

LIMITER LA NON-RÉPONSE DANS L'ENQUÊTE ELIPA 2

*L'attrition (**Définitions**) de la deuxième interrogation de l'enquête Elipa 2 (enquête longitudinale sur l'intégration des primo-arrivants - personnes admises au séjour -) a été réduite par l'élaboration d'un modèle prédictif de la non-réponse basé sur l'expérience de la première interrogation. Le modèle élaboré est un modèle polytomique, la non-réponse ayant été distinguée en trois catégories de risques, le modèle attribue à un individu la probabilité de présenter chaque risque ou de n'en présenter aucun. L'identification des individus les plus à risque a été validée mais le caractère prioritaire de ces individus devrait être mieux pris en considération.*

Loreline COURT et Jade HENRY, DSED

En 2018, environ 120 000 primo-arrivants (**Définitions**) majeurs – hors motif étudiant –, ressortissants des pays tiers à l'Union Européenne ont obtenu leur premier titre de séjour d'une durée d'au moins 1 an et résidaient en France métropolitaine. L'enquête Elipa 2 interroge à trois reprises (en 2019, 2020 et 2022) un échantillon de ces primo-arrivants, représentatif des dix départements où ils sont les plus nombreux. Cela revient à considérer l'Île-de-France (hors Seine-et-Marne -77-), le Nord (59), le Rhône (69) et les Bouches-du-Rhône (13). Ce champ représente environ la moitié des primo-arrivants de 2018 [1].

Cette enquête réalisée en face-à-face est proposée en dix langues (français, arabe, anglais, bengla, chinois, espagnol, russe, soninké, turc et tamoul) afin de capter un maximum de profils de primo-arrivants. La langue d'enquête est précisée lors d'un pré-recrutement téléphonique (*Computer-Assisted Telephone Interview* -CATI-), au

cours duquel la personne contactée donne également son accord de principe pour participer, dans un second temps, à l'entretien en face-à-face. L'enquête Elipa 2 permet de suivre le parcours migratoire des primo-arrivants et surtout leur parcours d'intégration au travers de l'acquisition de la langue française, de leur évolution professionnelle, financière, sociale, de leur accès au logement... L'enquête s'intéresse également au parcours administratif des primo-arrivants, par exemple, l'acquisition et le renouvellement de leur titre de séjour, et évalue le dispositif d'accompagnement des étrangers pour les personnes ayant signé un Contrat d'Intégration Républicaine (CIR).

Elipa 2 est une enquête longitudinale en trois vagues d'interrogation où seuls les répondants de la vague précédente peuvent être interrogés pour la suivante. Entre chaque vague, l'objectif fixé est de conserver au moins 75 % des enquêtés. Lors de la première vague en 2019,

6 547 primo-arrivants ont été interrogés, 5 021 ont pu être réinterrogés en 2020 (76,7 %). Pour la vague 3, prévue au printemps 2022, l'objectif est le même : au minimum 75 % de répondants de la vague 2 soit environ 3 750 individus. Cet objectif pourrait être plus difficile à tenir car deux ans se seront écoulés depuis la vague précédente, toutefois, par continuité, les individus à risque de non-réponse sont moins nombreux. Lors de la première édition d'Elipa, le même taux de 75 % visé avait été atteint, notamment grâce à la mise en place de certains outils pour limiter l'attrition. Un suivi inter-vagues avait été mis en place pour réactualiser les coordonnées téléphoniques et postales des répondants afin de ne pas perdre le contact avec eux. La non-réponse avait également été analysée pour identifier et cibler les potentiels « mauvais répondants » à travers deux modèles distincts déterminant pour l'un, une probabilité de contact, et pour l'autre, une probabilité d'accepter de répondre [2].

