

Direction des Statistiques Démographiques et Sociales

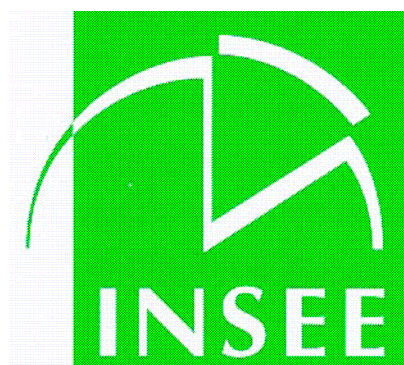
N° F1304

À L'ORIGINE DE L'ENQUÊTE TEO

ENJEUX DE L'ÉCHANTILLONAGE, COLLECTE ET  
PONDÉRATIONS DE L'ENQUÊTE

Elisabeth Algava et Bertrand Lhommeau

Document de travail



**Institut National de la Statistique et des Études Économiques**

**INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES**  
**Série des Documents de Travail**  
**de la**  
**DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES**

**N°F1304**

**À L'ORIGINE DE L'ENQUÊTE TEO**  
**ENJEUX DE L'ÉCHANTILLONAGE, COLLECTE**  
**ET PONDÉRATIONS DE L'ENQUÊTE**

**ÉLISABETH ALGAVA ET BERTRAND LHOMMEAU\***

(CELLULE STATISTIQUES ET ÉTUDES SUR L'IMMIGRATION)

**Document de travail**

**Avril 2013**

\* au moment de la rédaction de ce document les deux auteurs faisaient partie de la Cellule Statistiques et Études sur l'Immigration.

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.  
Working-papers do not reflect the position of INSEE but only their authors'views.

## **Résumé**

L'enquête « Trajectoires et Origines » (TeO) de l'Ined et de l'Insee cherche à appréhender dans quelle mesure les origines migratoires (géographiques, nationales) sont susceptibles de modifier par elles-mêmes les chances d'accès aux principaux biens qui définissent la place de chacun dans la société comme l'éducation, l'emploi... Pour réaliser cet objectif, il a fallu relever plusieurs défis. En premier, celui de la création d'une base de sondage permettant la construction d'un échantillon d'individus qui devait comprendre une surreprésentation des immigrés, des descendants d'immigrés, des natifs d'un DOM et des descendants d'un natif d'un DOM et pour des origines de populations issues de l'immigration peu présentes en France. Cette phase préalable à la collecte d'entretiens a constitué en soi une opération statistique lourde avec la saisie de 300 000 noms de personnes recensées en 2007 et la recherche dans les registres de l'état civil des mairies du lieu de naissance des parents pour près de 100 000 personnes. Ensuite, l'enquête proprement dite a été organisée pour résoudre plusieurs difficultés : une forte concentration des entretiens en zone urbaine, une population jeune très mobile d'où un suivi des adresses en cas de déménagement et parfois des interviewés pour lesquels un entretien en français aurait été impossible. Enfin, la pondération des données a demandé des calculs particuliers liés à la complexité de la construction de l'échantillon.

## **Summary**

The "Trajectories and origins" (TeO) survey, carried out by INED and INSEE seeks to understand the extent to which migrant origin (geographical and national) influences access to sectors such as education and employment that play a key role in individuals' place in society. To investigate this question, several obstacles had to be overcome. First, it was necessary to create a database for the construction of a sample which would include an over-representation of immigrants, descendants of immigrants, along with natives of French Overseas Territories (DOM) and their descendants. The database also needed to allow an overrepresentation of numerically small national groups such as Turks. This preliminary phase was in itself a complex statistical operation, involving the data capture of 300 000 individuals recorded in the 2007 yearly census survey, and the identification in the municipal registers of the parental place of birth of nearly 100 000 people. The survey itself also necessitated the resolution of several problems: the high concentration of interviewees in urban areas; a very young, mobile population, frequently changing address; plus the logistical problems involved in interviewing non-French speakers. Finally, the weighing of the data required specific calculations related to the complexity of the sample construction process.

À l'origine de l'enquête :  
enjeux de l'échantillonnage, collecte et pondérations de l'enquête TeO

Élisabeth Algava et Bertrand Lhommeau<sup>1</sup>

## Introduction : TeO un outil statistique pour l'étude des questions d'immigration et d'intégration

Avec l'enquête Trajectoires et Origines (TeO), et ce dès l'amorce du projet, l'Ined et l'Insee se sont donnés l'objectif de remédier au décalage important entre la place grandissante des questions d'immigration et d'intégration dans le débat public et le niveau des informations statistiques disponibles (voir encadré 1 sur les acteurs du projet). L'enquête cherche à appréhender dans quelle mesure les origines migratoires (géographiques, nationales) sont susceptibles de modifier par elles-mêmes les chances d'accès aux principaux biens qui définissent la place de chacun dans la société : le logement, l'éducation, la maîtrise de la langue, l'emploi et la promotion, les services publics et les prestations sociales, la vie du couple et la famille et la santé et le réseau de relations, le marché matrimonial, la nationalité et la citoyenneté...

L'information disponible dans les recensements ou dans l'enregistrement administratif des flux ne suffit pas, car elle reste instantanée et limitée à des individus, alors qu'il importe de retracer des trajectoires et de reconstituer les unités de décision réelles en préservant les liens entre conjoints et entre générations. Il n'existe qu'un seul moyen d'obtenir de telles informations à grande échelle : interroger directement les intéressés.

L'enquête Mobilité Géographique et Insertion Sociale (MGIS) de 1992 abordait déjà ces thématiques et l'enquête TeO se situe dans son prolongement. Cependant, la question de l'intégration est aujourd'hui davantage posée en termes de discriminations. L'enquête TeO les aborde par la mesure d'inégalités de situation dans différentes sphères de la vie sociale (logement, emploi, vie familiale...) et la perception des inégalités de traitement, tout particulièrement liées à l'origine ou la couleur de peau.

Cette perspective a dicté les principaux choix en matière de populations à enquêter :

- une importance au moins égale accordée aux « secondes générations », descendants d'un ou deux parents immigrés, qu'aux immigrés eux-mêmes. Les questions se posent

---

<sup>1</sup> Les auteurs remercient particulièrement Catherine Borrel et Pascale Breuil pour leurs conseils et remarques.

en effet différemment mais avec tout autant d'acuité pour ces « secondes générations » nées et socialisées en France ;

- l'extension du champ de l'enquête aux populations issues des départements d'Outre-mer et leurs descendants vivant en France métropolitaine;
- la représentation de l'ensemble des immigrés et de leurs descendants vivant en France métropolitaine, quel que soit leur pays d'origine, afin de décrire un paysage complet, tout en préservant la capacité à faire des analyses origine par origine.

## Un échantillon complexe à bâtir

La traduction concrète de ces attentes n'allait pas de soi et de nombreuses difficultés sont venues émailler la mise en place de l'enquête et de son échantillon : les populations visées par l'enquête sont relativement rares et pour certaines difficiles à identifier.

## Des objectifs initiaux ambitieux

Tout d'abord, s'agissant d'une enquête qui cherche à appréhender dans quelle mesure les origines migratoires (géographiques, nationales) sont susceptibles de modifier par elles-mêmes les chances d'accès aux principales ressources qui permettent l'insertion sociale et professionnelle (logement, l'éducation, la maîtrise de la langue, l'emploi...), son champ exclut les mineurs.

Des enquêtés en ménage ordinaire, âgés de 18 à 60 ans (50 ans pour les descendants).

Le champ couvert par l'enquête a été limité aux personnes âgées de 18 à 60 ans vivant en ménage ordinaire. Pour les descendants d'immigré(s) ou de natif(s) d'un DOM, le seuil a été abaissé à 50 ans.

L'enquête MGIS portait sur les immigrés de 20 à 59 ans, et sur les enfants d'immigrés ayant entre 20 et 29 ans. De la même manière, l'enquête TeO étant axée sur les questions d'insertion sociale et professionnelle, il semblait justifié de se limiter aux tranches d'âge actif. De plus, une enquête menée en 2002 de la CNAV a porté sur les « seniors immigrés », abordant tout particulièrement les questions liées au passage à la retraite.

En ce qui concerne les descendants directs d'immigrés, le champ a été étendu par rapport à celui de l'enquête MGIS : les descendants directs d'immigrés sont plus nombreux aujourd'hui qu'en 1994 à avoir dépassé la trentaine. Le déroulement des carrières professionnelles et de la vie familiale par exemple peut être étudié de façon plus intéressante chez ceux qui ont eu une

vie adulte plus longue. Cependant, les natifs descendants d'immigrés sont relativement plus rares au-delà de 50 ans (et de ce fait leur échantillonnage eut été trop coûteux pour un effectif restreint).

Le champ a également été restreint aux personnes vivant en ménages ordinaires. La passation de l'enquête MGIS dans les collectivités avait en effet été particulièrement difficile : en dépit d'efforts importants, très peu de questionnaires avaient été collectés. L'exclusion des collectivités a semblé une limite acceptable car entre 1990 et 2008, la proportion d'immigrés vivant en collectivité a diminué (elle est de 3,6 % en 2008) et l'âge moyen de ceux qui vivent en collectivité a augmenté : près de 30% d'entre eux ont 60 ans ou plus et auraient été en dehors du champ de l'enquête.

Cinq sous échantillons : les immigrés, les natifs d'un DOM, les descendants d'immigrés, les descendants d'un natif de DOM et un sous échantillon « témoin »

Ensuite, pour l'enquête MGIS, la stratégie retenue était le choix d'un nombre limité d'origines : seuls les immigrés originaires de sept pays ou groupes de pays (Afrique subsaharienne, Algérie, Espagne, Maroc, Portugal, Turquie, Asie du Sud-est) entraient dans le champ de l'enquête, et trois origines avaient été retenues pour les descendants. Cette stratégie avait l'avantage de permettre l'obtention de groupes homogènes du point de vue de l'origine et de taille suffisante pour pouvoir comparer les immigrés selon leur origine. Elle conduisait en revanche à exclure complètement du champ de l'enquête les origines plus rares.

Compte tenu des principes rappelés plus haut, il était souhaitable que TeO soit constituée d'échantillons représentatifs de l'ensemble des populations immigrée et descendante d'immigrés, quel que soit leur pays d'origine. Mais la comparaison de groupes d'origine de taille suffisante restait un objectif important de l'enquête, tout comme la possibilité de comparer les résultats de l'enquête TeO à ceux de l'enquête MGIS origine par origine.

La solution retenue permet de tenir tous ces objectifs : les échantillons visés d'immigrés et de descendants d'immigrés sont de taille importante de manière à être représentatifs de l'ensemble de la population immigrée, tout en étant stratifiés pour assurer des effectifs suffisants pour les origines trop rares.

L'objectif était ainsi d'interroger parmi les adultes vivant en ménages ordinaires en France métropolitaine :

1. 9 600 immigrés, personnes nées étrangères à l'étranger (frontières actuelles) ;

2. 800 natifs d'un département d'Outre-mer (**DOM : GUYANE, GUADELOUPE, MARTINIQUE ET RÉUNION**) ;
3. 9 600 descendants d'immigré(s), personnes nées en France métropolitaine et dont au moins l'un des parents est immigré.
4. **800 DESCENDANTS DE NATIF(S) D'UN DOM : PERSONNES NÉES EN FRANCE MÉTROPOLITAINE DONT AU MOINS UN PARENT EST NÉ DANS UN DOM.**
5. 2 400 personnes de la « **POPULATION MAJORITAIRE** ». **CE TERME DÉSIGNE L'ENSEMBLE DES PERSONNES, NUMÉRIQUEMENT LES PLUS NOMBREUSES, QUI RÉSIDENT EN FRANCE MÉTROPOLITAINE ET QUI NE SONT NI IMMIGRÉES, NI NATIVES D'UN DOM, NI DESCENDANTES DE PERSONNE(S) IMMIGRÉE(S) OU NATIVE(S) D'UN DOM<sup>2</sup>.**

Au sein des immigrés et des descendants d'immigrés, le choix a été fait de surreprésenter quatre pays ou groupes de pays d'origine :

1. L'Asie du Sud-Est (Cambodge, Laos et Vietnam) ;
2. La Turquie ;
3. L'Afrique sahélienne (Sénégal, Mauritanie, Gambie, Guinée Bissau, Guinée, Mali, Burkina Faso, Niger et Tchad) ;
4. L'Afrique guinéenne ou centrale (Côte d'Ivoire, Ghana, Togo, Bénin, Nigeria, Cameroun, République Centrafricaine, Gabon, Congo, République Démocratique du Congo et Guinée Équatoriale) ;

Ces objectifs permettent à TeO de se situer dans la lignée de MGIS, tout en élargissant considérablement la population couverte par l'enquête.

