

Appendix: Trajectories and Origins: Survey Methodology

Elisabeth Algava, Bertrand Lhommeau, and Cris Beauchemin

A.1. The Complex Construction of the TeO Sample

A.1.1. Objectives

The TeO survey was designed to achieve the following objectives:

- to be representative of all immigrants¹ and their descendants living in metropolitan France, whatever their country of origin, while maintaining the ability to perform analyses by origin for the principal communities (with an objective of 800 persons for each targeted origin group);

This chapter is a summary, written by Cris Beauchemin, of Algava E., Lhommeau B., 2015, Echantillonnage, collecte et pondération de l'enquête Trajectoires et origines, in Beauchemin C., Hamel C., Simon P. (eds.), 2015, *Trajectoires et origines: enquête sur la diversité des populations en France*, Paris, INED (Grandes Enquêtes book series), pp. 585–606. Complete documentation on the survey is available from the website of the Centre Quetelet at: <https://www.cmh.ens.fr/greco/enquetes/XML/lil-0494.xml>

¹In France, immigrants are defined as persons born non-French outside France.

E. Algava
Institut national de la statistique et des études économiques (INSEE), Paris, France
e-mail: elisabeth.algava@insee.fr

B. Lhommeau
Direction de l'animation de la recherche, des études et des statistiques (DARES), Paris, France
e-mail: bertrand.lhommeau@travail.gouv.fr

C. Beauchemin (✉)
Institut national d'études démographiques (INED), Paris, France
e-mail: cris.beauchemin@ined.fr

- to attach at least as much importance to the “second generation” – descendants of one or two immigrant parents – as to immigrants themselves (a target of 9600 individuals in the first generation and as many in the second);
- to specifically identify populations from the overseas *départements*, or DOMs (French Guiana, Guadeloupe, Martinique, and La Réunion) and their descendants living in metropolitan France (a target of 800 individuals in each category);
- to be representative of the entire population residing in metropolitan France, and consequently to also survey members of the “majority population” (target: 3400 individuals), i.e. individuals who are neither immigrants, nor individuals native-born in a DOM, nor children of a member of one of these groups (see Box 2.1 in Chap. 2).

The population covered by the survey was limited to persons aged 18 to 60 years² living in private households. The exclusion of institutional households was considered an acceptable limitation as the proportion of immigrants living in an institutional household was low (3.6% in 2008). Moreover, given their relatively high age, nearly 30% of immigrants in institutional households would have been out of scope.

A.1.2. Specific Difficulties

To constitute the TeO survey sample, two principal difficulties had to be overcome:

- the populations covered are relatively small;
- the descendants of immigrants or DOM native-borns are rarely identifiable in sampling frames.

In the 1999 census, the reference census when the survey was being prepared, immigrants made up 7.4% of the total population (INSEE 2005). Direct descendants of immigrants born in France made up only 8% of persons aged 17–65 years (1 in 12 persons in this age range): half with one immigrant parent and half with two (INSEE 2005). To constitute a sample of adequate size, a sampling frame approximately 12 times larger than the sample size was required. Based on an average response rate of two-thirds among selected individuals,³ a general population sam-

²Due to sampling constraints, the upper age limit had to be set at 50 years for descendants of immigrants or of DOM native-borns. However, as the proportion of descendants above this age is relatively low, the consequences are limited. In 2008, around 3 in 10 adult descendants of immigrants were aged 55 or older. This proportion was even lower for descendants from the most recent migration streams: only one in 6 among those with a parent born in Asia, and 1 in 50 for those with a parent born in sub-Saharan Africa.

³Some are unreachable, others have moved, refuse to respond, etc. A survey sample always contains more address files (a file identifying the person to be surveyed which is given to the interviewer) than the expected number of respondents.

pling frame around 18 times larger than the final number of respondents who were descendants of immigrants was required. The need to increase the initial size to enable over-representation of certain origins meant that the sampling frame had to be even larger: for example, the descendants of DOM native-borns in metropolitan France represent less than 1% of the population. To survey 100 people in this group, a target population of 10,000 persons thus had to be constituted.

