

---

N° 3 | 2025  
2025

---

# Présentation de l'enquête électorale française de 2017 : méthodologie, données, documentation et diffusion

*Jieun JEONG*

**Flora CHANVRIL** Ingénieure de recherche  
CEVIPOF  
Sciences Po, Paris

---

**Édition électronique :**

**URL :**

<https://demc-journal.org/articles/revue-3/3355-presentation-de-lenquete-electorale-francaise-de-2017-met-hodologie-donnees-documentation-et-diffusion>

**ISSN :** 3036-5295

**Date de publication :** 19/12/2025

**CertiScience**® *Certifié évalué par les pairs*

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

---

Pour **citer cette publication** : JEONG, J., CHANVRIL, F. (2025) Présentation de l'enquête électorale française de 2017 : méthodologie, données, documentation et diffusion. *DEMC Journal*, (3).  
<https://doi.org/10.34745/>

**Mots-clés :**

---

## **Introduction**

L'utilisation de panels demeure particulièrement pertinente pour les analyses en sociologie électorale. À titre d'exemple, deux études récentes illustrent leur intérêt : la première, basée sur le *British Election Study Panel*, examine le lien entre autoritarisme, attitudes politiques et choix de vote (Vasilopoulos, Robinson, 2025) ; la seconde mobilise les données du *Comparative Attitudes Under COVID-19 Project* pour analyser l'impact de la pandémie sur le comportement électoral (Brouard, Michel, 2025). Elle reste également discutée : si l'usage des panels en science sociale remonte aux années 1930, grâce aux travaux pionniers de Paul Lazarsfeld et de l'école de Columbia, leurs avantages ne doivent pas occulter les désagréments liés notamment à l'attrition et aux durées souvent courtes des panels qui en découlent (Cohen, Mayer et Marx, 2023).

L'enquête électorale ENEF 2017 s'inscrit dans un long historique d'enquêtes électorales longitudinales (Le Hay, 2009 ; Le Hay, 2017). Ces enquêtes ont d'abord été archivées et diffusées par le CIDSP (Centre d'informatisation des données socio-politiques) à Grenoble, puis par le CDSP (Centre de données socio-politiques) à partir de 2005.

Différents partenaires institutionnels ont participé au financement, en grande partie publique, de l'enquête ENEF 2017, c'est notamment le cas du Ministère de l'intérieur, en tant qu'organisateur des scrutins électoraux, de Sciences Po et de la Fondation Jean Jaurès. Menée par les chercheurs électoralistes du CEVIPOF en partenariat avec IPSOS et *Le Monde* - et la Fondation Jean Jaurès à la dernière vague -, l'enquête électorale ENEF 2017 représente une initiative ambitieuse pour suivre les dynamiques électorales en France sur une période étendue. Son objectif principal est d'étudier les comportements et attitudes des électeurs français en amont, pendant et après les scrutins de 2017, en s'appuyant sur un panel large et diversifié.

Cette enquête se distingue par son caractère innovant et sa robustesse méthodologique. Il s'agit d'un panel de 23 vagues permettant d'interroger à plusieurs reprises les mêmes individus afin de saisir l'évolution de leurs opinions, attitudes et comportements. L'ensemble des thématiques abordées dans les questionnaires n'était pas figé au lancement de l'enquête : des modules thématiques, parfois liés à l'actualité, ont pu être ajoutés en cours de dispositif ou

proposés par des équipes de recherche extérieures au CEVIPOF, enrichissant ainsi la portée analytique du panel.

L'échantillon initial comprenait 24 369 participants, incluant un échantillon principal de la population française en âge de voter et un échantillon secondaire de primo-votants âgés de 15 à 17 ans lors de la première vague. ENEF 2017 a également été conçu avec une souplesse d'interrogation : les panélistes pouvaient choisir de ne pas répondre à certaines vagues puis de revenir par la suite, ce qui a permis de limiter les effets d'attrition tout en maintenant un noyau dur de répondants constants.

La première partie de l'article porte sur la phase de production et d'exploitation scientifique des données par l'équipe de recherche chargée de l'enquête. Nous y expliquons dans un premier temps le dispositif d'échantillonnage et le déroulement général du dispositif. Nous détaillons ensuite le phénomène d'attrition associé au panel, ses éventuels impacts sur la qualité de l'échantillon et les moyens d'y remédier. Nous présentons également un exemple d'expérimentation parmi celles développées dans cette enquête.

La seconde partie de l'article porte sur les phases de préparation des données avant diffusion et de diffusion proprement dite. Nous y détaillons d'abord les traitements effectués pendant la phase de curation à partir de plusieurs exemples de types de variables. Puis nous expliquons la phase de documentation au standard Data Documentation Initiative. S'ensuit la présentation des données disponibles pour l'enquête ENEF 2017, et de l'accès via la plateforme de diffusion du CDSP. Enfin, nous décrivons comment la valorisation des données a été pensée pour favoriser leur analyse longitudinale<sup>[1]</sup>.

## **1. Présentation de l'enquête**

### **1.1. Méthodologie de l'enquête et échantillonnage**

L'échantillonnage de l'enquête repose sur une stratification géographique (grandes régions et catégorie d'agglomération) et des quotas sociodémographiques classiques (âge, sexe, profession de la personne de référence du ménage) afin de garantir la représentativité de la population française. L'intégration du niveau de diplôme comme variable de quotas, très fortement biaisée dans les enquêtes d'opinion, aurait conduit à des coefficients de pondération d'amplitude excessive et risqué de déformer la structure d'autres variables. Ce choix a donc été écarté, et les utilisateurs des données sont invités à prendre en compte ce biais résiduel dans leurs analyses.

La collecte des données a été réalisée par IPSOS avec la méthode CAWI (Computer Assisted Web Interviewing). Le terrain d'enquête s'est déroulé sur 18 vagues principales et plusieurs vagues intermédiaires spécifiques, de novembre 2015 à mai 2018. L'analyse de la construction des opinions sur le temps long, dans le sens d'un temps d'observation étendu couvrant plusieurs scrutins successifs (Le Hay, 2017 ; Muxel, 2018), est une des spécificités marquantes de cette enquête. Cette temporalité étendue a permis de suivre différentes élections : les élections régionales de 2015, les deux tours de l'élection présidentielle de 2017, les élections législatives de 2017. Il a ainsi été possible de couvrir différentes périodes préélectorales, les formations des intentions de vote pour chaque scrutin, les phases de primaires précédant l'élection présidentielle, tout en mesurant l'évolution de l'espace idéologique des personnes interrogées ainsi que l'entrée dans le possible électoral d'une nouvelle génération de primo-votants.

Tableau 1 : Calendrier de l'enquête électorale française de 2017 (vagues spécifiques en grisé)<sup>[2]</sup>

Vague	Dates du terrain
V1	20-29/11/2015
V1bis	16/12/2015-06/01/2016
V2	22-31/01/2016
V3	11-20/03/2016
V3bis	15-25/04/2016
V4	13-23/05/2016
V5	17-26/06/2016
V6	09-18/09/2016
V7	14-19/10/2016
V8	08-13/11/2016
V9	02-07/12/2016
V10	10-15/01/2017
V11	07-12/02/2017
V11bis	01-05/03/2017
V12	14-19/03/2017
V12bis	31/03-04/04/2017
V13	16-19/04/2017
V14	30/04-03/05/2017
V15	27/05-02/06/2017
V16	27/06-09/07/2017
V17	21-27/11/2017
V18	25/04-02/05/2018

L'échantillon initial en vague 1 comporte un échantillon principal de 24139 français et françaises en âge de voter, ainsi qu'un échantillon secondaire de 230 primo-votants, âgés de 15 à 17 ans au moment de cette première interrogation et en âge de voter pour les élections de 2017. Pour atteindre cet échantillon total de 24369 personnes, il a fallu inviter 205309 personnes en tout. Parmi elles, 47646 ont cliqué sur le lien de l'invitation. Le taux de réponses prend en compte trois situations :

- les réponses partielles (c'est-à-dire les participants n'ayant pas répondu à certaines questions tout en allant au bout du questionnaire) ;

- les réponses totales ;
- celles se terminant par une sortie anticipée du questionnaire : cela correspond à une sortie rapide lorsque les quotas correspondants étaient déjà remplis ou à une sortie identifiée en cours de questionnaires dans le cas de répondants identifiés comme “fraudeurs” (répondant trop rapidement par exemple).

