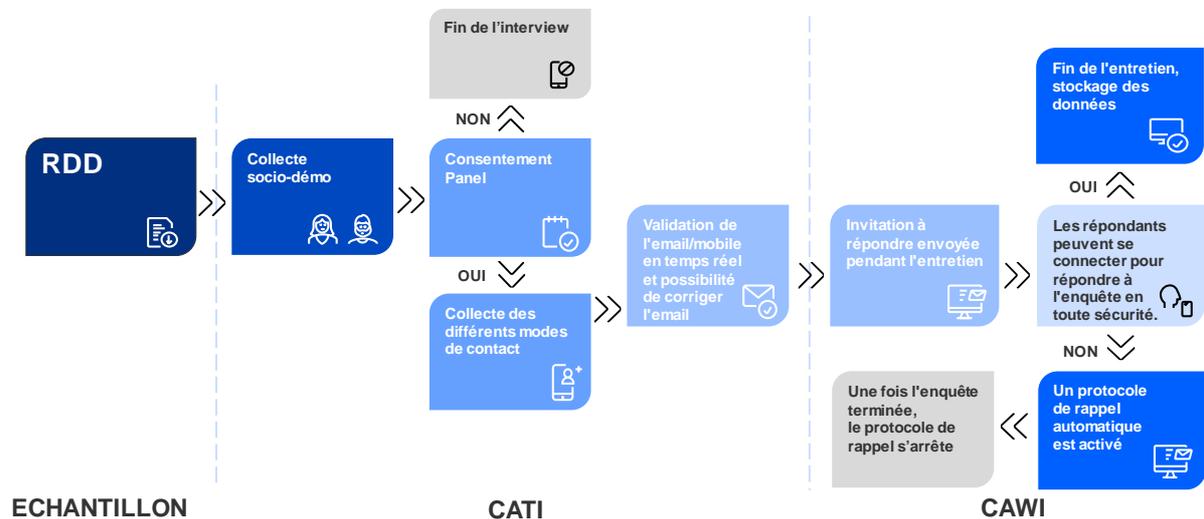
227 recrutements ont été réalisés à partir de 5 437 numéros.

Le résultat en termes de structure des panélistes recrutés est plutôt meilleur que le face à face, en particulier pour les 16-24 ans (17 % vs 14 % dans la population).