9 600 immigrés

Parmi les enquêtés immigrés, les origines surreprésentées ont été choisies pour permettre la comparaison avec l'enquête MGIS.

Les deux premières (Turquie et Asie du Sud-est) reprennent la classification de l'enquête MGIS. Quant aux deux dernières, elles constituent une subdivision de l'ensemble « Afrique subsaharienne ». Cette subdivision paraissait intéressante car la population de migrants issue de cette région a augmenté depuis 1992 et les deux groupes de pays se distinguent par leur

---

<sup>2</sup> La plupart sont des personnes nées en France de parents tous deux nés en France, dites aussi natives descendants de natifs. Cette « population majoritaire » comprend toutefois également les rapatriés, personnes nées françaises dans l'un des anciens territoires coloniaux avant les indépendances, leurs descendants ainsi que les autres personnes nées françaises à l'étranger ainsi que leurs descendants.

histoire migratoire en France : les migrants venus d'Afrique centrale ou du Golfe de Guinée (Côte d'Ivoire, République démocratique du Congo, République populaire du Congo, Cameroun...) sont en moyenne plus jeunes, arrivés plus jeunes en France et plus diplômés que ceux en provenance d'Afrique sahélienne.

Pour leur part, les personnes originaires d'Algérie, d'Espagne ou d'Italie, du Maroc ou de Tunisie, du Portugal sont assez nombreuses dans la population pour qu'un échantillon aléatoire de 9 600 immigrés en comprenne un nombre suffisant pour autoriser des analyses statistiques distinctes.

9 600 descendants d'immigré(s)

Fait nouveau par rapport aux enquêtes antérieures, les mêmes origines étaient recherchées pour les descendants d'immigrés. L'enquête MGIS ne portait que sur les personnes dont le chef de famille était immigré et né en Espagne, en Algérie ou au Portugal. Comme pour les immigrés, il était souhaitable d'avoir pour les descendants d'immigrés des origines comparables. Considérant toutefois que la demande sociale d'informations était particulièrement forte concernant les descendants d'immigrés, l'enquête TeO s'est donnée des objectifs beaucoup plus ambitieux : être représentative de l'ensemble des origines présentes en France, tout en ayant des groupes homogènes suffisamment importants. À cela s'ajoutait la volonté de pouvoir confronter la situation, les pratiques et les conceptions des immigrés et des descendants d'immigrés de même origine.

L'échantillon des descendants d'immigrés a donc été construit pour avoir autant que possible les mêmes groupes d'origine que celui des immigrés. Cela était particulièrement ambitieux pour les groupes migratoires dont l'arrivée est récente et dont la descendance est en grande partie constituée pour le moment de mineurs : Turquie, Afrique subsaharienne notamment.

Le champ de l'enquête a également été étendu aux personnes nées d'une mère immigrée et d'un père non immigré, ce qui permet de distinguer lors de l'analyse les descendants de deux immigrés et les descendants de couples mixtes. Ces élargissements augmentent l'hétérogénéité de la population et rendent nécessaire un échantillon bien plus important que celui de l'enquête MGIS (1921 entretiens exploitables).

Par cohérence avec la définition d'une personne immigrée, une personne dont un parent est né Français à l'étranger (rapatrié d'Algérie, Français vivant à l'étranger...) n'est pas descendante d'immigré. Cette restriction réduit fortement le champ car environ un tiers des personnes ayant au moins un parent né à l'étranger n'est pas descendant d'immigré. Elle impose



également la connaissance de la nationalité de naissance de chacun des parents, information relativement complexe à obtenir.

### **800 NATIF(S) D'UN DOM ET 800 DESCENDANTS DE NATIF(S) D'UN DOM**

Autre nouveauté majeure, il paraissait essentiel d'intégrer dans le champ de l'enquête les ultra-marins et leurs descendants, compte tenu des obstacles que ces populations, pourtant sans origine étrangère, peuvent rencontrer du fait de leur couleur de peau en métropole.

Le Conseil national de l'information statistique (CNIS) a d'ailleurs soutenu cette initiative d'échantillonner spécifiquement les personnes originaires d'un département d'Outre-mer pour l'enquête TeO. L'avis d'opportunité accordé le 15 mai 2006 précisait en effet : « L'enquête portant notamment sur l'effet de l'origine, la formation exprime son intérêt pour qu'un complément d'échantillon de personnes originaires des DOM puisse être interrogé. » La connaissance des conditions de vie des originaires des DOM en métropole et surtout de leurs descendants est en effet très parcellaire. La constitution de cet échantillon s'est faite de façon parallèle aux échantillons d'immigrés et de descendants d'immigrés : l'interrogation de personnes ayant un ou deux parents né(s) dans un département d'Outre-mer nécessitait la collecte d'information sur le département de naissance de chacun des parents.

2 400 personnes de la « population majoritaire »

Dans l'enquête MGIS, un échantillon témoin était constitué de 1900 personnes, sans aucune sur-représentation particulière. Pour TeO, l'objectif a été quelque peu relevé pour s'approcher de 3 000 personnes. Cet échantillon témoin est construit pour être représentatif de l'ensemble de la population de 18 à 60 ans vivant en France métropolitaine en ménage ordinaire. Il comprend donc des immigrés, des descendants d'immigrés, etc.

L'objectif était d'obtenir, après réaffectations, environ 2 400 personnes n'appartenant à aucun des groupes spécifiquement échantillonnés pour TeO (immigrés, descendants d'immigrés, natifs d'un DOM ou descendants de natifs des DOM) et constituant la population dite majoritaire.

**Les difficultés : une information non disponible, une population rare**

La constitution des échantillons de l'enquête TeO a dû composer avec deux difficultés principales :

- les populations couvertes sont des populations relativement rares ;

- les descendants d'immigrés ou de natifs d'un DOM sont rarement identifiables dans les bases de sondage.

L'enquête vise une population relativement « rare ». Au recensement de 1999, les immigrés constituaient 7,4 % de la population totale (Insee, 2005). Quant aux descendants directs d'immigrés nés en France, ils représentaient seulement 8 % des personnes âgées de 17 à 65 ans, pour moitié ayant un seul parent immigré, pour moitié en ayant deux (Insee, 2005). Pour constituer un échantillon suffisant, si on n'est pas capable de sous échantillonner il fallait disposer d'une base de sondage approximativement 12 fois plus importante que la taille de l'échantillon : par exemple, si on sélectionne 1 200 personnes au hasard, environ 8%, soit 100, seront immigrées. On peut escompter qu'environ les deux tiers des personnes sélectionnées répondront véritablement à l'enquête. Certaines sont injoignables, d'autres ont pu déménager si on échantillonne les personnes par leur logement et d'autres refusent de répondre, etc. L'échantillon d'une enquête comprend toujours plus de fiches-adresses (FA, fiche identifiant la personne à enquêter et confiée à l'enquêteur) que de répondants escomptés. Cela fait qu'il faut une base de sondage en population générale à peu près 18 fois plus importante que le nombre final de répondants descendants d'immigrés. Et cela sans même prendre en compte la nécessité d'augmenter la taille initiale pour permettre les surreprésentations de certaines origines : par exemple, les descendants de natifs de DOM en métropole représentent moins de 1 % de la population. Pour en trouver 100, il faut donc une population de plus de 10 000 personnes. Cela a pour conséquence de limiter les scénarii d'échantillonnages possibles.

En outre, les descendants directs d'immigrés ou de natifs d'un DOM sont difficilement repérables dans les bases de sondage disponibles. En effet, deux questions suffisent pour définir si une personne est immigrée ou non : Est-elle née en France ? Est-elle française de naissance ? Elles sont bien standardisées et présentes dans la quasi-totalité des enquêtes menées par l'Insee, y compris le recensement de la population (voir figure 1).

Figure 1 : Extrait du bulletin individuel de recensement (2004)

**2 Date et lieu de naissance**

Né(e) le :        
jour mois année

à :   
commune

département n° DOM pays pour l'étranger, territoire pour les TOM

**Si vous êtes né(e) à l'étranger, en quelle année êtes-vous arrivé(e) en France ?**      
année

**3 Quelle est votre nationalité ?**

- Française
  - Vous êtes **né(e) français(e)** .....  1
  - Vous êtes **devenu(e) français(e)** (par exemple : par naturalisation, par déclaration, à votre majorité) .....  2
  - ↳ Indiquez votre nationalité à la naissance :
- Étrangère .....  3
  - ↳ Indiquez votre nationalité :

En revanche, les questions permettant de repérer les descendants directs d'immigrés portent à la fois sur la personne enquêtée et sur chacun de ses parents. Il faut savoir pour l'enquêté, son père et sa mère, le lieu de naissance et la nationalité de naissance. Les questions sur les parents ne font pas partie des questions standards : leur introduction dans une enquête doit être motivée par les thématiques abordées. De ce fait, elles ne sont pas toujours présentes.

Ces questions ne figurent pas dans les bulletins individuels des recensements. L'enquête Étude de l'histoire familiale (EHF) de 1999, associée au recensement, ne permettait pas de définir convenablement les descendants directs d'immigrés. Seule une population plus large pouvait être définie à partir du pays de naissance des deux parents, appelée les « descendants de migrants », et composée des personnes dont au moins un des deux parents est né à l'étranger (mais pas nécessairement immigré, s'il est né français). Cela figure dans l'enquête Famille et logement 2011 associée au recensement. Ces questions figurent toutefois dans un nombre croissant d'enquête. Par exemple, l'enquête « Formation et qualification professionnelle » de 2003 comprenait l'ensemble des questions nécessaires et l'enquête Emploi en continu les comporte à compter de janvier 2005<sup>3</sup> et l'enquête Logement 2006.

<sup>3</sup> Cette introduction est progressive puisque seul le sixième entrant est interrogé. L'information est donc disponible sur l'ensemble de l'échantillon à partir de juin 2006.

## Les solutions retenues

Quelle base de sondage initiale ?

Pour la plupart des enquêtes auprès des ménages réalisées par l'Insee, le tirage se fait dans l'échantillon maître (EM). Celui-ci est une sélection de logements au sein du recensement<sup>4</sup>. Ces logements sont sélectionnés dans les zones d'action enquêteur (ZAE), zones géographiques choisies pour être représentatives de l'ensemble du territoire et qui constituent les zones de travail d'un enquêteur. Un échantillon tiré dans l'échantillon-maître est donc constitué de logements dont on interroge les occupants au moment de l'enquête. C'est un outil qui permet d'avoir un échantillon représentatif des logements et donc des habitants, de façon bien rodée et routinière.

Cette solution a pourtant été d'emblée écartée. En effet, pour une collecte prévue en 2008, l'échantillon maître s'appuyait sur le recensement de 1999. Dans une enquête classique, ce sont des logements qui sont échantillonnés et peu importe que les occupants aient changé. Pour TeO, les échantillons étaient constitués d'individus et les sélectionner d'après des informations datant de presque 10 ans exposait l'enquête au risque qu'une majorité ait déménagé. Seul l'échantillon témoin pouvait a priori être constitué à partir de cette base de sondage. Ce sont donc pour les autres échantillons que différentes solutions ont été évoquées.

L'utilisation comme base de sondage des sortants de l'enquête Emploi en continu (EEC) était une autre possibilité simple à mettre en œuvre, reprenant le principe des enquêtes complémentaires. Mais pour les immigrés par exemple, il aurait fallu les sortants de cinq années environ pour parvenir à l'effectif envisagé de 9 600 personnes. Le nombre d'années aurait été encore plus élevé car le plan d'échantillonnage prévoyait des taux de sondage différenciés selon le pays de naissance. Ainsi, pour constituer un échantillon de 1 000 personnes nées en Turquie, il aurait fallu utiliser les sortants de 10 ou 11 années d'enquête Emploi. Par ailleurs, cela aurait empêché l'utilisation de l'enquête Emploi pour d'autres enquêtes complémentaires.