Pour cette nouvelle édition, les mêmes outils ont été repris et adaptés mais, cette fois-ci, dès la deuxième vague. Le suivi entre la première et deuxième vague fait toujours l'objet d'une attention particulière pour contrôler et minimiser l'attrition. Néanmoins, le modèle mis en place diffère de celui établi lors de la première édition, puisqu'il s'agit d'un seul et même modèle, déterminant quatre probabilités correspondant à quatre types de réponse/non-réponse (modèle polytomique). Ce type d'analyse sera également appliqué entre la deuxième et troisième vague.

Cibler les mauvais répondants pour mieux prévenir la non-réponse

Afin de prévenir la non-réponse en vague 2, nous cherchons à caractériser les profils des mauvais répondants de la première vague. La non-réponse est classifiée en trois catégories distinctes : « refus » lorsque l'individu exprime clairement son refus de participer à l'enquête, « problème de téléphone » lorsque le numéro renseigné est incorrect et enfin les autres causes de non-réponse « problème de contact » incluant, entre autres, les problèmes de langues, les déménagements, les décès etc.

Pour cette modélisation, la propension que nous cherchons à estimer est celle d'accepter de répondre à l'enquête lors du premier contact : les données de CATI sont utilisées. En première vague, à l'issue de la collecte CATI, 9 021 individus acceptent de répondre à l'enquête, 516 refusent, 7 747 ont un problème de téléphone et enfin 2 676 sont notés en « problème de contact ».

Les variables disponibles en vague 1 pour expliquer le comportement de réponse proviennent de la base de sondage Agdref (Application de Gestion des Dossiers des Ressortissants Etrangers en France) qui gère

1 Répartition des non répondants en vague 1

		Population "non réponse" (en %)			Population échantillon	
		Non contact	Problèmes téléphone	Refus	Effectif total	Pourcentage
Age	Moins de 24 ans	19,51	20,80	32,75	4 165	20,8
	25 à 54 ans	66,52	71,67	58,14	14 213	71,1
	55 ans et plus	13,98	7,54	9,11	1 622	8,1
Origine	Francophone	35,05	58,22	51,55	11 182	55,9
	Afrique	5,19	6,83	2,91	1 336	6,7
	Asie	46,52	26,62	31,98	5 655	28,3
	Amérique	6,76	4,20	6,01	931	4,7
	Europe hors EEE et AELE	5,94	3,70	6,98	820	4,1
	Autres	0,52	0,44	0,58	76	0,4
Motifs	Economique	21,23	20,76	16,47	4 118	20,6
	Membre de famille	15,10	15,86	17,44	3 017	15,1
	Famille de Français	15,25	21,69	23,26	4 307	21,5
	Liens personnels et familiaux	9,04	10,24	9,30	2 169	10,9
	Réfugiés, Protection internationale	28,66	21,87	15,89	4 350	21,8
	Autres motifs	10,72	9,59	17,64	2 039	10,2
Sexe	Femme	45,44	43,11	54,84	8 714	43,6
	Homme	54,56	56,89	45,16	11 286	56,4
Ile de France	Non	18,95	19,07	22,29	3 829	19,2
	Oui	81,05	80,93	77,71	16 171	80,9
NPAI	A reçu la lettre-avis	91,67	90,81	90,31	18 488	92,4
	N'a pas reçu la lettre-avis	8,33	9,19	9,69	1 512	7,6
<i>Effectif total</i>		2 676	7 747	516	20 000	

Champ : Non répondants en vague 1 et population initiale de l'échantillon tiré de la base Agdref.

Lecture : 32,8 % des entretiens sont refusés par des jeunes de moins de 24 ans. Cette sous-population représente 20,8 % de l'échantillon.