Ce taux de réponse, calculé en rapportant le nombre d'ouvertures du lien d'enquête au nombre d'invitations, est de 23 %. Ce niveau est habituel pour une enquête CAWI avec contact uniquement par internet (voir par exemple Charrance, Cochet, et al., 2022).

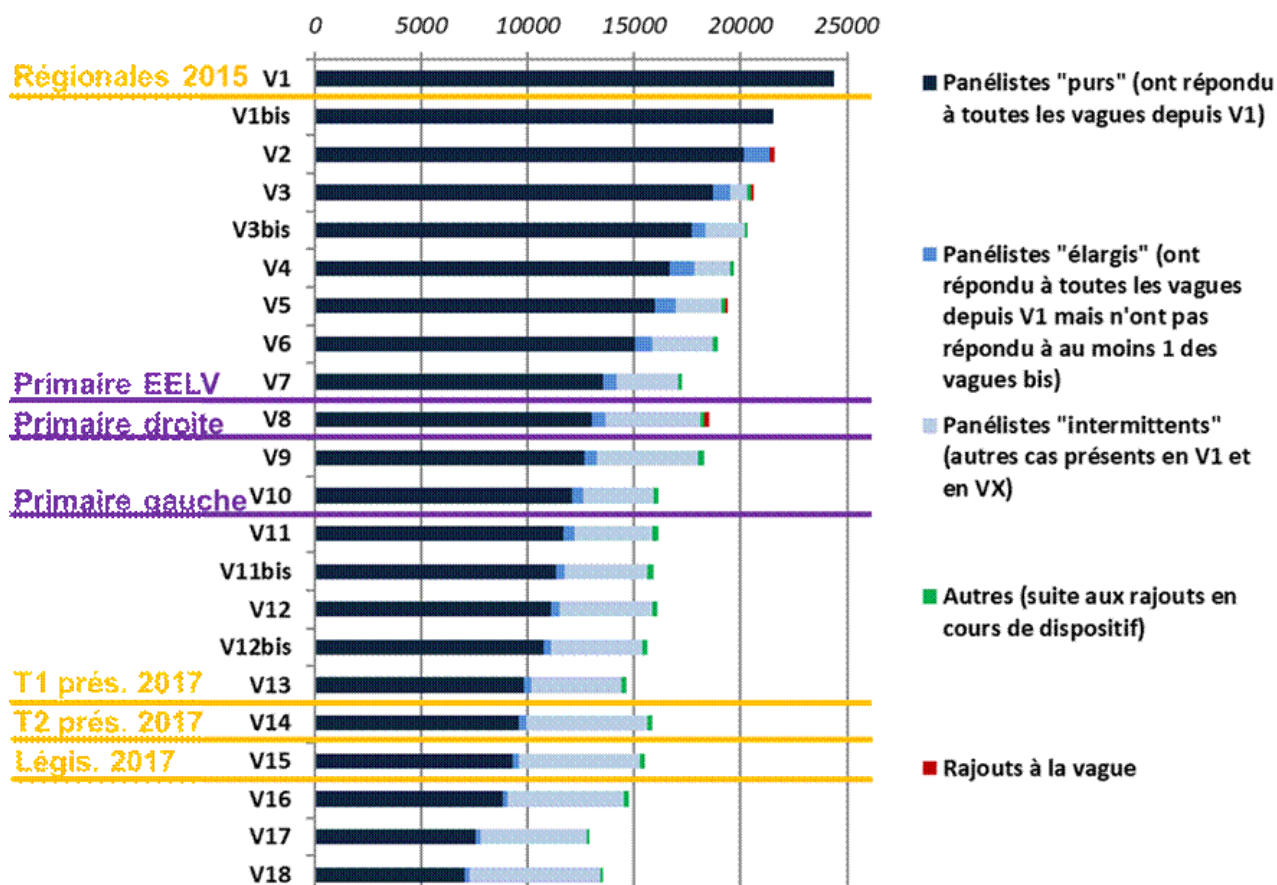
Le panel disposait également de deux autres règles de conception importantes qu'il faut prendre en compte pour l'analyse des résultats : des ajouts de nouveaux panélistes ciblés ont eu lieu dans certaines vagues ; les panélistes pouvaient quitter temporairement le panel et revenir plus tard répondre à des vagues ultérieures. La composition de l'échantillon a donc évolué au fil des vagues, intégrant des rafraîchissements ponctuels pour pallier les pertes liées à l'attrition et assurer la continuité des analyses longitudinales.

## 1.2. Attrition du panel

Dans une enquête par panel, l'attrition désigne la diminution progressive du nombre de participants au fil du temps, en raison de divers facteurs comme le désintérêt, le manque de disponibilité, des changements de situation ou une difficulté à les recontacter. Ce phénomène peut entraîner des biais si les personnes qui quittent l'étude ont des caractéristiques systématiquement différentes de celles qui restent, compromettant ainsi la représentativité et la validité des résultats. Le suivi de l'attrition et la stratégie mise en place pour pallier ses effets constituent un défi fondamental pour les enquêtes longitudinales. L'expérience du panel Elipss montre ainsi qu'une stratégie possible, non utilisée dans le cadre de l'ENEF 2017, consiste en des relances téléphoniques (Palat, Elie, et al., 2023). Pour décrire dans le détail le phénomène d'attrition dans ENEF 2017, nous avons catégorisé les panélistes selon leur degré de participation à l'enquête, et ce pour chacune des vagues :

- Panélistes “purs” : répondants ayant participé à toutes les vagues, y compris les vagues bis, jusqu'à la vague considérée
- Panélistes “élargis” : répondants ayant participé à toutes les vagues hors vagues bis, mais n'ayant pas répondu à au moins une des vagues bis, jusqu'à la vague considérée
- Panélistes “intermittents” : autres cas de répondants, en dehors des panélistes “purs” et “élargis”, présents en vague 1 et à la vague considérée
- Rajouts à la vague : répondants ajoutés à la vague considérée
- Autres cas : répondants issus des ajouts précédents

Figure 1 : Attrition du panel de l'enquête ENEF 2017 au fil des vagues, données non pondérées



Les ajouts réalisés en cours de dispositif concernent exclusivement les primo-votants, dont l'échantillon initial a connu une attrition plus rapide. L'attrition des jeunes répondants soulève des problématiques spécifiques : la petite taille de cet échantillon secondaire et ses caractéristiques sociales - composé de jeunes plus volatiles sur le long terme et donc particulièrement sensibles à la contrainte de répondre régulièrement à un panel électoral - imposent un rafraîchissement ciblé de l'échantillon des primo-votants au fil du temps. La littérature montre par ailleurs que certaines variables influencent plus fortement l'attrition dans ce public spécifique, comme le contexte socio-culturel ou l'origine migratoire (voir Malschinger, Vogl et al., 2023).

En vague 18, la dernière de l'enquête, l'échantillon comprenait au total encore 13540 participants, soit 55,5 % de l'échantillon initial, témoignant de la solidité du dispositif après deux ans et demi de terrain et ce malgré les inévitables déperditions.

Quels sont les effets de l'attrition sur la structure géographique et sociodémographique du panel ? Deux types de biais ont été identifiés. Le premier concerne les biais présents dès le

recrutement initial, sur lesquels un contrôle a été exercé pour éviter que l'attrition ne les accentue. Le second correspond aux biais apparus au fil du panel, c'est-à-dire générés par l'attrition elle-même. Une attention particulière a permis de les contenir grâce à diverses mesures, telles que le rafraîchissement ciblé de l'échantillon, les relances adaptées et le suivi des répondants à risque de décrochement.