Moreover, direct descendants of immigrants and DOM native-borns were (and continue to be) difficult to identify in the available sampling frames. Two questions suffice to define whether or not a person is an immigrant: Was the person born in France? Is the person a French citizen by birth? These questions are standardized and are included in virtually all surveys carried out by the National Institute for Statistics and Economic Studies (INSEE), and notably in the population census. The questions used to identify direct descendants of immigrants, on the other hand, concern not only the respondent but both of his or her parents, whose place and nationality of birth must be known. These questions are not standardized, and are rarely present in data sources. In particular, they are not included in individual census questionnaires. They are included in the Labour Force survey and the Housing survey, but the samples are too small to serve as a sampling frame for TeO. An innovative solution was needed to sample members of the second generation.

A.1.3. A Sampling Frame with Multiple Sources

The 2007 annual census survey was used to obtain an initial sampling frame of named individuals in which three strata could be distinguished on the basis of place of birth and nationality at birth: immigrants, DOM native-borns, and individuals born in metropolitan France. The task was then to distinguish, among persons in the latter group, members of the majority population, descendants of immigrants, and descendants of DOM native-borns. To do this, we drew on another information source, that of civil registration, where parents' place of birth is recorded. The complete identification process is described in Table A.1. For each of the individuals born in metropolitan France drawn from the census (240,000 persons, step 2 in Table A.1), we then had to match census data with civil registration data (step 3). INSEE's central database of vital records presents two drawbacks for this operation. The first is that it contains only persons born after 1967. It was thus impossible to match records from the two sources – census and civil registration – for anyone born before that year (hence the different options at step 3). For these individuals, there was no other choice but to match the data manually by consulting records stored at municipal level (step 7). The second drawback is that the central database of vital records is anonymous. Individuals' census data thus could not be matched with vital records based on their names. In the absence of any other unique identifier, for a given individual selected in the census, we thus searched in the database of vital records for all persons born on the same day, in the same municipality, of the same sex (step 4) to obtain “anonymous” or “echo” correspondences. This matching

Table A.1 Steps involved in selecting the sample of descendants of immigrants

1. Selection of the survey population from the 2007 annual census survey: day, month, and year of birth, residence in metropolitan France, and private household	=> 300,000		
2. Selection of persons born in metropolitan France: constitution of the file of named individuals	=> 240,000		
3. Differential treatment of cohorts born before and after 1967	Born between 1957 and 1967 => 33,000 Not applicable. No anonymous INSEE file	Born between 1968 and 1990 => 210,000	
4. Matching with the anonymous INSEE file by sex, date and place of birth		Matching	
5. Result of matching		At least 1 echo (born the same day in the same municipality) with 1 parent born abroad => 127,500	All echoes had 1 parent born abroad => 8,500
6. Drawing before searches in municipal records, with over-representation of rare origins		Manual search in register needed => 65,000	None with a parent born abroad: Out of scope => 74,000
7. Search in municipal registers	=> Systematic manual search in register => 33,000	Manual search in register needed => 65,000	Manual search in register not needed
8. Selection of persons born to a parent who was born abroad	=> 25,000		=> 8,500
9. Drawing of the sample, with over-representation of rare origins	=> 18,400*		

* a reserve of 1,400 address files was also created in this step

operation yielded one of three results (step 5), each leading to specific subsequent sampling steps:

- if, for a selected individual, there was no echo corresponding to a descendant of immigrants or of DOM native-borns, the individual then entered the stratum for the majority population;
- if all of the echo correspondences were descendants of immigrants or DOM native-borns, we could integrate the individual into the second-generation sampling frame with no further searches;
- if an individual selected from the census had at least one echo with a parent born abroad or in a DOM and at least one echo with no parents born abroad or in a DOM,⁴ it remained impossible to include that individual in the stratum of descendants of immigrants or of DOM native-borns. This was true of 127,500 persons in the sampling frame constructed from the census. For them, as for individuals born before 1968, there was no alternative but to perform searches at the relevant municipal registry. As municipal registers are nominative, among individuals native-born in metropolitan France drawn from the census, it was possible to identify those with at least one parent born abroad or in a DOM (step 8).