Source : Elipa 2 - vague 1, DSED

LIMITER LA NON-RÉPONSE DANS L'ENQUÊTE ELIPA 2

les délivrances de titres de séjour en France. Elle dispose d'informations sur les titres de séjour (dates, type, motif de délivrance etc.) ainsi que sur leur titulaire (sexe, date de naissance, adresse postale, numéro de téléphone, nationalité et parfois langues parlées). Est également disponible une base de gestion fournie par le prestataire chargé de la mise en place et de la passation des entretiens. Cette base inclut différentes variables de suivi dont certaines relatives aux coordonnées et aux différentes prises de contact comme un indicateur sur la réception d'une lettre-avis informant de l'approche des entretiens et permettant ainsi de valider l'adresse postale.

Une variable type de « réponse » en quatre modalités est créée : une modalité d'accord (accepte) et les 3 modalités de non-réponse (refus / problème de téléphone / problème de contact). La répartition des trois modalités de non-réponse par rapport à l'ensemble des variables de la base de sondage Agdref ainsi que de la base de gestion est analysée afin de mettre en avant les caractéristiques différentes de ces trois types de non-réponse. Des profils différents se distinguent (**Figure 1**). La modalité « *refus* » ressort plus nettement pour les jeunes de moins de 24 ans puisqu'ils représentent près de 33 % des refus alors qu'ils sont moins de 21 % dans l'échantillon. De même pour les femmes qui représentent près de 55 % des refus alors qu'elles ne sont que 44 % dans l'échantillon. Le risque de non-contact est plus important chez les 55 ans et plus (14% de non-contact pour 8 % dans l'échantillon). La variable origine a une influence importante sur la modalité « *non-contact* ». En particulier, les personnes venant de

2 Modèle polytomique de la non-réponse en vague 2

	Modalités	Type de Non réponse	Rapport de côte	Intervalle de confiance à 95 %
Age	25 à 54 ans	Non contact	0,99	0,88 - 1,12
	25 à 54 ans	Pb Tel	0,96	0,89 - 1,05
	25 à 54 ans	Refus	0,54	0,43 - 0,68
	55 ans et plus	Non contact	2,41	2,04 - 2,85
	55 ans et plus	Pb Tel	1,14	0,99 - 1,3
	55 ans et plus	Refus	0,83	0,59 - 1,17
	Moins de 24 ans	REF		
Origine	Afrique	Non contact	1,27	1,03 - 1,56
	Afrique	Pb Tel	0,95	0,84 - 1,09
	Afrique	Refus	0,66	0,39 - 1,14
	Asie	Non contact	3,29	2,94 - 3,68
	Asie	Pb Tel	1,11	1,03 - 1,21
	Asie	Refus	2,05	1,64 - 2,57
	Amérique	Non contact	2,61	2,16 - 3,17
	Amérique	Pb Tel	0,99	0,85 - 1,15
	Amérique	Refus	1,73	1,17 - 2,56
	Europe hors EEE	Non contact	2,59	2,11 - 3,18
	Europe hors EEE	Pb Tel	1,01	0,86 - 1,19
	Europe hors EEE	Refus	2,22	1,53 - 3,21
	Autres	Non contact	3,12	1,61 - 6,07
	Autres	Pb Tel	1,59	0,94 - 2,67
	Autres	Refus	2,97	0,88 - 9,99
Pays francophones	REF			
Motif	Membre de famille	Non contact	1,20	1,03 - 1,41
	Membre de famille	Pb Tel	1,12	1 - 1,24
	Membre de famille	Refus	1,30	0,94 - 1,8
	Famille de Français	Non contact	0,89	0,76 - 1,04
	Famille de Français	Pb Tel	0,94	0,85 - 1,04
	Famille de Français	Refus	1,31	0,96 - 1,78
	Liens personnels et familiaux	Non contact	0,77	0,64 - 0,91
	Liens personnels et familiaux	Pb Tel	0,85	0,75 - 0,95
	Liens personnels et familiaux	Refus	0,84	0,58 - 1,22
	Réfugiés, Protection internationale	Non contact	1,12	0,97 - 1,28
	Réfugiés, Protection internationale	Pb Tel	1,05	0,95 - 1,16
	Réfugiés, Protection internationale	Refus	0,74	0,53 - 1,02
	Autres motifs	Non contact	1,13	0,94 - 1,35
	Autres motifs	Pb Tel	0,92	0,81 - 1,05
	Autres motifs	Refus	1,49	1,05 - 2,1
Economique	REF			
Genre	Féminin	Non contact	1,14	1,04 - 1,26
	Féminin	Pb Tel	1,01	0,95 - 1,08
	Féminin	Refus	1,50	1,24 - 1,81
	Masculin	REF		
Ile-de-France *	-	-	-	-
NPAI	A reçu la lettre-avis	Non contact	0,67	0,57 - 0,79
	A reçu la lettre-avis	Pb Tel	0,61	0,54 - 0,69
	A reçu la lettre-avis	Refus	0,56	0,41 - 0,76
	N'a pas reçu la lettre	REF		