Les tableaux 2 et 3 présentent l'évolution des structures géographiques et sociodémographiques des panélistes "purs" de la vague 1 à la vague 18. En grisé sont représentés les écarts supérieurs ou égaux à 5 en valeur absolue entre la proportion parmi les panélistes "purs" dans la vague considérée et la proportion dans la cible telle que fournie par les données de recensement.

Tableau 2 : Attrition et critères géographiques, base panélistes "purs", données non pondérées

	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	Ecart	
	V1 - cible	V2 - cible	V3 - cible	V4 - cible	V5 - cible	V6 - cible	V7 - cible	V8 - cible	V9 - cible	V10 - cible	V11 - cible	V11b - cible	V12 - cible	V12b - cible	V13 - cible	V14 - cible	V15 - cible	V16 - cible	V17 - cible	V18 - cible	
Zones d'habitat	Zone d'habitat	0,2	0,32	0,91	0,12	0,02	0,55	0,52	0,21	0,17	0,12	0,03	0,32	0,24	0,01	0,02	0,21	0,11	0,2	0,05	0,07
	Zone d'habitat dense	0,1	0,14	0,35	0,31	0,30	0,21	0,23	0,21	0,25	0,25	0,2	0,21	0,34	0,21	0,27	0,27	0,19	0,31	0	0,05
	Zone d'habitat dispersé	0,1	0,16	0,16	0,01	0,01	0,06	0,19	0,04	0,12	0,17	0,12	0,11	0,12	0,16	0,11	0,11	0,11	0,12	0,16	0,04
	Zone d'habitat rural	0,2	0,16	0,14	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
	Zone d'habitat très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17
Zone d'habitat très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très urbain	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très très dense	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très très dispersé	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	
Zone d'habitat très très très très très très très très très très très très rural	0,0	0,16	0,16	0,17	0,14	0,08	0,08	0,11	0,14	0,14	0,14	0,12	0,12	0,17	0,17	0,17	0,17	0,17	0,18	0,17	

	Ecart V1-cible	Ecart V2-cible	Ecart V3-cible	Ecart V4-cible	Ecart V5-cible	Ecart V6-cible	Ecart V7-cible	Ecart V8-cible	Ecart V9-cible	Ecart V10-cible	Ecart V11-cible	Ecart V11b-cible	Ecart V12-cible	Ecart V12b-cible	Ecart V13-cible	Ecart V14-cible	Ecart V15-cible	Ecart V16-cible	Ecart V17-cible	Ecart V18-cible		
Société	Homme/20-24 ans	0,00	0,00	0,00	-1,08	-1,08	0,00	-0,38	-0,38	0,00	-1,00	1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	
	Homme/25-34 ans	1,00	0,00	0,00	-2,00	-2,00	0,00	-1,00	-1,00	-2,00	0,00	0,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	
	Homme/35-44 ans	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	Homme/45-54 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	Homme/55-64 ans	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	
	Homme/65-74 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Homme/75-84 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Homme/85-94 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Homme/95-104 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Homme/105-114 ans	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Niveau de diplôme	Apprentissage/18-24 ans	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	-2,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	18C.608854-4-2008	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
Niveau de revenu	19.1	-19,01	-19,04	-19,03	-19,09	-19,1	-19,13	-19,16	-19,17	-19,12	-19,12	-19,09	-19,1	-19,09	-19,13	-19,14	-19,09	-19,05	-19,09	-19,14	-19,14	
	19.2	-7,24	-7,35	-7,4	-7,37	-7,48	-7,91	-7,68	-7,51	-7,46	-7,35	-7,28	-7,31	-7,18	-7,53	-7,58	-7,73	-7,81	-7,46	-7,15		
	19.3	1,12	1,10	1,12	1,12	1,1	1,12	1,11	1,09	1,08	1,07	1,05	1,05	1,05	1,02	1,02	1,02	1,02	1,02	1,02		
	19.4	24,7	24,97	25,28	25,3	25,37	25,43	25,94	25,7	25,5	25,4	25,18	25,16	25,23	25,18	25,67	25,61	25,67	25,54	25,45		
	19.5																					
	19.6																					
	19.7																					
	19.8																					
	19.9																					
	19.10																					

Pour les données sociodémographiques, le croisement genre x âge ne présente pas de biais de départ et sa distribution reste très bonne au fil des vagues. La répartition des professions de la personne de référence présente différents biais. Le premier apparaît dès la première vague en 2015 : les ouvriers et les retraités sont sous-représentés dès le début du dispositif, avec un biais relativement limité pour les ouvriers mais plus fort pour les retraités. Si celui des ouvriers s'accroît un peu avec l'attrition du panel, celui des retraités s'améliore pour être maîtrisé aux alentours de la vague 10. Le second biais pour la profession concerne les cadres supérieurs et apparaît en cours de dispositif, avec une surreprésentation qui apparaît en vague 7 et se résorbe en fin de panel.

Enfin le niveau de diplôme, qui n'est pas une variable de quotas, présente un biais très fort dès le démarrage de l'enquête, avec une sous-représentation des moins diplômés et une surreprésentation des diplômés de l'enseignement supérieur qui reste stable au cours du temps. Ce biais, très important et cohérent avec l'absence de quota de niveau de diplôme, peut être partiellement géré grâce à la grande taille de l'échantillon, qui limite l'impact de ce biais sur les analyses. Cependant, il ne faut pas oublier que la taille d'échantillon seule ne suffit pas à corriger tous les biais : il reste essentiel de rester vigilant quant à la représentativité, même dans de grands panels (Kaplan, Chambers et Glasgow, 2014).

Les autres biais présentés montrent que l'attrition au sens strict affecte seulement à la marge la qualité de l'échantillon de départ et qu'un travail d'analyse uniquement sur les panélistes "purs" est tout à fait possible.

**1.3. Contenu des questionnaires et exemple d'expérimentation**

ENEF 2017 se distingue aussi par la richesse de ses contenus, avec plus de 50 modules thématiques différents explorant des dimensions variées : des sujets directement liés aux élections (primaires, intentions ou reconstitutions de vote, engagement politique, vote stratégique, etc.) à d'autres liés à des sujets de recherche spécifiques (génétique, relations de travail, aides au développement, etc.) en passant par les modules fondamentaux ayant trait aux attitudes et comportements politiques ou encore au populisme ou à l'Europe, sans oublier les

variables techniques (pondérations, identifiant, etc.)<sup>[3]</sup>. Si l'essentiel des questionnaires ont été conçus par les électoralistes du CEVIPOF, une partie des modules thématiques est issue d'appels à proposition et des équipes de recherche externes au CEVIPOF ont ainsi pu participer à la rédaction des questionnaires, constitués à la fois de modules barométriques posés dans plusieurs vagues et d'autres posés une unique fois.

L'enquête dispose, notamment dans la première vague d'enquête, de variables sociodémographiques beaucoup plus détaillées que dans les enquêtes électorales françaises habituelles. On y trouve par exemple tout un module sur la composition détaillée du foyer, avec l'âge des enfants, la notion d'enfant vivant ou non au foyer, issue du couple actuel ou d'un précédent, qu'il s'agisse de ses propres enfants ou de ceux de son ou sa conjoint(e). Ces informations complexes adossées au nombre de personnes vivant au foyer et aux âges de chacun permettent de reconstituer le revenu par unité de consommation de manière beaucoup plus riche, ce qui était un besoin de l'équipe de recherche de l'enquête. Les questionnaires des vagues, avec les formulations détaillées de ces variables, sont accessibles sur la Banque de données du CDSP.

Des indicateurs politiques et électoraux sont bien entendu présents tout au long des vagues, qu'il s'agisse des mesures d'intentions ou de reconstitutions de vote, avec un aspect évolutif en fonction de l'avancée dans le calendrier électoral ou des modifications de l'offre à l'approche des scrutins ; des mesures d'opinion sur tel ou tel sujet, qu'il s'agisse par exemple d'attitudes au sens d'Alain Lancelot (Lancelot, 1985) ou de positionnement sur des éléments politiques contextuels.