1.3 Résultat de l'expérimentation sur le questionnaire

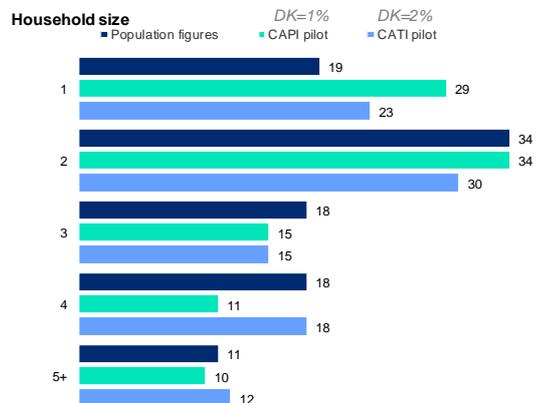
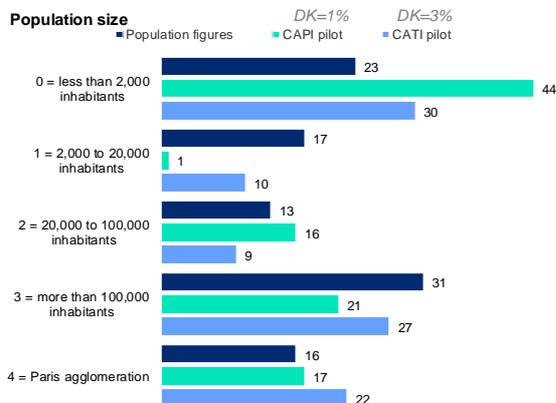
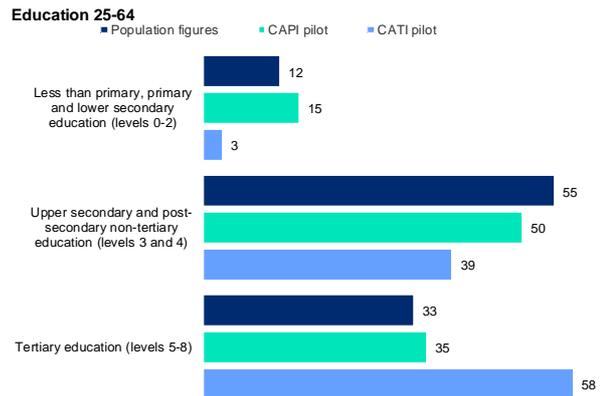
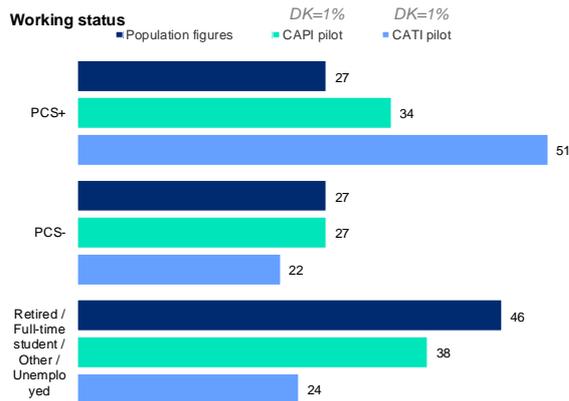
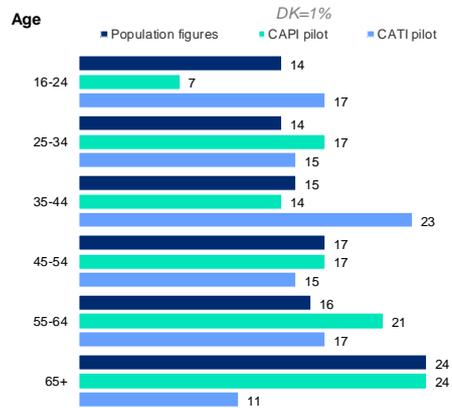
Pour chaque mode, nous avons testé deux versions de questionnaire A et B.

La version A proposait d'entrer dans le panel dès la phase de contact. La version B proposait de répondre à un questionnaire, venaient ensuite les questions de socio démo, puis enfin seulement, après quelques minutes, la question de recrutement dans le panel.



1.4 Quelques graphiques

Nous présentons ici la comparaison entre la structure de la population et le résultat de nos deux pilotes pour les variables sexe, âge, PCS, niveau de diplôme, catégorie d'agglomération, taille du ménage.



1.5 Conclusions des pilotes

La décision a été prise de recourir au CATI pour la phase du recrutement principal.

Les raisons sont les suivantes :

- On obtient globalement une meilleure structure avec le CATI.
- Le coût de collecte est deux fois moins élevé et deux fois plus rapide.

La seule modification apportée au protocole est d'augmenter la part des téléphones fixes, parce que nous avons constaté lors du pilote que quasiment l'intégralité des personnes non connectées à internet ont été recrutées à partir d'un numéro de téléphone fixe.

Par ailleurs, c'est la version B du questionnaire, approche plus « soft », qui a été choisie, parce que plus efficace. L'autre raison étant qu'en procédant ainsi on a la possibilité de comparer les refusants et les acceptants d'entrer dans le panel.