⁴That is, at least two persons were born on the same day in the same place and had parents with different places of birth.

After matching with vital records (using either central INSEE data or municipal registers) was complete, we now had a sampling frame to select second-generation individuals, taking into account the need for over-representation of certain origin groups (step 9).

Given its scale and the sensitivity of the information collected, this operation was a preliminary survey in itself. In total, 98,000 files were manually consulted in municipal records for named individuals to recover the parents' place of birth (step 7).⁵ This search required not only the approval of the French data protection agency (CNIL) and the Ministry of Justice, but also personal authorizations for access to civil registers from senior magistrates (*procureurs de la République*) for each of the interviewers going to municipal registries to perform searches.

A final selection stage took place at the time of the survey. Vital records do not include information on parents' nationality at birth. The sampling frame for second-generation individuals thus included persons whose parents were born as French citizens abroad. This group notably includes the sons and daughters of individuals repatriated from the former colonial empire (notably Algeria). These individuals were identified in the field, during administration of the questionnaire, which includes precise questions on the origins of the respondent's parents. To save time and money, the interview was ended if the respondent proved not to be a descendant of immigrants or of DOM native-borns.⁶

A.2. Data Collection

Some particularities of the population targeted by the TeO survey had to be taken into account in the data collection protocol: geographical concentration in the Île-de-France region, mobility, and, in some cases, limited command of French. Three preliminary tests, carried out in 2007, were used to validate all data collection procedures. Data collection took place between September 2008 and February 2009. The face-to-face interviews were performed by 566 interviewers. The respondents were informed of the survey by a notification letter, translated into 10 languages. They were then directly contacted by an interviewer, in most cases by telephone, in order to schedule a time for the interview. The interviews generally took place in the respondent's home using an electronic questionnaire, with the interviewer entering the responses directly into a laptop (computer-assisted personal interviewing, or CAPI).

⁵ Among the sampled individuals, only 2% were not found.

⁶ Around a third of individuals with at least one parent born abroad are not descendants of immigrants.

A.2.1. Geographical Concentration in Île-de-France

The population of immigrants and their descendants is highly concentrated in urban areas, and particularly in the Île-de-France region⁷ (37% of all address files in a region that totals 18% of the metropolitan French population). To cope with this concentration, new recruits were brought in to reinforce INSEE's usual pool of interviewers. Moreover, members of the target population are more likely to live in disadvantaged neighbourhoods, which can be more difficult for INSEE interviewers to access. A specific bonus for interviews carried out in a ZUS (sensitive urban area) was thus added to the interviewers' pay (16% of address files in the sample). In Île-de-France, interviewers followed a specific complementary training course on interviewing in sensitive urban areas.

A.2.2. A Mobile Population

The sample population proved to be highly mobile: between the annual census survey of January 2007, which served as the sampling frame, and the TeO survey, which began in September 2008, 28% of persons in the sample had moved. Among those who had moved, only one in two persons could be found for interview. This proportion varied by place of residence: in rural areas 3 out of 4 addresses were recovered, but the figure for the Parisian agglomeration was only 4 in 10. The interviewers used two different methods, on average, to find the respondents (Algava et Lhommeau 2009). In order of frequency, these included neighbours (53%), mayor's office (44%), the post office register of address changes (40%), directory searches, and Internet searches. Other means (family in cases where the person had moved out of the family home, ex-partner in cases of separation, or new occupants of the dwelling) were used in 30% of cases.