Les difficultés d'un tirage dans l'enquête emploi étaient identiques pour l'échantillon d'immigré et celui de descendants d'immigrés : les effectifs sont faibles et il aurait fallu soit étaler la collecte, soit échantillonner sur des vagues anciennes, moins efficaces et plus dispersées géographiquement. À cela s'ajoute le fait que l'EEC ne permettait de repérer les

---

<sup>4</sup> Jusqu'à la fin des années 2000, l'échantillon-maître s'appuyait sur le recensement de 1999. Désormais, rebaptisé Octopusse, il s'appuie sur le recensement annuel.

descendants d'immigré qu'à compter des entrants de janvier 2005. Pour constituer un échantillon utilisant les sortants de 5 années, il aurait donc fallu soit étaler la collecte de 2007 à 2010 ou 2011, soit enquêter par exemple en 2009 les sortants 2005-2009. Cette option n'a pas paru envisageable.

A fortiori, les autres enquêtes auprès des ménages dont l'effectif enquêté est inférieur, constituaient une base de sondage insuffisante. L'enquête Logement de 2006 permettait aussi de repérer les descendants directs d'immigrés. Cependant, même l'adjonction de l'échantillon ainsi constitué à celui de l'EEC ne permettait pas d'augmenter suffisamment la taille de l'échantillon et ajoutait d'autres inconvénients : épuisement d'une catégorie de population, les descendants d'immigrés, ce qui interdisait toute autre enquête sur les sortants de l'enquête Logement, et constitution d'un échantillon provenant de sources disparates avec des conséquences sur le redressement.

À l'issue de cet inventaire, il paraissait indispensable d'utiliser une ou plusieurs vagues du recensement rénové de la population (les enquêtes annuelles du recensement ou EAR) comme base initiale d'échantillonnage. Cela permettait de concilier fraîcheur des informations individuelles et taille importante de la base de sondage. Tous les échantillons ont donc été tirés à partir d'une des enquêtes annuelles du recensement, la plus récente possible : celle de janvier 2007 pour une enquête réalisée au dernier trimestre 2008.

Mais surtout, il restait à trouver un moyen de collecter l'information sur le lieu de naissance des parents, absente du bulletin individuel du recensement et nécessaire au repérage des descendants d'immigrés ou de natifs de DOM. Le recensement ne renseigne ni la nationalité des parents, ni leur lieu de naissance (pays et département). C'est grâce au recours à d'autres sources de données que les échantillons ont été constitués.

Où trouver l'information complémentaire sur les parents ?

Pour l'enquête MGIS, le tirage des descendants d'immigrés s'est fait dans l'échantillon démographique permanent. Il s'agit d'une base d'étude constituée par l'Insee pour suivre les individus nés 4 jours donnés de chaque année. Pour chacun de ces individus, elle permet de disposer (de façon anonyme) des informations collectées chaque fois qu'il est recensé ainsi que des principaux actes d'état civil le concernant (naissance, mariage, décès, divorce). Étaient ainsi sélectionnées pour MGIS des personnes qui étaient recensées en 1990 (pour prendre leur adresse à cette date) et qui étaient repérées comme enfants d'immigré en 1975

grâce aux informations du recensement de cette année-là : ils étaient alors enfants dans un ménage dont le chef de famille était immigré). L'échantillon des enfants immigrés était constitué de 2 768 fiches adresses, qui ont permis d'interroger in fine 1 956 personnes, soit 71 %. Un tiers avaient déménagé depuis le dernier recensement (1990), mais trois fois sur quatre, la nouvelle adresse a pu être retrouvée. Toutefois, cette solution était insuffisante dans le cas de TeO. En l'absence de recensement exhaustif une année donnée, il aurait fallu cumuler plusieurs vagues de recensement pour obtenir un échantillon comparable à l'enquête MGIS, de surcroît bien inférieur à nos objectifs tant par sa taille globale que par la nécessité de surreprésenter certaines origines.

La première solution alternative envisagée fût une enquête de pré-filtrage adossée à une collecte du recensement, sur le modèle de l'enquête Étude de l'histoire familiale en 1999, c'est-à-dire en incluant pour une partie des personnes recensées un volet supplémentaire collecté directement par les agents recenseurs. Une telle enquête aurait présenté un intérêt bien plus large que le seul recrutement de l'échantillon de l'enquête TeO. Cependant, pour des raisons institutionnelles et de calendrier, cette solution a été abandonnée.

La seconde possibilité, dans la lignée de la précédente était de faire une enquête de pré-filtrage auprès de personnes recensées mais de façon indépendante et quelque mois après la collecte du recensement. Cette opération était alors uniquement dédiée à l'échantillonnage. Qu'elle ait lieu par courrier ou téléphone, le coût était exorbitant puisqu'il fallait contacter entre 300 000 et 400 000 personnes (toujours sans prendre en compte les surreprésentations). Le coût restait prohibitif même en faisant une pré-enquête commune à celle réalisée pour constituer l'échantillon de l'enquête Handicap-Santé pour mutualiser les charges. De plus, la rédaction d'un questionnaire auto-rempli très court et qui permette à la fois de repérer de façon fiable les situations de handicap ou d'invalidité et l'origine géographique des parents paraissait pour le moins très délicate.

Après avoir examiné la possibilité d'interroger les enfants des immigrés interrogés, cette solution a été abandonnée car elle comportait trop de biais. Pour ne donner qu'un exemple pour chaque solution, interroger les enfants des immigrés enquêtés exclut ceux des descendants d'immigrés qui vivent en France mais dont les parents sont décédés ou ne vivent plus en France.

La solution retenue finalement a consisté à jumeler l'information sur la résidence actuelle issue du recensement et l'information sur la filiation tirée des bulletins de naissance. Elle est

en principe assez peu biaisée car fondée sur une source exhaustive lorsque l'on s'intéresse à des descendants d'immigrés ou de natifs des DOM, par définition tous nés en métropole. Elle permettait par ailleurs de réaliser la surreprésentation de certaines origines. Toutefois, cette méthode étant inédite, elle soulevait des interrogations quant à sa faisabilité juridique, technique et organisationnelle.

## La constitution des échantillons

### La constitution d'une base de sondage nominative à partir de l'EAR 2007

Cette procédure de recours au recensement rénové était alors relativement nouvelle. Plusieurs difficultés se sont donc présentées qui ont nécessité des aménagements.

#### Les limites à la dispersion géographique des enquêtés

En premier lieu, la base de sondage de l'enquête ne pouvait être constituée qu'à partir des personnes recensées en 2007 résidant dans une commune de l'échantillon-maître. Il s'agit des communes qui figurent dans une zone d'action d'enquêteur (pour que ceux-ci ne soient pas amenés à enquêter sur un territoire trop large). Or, depuis 2004, une fraction seulement de la population est recensée chaque année : l'ensemble des personnes qui vivent dans les communes de moins de 10 000 habitants recensées cette année-là (une commune sur cinq) et environ 7% de la population vivant dans des communes de plus de 10 000 habitants. Cette fraction recensée n'est pas automatiquement compatible avec les zones d'action enquêteurs (ZAE), zones couvertes par un enquêteur du réseau Insee, soit presque toutes les zones urbaines mais une partie seulement des zones rurales. Dans certaines zones d'enquêteurs, les communes recensées lors du recensement 2007 étaient peu peuplées<sup>5</sup> et cela a nécessité l'ajout de communes supplémentaires recensées mais hors ZAE. Enfin, pour assurer une bonne représentation des habitants des zones urbaines sensibles, et notamment parmi les témoins, les communes comprenant une ZUS mais situées en dehors d'une grande unité urbaine ont été incluses dans l'échantillon.

N'ont donc été incluses dans la base de sondage que les personnes résidant dans une commune de l'échantillon-maître, une commune ajoutée ou une commune hors échantillon-maître comportant une ZUS.

---

<sup>5</sup> Par exemple, pour un enquêteur en zone rurale, dont le champ d'action est composé d'une vingtaine de communes, il se peut que les quatre communes recensées cette année-là soient les moins peuplées et que la plus importante ne soit pas recensée. Il se trouve alors presque sans travail. Cette mauvaise répartition nuit par ailleurs à la représentativité de l'échantillon.

La saisie au recensement : 302 584 individus sélectionnés

Habituellement, les noms et prénoms qui figurent sur les bulletins individuels du recensement ne sont pas saisis dans le cadre de l'exploitation statistique de ce recensement, qui doit rester anonyme. Dans la procédure d'enquête standard utilisant l'échantillon-maître (lui-même extrait du recensement), les enquêteurs disposent pour contacter les ménages d'une image scannée du nom et de l'adresse de l'occupant principal au recensement, écrite par celui-ci à la main. Pour les besoins de l'enquête TeO, il était indispensable de disposer du nom et du prénom des personnes, pour permettre la consultation des actes de naissance et lors de la collecte la prise de contact avec une personne nommément désignée.

Il a donc fallu demander l'accord de la Cnil et mettre en place une saisie du nom et du prénom de la personne à interroger (pas forcément l'occupant principal).

Afin d'obtenir les effectifs recherchés tout en ayant des critères de sélection relativement simples, le choix s'est porté sur les jours de naissance, ainsi que le détaille le tableau 1. Cette procédure de saisie des noms et prénoms était déjà mise en place pour l'échantillon démographique permanent, ce qui représentait une petite économie de saisie (environ 19 000 personnes) dont l'enquête a profité. Les personnes nées un 1<sup>er</sup>, 2, 3 ou 4 octobre ont donc également été intégrées à la base de sondage dans chacun des quatre groupes.

Au total 302 584 individus étaient concernés par cette saisie.

Tableau 1 : sélection des bulletins individuels du recensement dont les noms et prénoms ont été saisis

		Lieu de naissance	Nationalité	Années de naissance (bornes incluses)	Mois de naissance	Jour de naissance	Effectif
1	Descendants potentiels sans appariement anonyme	En France métropolitaine	Toutes	1958 à 1967	Tous	20 ou 21	33 000
2	Descendants potentiels avec appariement anonyme	En France métropolitaine	Toutes	1968 à 1990	Tous	20, 21, 22, 23, 24 ou 25	210 000
3	Natifs de DOM	Dans un département d'Outre-mer (Martinique, Guadeloupe, Réunion, Guyane)	Toutes	1948 à 1990	Tous	20, 21, 22, 23, 24 ou 25	5 000
4	Immigrés	Dans un pays étranger	Non renseignée ou français(e)s par acquisition ou étranger(e)s	1948 à 1990	Tous	20, 21, 22, 23, 24 ou 25	53 000

Dans la mesure où cette étape précédait la consultation des actes de naissance, cette saisie a concerné un nombre très important de personnes nées en métropole, potentiellement



descendantes d'immigré(s) ou de natif(s) d'un DOM. L'objectif qui a présidé au calibrage de cette saisie était qu'au sein de ces 240 000 personnes (groupes 1 et 2 du tableau 1) figure après vérification un nombre suffisant de descendants d'immigré(s) ou de natif(s) d'un DOM, y compris issus des origines surreprésentées les plus rares (notamment par exemple de descendants d'immigrés venus de Turquie).

Pour les immigrés et les natifs de DOM, il s'agissait de constituer une base de sondage suffisamment large pour permettre les surreprésentations. Il aurait été trop complexe de faire figurer le pays de naissance au rang des critères de saisie. C'est pourquoi un peu plus de 50 000 noms et prénoms de personnes nées à l'étranger, étrangères ou françaises par acquisition ont été saisis<sup>6</sup>.

A l'examen plus détaillé des réponses qui figuraient sur les bulletins individuels du recensement, il est apparu que la sélection sur les critères de lieu de naissance et de date de naissance risquait d'introduire des biais pour les immigrés. Deux cas de figures notamment ont justifié l'ajout a posteriori de 4 791 personnes dans le groupe 4 de la base de sondage :

- Des immigrés qui avaient omis leur date de naissance ou déclaré être nés un 1<sup>er</sup> janvier, car dans certains pays l'état civil n'est pas aussi prégnant qu'en France et dans plusieurs préfectures le 1<sup>er</sup> janvier est alors affecté comme jour de naissance, jours qui n'avaient pas été retenus pour la saisie. Cet oubli aurait exclu du champ une proportion importante d'immigrés en provenance de certains pays (notamment le Maroc ou des pays d'Afrique subsaharienne). 3 840 sont ajoutés à ce titre<sup>7</sup>.