Champ : Répondants de la première interrogation d'Elipa 2

Lecture : Le risque de ne pas être contacté est 2,41 fois plus grand chez les personnes âgées de 55 ans et plus que les jeunes de moins de 24 ans.

* La variable Ile-de-France est non significative dans le modèle polytomique.

Source : Elipa 2 - vague 2, DSED

pays francophones ont moins de risque de non-contact (35 % par rapport 56 %) alors que celle venant d'Asie ont un fort risque de non-contact (47 % par rapport à 28 %). Concernant le motif du titre de séjour, on peut noter que les bénéficiaires d'une protection internationale ont peu tendance à refuser (16 % par rapport à 22 %) mais présentent un fort risque de non-contact (29 % par rapport à 22 %). Enfin, les « problèmes de téléphone » ne présentent aucune caractéristique spécifique par rapport à l'échantillon de départ.

Plusieurs modèles d'explication de la non-réponse sont ainsi testés. Le modèle retenu est un modèle polytomique non ordonné incluant les variables âge, origine (pays francophone/pays non-francophones répartis par continent), motif de titre de séjour, sexe ainsi qu'une indicatrice estimée de réception de la lettre-avis et une indicatrice de résidence en Ile-de-France (Figure 2).

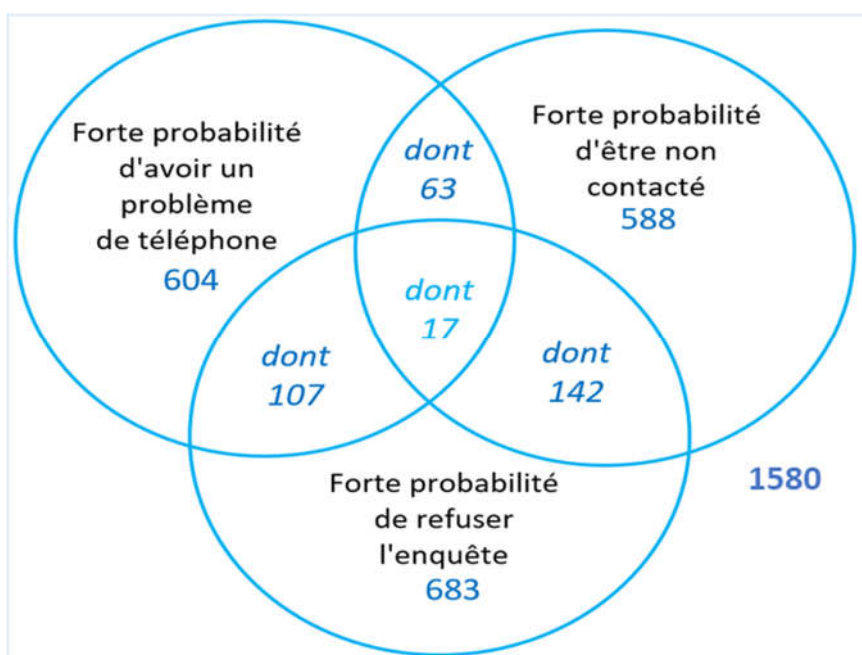
Ensuite, est estimée pour chaque répondant de la vague 1 une probabili-