2. La phase de recrutement

2.1 Résultat du CATI

En France, il existe trois approches principales pour constituer un échantillon probabiliste :

- 1) La génération aléatoire de numéros de fixes et mobiles,
- 2) Les annuaires,
- 3) Les bases administratives (type FIDELI)

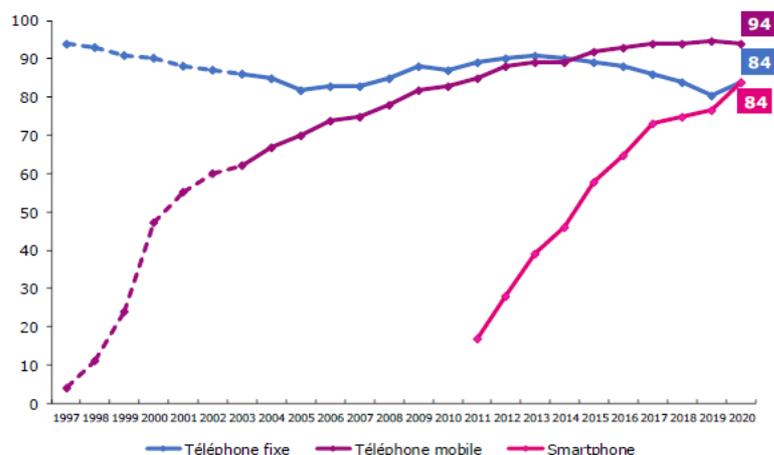
Chacune de ces trois méthodes constitue une option appropriée pour un échantillon probabiliste, à condition qu'elle puisse répondre à certaines exigences minimales en matière de couverture et de représentation.

Au minimum, la base de sondage doit offrir une couverture de 95 % de la population cible, ce qui exclut de fait le recours aux annuaires, de moins en moins exhaustifs.

N'ayant pas accès à FIDELI, notre choix se porte naturellement sur la génération aléatoire de numéros.

Stratification : Pour la génération de mobiles, nous recommandons que la base de sondage soit stratifiée par opérateur sur la base du préfixe du numéro et pour la génération de fixes par région (NUTS3). La stratification est un bon moyen de vérifier que l'échantillon a été tiré correctement, même en tenant compte du phénomène grandissant des numéros portés.

Mix fixe/mobile



Bien que le taux de possession de téléphones mobiles soit de 94 % en France, nous préférons toujours utiliser une part de téléphones fixes, d'une part pour inclure les personnes non connectées à Internet, d'autre part pour assurer une couverture plus large de la population de façon générale.

Sélection des répondants à l'intérieur du ménage

Lorsque la sélection d'un individu au sein du ménage est nécessaire, tous les individus éligibles qui résident dans le ménage, qu'ils soient au moment de l'appel à la maison ou non, doivent avoir une probabilité de sélection non nulle.

Pour Public Voice, nous avons utilisé la méthode Rizzo :

Pour cette méthode, l'enquêteur demande d'abord à parler à un membre éligible du ménage. Il demande ensuite combien de personnes éligibles vivent dans le ménage. S'il y en a deux, il choisit soit l'interlocuteur (la personne qui répond au téléphone), soit l'autre personne éligible du ménage. Le script effectue une sélection aléatoire et l'enquêteur reçoit l'instruction de continuer avec l'interlocuteur ou de demander à parler à l'autre membre éligible du ménage. Pour les ménages comptant trois membres éligibles, le script sélectionne soit l'interlocuteur, soit les deux autres membres éligibles du ménage. La probabilité de sélectionner l'interlocuteur serait de $1/3$, et pour les deux autres membres de $2/3$. Si les deux autres membres sont sélectionnés, l'enquêteur sera chargé de choisir l'un d'entre eux en suivant la méthode anniversaire (dernier/prochain anniversaire).