- Des immigrés qui n'ont pas donné de pays mais simplement une ville ou une région à l'étranger, car ce type de déclaration peut correspondre à une perception particulière de l'unité nationale de leur pays d'origine (régime plutôt fédéral ou mouvement autonomiste...). 949 sont ajoutés à ce titre<sup>8</sup>.

---

<sup>6</sup> Les personnes qui n'ont pas répondu sur leur nationalité ont également été intégrées dans la base de sondage, si elles étaient nées à l'étranger.

<sup>7</sup> Il s'agit d'un échantillon aléatoire de personnes nées à l'étranger un 1<sup>er</sup> janvier ou sans date de naissance. Il est donc représentatif de ce groupe particulier et comprend en proportions plus importantes des personnes venues de pays où l'état civil est moins stabilisé.

<sup>8</sup> A noter que la déclaration des lieux de naissance (pays ou à défaut commune) était indéfinie pour 18 000 bulletins sur 300 000. Or aussi bien pour affecter la personne dans un des quatre groupes que pour consulter en mairie le bulletin de naissance, il était indispensable de reprendre et recoder ces déclarations. Cette opération était parfois facile (plusieurs noms pour un même pays, ou le pays déclaré à la place de la commune) et parfois plus complexe. Par exemple, sans indication sur le pays de naissance, comment trancher entre Rabat-Les Trois-Seigneurs, village de l'Ariège, et Rabat, capitale du Maroc ?

Leurs noms et prénoms n'avaient pas été saisis. Lorsque ces individus ont été ensuite enquêtés, leurs noms et prénoms ont été cherchés manuellement dans leur bulletin individuel de recensement dans chaque direction régionale.

Parmi ces 55 000 personnes, un peu moins de 13 000 fiches-adresses ont été échantillonnées, ainsi qu'une réserve de 1400 supplémentaires (tableau 3). L'objectif était d'obtenir ainsi 9 300 immigrés interrogés. Environ 300 questionnaires « immigrés » supplémentaires devaient provenir de l'échantillon témoin.

### Récupérer le lieu de naissance des parents pour l'échantillon des descendants

Nécessaire à la construction des échantillons de descendants, le lieu de naissance des parents a été trouvé en consultant les actes de naissance de l'état-civil ou bulletins statistiques anonymisés.

Plus de 240 000 personnes ont fait l'objet de cette recherche (groupes 1 et 2).

Il s'agissait de limiter les recherches directes dans les registres de la commune de naissance qui constituent une opération coûteuse. C'est pour cette raison qu'un appariement anonyme a été réalisé au préalable.

Pour les personnes nées entre 1958 et 1967 (groupe 1, 33 000), aucune information préalable n'était disponible et la consultation des registres en mairie a été systématique (cf. tableau 2 étape 3).

Pour les personnes nées après 1967 (groupe 2), il était possible de consulter préalablement les bulletins anonymes de naissance, collectés par l'Insee depuis 1968, cela représente 210 000 individus sélectionnés (groupe 2 du tableau 1). Pour une personne donnée, l'ensemble des correspondants « anonymes » ou « échos » ont été recherchés, c'est-à-dire l'ensemble des personnes nées le même jour, dans la même commune, de même sexe (cf. tableau 2 , étapes 3 et 4).

Tableau 2 : Étapes du tirage de l'échantillon pour les descendants d'immigrés

1 - sélection du champ de l'enquête dans l'EAR 2007 : jour, mois et année de naissance, résidence France métropolitaine et ménage ordinaire	⇒ 300 000		
2 - sélection des personnes nées en métropole : constitution du fichier nominatif	⇒ 240 000		
3 - traitement différent des générations avant et après 1967	Nés entre 1957 et 1967 ⇒ 33 000	Nés entre 1968 et 1990 ⇒ 210 000	
4 - Appariement avec le fichier anonyme de l'Insee selon le sexe, la date et le lieu de naissance	Sans objet Pas de fichier anonyme Insee	appariement	
5 - résultat de l'appariement		Au moins 1 écho (né le même jour dans la même mairie) avec 1 parent né à l'étranger ⇒ 127 500	Tous ont 1 parent né à l'étranger ⇒ 8 500
6 - tirage avant recherche dans les mairies avec surreprésentation des origines rares			Aucun n'a de parent né à l'étranger : hors champ ⇒ 74 000
7 - recherche dans les registres en mairie	⇒ recherche systématique dans les registres en mairie ⇒ 33 000	Recherche dans les registres en mairie nécessaire ⇒ 65 000	Recherche dans les registres en mairie inutile
8 - sélection des personnes nées d'un parent né à l'étranger	⇒ 25 000		⇒ 8 500
9 - tirage de l'échantillon avec surreprésentation des origines rares	⇒ 18 400*		

\* une réserve de 1 400 Fiches adresses a également été constituée à cette étape

Lors de cet appariement, si aucun de ces échos n'était descendant d'immigrés ou de natifs de DOM, il était inutile d'organiser une recherche en mairie (74 000). À l'opposé, si tous les échos étaient descendants d'immigrés ou de natifs de DOM, nous pouvions intégrer l'individu dans la base de sondage sans plus de recherche (8 500) (cf. tableau 2, étape 5).

Restaient 127 500 personnes nées en France entre 1968 et 1990 qui avaient au moins un écho avec un parent né à l'étranger et au moins un écho sans parent né à l'étranger. Pour ces personnes, il fallait effectuer une recherche en mairie.

Cette opération préalable a également permis de surreprésenter certaines origines (cf. tableau 2 étape 6). Par exemple, la recherche en mairie a été effectuée systématiquement s'il existait au moins un descendant d'immigré(s) d'origine turque parmi les échos. Elle n'a été effectuée que pour une fraction des personnes qui avaient un correspondant ayant des parents immigrés d'origine algérienne. Ainsi, 65 000 personnes ont été tirées parmi ces 127 500, en surreprésentant les origines des parents les plus rares.

Au total, ce sont 98 000 fiches qui ont ainsi été consultées dans les registres nominatifs en mairie pour récupérer le lieu de naissance des parents (cf. tableau 2 étape 7).

Par son ampleur et la sensibilité des informations recueillies, cette opération a presque représenté une pré-enquête :

- Après l'aval de la CNIL et avec l'appui du ministère de la Justice, il était nécessaire d'obtenir de chaque procureur de la République l'autorisation nominative d'accès aux registres d'état civil pour les « enquêteurs » (des agents des différentes directions régionales de l'Insee et de la direction générale). Tous les procureurs de la République ont donné les autorisations nécessaires, permettant ainsi la constitution d'une base de sondage représentative de la France métropolitaine.
- Le protocole de collecte a combiné différentes possibilités selon la taille des échantillons à consulter et les possibilités d'organisation : déplacement en mairie lorsque l'échantillon à consulter était supérieur à 20, courriers lorsqu'il était inférieur. Les informations collectées étaient ensuite saisies informatiquement et centralisées.

2 % des personnes échantillonnées n'ont pu être retrouvées, et pour 9 %, l'état civil issu du recensement était erroné, généralement c'est l'orthographe du nom ou du prénom qui a été modifiée. À l'issue de ces opérations, 25 000 individus nés en France Métropolitaine dont un parent était né à l'étranger ont été identifiés. Ces 25 000 individus et les 8 500 individus dont tous les échos avaient un parent étranger ont constitué la base (cf. tableau 2 étape 8) dans laquelle a été tiré l'échantillon final : 18 400 fiches adresses pour les descendants de parents nés à l'étranger ainsi qu'un échantillon de réserve d'environ 1 400 fiches-adresses. Selon la même méthode un échantillon de 1 200 individus descendants de parents nés dans un DOM<sup>9</sup> a été constitué, avec une réserve de 124 fiches-adresses

Pour les descendants d'immigrés, c'est ensuite le questionnaire de l'enquête qui va permettre de distinguer parmi cet échantillon de descendants d'une personne née à l'étranger, ceux qui sont nés d'un parent de nationalité étrangère à la naissance (descendant d'immigré) de ceux qui sont nés d'un parent né à l'étranger mais de nationalité française (descendants de rapatriés de l'Algérie, descendants d'expatriés...). Dans ce dernier cas, pour des raisons de coût, le protocole de l'enquête prévoit que l'enquête est interrompue (l'entretien est réalisé « court »). Pour cette raison, le nombre de fiches-adresses incluses dans l'échantillon est plus élevé que pour les immigrés. À la prévision des échecs habituels (impossibles à joindre, refus, absent de longue durée, inaptitude à répondre, etc.) s'ajoutait celle des entretiens dits « courts »

---

<sup>9</sup> Comme pour les descendants de parents nés à l'étranger, une réserve de 120 FA supplémentaires est également tirée à cette étape.

interrompus au bout d'une quinzaine de minutes car aucun des parents de l'enquêté n'est immigré. Il était estimé qu'un tiers des questionnaires collectés seraient ainsi écourtés.

Au total, ce sont ainsi environ 43 000 fiches-adresses, en comptant la réserve, qui ont été livrés aux directions régionales de l'Insee afin de réaliser la collecte.

Tableau 3 : nombre de fiches adresses et objectifs de répondants par sous échantillons

	Fiches adresses initiales	Fiches adresses en réserve	Nombre de répondants attendus	Nombre de répondants attendus après réallocation	
Immigrés	12982	1374	9300	9600	Immigrés
Natifs de DOM	1070	126	800	800	Natifs de DOM
Descendants d'immigrés (d'un parent né à l'étranger)	18430	1379	9300*	9600*	Descendants d'immigrés (d'un parent né étranger à l'étranger)
Descendants d'un natif de DOM	1207	121	800	800	Descendants d'un natif de DOM
Témoin	5598	1001	3000	2400	Population majoritaire
Ensemble	38988	4300	23600	23600	Ensemble

\* et 4500 descendants de rapatriés attendus : nés français à l'étranger : questionnaires courts

Au sein de l'échantillon témoin, il était attendu qu'environ 300 personnes soient immigrés et 300 descendants d'immigrés, ce qui laissait 2400 personnes de la « population majoritaire » et portait à 9600 les effectifs des échantillons d'immigrés et de descendants d'immigrés.

## Le déroulement de la collecte

La spécificité de la population qui était visée par l'enquête a donc nécessité des modalités très originales pour la construction de l'échantillon. Par rapport à la plupart des enquêtes menées par l'Insee, elle a également impliqué des aménagements particuliers sur divers points du protocole habituel.

La première contrainte était de limiter la dispersion géographique des enquêtés, pour des raisons de coût mais aussi pour éviter de recruter des enquêteurs supplémentaires, moins rodés aux enquêtes ménages que ceux qui travaillent couramment pour l'Insee. Compte tenu de la taille de l'enquête, des difficultés spécifiques d'échantillonnage qui risquaient d'augmenter le nombre des fiches-adresses à exploiter pour obtenir un nombre donné de questionnaires, il y avait un risque que l'enquête représente une charge trop importante, surtout si elle était réalisée en même temps qu'une autre enquête. C'est pour cette raison, que la collecte s'est déroulée en deux « vagues », la première auprès des immigrés, des natifs d'un DOM et des témoins, la seconde auprès des descendants.

La population visée par l'enquête n'est pas répartie sur le territoire métropolitain comme l'ensemble de la population : elle est beaucoup plus souvent urbaine et plus concentrée dans quelques régions (schématiquement l'Île-de-France et les régions de l'Est et du Sud-Est).

En outre, l'échantillon comporte davantage d'enquêtes réalisées dans les zones urbaines sensibles.

Enfin, certains immigrés, notamment ceux arrivés en France récemment, n'ont pas un niveau en français suffisant pour permettre la réalisation de l'enquête dans de bonnes conditions. Un dispositif de recours à des interprètes a été mis en place.