té, pour la vague 2, d'être en « refus », d'être en « problème de téléphone » et d'être en « non contact ». La probabilité de donner son accord de principe pour répondre à l'enquête est aussi mécaniquement estimée mais n'est pas mobilisée pour ce travail. Une probabilité est notée comme forte si elle est parmi les 10 % les plus élevées de sa catégorie de non-réponse, et de là découle la qualification d'individu à risque. Un individu avec une probabilité forte de refuser peut aussi avoir une probabilité forte d'avoir un problème de téléphone. De plus, 261 individus (3,99 %) cumulent deux risques et 17 individus (0,26 %) cumulent les 3 risques. Plus explicitement, les individus à risque dans chaque catégorie se répartissent comme suit (Figure 3).

Un individu est défini « prioritaire » s'il présente au moins un risque. Ainsi, le modèle identifie 1 580 individus prioritaires.

Méthodologie mise en place pour valider le modèle

③ Répartition des différents types de non-réponses



Champ: Individus classés prioritaires par le ciblage réalisé.

Source : Elipa 2 - vague 1, DSED.

Afin de mesurer l'impact d'un événement, on constitue usuellement un groupe de contrôle (« groupe témoin ») ne recevant pas le traitement administré aux autres individus. Le groupe auquel le traitement est appliqué est appelé quant à lui « groupe test ». La population test doit présenter les mêmes caractéristiques que celle de contrôle afin de les rendre comparables. A la fin de l'expérience, la comparaison entre les deux groupes permet d'estimer l'effet du mécanisme étudié.

Ici, le traitement appliqué aux individus de la *population de test* consiste à contacter en priorité ces individus : les enquêteurs ont comme consigne de les appeler dès le début du terrain pour passer l'entretien de la deuxième vague afin d'avoir plus de temps pour parvenir à les interroger. Ce mécanisme de priorisation de certains individus a également été présenté lors des formations enquêteurs et enquêtrices pour la vague 2 afin de bien répondre à certaines questions. Aucune incitation financière n'a été proposée pour prioriser ce traitement.

Les deux groupes sont constitués aléatoirement : tous les individus ont la même probabilité d'être sélectionnés dans un groupe ou dans l'autre. Ainsi, sur les individus prioritaires, 785 sont notés comme *population de test* et sont envoyés au prestataire comme prioritaires. Le reste (795 individus) est utilisé comme *population de contrôle* en interne.

Les caractéristiques des deux groupes par rapport au nombre de risques sont les suivantes (Figure 4).

Les profils sont ensuite uniformisés pour ne garder que les profils « est prioritaire » et « n'est pas prioritaire », quel que soit le nombre de

risques ou le type de risque de l'individu. Seuls les identifiants des individus de la population de test ont été communiqués aux enquêteurs, tandis que ceux du groupe de contrôle ont été conservés au sein du département.

Un modèle global qui s'avère pertinent

La durée de collecte de la seconde vague, en 2020, fut bien sûr marquée par la crise sanitaire et le confinement (Figure 5). La collecte s'est ainsi effectuée sur deux périodes : une première de 23 jours avant le 17 mars puis une deuxième d'environ cinq mois du 6 juin au 7 novembre 2020. Sur les 785 individus transmis en fiches prioritaires au prestataire, 589 ont finalement été interrogés (soit 75 %). Plus spécifiquement, ils ont été répartis de la manière suivante : 63 entretiens durant la première phase de l'enquête (1,3 % de l'échantillon total) et 526 à partir de juin (10,5 % de l'échantillon total).