Un point relativement nouveau dans les enquêtes électorales françaises est le développement de nombreuses expérimentations tout au long du dispositif. Les vagues d'enquête ont intégré ces dispositifs expérimentaux pour tester des hypothèses sur les comportements électoraux et en particulier sur deux thématiques : le vote stratégique ; l'évaluation de l'exécutif et son usage du 49.3. Ces expérimentations ont permis de produire des données originales, particulièrement précieuses pour comprendre les déterminants des choix électoraux et les dynamiques d'opinion. Un des premiers exemples d'expérimentation propose des scénarios électoraux hypothétiques pour analyser les réactions des panélistes face à une configuration de triangulaire entre le Parti Socialiste, Les Républicains et le Front National (devenu le Rassemblement National par la suite) dans le cadre d'un scrutin législatif et évaluer la réaction de ces mêmes panélistes sur le second tour du même scrutin en fonction des résultats hypothétiques du premier tour. Cette expérimentation vise donc à analyser à la fois les préférences de vote initiales des électeurs et leur comportement stratégique au second tour.

Figure 2 : Formulation de la première partie de l'expérimentation, vague 1

TO ALL

EXP1

Imaginez qu'une élection ait lieu dans votre circonscription où seuls trois candidats sont en concurrence au premier tour : un candidat du Parti Socialiste, un candidat Les Républicains (ex-UMP), un candidat du Front National.

Note : Un candidat rassemblant plus de 50% des voix est élu dès le 1er tour. Seuls les candidats rassemblant plus de 10% des voix sont qualifiés pour le second tour.

Dans cette hypothèse, quel serait votre choix au premier tour ?

ONE ANSWER POSSIBLE

1. Je vote pour le candidat du Parti Socialiste
2. Je vote pour le candidat Les Républicains
3. Je vote pour le candidat du Front National
4. Je vote blanc ou nul
5. Je m'abstiens / je ne participe pas

Figure 3 : Formulation de la seconde partie de l'expérimentation, vague 1

TO ALL

EXP2

A l'issue du premier tour, chacun des candidats a recueilli suffisamment de voix pour se maintenir au second tour. Le tableau ci-dessous vous donne le score pour chacun des candidats au 1er tour.

PLEASE INSERT ONE HYPOTESIS (RANDOM) FROM EXCEL DOCUMENT UNTITLED « HYPOTHESES RESULTATS EXPERIENCE »

	Résultats 1er tour
Candidat du Parti Socialiste	
Candidat Les Républicains (ex-UMP)	
Candidat du Front National	

Note : Seul le candidat arrivé en tête à l'issue du second tour sera élu

Dans cette situation, quel serait votre choix au SECOND TOUR ?

ONE ANSWER POSSIBLE

1. Je vote pour le candidat du Parti Socialiste
2. Je vote pour le candidat Les Républicains
3. Je vote pour le candidat du Front National
4. Je vote blanc ou nul
5. Je m'abstiens / je ne participe pas

Les résultats sur une triangulaire au premier tour de cette élection législative hypothétique montrent trois choses : les votes sont répartis de manière assez équilibrée entre les candidats des trois formations politiques avec un léger avantage au candidat identifié au Parti Socialiste ; la déclaration d'un potentiel vote blanc ou nul est importante ; ce type d'enquête mesure mal l'abstention potentielle.

Tableau 4 : Résultats pour la première partie de l'expérimentation, vague 1, données pondérées par la pondération "décision" combinant redressement sociodémographique et politique (code en STATA)

```
. tab AEXP1 [iweight=APPUB]
```

EXP1. Imaginez qu'une élection ait lieu dans votre circonscription où seuls 3can	Freq.	Percent	Cum.
Je vote pour le candidat du PS	6,781.048	28.09	28.09
Je vote pour le candidat LR	6,196.2597	25.67	53.76
Je vote pour le candidat du FN	6,063.9908	25.12	78.88
Je vote blanc ou nul	3,895.0851	16.14	95.02
Je m'abstiens / je ne participe pas	1,202.6164	4.98	100.00
Total	24,139	100.00	

Le tableau 5 présente ensuite un extrait des différentes hypothèses présentées aux répondants. Elles font varier les votes exprimés sur ce premier tour hypothétique et par conséquent le rapport de force entre les trois candidats : de 10% à 49% pour chaque candidature, avec toutes les combinaisons possibles aboutissant à un total valant 100% aux exprimés.

Tableau 5 : Extrait de la distribution des hypothèses de résultats du premier tour affichées avant la mesure de l'intention de vote potentielle sur le second tour en découlant, vague 1, données pondérées par la pondération "décision" combinant redressement sociodémographique et politique (code en STATA)

. tab AEXP2HYP [iweight=APPUB]

EXP2 - Hypothèse affichée	Freq.	Percent	Cum.
10-41-49	24.2226959	0.10	0.10
10-42-48	32.1944737	0.13	0.23
10-43-47	15.728063	0.07	0.30
10-44-46	19.5001687	0.08	0.38
10-45-45	20.5159843	0.08	0.46
10-46-44	23.2386365	0.10	0.56
10-47-43	22.4964916	0.09	0.65
10-48-42	22.4157774	0.09	0.75
10-49-41	19.7273504	0.08	0.83
11-40-49	23.9839156	0.10	0.93
11-41-48	24.5280439	0.10	1.03
11-42-47	29.3619993	0.12	1.15
11-43-46	22.7654365	0.09	1.25
11-44-45	24.9559789	0.10	1.35
11-45-44	24.4423872	0.10	1.45
11-46-43	20.5411891	0.09	1.54
11-47-42	19.402328	0.08	1.62
11-48-41	23.1646993	0.10	1.71
11-49-40	22.9787895	0.10	1.81
12-39-49	25.4115073	0.11	1.91
12-40-48	24.7828396	0.10	2.01
12-41-47	19.2200986	0.08	2.09
12-42-46	20.2264895	0.08	2.18

Enfin le tableau 6 fournit les résultats de la deuxième question portant sur l'expression d'un vote hypothétique de second tour, exprimé après s'être positionné sur un premier tour hypothétique puis visualisé l'hypothèse de résultats sur ce premier tour hypothétique. Il est aussi clairement indiqué que ce second tour intervient dans un cadre de maintien de la même triangulaire au second tour. Les résultats montrent une répartition différente entre les trois candidats : + 1 point pour le candidat PS ; + 3,5 points pour le candidat LR ; - 2 points pour le candidat FN. Les électeurs potentiels ajustent ici leur vote en fonction du résultat hypothétique présenté pour le premier tour. Le vote stratégique intervient plus favorablement au profit du candidat LR que du candidat PS et s'inscrit dans une dynamique de "front républicain" pour contrer la victoire possible du FN.

Tableau 6 : Résultats pour la deuxième partie de l'expérimentation, vague 1, données pondérées par la pondération "décision" combinant redressement sociodémographique et politique (code en STATA)

```
. tab AEXP2 [iweight=APPUB]
```

EXP2. A l'issue du 1er tour, chacun des candidats a recueilli suffisamment de vo	Freq.	Percent	Cum.
Je vote pour le candidat du PS	7,064.7367	29.27	29.27
Je vote pour le candidat LR	7,048.2314	29.20	58.47
Je vote pour le candidat du FN	5,541.3006	22.96	81.42
Je vote blanc ou nul	3,414.3232	14.14	95.57
Je m'abstiens / je ne participe pas	1,070.4081	4.43	100.00
Total	24,139	100.00	