Malgré ces particularités, le repérage, le contact puis l'entretien ont été assurés par le réseau des enquêteurs de l'Insee dans le cadre général habituel d'un entretien en face à face. Une lettre avis annonçant l'enquête a été envoyée à tous les enquêtés dans un premier temps. Puis l'enquêteur a cherché à prendre contact directement avec l'enquêté, la plupart du temps par téléphone afin de lui fixer un rendez vous pour un entretien en face à face. Ce dernier s'est déroulé alors en règle générale chez l'enquêté avec l'aide d'un questionnaire électronique et une saisie directe des réponses sur ordinateur portable par l'enquêteur (CAPI, collecte assistée par informatique).

## Les difficultés attendues et les solutions apportées

### Une concentration géographique en Île-de-France

La population des immigrés et des descendants est fortement concentrée dans les zones urbaines et particulièrement dans la région Île-de-France (37% de l'ensemble des fiches adresses). Cela constituait une charge de travail très importante pour le réseau d'enquêteurs de l'Insee, compte tenu d'un calendrier initialement réduit à trois mois. Tout d'abord, les fiches adresses ont été livrées en deux vagues : la première vague composée des fiches adresses des sous échantillons d'immigrés, de natifs d'un DOM et des témoins a été livrée le 1<sup>er</sup> septembre 2008 aux directions régionales. La seconde vague regroupant les fiches adresses des descendants a été livrée le 6 octobre. Pour les régions les plus sollicitées (particulièrement l'Île de France) le réseau habituel des enquêteurs de l'Insee a été renforcé avec le recrutement de nouveaux enquêteurs formés spécifiquement pour la réalisation de TeO. En outre, la population recherchée vit davantage dans les quartiers défavorisés, dans les zones sensibles qui peuvent être plus délicates pour les enquêteurs de l'Insee. Cependant, des précédents existent. Certaines enquêtes comprennent même des échantillons ZUS supplémentaires (enquête Histoire de vie, enquête Information et vie quotidienne) et la rémunération des

enquêteurs est relevée d'une prime spécifique pour les entretiens réalisés dans une ZUS (16% des fiches adresses de l'échantillon), afin de prendre en compte les difficultés spécifiques à cette situation. En Île de France, les enquêteurs ont suivi un module complémentaire de formation spécifique pour enquêter dans une zone urbaine sensible.

Une enquête individuelle auprès d'une population plutôt mobile

S'agissant d'une enquête auprès d'individus avec des caractéristiques très spécifiques, TeO présente une difficulté supplémentaire par rapport aux enquêtes plus fréquentes auprès de ménages. Il ne « suffit » pas de repérer un logement, ni même un ménage mais bien un individu en particulier dans le ménage. La concentration géographique de l'échantillon dans des zones fortement urbanisées et la jeunesse des enquêtés (en particulier des descendants d'immigrés) ne facilitent pas non plus le repérage des enquêtés : en milieu urbanisé, la sociabilité de voisinage est plutôt plus faible et les dispositifs de sécurité à l'entrée des immeubles sont quasiment généralisés. En outre, les jeunes actifs passent peu de temps dans leur logement, et en changent plus souvent. En effet, entre l'adresse à l'enquête annuelle du recensement 2007, qui a servi de base de sondage et l'enquête TeO qui a commencé en septembre 2008, la personne qui doit être enquêtée a pu déménager, et parfois même plusieurs fois.

C'est pourquoi le suivi des adresses en cas de déménagement a fait l'objet d'un protocole plutôt innovant utilisant un fichier d'adresses fourni par La Poste qui permettait de retrouver les enquêtés qui ont fait suivre leur courrier à la suite d'un déménagement depuis leur recensement (en janvier 2007).

Si lors de la prise de repérage, l'enquêteur recueillait des informations laissant penser que la personne avait déménagé, il devait essayer de trouver sa nouvelle adresse (tableau 4). Deux cas se présentaient alors :

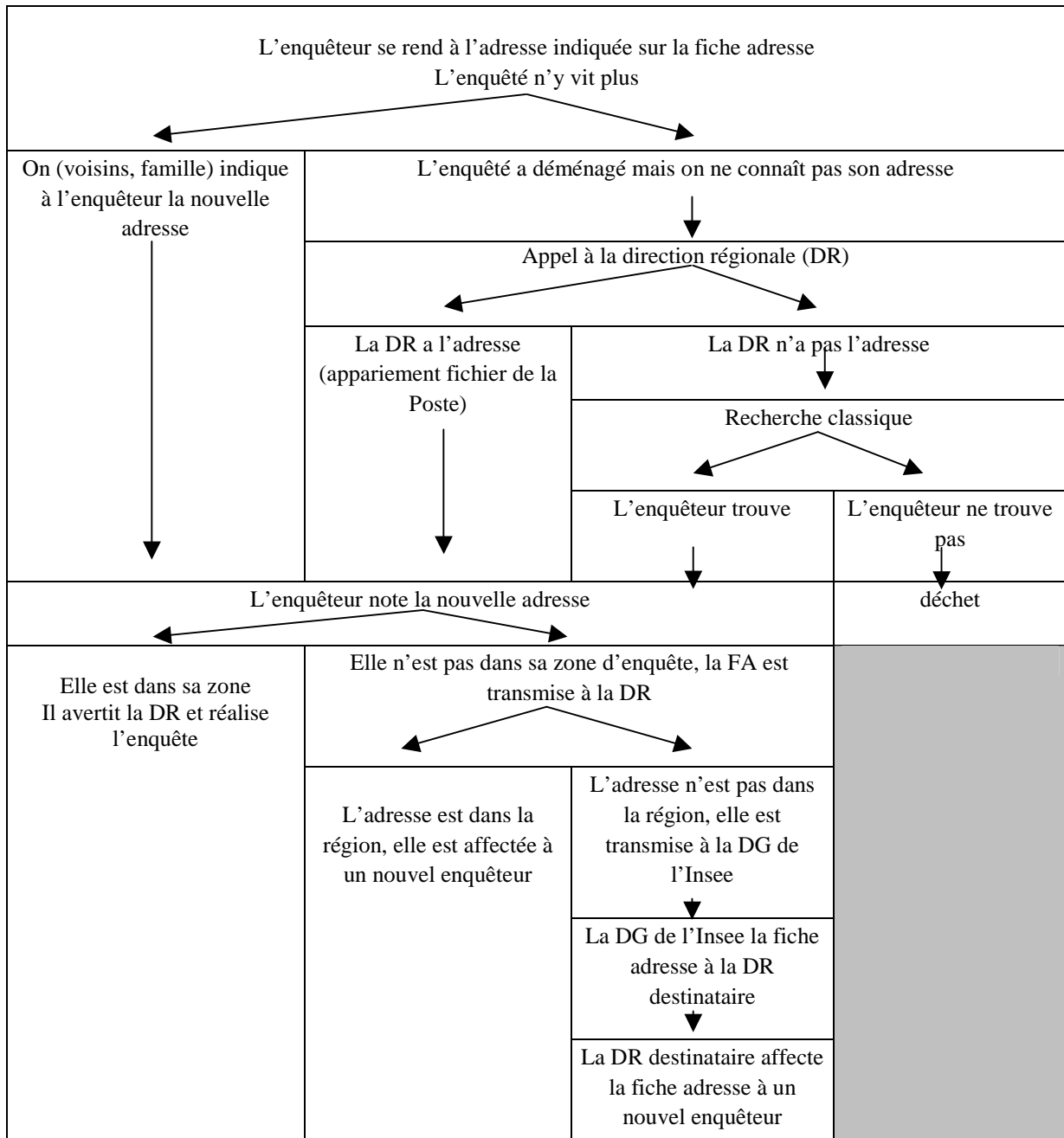
- La personne avait déménagé mais l'enquêteur avait obtenu sa nouvelle adresse en se rendant à son ancienne adresse : par exemple l'enquêté jeune avait quitté le logement des parents pour prendre un logement indépendant et les parents avaient fourni la nouvelle adresse.
- l'enquêteur avait seulement des présomptions de déménagement : le nom n'était pas sur les boîtes à lettres, informations glanées dans le voisinage... Dans ce cas, l'enquêteur avait pour consigne de se renseigner en première intention auprès de sa direction régionale, pour vérifier si une nouvelle adresse n'était pas disponible dans le fichier de La Poste. Si le recours à ce fichier était infructueux, l'enquêteur devait essayer de retrouver la nouvelle adresse par les

moyens habituels (en insistant auprès du voisinage, en se renseignant auprès de la mairie... et aussi en recherchant sur internet ).

Si l'enquêteur retrouvait la nouvelle adresse, le traitement de la fiche adresse dépendait de la localisation : soit la nouvelle adresse était toujours dans la zone de l'enquêteur et après signalement à la direction régionale de l'Insee, l'enquêteur allait l'interroger à la nouvelle adresse ; soit la nouvelle adresse était hors de la zone de l'enquêteur, la fiche adresse était alors transférée à la direction générale qui assurait la transmission à la direction régionale correspondante pour être finalisée par un autre enquêteur.



Tableau 4 : protocole pour retrouver l'adresse d'un enquêté qui a déménagé depuis le recensement



#### Les résultats mitigés du suivi des adresses des enquêtés en cas de déménagement

Dans les faits, la mobilité des personnes de l'échantillon a constitué un facteur important de déperdition. D'une part, entre le recensement et l'enquête TeO, soit au plus 18 mois, environ 28% des personnes ont changé d'adresse et d'autre part, le protocole prévu pour le suivi des adresses n'a permis de retrouver l'adresse que dans un cas sur deux. Cette difficulté d'entrer en contact avec la personne désignée a d'ailleurs été exprimée largement par les enquêteurs.

Elle constitue la principale cause de déperdition d'entretiens. Sur le seul champ des sédentaires (qui n'ont pas déménagé entre janvier 2007 et septembre/février 2009) la part des entretiens réalisés rapportés au nombre des fiches adresses traitées dans le champ de l'enquête (le taux de réponse) atteint 70%. Il atteint seulement 36% pour ceux qui ont déménagé mais il s'approche de 79% lorsque l'adresse retrouvée est dans le champ (3814/(5947-1147) dans le tableau 5).

Tableau 5 : Taux de réponse à l'enquête TeO selon la mobilité de l'enquêté

	Effectifs	Adresse retrouvée	Hors champ*	Questionnaires réalisés**	Réalisés/dans le champ**
Fiches adresses non traitées**	782				
Fiches adresses traitées, dont	42506	36758	1147	25266	61%
- Personnes sédentaires	30 811	30811	0	21452	70%
- Personnes mobiles	11 695	5947	1147	3814	36%

Source : enquête Trajectoires et Origines, 2008, Ined-Insee.

\*sont exclus ceux qui sont partis à l'étranger ou en institution, décédés, les logements impossibles à identifier.

\*\* questionnaires longs et courts.

\*\*\* fiches adresses qui n'ont pas été traitées faute de temps (voir calcul de la pondération).

Le taux de suivi des adresses apparaît corrélé avec la densité urbaine. Lorsqu'une personne « mobile » vivait en 2007 dans une commune rurale, son adresse a pu être retrouvée 3 fois sur quatre. Cela n'a été le cas que 6 fois sur 10 dans les communes moyennes et seulement 4 fois sur 10 dans l'agglomération parisienne. Cela explique en partie les différentiels de taux de réponse global selon la densité urbaine (tableau 6) .

Tableau 6 : déménagement, adresse retrouvée et taux de réponse

Taux de réponse : entretiens réalisés/ fiches adresses traitées dans le champ*	rurales	<20000 h.	De 20000 h. à 200000 h.	>200000 h.	Agglomération parisienne	France métropolitaine
Taux de réponse global	77%	72%	67%	62%	53%	61%
Parmi les sédentaires	82%	77%	74%	74%	61%	70%
Parmi ceux qui ont déménagé	59%	52%	45%	37%	24%	36%
Part des déménagements**	22%	20%	27%	33%	24%	28%
Proportion d'adresses de destination** retrouvées en cas de déménagement	75%	70%	59%	50%	42%	51%
Entretiens réalisé / adresses de destination retrouvées dans le champ	83%	80%	82%	81%	72%	79%

\* champ : personnes qui sont restées en ménages ordinaires en France métropolitaine

\*\*yc les adresses sorties du champ

Champ : fiches adresses traitées et adresse dans le champ de l'enquête

Source : enquête Trajectoires et Origines, 2008, Ined-Insee.