A.2.3. Language Barriers in the Survey Population

A portion of the immigrant population, and notably the most recent arrivals in France, are not perfectly fluent in the French language. Several measures were used to allow the least fluent persons to participate (and thus to avoid bias), while ensuring that they understood the questions (and thereby to ensure the quality of their responses). Various documents were systematically translated into 10 languages (Arabic, English, German, Mandarin, Portuguese, Russian, Serbo-Croat, Spanish, Turkish, Vietnamese): the notification letter announcing the survey; a form with keywords, important concepts, more complicated questions; and code cards detailing the categories of answers for some questions. Furthermore, the

⁷The Île-de-France region comprises Paris and its surrounding suburbs.

interviewers could obtain help in translating the interview. In total, 1084 interviews required at least partial translation of the questionnaire. When language difficulties were identified, interviewers were advised to make use of an accredited interpreter with no connection to the respondent to make it easier for him or her to answer the most sensitive questions. Despite this recommendation, more than 900 interviews were carried out with the help of spontaneous interpreters from among the family or neighbours. The complexity of scheduling a three-way meeting (between respondent, interviewer, and interpreter) made it difficult to rely on professional interpreters, who were only actually used in 128 cases (1.5% of immigrant respondents).

A.2.4. The Final Sample: Objectives and Results

Despite these different measures, as well as recourse to reserve samples and a two-month extension of the collection period to the end of February 2009, the survey did not fully reach its objectives: 21,761 “long” interviews were carried out, or 92% of the initial target (Table A.2). An additional 3328 “short” interviews were performed with descendants of French citizens born abroad. Non-response, which occurred with 4 in 10 address files, was mainly due to the mobility of the target population. The mainstream population sub-sample had the highest response rate. For immigrants, the initial objectives were generally attained. However, reaching descendants was much more difficult than expected, and in particular those from the most

Table A.2 Expected and actual sample sizes

	Objective	Actual	Actual/objective (%)
Immigrants, of which	9600	8456	88%
Southeast Asian	800	774	97%
Turkish	800	830	104%
Sahelian African	800	665	83%
Guinean and Central Africa	800	736	92%
Descendants of immigrants, (18–50 years) ^a of which	9600	8110	84%
Southeast Asian	800	573	72%
Turkish	800	447	56%
Sahelian African	800	480	60%
Guinean and Central Africa	800	333	42%
DOM native-borns	800	712	89%
Descendants of DOM native-borns (18–50 years) ^a	800	650	81%
Mainstream population	3400	3781	112%
Total ^a	23,600	21,761	92%

Note: The list of countries in each geographical entity is given in Chap. 2, Table 2.2

Source: TeO survey, INED-INSEE 2008

^aThere were also 52 observations corresponding to descendants of immigrants or of DOM native-borns aged 51–60 years

recent migration streams, i.e. Turkey and sub-Saharan Africa. These younger groups are both less often present in the home and more prone to move. In sum, the largest source of lost interviews was the inability to contact members of a young, mobile population living in highly urbanized areas. Once the respondents had consented to the interview, it continued to completion in the great majority of cases, despite the length of the questionnaire (63 minutes on average) and the sometimes sensitive topics covered (religion, contraception, and political opinions, for example). Weightings were calculated to correct the biases resulting from non-response (INSEE 2010).

References

- Algava É., & Lhommeau B. (2013). “Enjeux de l'échantillonnage, collecte et pondérations de l'enquête Trajectoires et Origines”, INSEE, Document de travail, série DSDS, F1304.
- Algava É., & Lhommeau B. (2009). “T'es où TeO? À la recherche de la 2^e génération pour l'enquête Trajectoires et Origines”», *Actes des journées de méthodologie statistiques*, INSEE.
- INSEE. (2010). Note sur la “pondération de l'enquête TeO”, appendix to the dictionary of TeO survey codes. <https://www.cmh.ens.fr/adisp/documents/lil-0494/lil-0494pond.pdf>.
- INED/INSEE. (2007). “Enquête Trajectoires et Origines, La diversité des populations en France (2008)”, Presentation document submitted to the Comité du Label, request examined on Thursday 31 May 2007.