## 2. Rendre les données compréhensibles et accessibles

### 2.1. Traitement des données

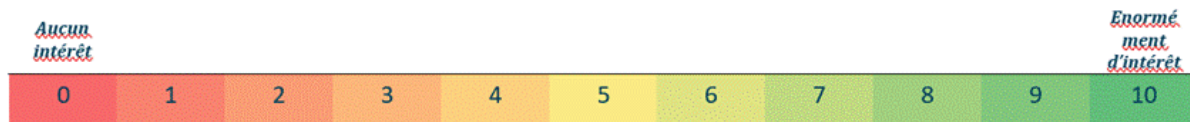
Nous avons appliqué plusieurs étapes de traitement à la base de données d'origine afin de la rendre plus exploitable. La base initiale était cumulative et comportait de nombreuses variables recodées, ce qui alourdissait considérablement le fichier. Au final, il y avait au total 2864 variables dans la base de données. Un travail de nettoyage et d'apurement des données a donc été indispensable avant toute documentation. Nous avons segmenté les fichiers selon les vagues de l'enquête et sélectionné les variables pertinentes pour la diffusion en vue d'analyses secondaires. Concernant le critère de sélection, nous avons gardé uniquement des variables d'origine et supprimé des variables recodées. Par exemple, les questions aux réponses en échelle, telles que la question Q0, ont été automatiquement recodées en sous-totaux, comme illustré à la figure 4. Ces variables recodées en "sous-totaux" ne figurent pas dans la base de données diffusée pour ne pas alourdir le fichier. De même, certaines variables sociodémographiques du questionnaire ne sont pas incluses, ou seules leurs versions recodées sont présentes. Par exemple, dans la vague 1, le pays de naissance n'a été demandé qu'aux personnes nées hors de France métropolitaine ou d'outre-mer (figure 5). Dans cette même vague, les questions sur le lieu de résidence principale étaient posées au niveau du département et de la commune (figure 6). La variable relative au pays de naissance n'est pas présente dans la base diffusée, tandis que les deux dernières sont disponibles sous forme recodée par région. Ces choix visent à préserver l'anonymat des répondants tout en offrant aux utilisateurs potentiels l'information la plus détaillée possible.

Figure 4 : Question Q0 issu de questionnaire ENEF2017 de vague 2

Q0

Sur une échelle de 0 à 10, où 0 signifie aucun intérêt et 10 signifie énormément d'intérêt, quel est votre niveau d'intérêt pour la prochaine élection présidentielle de 2017 ?

**ONLY 1 ANSWER FROM 0 TO 10**



+ 99. (NR)

Ajout de Q0ST

1. ST Pas intéressé (0 à 3)
2. ST Ni intéressé, ni pas intéressé (4 à 6)
3. ST Intéressé (7 à 10)
4. (NR)

Figure 5 : Question NAIS3 issu de questionnaire de vague 1 de ENEF2017

NAIS3

Dans quel pays êtes-vous né(e)?

**ONE ANSWER POSSIBLE**

Une seule réponse possible

**DISPLAY COUNTRY LIST**

Figure 6 : Question LOC1 et LOC2 issu de questionnaire de vague 1 de ENEF2017

## LOC1 COMPULSORY QUESTION

**Veillez indiquer le département de votre résidence principale :**

**ONE ANSWER POSSIBLE**

*Une seule réponse possible*

**SINGLE PUNCH DROPDOWN PREQUAL**

**DISPLAY FRDEPARTMENT LIST**

**TO ALL**

DOCUMENT CONFIDENTIEL

Page 4 sur 38

SOS

Enquête électorale

15-050190-01

## LOC2 COMPULSORY QUESTION

**Veillez indiquer la commune de votre résidence principale :**

**ONE ANSWER POSSIBLE**

*Une seule réponse possible*

**SINGLE PUNCH DROPDOWN PREQUAL**

**DISPLAY FRMUNICIPALITY LIST**

Dans la banque de données du CDSP (BDSP), les données diffusées suivent un format standardisé, également appliqué aux bases de l'ENEF 2017. Les non-réponses sont classées en trois catégories distinctes : 9999 pour les personnes qui n'ont pas répondu ; 6666 pour les « non concerné-e-s », lorsque la question ne s'appliquait pas à une partie des répondants en raison d'un filtre ; et 9996 pour les individus absents à la vague concernée bien qu'ils fassent partie du panel, c'est-à-dire ceux ayant participé à une ou plusieurs vagues précédentes ou ultérieures mais non à celle du fichier.

Les variables du fichier de données sont également réorganisées selon un ordre précis. L'identifiant individuel figure en première position, suivi des variables d'enquête correspondant aux questions posées dans les questionnaires. Viennent ensuite les variables sociodémographiques relatives au profil des répondants. Les variables techniques, notamment celles liées à la pondération, sont placées en fin de fichier.

## 2.2. Documentation des données

Une fois la base de données finalisée, nous l'avons documentée selon le standard DDI, largement utilisé pour les enquêtes en sciences sociales. Le DDI (Data Documentation Initiative) permet de documenter et gérer différentes étapes du cycle de vie des données de recherche, telles que la conception, la collecte, le traitement, la diffusion, la découverte et l'archivage (Danciu & Mairot, 2019). Parmi les différents produits DDI, nous avons retenu le DDI codebook, comme pour d'autres enquêtes disponibles sur la BDSP, pour une documentation descriptive enquête par enquête. La documentation s'effectue principalement à deux niveaux : celui de l'enquête et celui de la variable. Des informations générales telles que l'auteur, la période de

collecte, le sujet ainsi que la méthodologie sont renseignées au niveau de l'enquête. Au niveau de la variable, chaque variable est décrite individuellement, avec la question posée, les filtres éventuels et les notes explicatives, notamment pour détailler l'origine des variables techniques ou recodées.

DDI est disponible dans un schème XML mais pour éviter d'écrire la documentation en XML, nous avons procédé en trois étapes différentes (Danciu & Sauger, 2022). Pour la partie documentation de l'enquête, nous la saisissons directement sur la BDSP, en remplissant le champ librement ou en choisissant parmi les vocabulaires contrôlés (Figure 7). Pour la deuxième étape, le premier travail de la documentation au niveau variable se fait sur R, en même temps que l'apurement (Figure 8). Il s'agit principalement de corriger et ordonner les modalités de la variable en adressant un label pour chaque modalité. Ensuite, nous utilisons Nesstar pour finir la documentation qui permet de générer le fichier XML et le codebook (Figure 9).

Figure 7 : Study documentation dans le BDSP de la vague 10 de l'ENEF 2017

Dans cette partie, nous documentons l'enquête en général : sujet, information concernant l'échantillon et échantillonnage, méthode de collecte, etc.

The screenshot shows the 'SciencesPo' metadata form for the 2017 ENEF survey. The form is titled 'Métadonnées sur les sciences sociales et les sciences humaines' and includes the following fields:

- Unité d'analyse**: Individu
- Univers**: Électeurs en France métropolitaine: - ayant 18 ans et plus inscrits sur les listes électorales et juridiquement éligibles à voter lors des élections régionales de 2015. ; - lisant suffisamment le français pour répondre à un questionnaire auto-administré.
- Méthode temporelle**: Longitudinale : panel
- Responsable de la collecte de données**: Nom, prénom ou nom de l'organisation
- Formation du responsable de la collecte de données**: (empty field)
- Fréquence**: Les enquêtes ENEF sont généralement bimestrielles.
- Méthode d'échantillonnage**: Non probabiliste : quota
- Taille de l'échantillon cible**: Réelle (Entrez un nombre entier...) and Formule (empty field)
- Écarts importants pour le plan d'échantillonnage**: (empty field)
- Mode de collecte**: Questionnaire auto-administré : par Internet (CAWI)
- Type d'instrument de recherche**: Questionnaire structuré
- Caractéristiques de la collecte de données**: Durée du questionnaire estimée à 20 minutes. Les panélistes n'ayant pas répondu ou ayant commencé le questionnaire ont reçu des relances deux jours après le démarrage du terrain d'enquête par courrier électronique.

Figure 8 : Extrait du code de documentation de V1 de l'ENEF 2017

Cet extrait est issu d'un script écrit sur markdown ayant pour but de nettoyer et documenter la première vague de l'ENEF 2017.