Compte tenu du caractère inhabituel du dispositif, les enquêteurs devaient décrire les moyens mis en œuvre pour retrouver l'adresse des personnes mobiles. Ils disent avoir mobilisé en moyenne près de 2 moyens différents. Les enquêteurs ont recouru au fichier de la poste dans 40% des déménagements. Dans 53% des cas, ils ont cherché dans le voisinage, dans 14% ils ont contacté la mairie et dans 44 % ils ont recherché dans un annuaire ou bien sur internet<sup>10</sup>. D'autres moyens (famille en cas de décohabitation, ex-conjoint en cas de séparation ou nouveaux occupants du logement) ont été mobilisés dans 30% des cas.

Le nombre de moyens mis en œuvre est le plus élevé pour les villes moyennes et grandes et le recours au fichier de la poste et à la recherche dans les annuaires et sur internet est croissant avec la densité urbaine de l'adresse. En revanche, les enquêteurs prennent plus souvent contact avec la mairie dans les communes rurales ou les petites villes.

En cas de recours aux différents moyens, le taux de réussite (l'adresse a été retrouvée grâce à ce moyen) atteint 11% pour le fichier de la poste, 25% pour le voisinage et 18% pour le contact en mairie et la recherche en annuaire ou sur internet. Quel que soit le moyen mobilisé, le taux de réussite diminue avec l'urbanisation de la zone d'enquête.

Au total, lorsque l'adresse a été retrouvée à la suite d'un déménagement (50% des cas), c'est pour 9% seulement des cas grâce au fichier de La Poste. Le voisinage a fourni la nouvelle

<sup>10</sup> Sur le fichier des enquêtés au 31 janvier 2009 (Algava et Lhommeau 2009).

adresse dans 25% des cas, la mairie pour 5 % et la recherche sur internet ou dans les annuaires a permis de pister 14% des déménagés. Dans 25% des cas, l'adresse a été retrouvée par un autre moyen. Enfin pour 22% des adresses retrouvées, l'enquêteur n'a pas précisé le moyen mobilisé.

Plusieurs facteurs peuvent expliquer cette relativement faible performance du fichier de La Poste. Cela peut être dû à un effet de sélection de la population suivie (faire suivre son courrier est coûteux). Ensuite, l'appariement nécessite une orthographe strictement identique des noms (alors que l'orthographe des noms d'origine étrangère fait beaucoup plus souvent l'objet d'erreurs de saisie). Enfin, le suivi des adresses s'arrêtait en juin 2008. Or, la collecte a été prolongée jusqu'en février 2009. Un rafraîchissement aurait été souhaitable. En dépit de son faible rendement, ce dispositif a été apprécié des enquêteurs d'après les bilans en directions régionales. Il constituait une forme d'encouragement de ceux-ci dans leur recherche d'une nouvelle adresse.

Une population enquêtée pas toujours parfaitement francophone

Pour une partie de la population immigrée et notamment pour les plus récemment installés en France, la langue française n'est pas toujours parfaitement maîtrisée, ce qui risquait de détériorer la qualité de l'enquête. La prise de contact peut être empêchée quand l'enquêté ne sait pas lire la lettre avis, il ne comprend pas l'enquêteur qui cherche à le joindre au téléphone. Mais aussi la réponse peut être biaisée lorsque l'enquêté ne comprend pas, ou bien comprend mal une question. Cela risque aussi d'augmenter la non-réponse et ce, d'autant que les questions abordent un registre sensible ou intime (comme par exemple la contraception) ou bien plus abstrait (identité).

Pour limiter ces biais, divers supports ont été traduits en dix langues (arabe, allemand, anglais, turc, espagnol, portugais, mandarin, russe, vietnamien, serbo-croate) :

- la lettre-avis annonçant la venue d'un enquêteur de l'Insee. Pour le test, la traduction apparaissait au verso de la lettre en français ; en définitive, suite aux tests, la lettre avis a été traduite dans plusieurs langues figurant sur une seule page. Cette feuille accompagnait toujours la lettre avis en français sans distinction des origines des enquêtés. Ceci a été apprécié par les enquêtés d'origine étrangère, mais a pu surprendre les Français sans ascendance migratoire directe ;
- une fiche avec les mots-clés, les concepts importants, les questions un peu compliquées afin de garantir une homogénéité des notions abordées ;

- les cartes codes, détaillant les modalités des questions fermées.

En outre, pour le seul échantillon des immigrés, l'enquêteur de l'Insee pouvait recourir à un service d'interprétariat. Ce service a été sous-traité à des associations ou des sociétés de service. Le recours à un interprète pouvait intervenir dès la prise de rendez-vous au téléphone ou bien lors du face à face. En effet, à la fin du module « langue », le niveau de l'enquêté dans la maîtrise du français est évalué avec les questions suivantes : « éprouvez-vous des difficultés en français pour donner votre nom, votre numéro de téléphone », « répondre à des questions simples sur vos activités de tous les jours », « demander un renseignement ou un service », « participer à une conversation », « raconter quelque chose en détail ». Si l'enquêté éprouvait des difficultés pour au moins un item, il était proposé à l'enquêté de recourir à un interprète. Dans ce cas, l'entretien se fait à trois, l'enquêté, l'enquêteur et l'interprète, soit un dispositif différent de l'enquête MGIS de 1992 pour laquelle des entretiens avaient été menés par des interprètes formés à l'enquête.

En fait, 128 interviews ont été réalisées avec l'appui d'un interprète d'un service d'interprétariat, soit 1,5 % des répondants immigrés. Ce recours concerne 4 % de ceux qui ont déclaré au moins une difficulté dans la maîtrise du français. Ce taux faible de recours a plusieurs explications : le repérage des difficultés en Français retient une définition plutôt extensive : ce filtrage retient 3 150 personnes alors que selon l'avis de l'enquêteur 1 911 enquêtés avaient des difficultés à comprendre le français, dont 1 583 issus du sous-échantillon des immigrés, le seul éligible à la mesure des difficultés en français dans le module langue. Ensuite la qualité du dispositif mis en place au niveau régional était hétérogène : on note un recours plus élevé dans les régions Rhône-Alpes, Basse-Normandie, Alsace et Île-de-France. Enfin, bien que le protocole prévoyait ce recours à un interprète assermenté sans lien avec l'enquêté afin de faciliter les réponses aux questions les plus sensibles, il y a eu aussi parfois l'appui d'interprètes appartenant à la famille ou au voisinage : 748 entretiens ont été réalisés avec l'aide d'une personne de la famille, 46 par un ami ou une connaissance de l'enquêté, 40 par une autre personne et enfin l'enquêteur a lui-même traduit tout ou une partie du questionnaire dans 122 cas. Au total, 1 084 entretiens ont nécessité une traduction des questionnaires (dont 355 pour l'ensemble du questionnaire). Dans tous les cas, les enquêteurs ont apprécié les cartes codes et les mots clés traduits qui ont permis de mener à bien l'enquête.

## L'échantillon final : objectifs et résultats

Malgré ces différents dispositifs mis en place pour tenir compte de la spécificité de l'échantillon enquêté, le taux de réponse des premières fiches adresses livrées s'est révélé inférieur aux attentes. Aussi, pour s'approcher de l'objectif de taille de l'échantillon répondant, deux mesures complémentaires ont été prises avec le déclenchement d'échantillons de réserve (le premier le 20 octobre, le deuxième le 4 décembre) et une prolongation de deux mois de la période de collecte jusqu'à fin février 2009 au lieu de fin décembre 2008 comme prévu initialement.

Tableau 7 : des fiches adresses livrées au fichier de diffusion

En entrée	43288 fiches adresses (FA) initiales			
Traitement des fiches	42 506 traitées	-782 (-1,8 % FA initiales) non traitées		
Fiches adresses dans le champ	41 359 dans le champ	-1147 (-2,7% FA initiales) hors champ		
Entretiens valides et complets	21 544 questionnaires attendus « longs »	+217 questionnaires « longs » attendus « courts »	-3 328 questionnaires attendus « courts »	-16 270 non répondants ou incomplets
Échantillon final	21 761 questionnaires exploitables			

Un peu moins de 2% des fiches adresses non traitées par le réseau d'enquêteurs

Sur le terrain, en dépit de la prolongation de la période de collecte au début de l'année 2009, 782 fiches adresses (FA) n'ont pas été traitées faute de temps. Cela signifie que l'enquêteur n'a pas même effectué un repérage de l'adresse de l'enquêté. Ces lacunes représentent 1,8% des lots livrés. Cette part de fiches adresses non traitées augmente significativement avec la tranche d'unité urbaine du lieu de résidence (tableau 8). Elle est aussi plus élevée pour les sous-échantillons d'immigrés et de natifs d'un DOM. Elle est quasiment deux fois plus élevée pour les FA en ZUS (3,1% contre 1,6%) et trois fois plus élevée en zone franche urbaine (4,9% contre 1,6%). Il est possible que les fiches adresses les plus faciles à traiter aient été exploitées en priorité. Toutefois cette part des fiches adresses non traitées ne dépasse 5% que pour une seule région, la Haute Normandie..

Tableau 8 : Part des fiches adresses non traitées selon la tranche d'unité urbaine de l'adresse initiale

%	France métropolitaine	Communes rurales	Communes < 20000 hab.	Communes de 20000 à 200000 hab.	Communes >200000 hab.	Agglomération parisienne
Non traitées	1,8	0,6	0,8	1,0	1,5	2,9

source : Enquête Trajectoires et Origines, Ined-Insee, 2008

Au total, 42 506 fiches adresses ont donc été traitées sur le terrain.

Pour 3% des fiches adresses, les individus, les logements sont sortis du champ de l'enquête. Parmi ces fiches adresse traitées, l'enquêté n'a pas été contacté pour un peu moins d'un millier (tableau 7) car il était sorti du champ de l'enquête. Il s'agit des personnes décédées, de celles parties en institution ou parties à l'étranger. Les fiches adresses qui correspondent à des logements impossibles à repérer sur le terrain sont également considérées comme hors champ et donc éliminées. 41 359 fiches adresses de l'échantillon ont donc été traitées et considérées dans le champ de l'enquête.

Un taux de réponse de 60% parmi les fiches adresses dans le champ

À l'issue du passage sur le terrain, on définit trois catégories selon la réussite de l'entretien :

- la non-réponse affecte 16 270 fiches adresses en fin de la collecte. Cela comprend pour l'essentiel des personnes impossibles à joindre, mais aussi des refus de réponse complets et une centaine de cas où l'entretien a été accepté mais l'enquêté l'a stoppé avant la fin. Ainsi, la non-réponse atteint 4 fiches adresses traitées dans le champ sur 10.
- Le tirage de l'échantillon a permis de repérer les descendants d'un parent né à l'étranger mais pas la nationalité à la naissance de ce parent. C'est le passage de l'enquête qui permet cette précision. Pour des raisons de coûts, le protocole initial de la collecte de l'enquête TeO prévoyait pour les sous-échantillons spécifiques des descendants d'une personne née à l'étranger (sous-échantillon 3) un questionnaire court lorsque l'enquêté n'était pas un descendant d'un immigré mais descendant d'un parent né Français à l'étranger. C'est le cas de 3 328 fiches adresses, pour lesquelles l'enquêté a accepté l'entretien mais celui-ci a été interrompu par le protocole de l'enquête à la fin du module 3 du questionnaire (nationalité et origine des parents) :
  - l'enquêté est bien celui recherché ;
  - l'enquêté est né entre 1958 et 1990 ;
  - l'entretien est un entretien « court » (sous-échantillon des descendants de personnes nées à l'étranger). Dans un premier temps les entretiens « courts » ne sont pas éliminés : ils

ont répondu mais leur questionnaire a été stoppé par le protocole de l'enquête (car leur parent né à l'étranger n'est pas immigré).

- l'entretien comporte les informations minimales permettant de le classer selon ses origines : sexe, l'année de naissance, le lieu de naissance, la nationalité à la naissance et pour ceux qui sont nés à l'étranger, le pays de naissance.