Pour 4 membres éligibles du ménage, l'approche est la même, mais la probabilité de sélection de l'interlocuteur devra être ajustée selon le tableau ci-dessous :

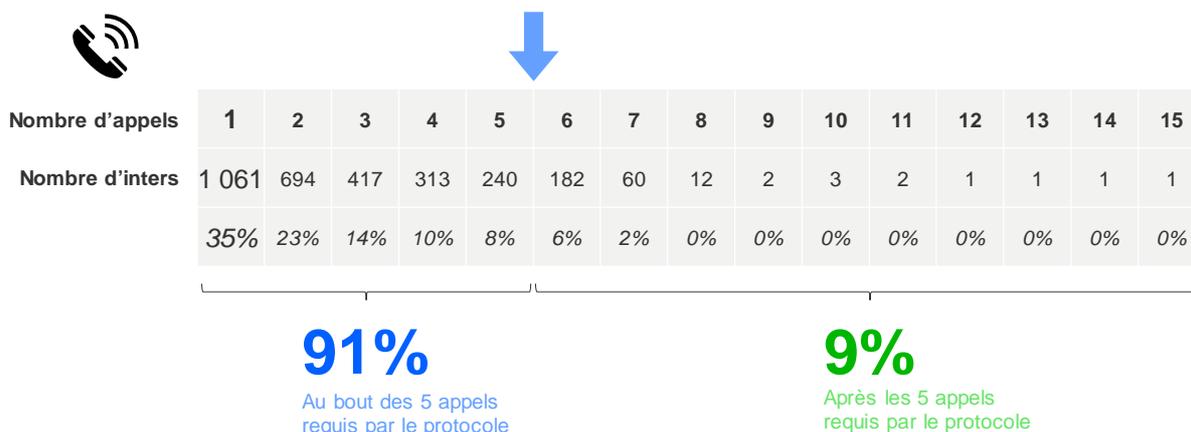
Taille du ménage	Probabilité de sélectionner l'interlocuteur
1	1
2	½
3	1/3
4	¼
5	1/5
N	1/n

Protocole d'appels

La qualité du travail sur le terrain est aussi importante que le processus de sélection d'un échantillon probabiliste. Nous avons fixé à 5 le minimum du nombre d'appels par numéros.