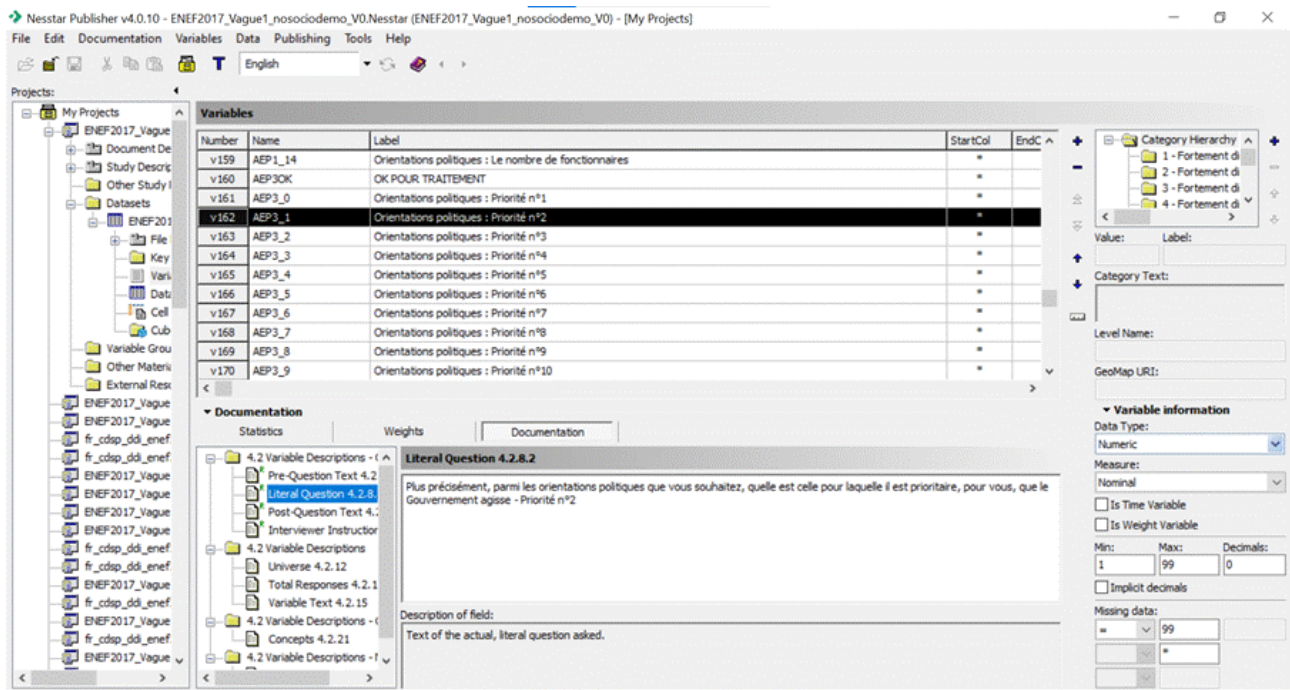
```
##
v1$AR20R11_0[v1$AREG13BIS != 11 | v1$AR19 != 2 ] <- 6666
v1$AR20R11_0[v1$AR20R11_0 == 99] <- 9999
v1$AR20R11_0 %<>%
  add_value_labels(c("Non concerné(e)" = 6666, "N'a pas répondu" = 9999))

kable(
questionr::freq(v1$AR20R11_0, digits=2, na.last = FALSE, total = TRUE)
) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),font_size = 12)%>%
  scroll_box(width = "500px", height = "200px")
```

	n	%	val%
[1] LO - N. Arthaud	51	0.21	0.21
[2]	0	0.00	0.00
[3]	0	0.00	0.00
[4]	0	0.00	0.00
[5] FG - P. Laurent	110	0.45	0.45
[6]	0	0.00	0.00

Figure 9 : Capture d'écran de la documentation des variables de la première vague de l'ENEF 2017

Comme indiqué précédemment, nous utilisons un logiciel Nesstar pour documenter chaque variable de la base de données.



### 2.3. Présentation de jeu de données disponibles

Nous produisons le fichier final après avoir passé ces trois étapes. Le jeu de données comprend

au total 23 vagues, y compris des vagues « bis » identifiées par l'ajout de la lettre « b » accolée au numéro de la vague. Au cours de l'enquête, certains participants se retirent donc le rafraîchissement du panel est réalisé comme décrit dans la partie 1.2 (Tableau 7).

Tableau 7. Description de fichier de données ENEF 2017

	<b>Nombre de lignes</b>	<b>Nombre de variables</b>	<b>Individus enquêtés</b>	<b>Nouvelles entrants</b>
V1	24369	294	24369	24369
V1b	24369	85	21555	
V2	24592	122	21574	223
V3	24730	168	20619	138
V3b	24730	86	20346	
V4	24730	197	19692	
V5	24827	136	19383	97
V6	24827	151	18909	
V7	24827	287	17268	
V8	25028	166	18513	201
V8b	25028	91	9433	
V9	25028	195	18273	
V10	25028	181	16166	
V11	25028	186	16124	
V11b	25028	172	15887	
V12	25028	145	16074	
V12b	25028	204	15623	
V13	25028	466	14655	
V14	25028	147	15807	
V15	25028	218	15493	
V16	25028	217	14748	

V17	25028	304	12875	
V18	25028	193	13540	

Chaque fichier de données débute par la variable « ID12 », correspondant à l'identifiant individuel. À l'exception de cette variable, les autres variables commencent par une lettre servant de préfixe pour indiquer la vague. ID12 est suivie par les variables issues du questionnaire de chaque vague d'enquête. Les variables d'enquête sont ensuite suivies systématiquement par les variables sociodémographiques, de « AGE\_r1 » à « ANAT4\_10 »<sup>[4]</sup>. Ces variables proviennent de la première vague ou de l'entrée dans le panel. Les modifications éventuelles de la situation peuvent être enregistrées dans la vague 3 ou la vague 18. Enfin, les variables techniques<sup>[5]</sup> sont placées à la fin du fichier.

## 2.4. Accès aux données

La BDSP, ayant obtenu la certification CoreTrustSeal en 2023 ([La Banque de données du CDSP, s. d.](#)), diffuse plus de 400 références en SHS auprès de la communauté académique sous licence CC BY-SA 4.0<sup>[6]</sup>. Elle propose plusieurs collections, notamment la collection Elipss et la collection « données de la recherche », qui inclut les données de l'ENEF. Ces données sont téléchargeables, soit en accès libre soit en accès restreint, après avoir accepté des conditions générales d'utilisation.



### Collection “Enquête électorale française 2017” déposée à la Banque de données du CDSP<sup>[7]</sup>

- vague 1 : doi:10.21410/7E4/N7HZLA
- vague 1bis : doi:10.21410/7E4/HEAXDG
- vague 2 : doi:10.21410/7E4/SPRKNV
- vague 3 : doi:10.21410/7E4/O2MURS
- vague 3bis : doi:10.21410/7E4/BZOEZN
- vague 4 : doi:10.21410/7E4/QZTEHO
- vague 5 : doi:10.21410/7E4/Q1IYIZ
- vague 6 : doi:10.21410/7E4/DQFYRV

- vague 7 : doi:10.21410/7E4/KW39HC
- vague 8 : doi:10.21410/7E4/YQMIY7
- vague 8bis : doi:10.21410/7E4/QL5I5H
- vague 9 : doi:10.21410/7E4/BU53Z6
- vague 10 : doi:10.21410/7E4/2OHZ4R
- vague 11 : doi:10.21410/7E4/JNDKUR
- vague 11bis : doi:10.21410/7E4/HYPAD0
- vague 12 : doi:10.21410/7E4/OUHSSH
- vague 12bis : doi:10.21410/7E4/A2TTY4
- vague 13 : doi:10.21410/7E4/SGT2SG
- vague 14 : doi:10.21410/7E4/JY3NYN
- vague 15 : doi:10.21410/7E4/TGKI82
- vague 16 : doi:10.21410/7E4/HTHN8G
- vague 17 : doi:10.21410/7E4/VXJKXG
- vague 18 : doi:10.21410/7E4/T85VW3