- Des questionnaire réussis « longs » : l'enquêté a été contacté, il a accepté l'entretien et le questionnaire a été passé dans son intégralité. 21 761 questionnaires remplissent ainsi les conditions suivantes:

- l'enquêté est bien celui recherché ;

- l'enquêté est né entre 1948 et 1990 ;

- l'entretien est complet ;

- l'entretien comporte les informations minimales permettant de classer l'enquêté selon ses origines : sexe, l'année de naissance, le lieu de naissance d'ego, la nationalité à la naissance et pour ceux qui sont nés à l'étranger, le pays de naissance.

Cela constitue l'échantillon exploitable et pondéré de l'enquête TeO. Le taux de réponse est le plus élevé pour les témoins et le plus bas pour les descendants de natifs d'un DOM pris dans leur ensemble. Les échantillons de réserve, déclenchés plus tardivement ont globalement des taux de réponse plus bas.

Il faut noter qu'une erreur de logistique dans un sous-échantillon de réserve n'a pas stoppé en questionnaires « courts » les entretiens de certains descendants d'un parent né à l'étranger non immigré. Sans cette erreur, 217 questionnaires du sous échantillon de réserve auraient connu le même sort que le sous échantillon initial et auraient été stoppés (attendus « courts » et réalisés « longs » dans le tableau 7).

In fine un échantillon de 21 800 entretiens

Au total environ 21 800 entretiens longs ont été recueillis soit 92 % de l'objectif fixé initialement. Si pour les immigrés, les objectifs initiaux ont été globalement atteints, il a été beaucoup plus difficile que prévu d'atteindre les descendants, en particulier ceux issus de courants migratoires plus récents : Turquie et Afrique Subsaharienne (tableau 9). Plus jeunes, ils sont à la fois moins souvent présents dans leur logement où on cherche à les joindre et à la fois plus mobiles et donc plus enclins aux déménagements.



Tableau 9 : taille des échantillons, attendue et réalisée

	Objectif	Réalisé	Réalisation / objectif En %
Immigrés, dont	9600	8456	88%
Asie du Sud-est	800	774	97%
Turquie	800	830	104%
Sahel occidental	800	665	83%
Afrique centrale et du Golfe de Guinée	800	736	92%
Descendants d'immigré(s), (18-50 ans)* dont	9600	8110	84%
Asie du Sud-est	800	573	72%
Turquie	800	447	56%
Sahel occidental	800	480	60%
Afrique centrale et du Golfe de Guinée	800	333	42%
Natifs de DOM	800	712	89%
Descendants de natifs de DOM (18-50 ans)*	800	650	81%
Groupe majoritaire	3400	3781	112%
Total*	23600	21761	92%

\* il y a également en plus 52 observations correspondant à des descendants d'immigrés ou de natifs d'un DOM âgés de 51 à 60 ans.

source : Enquête Trajectoires et Origines, Ined-Insee, 2008

## La pondération de l'échantillon

Entre le tirage de l'échantillon dans l'EAR2007 (43 288) et l'échantillon final (21 761 observations), la collecte s'est traduit, on l'a vu, par une forte déperdition que le calcul de la pondération a prise en compte pour maintenir la représentativité de l'enquête (tableau 10).

Tableau 10 : les principales étapes de la pondération

En entrée	43 288 fiches adresses tirées dans l'enquête annuelle du recensement (EAR) 2007	Poids bruts
Traitement des fiches	42 506	Affectation du poids des FA non traitées aux FA traitées
Fiches adresses dans le champ	41 539	Abandon du poids des FA hors champ
Non-réponse	25 089 questionnaires valides « courts et « longs »	Estimation de modèles de la non-réponse par sous échantillon et correction de la non-réponse
Traitements spécifiques au sous-échantillon des descendants d'un parent né à l'étranger	7 752 questionnaires courts et longs de ce sous échantillon	Calage sur les marges du sous-échantillon initial de l'EAR2007 Correction de l'erreur de protocole pour l'échantillon de réserve du sous échantillon de descendants d'un parent né à l'étranger
Partage des poids	21 761 questionnaires longs	Prise en compte de la possibilité pour un individu d'être tiré parmi deux sous-échantillon
Calage sur marges de l'EAR 2008	21 761 questionnaires longs	Modification de la pondération telle que quelques totaux (marges) d'une source de référence (l'EAR) se retrouvent dans l'enquête.

## Les poids bruts de l'échantillon initial

L'échantillon de l'enquête TeO a été tiré dans l'enquête annuelle du recensement (EAR) de la population de 2007 de manière à être représentatif de l'ensemble des personnes vivant en France métropolitaine dans un ménage ordinaire, nées entre 1948 et 1990 pour les immigrés, les natifs des DOM et le sous échantillon toutes origines. Il est représentatif des personnes vivant en France métropolitaine dans un ménage ordinaire, nées entre 1958 et 1990 pour les descendants d'immigrés ou d'un natif d'un DOM.

Les poids bruts des 43 288 fiches adresses constituant l'échantillon sont le produit des probabilités de tirage associées aux différentes étapes :

- La probabilité d'avoir été recensé en 2007<sup>11</sup>, dans un logement ordinaire d'une des communes couvertes par l'échantillon TeO.
- Celle d'appartenir à la « saisie TeO » c'est-à-dire d'être né un 20, 21, 22, 23, 24 ou 25 ou un 20 et 21 (selon les générations et le lieu de naissance) ou à la saisie « échantillon démographique permanent » pour les personnes concernées) (tableau 1).
- Celle d'avoir été recherché en mairie puis sélectionné dans l'échantillon pour les descendants selon leur origine (comme on l'a vu, les taux de tirages sont différenciés à ce stade pour obtenir la surreprésentation de certaines origines).
- Celle d'avoir été sélectionné pour les immigrés, puisque cela est également différencié selon l'origine.

## Correction du poids des fiches adresses non traitées sur le terrain

Cette étape consiste à réaffecter aux fiches adresses qui ont été traitées sur le terrain le poids de celles qui ne l'ont pas été. À l'exception de la région Haute Normandie, la part des fiches adresses non traitées est très faible et la correction du poids des fiches adresses non traitées est réparti uniformément sur les FA traitées (sans distinction de la région), à sous échantillon donné (immigrés, natifs d'un DOM, descendants d'un immigrés et descendants d'un natif de DOM). Pour la région haute Normandie, la correction est plus affinée. Elle croise les critères qui expliquent bien la probabilité pour une fiche d'être traitée dans cette région : le département, la qualification de la zone urbaine et le sous-échantillon. À l'issue de cette opération, les 42 506 fiches traitées redeviennent représentatives de la population initialement tirée dans l'EAR.

---

<sup>11</sup> Ou en 2006 pour la partie d'échantillon provenant de l'EDP.

## L'abandon du poids des fiches adresses sorties du champ de l'enquête

Parmi ces fiches traitées, 41 539 fiches adresses sont finalement considérées dans le champ de l'enquête. Celles qui sont considérées hors champ (parties en institution, à l'étranger,...) sont abandonnées. Elles ne sont plus représentatives des personnes vivant dans un ménage ordinaire en France métropolitaine.

## Estimation d'un modèle de la non-réponse totale en fonction des caractéristiques du logement et des individus

À l'issue de la collecte 25 089 entretiens (« courts » et « longs ») ont été validés. Il s'agit alors de corriger leur poids de façon à leur réaffecter le poids des personnes qui ont refusé de répondre à l'enquête. La correction de la non-réponse totale consiste à estimer sur le champ des répondants et des non répondants, la probabilité de répondre en fonction des caractéristiques disponibles dans la fiche adresse. Le poids des répondants est ensuite corrigé par l'inverse de la probabilité de réponse.

Ainsi le sous-échantillon de la base de sondage constitue un premier critère corrélé à la non-réponse. De nombreuses autres caractéristiques sont testées dans l'estimation du modèle logit de la non-réponse. C'est le cas de caractéristiques du logement au moment du recensement mais aussi de caractéristiques sociodémographiques des individus (année naissance, sexe, pays de naissance des parents...).

L'estimation de la probabilité de réponse sur l'ensemble des fiches adresses traitées dans le champ avec ces différents critères est menée séparément sur chacun des sous-échantillons : immigrés, natifs d'un DOM, descendants d'une personne née à l'étranger, descendants d'un natif de DOM et enfin échantillon témoin. Les dimensions explicatives de la non-réponse sont de différents ordres : pour certaines liées au dispositifs de l'enquête (échantillon initial ou échantillon de réserve), pour d'autres liées aux caractéristiques du logement (nombre de personnes habitant le logement, tranche d'unité urbaine et région) ou bien liées à des caractéristiques individuelles (âge et déménagement récent, sexe, diplôme, statut d'activité et origine géographique). L'échantillon des 25 089 répondants devient ainsi représentatif de l'ensemble du champ de l'enquête.

## Calage sur les marges de l'échantillon des descendants de parents nés à l'étranger

À la différence des échantillons des natifs d'un DOM ou des immigrés pour lesquels il existe une source de référence - le recensement de la population - qui permet d'étalonner l'enquête, il n'existe pas de marges de référence pour les échantillons des descendants. À défaut, un calage du poids des répondants des échantillons des descendants d'un parent né à l'étranger ou dans un DOM est mené en utilisant comme marges les sous-échantillons complets (répondants et non répondants) correspondants sur les poids bruts à l'EAR2007 selon un certain nombre de critères : la région de résidence au moment du recensement, la pyramide des âges, le pays de naissance du père, le pays de naissance de la mère, le diplôme, la nationalité, la conjugalité, le statut matrimonial, le logement HLM, le nombre de personnes du ménage, le nombre de pièces du logement, la situation principale par rapport à l'emploi, le type de contrat de l'emploi, la position de l'emploi occupé, la tranche d'unité urbaine du logement, le statut d'occupation du logement, la zone urbaine sensible.

## Traitement particulier des descendants d'un parent né Français à l'étranger

Comme cela a été exposé précédemment, une erreur de logistique a permis finalement de recueillir des questionnaires « longs » pour tous les répondants du sous-échantillon de réserve des descendants d'immigrés. Retenir ces (217) entretiens comme répondants revient à les considérer comme représentatifs de l'ensemble des personnes nées en France descendantes d'une personne née Française à l'étranger. Les conserver permet d'une part d'augmenter la taille de l'échantillon de la population majoritaire et d'autre part de constituer un échantillon spécifique aux descendants de rapatriés de taille suffisante pour quelques statistiques simples sur cette population peu étudiée. Pour les personnes qui sont descendantes d'un parent né Français à l'étranger du sous-échantillon des descendants d'une personne née à l'étranger (sous échantillon 3), la correction de la non-réponse (entretiens longs ou courts relativement aux non répondants) est suivie d'une réaffectation uniforme du poids des questionnaires courts abandonnés au poids des questionnaires longs conservés de cette catégorie.

## Le partage des poids pour éliminer les doubles comptes de l'échantillon

L'échantillon initial de l'enquête se décompose de cinq types de sous-échantillons. D'une part, 4 sous-échantillons distincts par leur règle de construction : les immigrés (initial et réserve), les natifs d'un DOM (initial et réserve), les descendants d'un parent né à l'étranger

(initial et réserve) et les descendants d'un natif de DOM (initial et réserve). D'autre part, un échantillon témoin représentatif de l'ensemble de la population française. Par construction, les contours de cet échantillon ne sont pas disjoints des quatre précédents. L'échantillon témoin compte aussi des immigrés, des natifs d'un DOM, des descendants d'immigrés et de natifs d'un DOM. Ainsi, la somme des poids bruts de l'ensemble de l'échantillon comporte des doubles comptes : une fiche adresse aurait pu être tirée dans un des sous échantillon spécifique ou bien dans l'échantillon témoin représentatif de l'ensemble de la population. L'opération du partage des poids consiste à repérer a posteriori (avec les informations de la collecte sur les seuls répondants), les individus qui potentiellement pouvaient être tirés dans deux sous-échantillons et à corriger leur poids. La somme des poids par catégorie concernée (par exemple, les immigrés) est divisée par deux et ensuite réaffectée au prorata des effectifs non pondérés. Ainsi la dispersion des poids à l'intérieur d'une catégorie donnée diminue significativement (le poids des individus issus de l'échantillon témoin est fortement abaissé à l'issue du partage). Une fois l'opération de partage terminée, la somme de l'ensemble des pondérations de l'enquête doit fournir l'effectif pondéré du champ global de l'enquête.