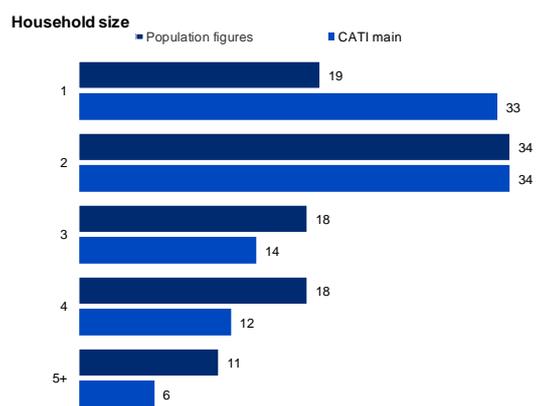
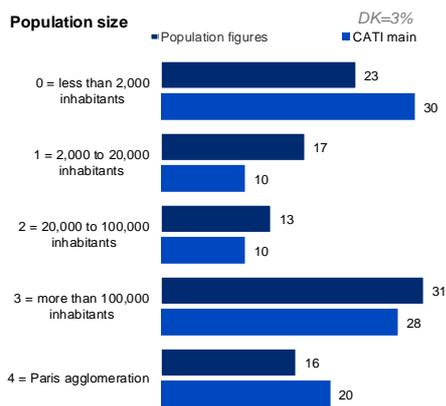
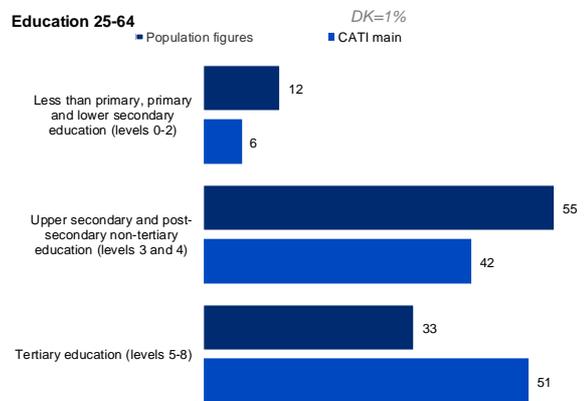
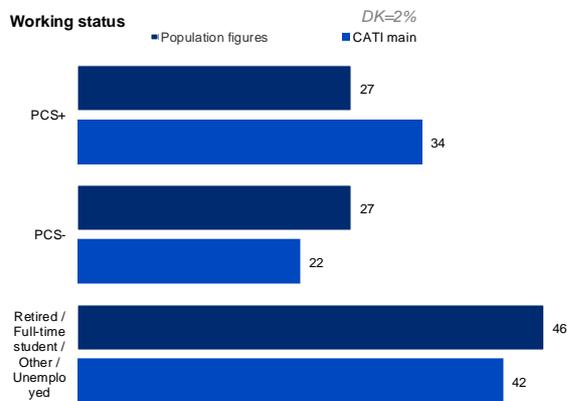
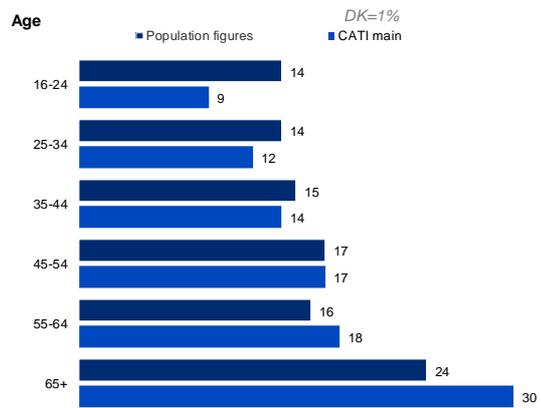
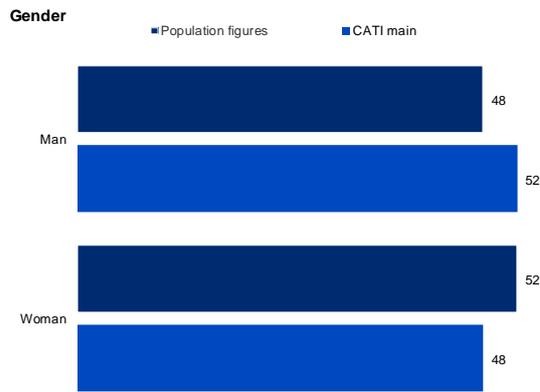
Les rappels doivent être effectués sur tous les numéros pour lesquels un résultat non final a été atteint lors de l'appel précédent. Les appels au même numéro doivent être répartis sur plusieurs jours et à différents moments de la journée pour maximiser le contact, avec un minimum de 50 % de tous les appels effectués après 17 heures un jour de semaine ou à n'importe quel moment du week-end.

Le terrain principal a été conduit entre septembre et décembre 2021. Nous avons mis sur le terrain 70 % de numéros de mobiles et 30 % de numéros de fixes divisés et exploités en 6 lots.



Au final, nous avons recruté 5 097 panélistes, âgés de 16 ans et plus vivant en France métropolitaine.

Voici la structure de l'ensemble de ces panélistes vs la population française :