Afin d'estimer si le modèle de non-réponse est pertinent, on compare le taux de réponse de la deuxième vague du groupe non-prioritaire, soit 77,5 %, à celui du groupe prioritaire de contrôle, soit 73,1 %. Il apparaît que la différence (4,4 points) est significative (Z-test), le modèle mis en place est donc pertinent. De plus, pour accréditer la pertinence du modèle, on observe que le taux de réponse diminue lorsque l'individu cumule les risques jusqu'à atteindre 70,6 % pour 3 risques.

Efficacité des démarches de priorisation

Après cette identification pertinente des individus prioritaires, une évaluation est faite sur l'efficacité

4 Répartition par nombre de risque(s) des groupes prioritaires de test et de contrôle

		Prioritaires test		Prioritaires contrôle	
		Effectif	Pourcentage	Effectif	Pourcentage
Nombre de risque(s)	1	647	82,4	655	82,4
	2	130	16,6	131	16,5
	3	8	1,0	9	1,1
<i>Total</i>		785	100	795	100

Champ : Individus classés « prioritaires » par le ciblage réalisé.

Source : Elipa 2 - vague 1, DSED.

du traitement appliqué par les enquêteurs sur la population prioritaire. Dans un premier temps, on regarde l'évolution du nombre d'entretiens prioritaires par rapport à la totalité des entretiens afin d'évaluer le travail de priorisation des enquêteurs. En moyenne, 12,7 % des individus interrogés avant le confinement étaient « prioritaires » contre 12,0 % de ceux interrogés après les individus « prioritaires » ont ainsi été un peu plus souvent traités au début qu'en fin d'enquête.

Dans un deuxième temps, on s'intéresse au gain en terme de taux de réponse entre les groupes prioritaires test et de contrôle, par rap-

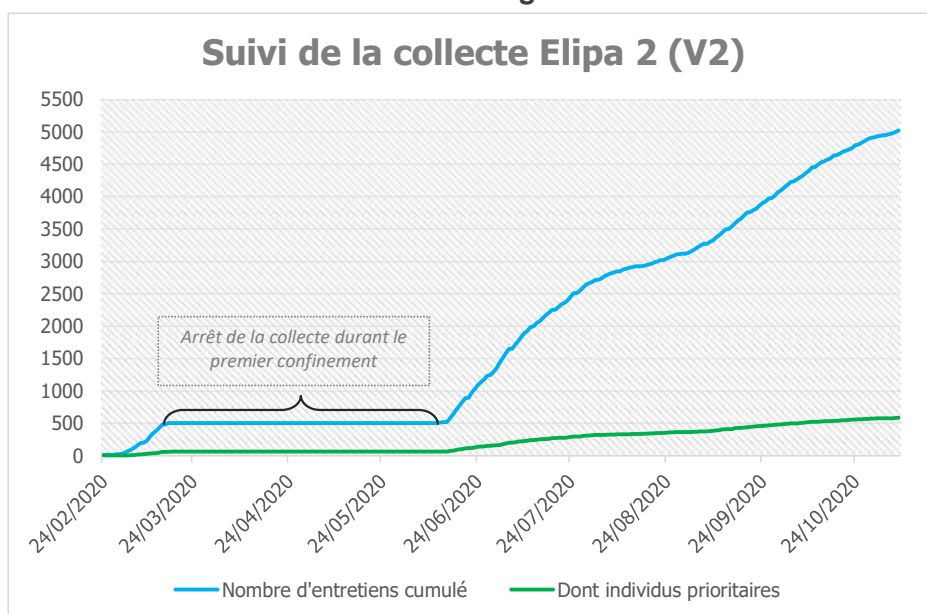
port aux individus non-prioritaires. Les résultats empiriques sont les suivants (Figure 6).

75,0 % de la *population de test* a répondu en vague 2, contre les 73,1 % de la *population de contrôle*. Le traitement a donc eu un effet positif, puisqu'il a permis de réduire de 2 points la différence de taux de réponse par rapport à la population non-prioritaire. Si l'écart est faible il est significatif avec un seuil de risque de 10 % (Z-test).