### Calage sur les marges de l'enquête annuelle du recensement 2008

Le calage cherche à corriger la non-réponse totale et permet en même temps de limiter l'erreur d'échantillonnage en utilisant des totaux connus sur l'ensemble de la population enquêtée. Cela nécessite donc de disposer d'une source externe qui fournisse les totaux connus de variables auxiliaires (les marges).

Le choix des marges se porte habituellement sur des variables explicatives de la non-réponse : sexe, âge, activité, type d'immeuble, taille d'agglomération et région.

L'enquête TeO s'est déroulée sur la fin de l'année 2008 et le début de l'année 2009. On retient donc comme source de référence pour le calcul des marges l'enquête annuelle du recensement de l'année 2008 (EAR2008). La taille de cette source (près de 5 millions d'observations sur le champ des personnes nées entre 1958 et 1990 vivant dans un ménage ordinaire en France métropolitaine) est le principal avantage de l'EAR. En particulier, elle permet de mesurer des caractéristiques sur les populations rares telles que les natifs d'un DOM (moins de 1% de la population). En contrepartie, le mode de questionnement est très différent de celui de l'enquête TeO. Le bulletin de l'EAR est pour l'essentiel auto administrée et parfois une seule personne répond pour l'ensemble du ménage. Ainsi de nombreuses caractéristiques ne peuvent pas être calées en raison d'un mode de questionnement trop

différent : par exemple le diplôme de l'individu fait l'objet d'une question dans l'EAR et d'un questionnaire détaillé dans TeO. Concernant les seuls immigrés, la question des équivalences des diplômes est nécessairement traitée différemment.

La procédure du calage traite séparément les trois populations suivantes :

- les immigrés , avec les marges d'âge et sexe, de vie en couple et de statut matrimonial, de pays de naissance, de nationalité à la naissance, de région de résidence, de tranche d'unité urbaine et de type d'habitat.
- Les natifs de DOM, avec les marges de sexe et d'âge, de vie en couple et de statut matrimonial, de région de résidence, de tranche d'unité urbaine.
- Les autres : ni immigrés ni natifs d'un DOM soit la population majoritaire et les descendants, avec les marges de sexe et âge, de vie en couple et statut matrimonial, de pays de naissance, de nationalité, de région de résidence, de tranche d'unité urbaine et type d'habitat. Il n'est en effet pas possible de repérer dans le recensement les descendants d'un parent immigré ou d'un natif de DOM.

## Conclusion

La mise à disposition des fichiers auprès des chercheurs via le réseau Quetelet et les premières publications de résultats (Borrel et Lhommeau, 2010, Beauchemin & alii 2010) est intervenue un an après la fin de la collecte et quatre ans après le début de la réflexion pour élaborer le questionnaire.

Coordonnée par l'Ined et l'Insee cette opération d'ampleur a nécessité un plan de sondage et un tirage d'échantillon très original pour interroger des personnes nées en France d'un parent né à l'étranger. Cette phase préalable à la collecte d'entretiens a constitué en soi une opération statistique lourde avec la saisie de 300 000 noms de personnes recensées en 2007 et la recherche dans registres de l'État civil des mairies du lieu de naissance des parents pour près de 100 000 personnes.

Pour les entretiens proprement dits, quelques ajustements ont été nécessaires pour approcher les objectifs initiaux : tous les échantillons de réserve ont dû être livrés et la période de collecte a été prolongée de deux mois supplémentaires. Au final, la taille de l'échantillon atteint 92% de l'objectif initial global avec un peu moins de 21 800 observations dans le fichier diffusé et pondéré mais ce taux de réussite est toutefois varié selon les origines ; en particulier, pour les descendants d'un parent venu de Turquie ou d'Afrique Centrale et du

Golfe de Guinée, la taille de l'échantillon diffusé est très inférieure à celle espérée respectivement 447 et 333 individus au lieu des 800 prévus.

C'est la prise de contact avec une population jeune, mobile et habitant des zones fortement urbanisées qui a constitué la plus grande source de déperdition d'entretiens. Une fois que l'enquêté avait accepté l'enquête, elle s'est déroulée jusqu'à son terme dans la très grande majorité des cas, malgré l'ampleur du questionnaire (63 minutes en moyenne) et les thèmes parfois sensibles abordés (religion, contraception, opinions politiques par exemple).

Avec un échantillon surreprésentant des populations encore rares et un questionnaire combinant des informations objectives (description de situations et de relations), subjectives (attitudes, perceptions et représentations) et contextuelles (informations sur l'environnement permettant de saisir les relations entre les intéressés et la société d'accueil), l'enquête TeO constitue un outil unique pour l'étude des chances d'accès des populations issues de l'immigration aux principaux biens et services qui définissent la place de chacun dans la société : logement, éducation, maîtrise de la langue, emploi et promotion, services publics et prestations sociales, planification familiale et santé, conditions de vie et réseau de relations, marché matrimonial, nationalité et citoyenneté...

## Bibliographie

É. Algava et B. Lhommeau « T'es où TeO ? À la recherche de la 2<sup>ème</sup> génération pour l'enquête trajectoires et origines », actes des journées de méthodologie statistiques, Insee, 2009

Beauchemin C., Hamel Ch. , Lesné M. , P. Simon et l'équipe TeO « les discriminations : une question de minorités visibles », populations et sociétés, n°466, Ined, 2010.

Borrel C. et B. Lhommeau « être né en France d'un parent immigré », Insee première, n°1287, mars 2010.

Insee, note sur la « pondération de l'enquête TeO », annexe associée au dictionnaire des codes de l'enquête Trajectoires et Origines, mars 2010.

Ined & Insee, « Enquête Trajectoires et Origines La diversité des populations en France (2008) », Dossier de présentation pour le Comité du Label , demande examinée le jeudi 31 mai 2007.

P. Pietri-Bessy et C. Ménard, « Cadre Organisationnel de Réalisation d'Enquête (CORE) pour l'enquête Trajectoire et Origine », mimeo, Insee, décembre 2009.

M. Tribalat, « De l'immigration à l'assimilation - enquête sur les populations d'origine étrangère en France », La découverte, 1996.



## **Encadré 1 : Les acteurs du projet**

La maîtrise d'ouvrage appartient conjointement à l'Insee (Unité des Études Démographiques et Sociales) et à l'Ined (l'unité Migrations Internationales et Minorités). La maîtrise d'œuvre statistique appartient à l'Unité méthodes statistiques de l'Insee.

Le groupe de conception du questionnaire ne s'est pas limité pas aux chercheurs et statisticiens de l'Ined et de l'Insee. Une grande représentativité de positions sur les thèmes abordés a été recherchée lors de la constitution d'un groupe de travail pluridisciplinaire (et pluri institutionnel), réuni dès la mi-2006, pour réfléchir au contenu du questionnaire. Ce groupe comprenait des chercheurs de plusieurs centres universitaires, tous spécialisés dans les études sur les migrations, l'intégration et les discriminations. Ils représentaient un large éventail de sciences sociales : démographie, sociologie, anthropologie, psychologie sociale, économie, et sciences politiques. Tous avaient mené des enquêtes quantitatives ou qualitatives sur les migrants ou leurs descendants.

Un comité de pilotage scientifique a suivi étape par étape la réalisation de l'enquête. Il comprenait des représentants des différents organismes partenaires dans la réalisation de l'enquête, ainsi que des chercheurs experts dans les domaines couverts par l'enquête.

Le comité de pilotage technique était pour sa part chargé du suivi technique de l'enquête : déroulement des différentes étapes, calendriers des tests, solutions techniques. Il a été composé de la maîtrise d'ouvrage déléguée (chef de division), du (ou des) concepteur(s), du chef de projet en organisation statistique, d'un représentant du département applications et projets et des équipes informatiques (responsables des équipes Blaise et Capi et de l'aval).

Le groupe d'exploitation a ensuite été constitué. Sa composition reprenait les membres du groupe de travail qui ont activement participé à la construction du questionnaire. Le coût total de l'enquête est estimé à 4 400 000€ dont 2 300 000€ de coûts internes (coûts liés à la conception, l'organisation et l'informatisation de l'enquête) et 2 100 000 € de coûts externes (essentiellement la rémunération des enquêteurs pour la collecte). Outre l'Ined et l'Insee, les organismes suivants ont participé au financement de l'opération :

- Le Ministère des Affaires sociales et der la Santé (Drees), celui du travail (Dares) ;
- La HALDE : Haute Autorité de Lutte contre les Discriminations et pour l'Égalité ;
- l'Acse : l'agence nationale pour la Cohésion Sociale et l'Égalité des chances
- l'IA-IdF : Institut d'Aménagement et d'Urbanisme de la région Ile de France
- le SG Ville : Secrétariat général à la Ville ;
- l'ONZUS : Observatoire des Zones Urbaines Sensibles ;
- l'AFD : Agence Française de Développement ;

## Encadré 2 : Les principales étapes du déroulement de l'enquête.

Entre le début de la réflexion pour élaborer le questionnaire et la publication des premiers résultats et la mise à disposition des fichiers aux chercheurs via le réseau Quételet, quatre ans se sont écoulés. Les principales phases de l'enquête sont détaillées dans le tableau suivant.

Étapes principales	Début	Fin	acteurs
Conception enquêt (questionnaire papier, échantillon...)	janvier 2006	février 2007	Insee - Ined
Passage CNIS : avis d'opportunité	mai-juin 2006		Insee - Ined
Test papier (terrain)	12 février 2007	12 mars 2007	Insee DR d'IdF et Languedoc-Roussillon
Dossier CNIL pour l' échantillonnage mairie	15 mai 2007		Insee - Division environnement Juridique
Réunion CNIS sur les catégories ethno raciales	24 mai 2007		CNIS
Passage au label	31 mai 2007		Maitrise d'Ouvrage Insee-Ined
Test de consultation de l'état civil en mairie	1 juin 2007	13 juillet 2007	Insee DR IdF, Aquitaine, Lorraine
2e réunion CNIS sur les catégories ethno-raciales	12 octobre 2007		CNIS
test CAPI (terrain)	26 novembre 2007	7 janvier 2007	directions régionales d'IdF, Aquitaine, Lorraine
Signature de convention entre l'Ined et l'Insee	1 décembre 2007		Insee - Ined
Dossier CNIL collecte	15 janvier 2008		Insee - Division environnement Juridique
Répétition générale	23 mai 2008	6 juin 2008	directions régionales d'IdF, Aquitaine, Lorraine
Recherche en mairie	1 mai 2008	15 août 2008	Insee
Tirage de l'échantillon	15 août 2008	31 août 2008	Insee
Livraison de l'échantillon aux directions régionales (DR) de l'Insee	11 juin 2008		Insee
Impression des fiches adresses	6 juin 2008	1 juillet 2008	Insee, Centre National Informatique d'Orléans
Formation des enquêteurs	août 2008	août 2008	Insee
Première vague d'enquêtes (immigrés, natifs de DOM, témoin)	1 <sup>er</sup> septembre 2008	1 Mars 2009	Insee
Deuxième vague d'enquêtes (descendants d'immigrés et de natifs de DOM)	6 octobre 2008	1 Mars 2009	Insee
Déclenchement de la première réserve d'échantillon	20 octobre 2008		Insee
Déclenchement de la deuxième réserve d'échantillon	4 décembre 2008		Insee
Apurement en Directions régionales	8 septembre 2008	15 mars 2009	Insee - directions régionales
Aval CAPI - création tables SAS	10 novembre 2008	1 juillet 2009	Insee
Livraison fichier brut de collecte	1 avril 2009		
Codification des professions	1 février 2009	1 mai 2009	Insee
Traitement post collecte (redressements, calage, recodifications...) et création des fichiers de diffusion	1 février 2009	1 septembre 2009	Insee
Documentation	8 septembre 2008	31 décembre 2009	Insee
Premiers résultats	Fin mars 2010		Insee et Ined
Accès aux fichiers sous contrôle	1 septembre 2009	Fin mars 2010	
Accès libre aux fichiers pour les chercheurs et les SSM : mise à disposition via le réseau Quételet	Fin mars 2010		

Source : CORE pour l'enquête TeO, Insee, 2010