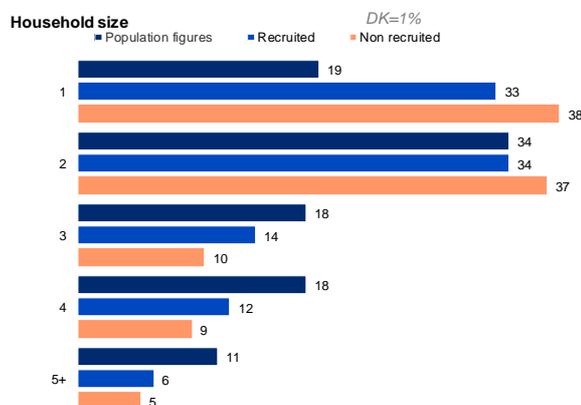
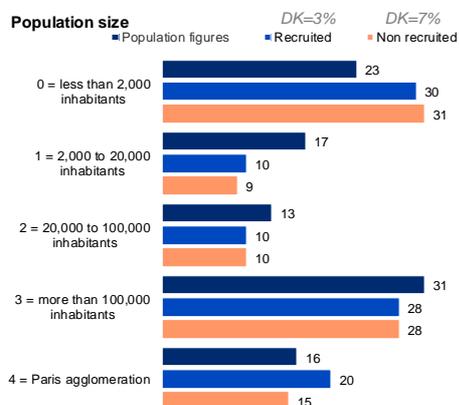
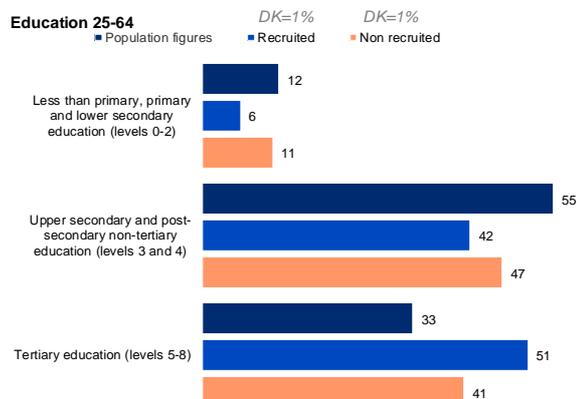
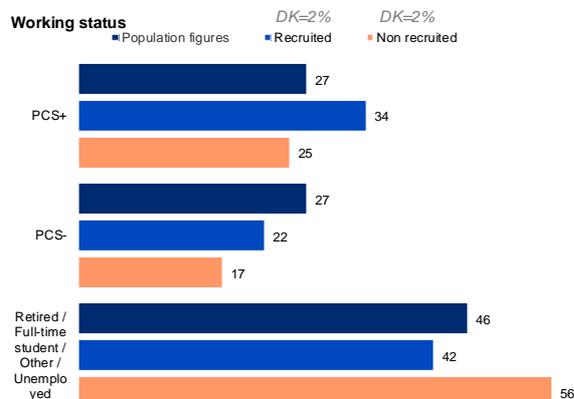
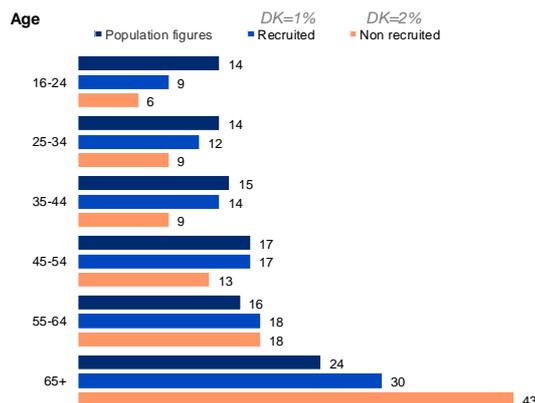
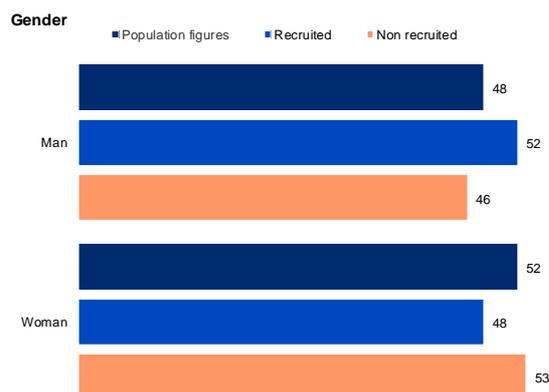


2.2 Comparaison panélistes/refusants

L'un des points intéressants que l'utilisation de la version B du questionnaire nous permet de documenter est la comparaison de la structure des panélistes.