Chaque vague de l'ENEF 2017 possède son propre DOI. Chaque jeu de données est diffusé sous quatre formats : dta, sav, SAS, ainsi qu'un format csv accessible à tous les utilisateurs. Les réponses aux questions ouvertes sont disponibles dans un fichier distinct intitulé « \_verbatim » le cas échéant (Figure 10). Les fichiers de documentation sont fournis par jeu de données, incluant un codebook au format pdf généré par Nesstar, un dictionnaire des codes, le questionnaire de l'enquête et un rapport technique présentant la distribution selon le redressement et l'échantillonnage. La documentation DDI est également disponible au format XML afin de garantir l'interopérabilité avec d'autres standards et la réutilisabilité. (Figure 11)

Figure 10 : Fichiers de données de vague 1

 **fr\_cdsp\_ddi\_enef2017\_v1-1.tab**  
Données tabulaires - 36.9 Mo  
Publié 30 oct. 2024  
7 téléchargements  
294 Variables, 24369 Observations UNF:6:H+Me...qWg== 



**Données** **STATA**

---

 **fr\_cdsp\_ddi\_enef2017\_v1.tab**  
Données tabulaires - 23.7 Mo  
Publié 30 oct. 2024  
6 téléchargements  
294 Variables, 24369 Observations UNF:6:H+Me...qWg== 



**Données** **SPSS**

---

 **fr\_cdsp\_ddi\_enef2017\_v1\_csv**  
Archive ZIP - 4.3 Mo  
Publié 30 oct. 2024  
8 téléchargements  
MD5: fe9...360 



**CSV** **Données**

---

 **fr\_cdsp\_ddi\_enef2017\_v1\_sas**  
Archive ZIP - 5.6 Mo  
Publié 30 oct. 2024  
5 téléchargements  
MD5: fb1...945 











**Données** **SAS**

---

 **fr\_cdsp\_ddi\_enef2017\_v1\_verbatims.zip**  
Archive ZIP - 275.2 Ko  
Publié 30 oct. 2024  
5 téléchargements  
MD5: 517...7e9 

**CSV** **Données**

Figure 11 : Fichiers de documentation de vague 1

	<p><b>Documentation_ENEF2017_V1.pdf</b>          Adobe PDF - 1.0 Mo          Publié 30 oct. 2024          38 téléchargements          MD5: cbf...c82 </p> <p><b>Documentation</b></p>
	<p><b>ENEF2017_V1aV18_Dictionnaire.zip</b>          Archive ZIP - 123.2 Ko          Publié 6 nov. 2024          22 téléchargements          MD5: 130...935 </p> <p>Dictionnaire variables</p> <p><b>Documentation</b></p>
	<p><b>fr_cdsp_ddi_enef2017_v1.xml</b>          XML - 1.0 Mo          Publié 30 oct. 2024          17 téléchargements          MD5: a67...63c </p> <p><b>Documentation XML</b></p>
	<p><b>Questionnaire_ENEF2017_V1.docx</b>          MS Word - 197.8 Ko          Publié 30 oct. 2024          25 téléchargements          MD5: bf5...899 </p> <p><b>Documentation Questionnaire</b></p>
	<p><b>RapportTechnique_ENEF2017_V1.pdf</b>          Adobe PDF - 746.9 Ko          Publié 23 oct. 2024          25 téléchargements          MD5: 9a5...0c0 </p> <p><b>Documentation</b></p>

## 2.5. Valorisation des données

À la suite de l'élection présidentielle de 2017, ont eu lieu les élections européennes de 2019 et les élections municipales de 2020. Le CEVIPOF et ses partenaires ont ainsi mené l'enquête ENEF 2019 portant sur ces deux scrutins <sup>[8]</sup>. Une partie des panélistes de 2017 a été conservée dans l'enquête de 2019, conservant le même identifiant. Cette continuité permet aux utilisateurs secondaires d'analyser les évolutions et comportements électoraux sur le long terme, de 2015 à 2020, en tenant compte de la répétition de certaines questions ou modules depuis le début de l'enquête en 2017, notamment le choix du vote pour chaque élection ou le degré de satisfaction envers le président .

Le CDSP a également publié un jeu de données apparié de l'ENEF 2017 en incluant toutes les vagues à l'exception de la vague 8bis (Centre de données socio-politiques (CDSP), 2025) <sup>[9]</sup>. Ce

Le jeu de données comprend 203 variables qui sont répétées au moins une fois tout au long de l'enquête. Cette publication vise à faciliter les analyses de séries temporelles de l'ENEF 2017. Le dictionnaire des codes et la documentation ainsi que le fichier de données sont disponibles au BDSP qui permet de mieux comprendre la construction et l'utilisation de cette base.

La bibliographie, non exhaustive, présente différentes utilisations des données, et est représentative de la richesse des usages possibles de ces données : l'étude d'indicateurs historiques de filiation politique (Muxel, 2018) ; des enjeux contemporains autour de la notion de populisme (Ivaldi, Akkerman, et al., 2017) ; le contexte particulier des outre-mer (Rafidinarivo, 2017) ; un usage des mixed methods en passant du côté de la théorie politique (Pélabay, Sénac, 2019) ; une autre des nombreuses expérimentations portant cette fois-ci sur l'impact de l'utilisation de l'article 49.3 par le Premier ministre (Becher, Brouard, 2020) ; ou encore une application en psychologie politique (Marcus, Valentino, et al., 2019). De nombreuses applications sont encore possibles en utilisant ces données.

Il est également possible d'utiliser cette enquête dans le cadre de l'enseignement comme ce qui est déjà fait dans l'atelier méthodologique "Méthodes quantitatives en sciences sociales avec R" à Sciences Po. L'enquête couvre une variété de thématique qui permet aux étudiant-e-s de trouver leur champ d'intérêts même pour ceux et celles qui n'intéressent pas à l'élection en tant que telle. Donc, il est plus facile pour qu'ils soient impliqués dans le cours. Le fait que le jeu de données est formaté et harmonisé constitue un autre avantage pour l'enseignement. Ceci permet de diminuer le temps de manipulation des données ou la préparation des données où les étudiants sont le moins accrochés.

## **Conclusion**

L'enquête électorale française de 2017 du CEVIPOF a fait l'objet de riches discussions lors de la journée d'études consacrée à l'ouverture publique de données provenant de différentes enquêtes électorales dont ENEF 2017 (Chanvriil, 2024 ; Jeong et Marie, 2024). Elle illustre l'importance des panels électoraux longitudinaux pour comprendre les dynamiques politiques contemporaines. Son dispositif méthodologique robuste, la rigueur de son traitement et la qualité de sa documentation en font une ressource de référence pour la recherche en sciences sociales.

La diffusion des données via la Banque de données du CDSP, dans le respect des standards internationaux, contribue à la transparence scientifique et à la reproductibilité des analyses. Par sa continuité avec les enquêtes ultérieures, notamment ENEF 2019, ENEF 2022 et ENEF 2023-2024, ce panel offre un cadre unique pour étudier les recompositions électorales et l'évolution du rapport au politique en France sur le temps long. Plusieurs publications récentes portent ainsi sur les évolutions de l'espace politique suite aux derniers scrutins électoraux de 2024 (Cautrès et Muxel, 2025 ; Revue française de science politique, 2025).

## Bibliographie

Becher, M. et Brouard, S. (2022). Executive Accountability Beyond Outcomes: Experimental Evidence on Public Evaluations of Powerful Prime Ministers. *American Journal of Political Science*, 66, 106-122.

<https://doi.org/10.1111/ajps.12558>

<https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12558>

Brouard, S. et Michel, E. (2025). Retrospective voting in times of pandemic: comparative evidence from panel data. *Political Science*, pp.1-25.

<https://www.tandfonline.com/doi/full/10.1080/00323187.2025.2500926>

Cautrès, B. et Muxel, A. (dir.) (2025). *Le vote sans issues. Chroniques électorales 2024*, Presses Universitaires de Grenoble, p.340.