Pistes d'amélioration

Les premiers résultats de cette méthodologie sont mitigés : le modèle a relativement bien identifié les personnes risquant de ne pas ré-

5 Evolution du nombre total d'entretiens et d'entretiens prioritaires de la vague 2



Source : Elipa 2 - vague 2, DSED.

6 Taux de réponse selon le caractère prioritaire

	Part de répondants (%)	Différence (en pts) avec les non-prioritaires
Dans la population totale	76,69	-
Dans la population prioritaire	74,05	
...dans la population de test	75,03	2,50
...dans la population de contrôle	73,08	4,45
Dans la population non prioritaire	77,53	-

Source : Elipa 2 - vague 2, DSED

pondre mais le traitement mis en place par la suite pour pallier la non-réponse n'a pas fortement limité l'attrition. Cette première procédure est cependant riche d'enseignements pour la suite et elle fera l'objet d'amélioration pour la prochaine interrogation.

Deux voies d'amélioration se dessinent, l'une sur le traitement à appliquer à la population ayant un risque de non-réponse, l'autre au modèle d'estimation de ce risque de non-réponse.

La première piste s'appuie d'une part sur une différenciation des profils de non-réponse dans la formation des enquêteurs. Alors que la différenciation de ces profils est mobilisée pour la mise en place du modèle de non-réponse, ils sont ensuite tous assimilés à un seul type de non-réponse lors de la transmission des fiches prioritaires. Il pourrait être intéressant d'explicitier les différents profils aux enquêteurs afin qu'ils sachent à quel type de risque ils seront confrontés et ainsi adapter leur prise de contact. Une solution complémentaire envisagée serait d'envoyer, en plus de la lettre-avis, un sms aux individus prioritaires pour doubler l'information de passage. Le sms très succinct pourrait

contenir un lien vers une page plus détaillée. Ce moyen a déjà été testé pour d'autres enquêtes longitudinales et avait permis de faire gagner 2 à 3 % de taux de réponse du groupe prioritaire.

La seconde amélioration envisagée est d'affiner le modèle d'estimation de la non-réponse avec l'intégration de nouvelles variables dans le modèle de non-réponse, notamment celles des répondants de la deuxième vague. Parmi les variables ressortant comme explicatives de la non-réponse de la deuxième vague, figurent notamment l'aisance en français, le fait de connaître des personnes d'origine française à son arrivée en France, le fait de percevoir des allocations et enfin le fait d'avoir eu le sentiment de n'avoir plus goût à rien. Ces variables, disponibles depuis la seconde vague et mobilisées pour en corriger la non-réponse, seront évidemment intéressantes à prendre en compte dans l'identification des individus prioritaires pour la troisième vague. Enfin, dans la construction de l'indicatrice de risque de non-réponse, la distribution des individus selon leur probabilité de non-réponse pourrait être interrogée afin d'optimiser le positionnement de la borne de « risque » (ici à 10%).

Définitions

Primo-arrivant : (ou primo-détenteur) : personne originaire d'un pays tiers à l'Union européenne des vingt-huit (UE28), l'Islande, la Norvège, le Lichtenstein et la Suisse disposant d'un premier titre de séjour d'au moins un an attribué en 2018. Dans cette étude, les primo-arrivants disposant d'un titre de séjour *étudiants* sont exclus de cette appellation.

Attrition : perte de représentativité d'un échantillon au fil du temps lié au phénomène de non-réponse.

Pour en savoir plus

[1] Jourdan V., Prevot M., « L'Enquête longitudinale sur l'intégration des primo-arrivants (Personnes admises au séjour) – Elipa 2 », *Infos migrations*, n°97, 2020.

[2] Jourdan V., « Méthodes pour limiter la non-réponse pour la vague 3 d'une enquête longitudinale », contribution associée aux *XIIèmes Journées de Méthodologie Statistique de l'Insee*, Paris, 2015.