Les personnes qui acceptent de répondre au questionnaire répondent d'abord à 6 questions sociodémographiques : sexe, âge, PCS, niveau de diplôme, catégorie d'agglomération et taille du ménage. On leur pose ensuite la question de recrutement, et ils acceptent à ce moment-là d'entrer ou non dans le panel. Nous avons donc les réponses de l'ensemble de ces personnes, acceptants et refusants pour ces 6 variables.

Voici le résultat :

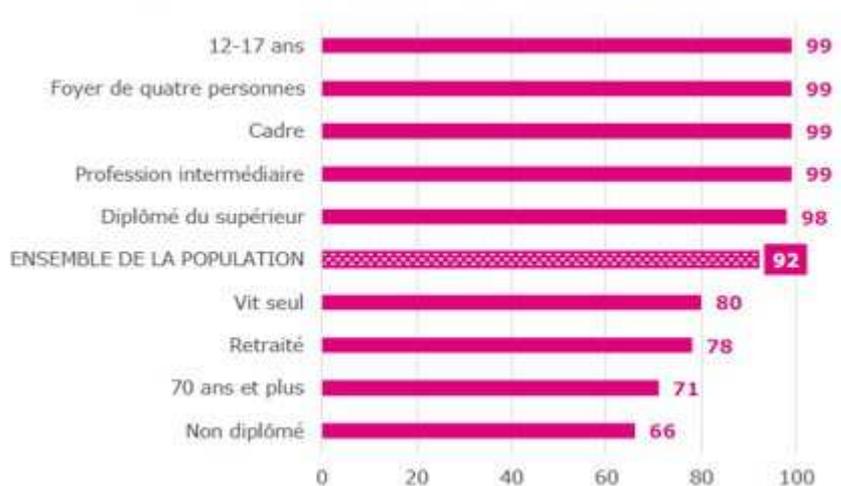


Les enseignements que nous tirons de ces résultats sont de deux ordres :

- La plus ou moins grande facilité à joindre les personnes est flagrante sur les dimensions d'âge et d'occupation : on le voit très nettement chez les plus de 65 ans/inactifs. C'est aussi le reflet de la surreprésentation délibérée des numéros de ligne fixe dans notre échantillon.
- Pour autant, et pour ces deux catégories (qui pointent vers les mêmes personnes), la proportion des refusants est majoritaire.
- A l'inverse, pour les autres tranches d'âge, et en particulier les plus jeunes (16-24), la proportion des acceptants est plus forte, ce qui signifie que les arguments utilisés sont efficaces auprès d'eux.
- Ces arguments portent aussi auprès des personnes avec le plus haut niveau de diplôme/PCS+. On doit encore travailler sur la façon de convaincre ceux qui ont un niveau de diplôme moins élevé/PCS-

2.3 Focus « offliners »

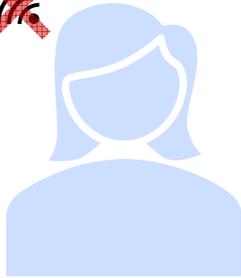
Graphique 38 – Les groupes les plus et les moins internautes en 2020
- Champ : ensemble de la population de 12 ans et plus, en % -



Source : CREDOC, Baromètre du numérique, édition 2021.

L'une des caractéristiques principales du positionnement de Public Voice est l'inclusion des « offliners ». C'est en cela que Public Voice n'est pas un panel CAWI, mais un panel majoritairement online et CATI pour deux catégories de personnes : celles qui n'ont pas de connexion Internet mais aussi celles qui en ont une qu'elles n'utilisent que très peu. C'est pour inclure ces deux catégories de la population que le choix du mode de collecte est laissé au panéliste. 479 d'entre eux ont ainsi choisi de répondre à nos enquêtes par téléphone.