<https://doi.org/10.3917/pug.cautr.2025.01>

Centre de données socio-politiques (CDSP), Sciences Po, CNRS. (2025). *Enquête électorale française 2017 (ENEF 2017), jeu de données apparié (Version V2) [Jeu de données]*. data.sciencespo.

<https://doi.org/10.21410/7E4/FIVFXR>

*La Banque de données du CDSP.* (s. d.). Centre de données socio-politiques. Consulté 22 octobre 2025, à l'adresse <https://www.sciencespo.fr/cdsp/fr/donnees/la-banque-de-donnees-du-cdsp/>

Chanvril, F. (2024). De ENEF 2017 à ENEF 2024 : une enquête riche et innovante. *Journée d'étude Ouverture des données : l'enquête électorale 2017 du CEVIPOF publiée par le CDSP.*

<https://hal.science/hal-04828602>

Charrance, G., Cochet, P., Leduc, A. et Bondon, M. (2022), Apprendre des paradonnées pour améliorer les protocoles de collecte : l'exemple d'EpicoV. *Journée de méthodologie statistique de l'Insee*, Paris.

<http://hdl.handle.net/20.500.12204/AYUuifD2Lg0aT10RuaVQ>

Cohen, S., Mayer, N. et Marx, P. (2023). Advantages and limitations of panel surveys. *Séminaire Les sciences sociales en question : grandes controverses épistémologiques et méthodologiques*, CERI-CEE.

<https://sciencespo.hal.science/hal-04724187/>

Danciu, A. et Mairot, A. (2019). Data Documentation Initiative (DDI), un standard de documentation des données. *Webinaires Tuto Mate*. <https://doi.org/10.5281/zenodo.6590698>

Danciu, A. et Sauger, N. (2022). Workflows after Nesstar. *European DDI User Conference, EDDI 2022, Paris, Sciences Po*.

<https://doi.org/10.5281/zenodo.7406108>

Ivaldi, G., Akkerman, A. et Zaslove, A. (2017). La France populiste ?. *Note ENEF 30, Sciences Po CEVIPOF*.

<https://shs.hal.science/halshs-01491961/>

Jeong, J. et Marie, L. (2024). La Banque de données du CDSP et la valorisation des Enquêtes électorales. *Journée d'étude Ouverture des données : l'enquête électorale 2017 du CEVIPOF publiée par le CDSP*.

<http://sciencespo.hal.science/hal-04834217/>

Kaplan, R.M., Chambers, D.A. et Glasgow, R.E. (2014). Big data and large sample size: a cautionary note on the potential for bias. *Clinical and Translational Science*, Aug;7(4):342-6.

<https://doi.org/10.1111/cts.12178>

Lancelot, A. (1985). L'orientation du comportement politique. Dans M. Grawitz et J. Leca (dir.), *Traité de Science Politique*, Paris, PUF, 1985, tome 3, p.368.

Le Hay, V. (2009). Le panel électoral français 2007 - Enjeux de méthode. Dans B. Cautrès et A. Muxel (dir.), *Comment les électeurs font-ils leur choix ? Le Panel électoral français 2007* (pp. 259-284), Paris : Presses de Sciences Po.

Le Hay, V. (2017). 70 ans d'enquêtes électorales et de révolution méthodologique. Dans Y. Déloye et N. Mayer (dir.), *Analyses électorales* (pp. 117-172), Bruxelles : Éditions Bruylant.

Malschinger, P., Vogl, S. et Schels, B. (2023). Drop in, drop out, or stay on: Patterns and predictors of panel attrition among young people. *Österreich Z Soziol* 48, 427-450 (2023). <https://doi.org/10.1007/s11614-023-00545-z>

Marcus, G. E., Valentino, N. A., Vasilopoulos, P. et Foucault, M. (2019). Applying the Theory of Affective Intelligence to Support for Authoritarian Policies and Parties. *Political Psychology*, 40: 109-139.

<https://doi.org/10.1111/pops.12571>

Muxel, A. (2018). La politique dans la chaîne des générations. Quelle place et quelle transmission ? *Revue de l'OFCE*, N° 156(2), 29-41.

<https://doi.org/10.3917/reof.156.0029>

<https://shs.cairn.info/revue-de-l-ofce-2018-2-page-29>

Palat, B., Elie, M., Bendjaballah, S., Garcia, G., et Sauger, N. (2023). Give Them a Call! About the Importance of Call-Back Strategies in Panel Surveys. *Survey Practice* 16 (1). <https://doi.org/10.29115/SP-2023-0009>.

Pélabay, J. et Sénac, R. (2019). French critical citizenship: between philosophical enthusiasm and political uncertainty. *French Politics, Volume 17*, 407-432.

<https://doi.org/10.1057/s41253-019-00095-5>

<https://link.springer.com/article/10.1057/s41253-019-00095-5>

Rafidinarivo, C. (2017)., Dynamique de la recomposition politique : le jeu électoral. Analyse comparée du vote Outre-mer et France entière de la présidentielle et des législatives 2017. *Note ENEF 42, Sciences Po CEVIPOF*. <https://sciencespo.hal.science/CEVIPOF/hal-02417108>

Revue française de science politique (2025). Élections françaises 2024. *Revue française de science politique*, Paris : Presses de Sciences Po, Vol 1 n° 75, p.208.

<https://shs.cairn.info/revue-revue-francaise-de-science-politique-2025-1?lang=fr>

Vasilopoulos, P. et Robinson, J. (2025). Authoritarianism, Political Attitudes, and Vote Choice: A Longitudinal Analysis of the British Electorate. *Political Behavior*, 47, 503-527. <https://doi.org/10.1007/s11109-024-09961-7>

---

[1] Nos rôles dans cette enquête ont été les suivants : accompagnement méthodologique et analyses statistiques (Flora Chanvril) ; curation, documentation et diffusion des données (Jieun Jeong).

[2] Une vague 8 bis spécifique a également été menée dans le cadre de post-premier tour de l'élection primaire des LR en novembre 2016, voir <https://data.sciencespo.fr/dataset.xhtml?persistentId=doi:10.21410/7E4/QL5I5H> (consulté le 4 novembre 2025)

[3] Liste des principaux modules thématiques tels que présentés dans les questionnaires et classés par ordre d'apparition : Techniques ; Socio démographiques ; Problèmes les plus importants ; IV Régionales ; Connaissance institution régionale ; Satisfaction région ; REC Vote ; Attitudes ; Politique ; Emotions ; Expérimentations ; Primaires droite ; Présidentielle 2017 ; Primaires gauche ; Media ; Mode de vie ; Actualité ; Sémantique ; Pluralisme politique ; Participation online ; Europe et globalisation ; Bien-être ; Contexte ; Psychologie ; Populisme ; Engagement politique ; Party liking PTV ; Génétique ; Risque ; Leader ; Démocratie ; Économie (non Socio démographiques) ; Rencontre ; Relations de travail ; Images ; Sophistication politique ; Need for cognition ; Ethnocentrisme ; Législatives 2017 ; Aide au développement

; Vote stratégique ; Political psychology ; Sondages ; Ownership ; Procuration ; Misinformation ; Mesures ; Protectionnisme ; Santé ; Vote personnel ; Mode de scrutin ; Conseil constitutionnel ; Sexisme ; Confiance ; Gauche-droite ; medialab ; E. Macron.

[4] La variable "AGE\_r1" signifie la variable d'âge recodée et la variable "ANAT4\_10" signifie la variable sur la nationalité des parents ou des grands-parents.

[5] Il s'agit notamment des variables de pondération, d'échantillonnage, rafraîchissement du panel et l'appareil utilisé pour répondre à l'enquête.

[6] <https://creativecommons.org/licenses/by-sa/4.0/>

[7] <https://data.sciencespo.fr/dataverse/ENEF2017>

[8] Les 5 vagues de l'ENEF 2019 sont disponibles sur la banque de données du CDSP : <https://data.sciencespo.fr/dataverse/ENEF2019>.

[9] [doi:10.21410/7E4/FIVFXR](https://doi.org/10.21410/7E4/FIVFXR)