Ceux qui n'ont pas de connexion internet



65%
87%
96%
71%
42%

Une femme
Âgée de 65 ans et +
Inactive
Vivant seule
Dans une zone rurale

Ceux qui ont une connexion internet mais préfèrent être interrogés CATI



58%
61%
71%
48%
30%

2.4 Focus « profiling »

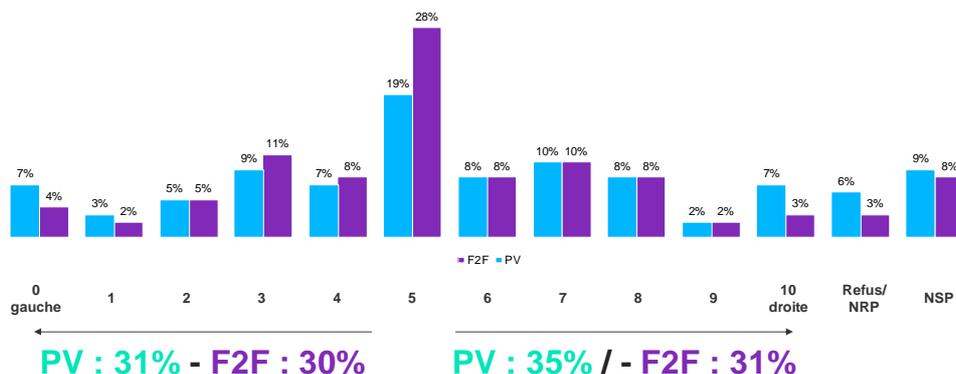
Comme nous l'avons dit précédemment, le questionnaire de recrutement se clôt par une partie « profiling », série de questions portant sur l'équipement des personnes, leurs usages numériques, la confiance dans les institutions, les opinions politiques, la propension à prendre des risques, l'état de santé, la pratique d'exercices physiques, la consommation de tabac, d'alcool, les modes de transport, l'attitude face à la vaccination contre le COVID-19, l'environnement, les valeurs personnelles.

Nous présentons ci-dessous les réponses à trois questions (sur la base des données brutes) comparées avec les données d'autres études publiées :

- Pour ce qui concerne l'équipement en ordinateur, nos panélistes déclarent en être équipés à 82 %.
Le taux d'équipement mesuré par le CREDOC dans l'édition 2021 de son baromètre du numériques est de 83 %.
- Pour ce qui est de la consommation de tabac, les personnes qui déclarent fumer (quelle que soit la fréquence) sont 27 %.
Pour Santé publique France, dans l'édition 2020 du Baromètre Santé, c'est 25,5 %.
- Enfin nous présentons ci-dessous la comparaison entre la question de l'appartenance gauche-droite, mesurée de la même façon dans ce questionnaire de profiling et dans une grande enquête de Statistique Publique produite en face à face en aléatoire fin 2021 :

En politique, on parle parfois de "gauche" et de "droite".

Sur une échelle de 0 à 10, où vous classeriez-vous si 0 est la gauche et 10 la droite ?



On le voit nettement : quand un enquêteur pose la question, on a moitié moins de répondants se déclarant d'extrême gauche/droite (les notes 0 et 10), et quasiment un tiers de répondants en plus sur la note médiane.

Bibliographie

[1] Building Pew Research Center's American Trends Panel, Scott Keeter, Director of Survey Research

[2] Setting Up an Online Panel Representative of the General Population: The German Internet Panel, Annelies G. Blom, Christina Gattmann, Ulrich Krieger

[3] Technical Overview of the Amerispeak Panel Norcs Probability Based Household Panel, J. Michael Dennis, Ph.D.

[4] Building a Probability Based Online Panel: Life in Australia, L. Kazcmirek, B. Phillips, DW Pennway, PJ Lavrakas, D. Neiger

[5] Developing the NatCen Panel, Curtis Jessop

[6] GESIS Panel, An Open Omnibus Access Panel for the Social Sciences, Prof. Dr. Michael Bosnjak