DEMOGRAPHIC RESEARCH

Research Article

Sampling international migrants with origin-based snowballing method: New evidence on biases and limitations

Cris Beauchemin

Amparo González-Ferrer

Table of Contents

Sampling international migrants with origin-based snowballing method: New evidence on biases and limitations

Cris Beauchemin[1]

Amparo González-Ferrer[2]

Abstract

This paper provides a methodological assessment of the advantages and drawbacks of the origin-based snowballing technique as a reliable method to construct representative samples of international migrants in destination areas. Using data from the MAFE-Senegal Project, our results indicate that this is a very risky method in terms of quantitative success. Besides, it implies some clear selection biases: it overrepresents migrants more strongly connected to their home country, and it tends to overestimate both poverty in households at origin and the influence of previous migration experiences of social networks on individuals' out-migration.

---

[1] Institut National d'Études Démographiques (INED), Paris, France. E-mail: cris.beauchemin@ined.fr.
[2] Spanish Council for Scientific Research (CSIC-IEGD), Madrid, Spain.
E-mail: amparo.gonzalez@cchs.csic.es.

1. Introduction

International migration has become a major concern for multilateral agencies as well as for origin and destination states. Surveys on the causes and consequences of international migration remain rare, however, especially in some parts of the world such as sub-Saharan Africa. In fact, studies on international migration are hampered by specific methodological issues not encountered in other fields of demographic research. International migration, by definition, involves several countries (origin and destination countries and, in some contexts, transit countries too). Accordingly a multi-country design is an option recommended by specialists to study the causes and consequences of international migration in a rigorous manner (Massey 1994; Bilsborrow et al. 1997; Stillman, McKenzie, and Gibson 2007; Rallu 2008). As stated by Parrado, McQuiston, and Flippen (2005:222), "to understand the selectivity of the migration flow, and to separate patterns of behaviour that arise from practices and attitudes that migrants bring with them from their communities of origin, and patterns of behaviour connected to migration itself" requires collecting data both in destination countries (on migrants) and in origin countries (on non migrants and return migrants). Several transnational surveys on migration (i.e., conducted both in sending and receiving regions) have been designed, some of which include a link between the samples at origin and destination (Massey 1987; Parrado, McQuiston, and Flippen 2005; Mazzucato 2008; Arenas et al. 2009). Mazzucato (2008) provides a good overview of the advantages of what she calls "simultaneous matched surveys (SMS)." In terms of data collection, these surveys rely on a transnational snowballing methodology that starts either at origin or at destination.

The need for multi-country surveys, combined with an objective of representativeness, raises tremendous sampling challenges. However as yet, methodological studies remain rare in this field of research. McKenzie and Mistiaen (2009) have already provided a very useful state of the art concerning the various methods used to sample migrants in destination areas, and also a comparative assessment of their cost and efficiency. However their work omitted any reference to surveys on international migration built upon transnational samples, and especially those based on a snowballing methodology. Groenewold and Bilsborrow (2008) provided an interesting overview of the different sampling experiments carried out as part of the Push-Pull project to collect data both in African countries (and Turkey) and Europe. Although this project was based on a transnational approach, samples at origin and destination were not linked: households interviewed in Europe had no relationship with those sampled in Africa and Turkey.

In sum while transnational surveys with linked samples are quite common, no systematic assessment of this methodology has yet been undertaken. The objective of this paper is precisely to provide a rigorous methodological assessment of the

advantages and drawbacks of the origin-based snowballing technique, i.e., the method consisting in collecting contacts in origin households to sample migrants at destination. We use data from the Senegalese part of the *Migration between Africa and Europe* (MAFE) project[3], for which a transnational survey was carried out both in Africa (Senegal) and in Europe (France, Spain, Italy), with partially linked samples. This article investigates whether this method is reliable, from both quantitative and qualitative viewpoints. Two important questions are at stake. First when planning a survey on international migration, under what conditions should researchers rely upon the origin-based snowballing method, and how can they be sure of achieving the expected number of interviews? Second to what extent are the migrant samples constructed with this method subject to biases that might affect the results of the statistical analyses on the causes and consequences of international migration?

2. Previous experience of transnational surveys using snowballing techniques

It may be difficult to obtain a representative sample of non migrants and return migrants in sending countries, depending on the national context (Bilsborrow et al. 1997). But building a representative sample of migrants at destination is often a real challenge for at least three reasons. First, except in a few countries, migrants are a rare population. Second, because they are vulnerable, especially when undocumented, they are difficult to reach. Non-response can thus be a major problem (Fawcett and Arnold 1987). Third, appropriate sampling frames are rarely available, especially if the survey aims at reaching undocumented migrants who, by definition, are usually not registered. Migrant selection is all the more complicated when the survey objective is to target a group from a specific origin: depending on the group, the sampling process may become a search for "a needle in a haystack." Snowballing techniques or "chain-referral" methods provide a potentially attractive solution to these difficulties. They consist of selecting a sample of "seed" individuals (or households) to start the survey, and then asking these "seeds" for additional contacts to reach other individuals (or households) in the population of interest.

---

[3] The Senegalese part of the Migration between Africa and Europe (MAFE) project is coordinated by INED (C. Beauchemin), in association with the Université Cheikh Anta Diop (P. Sakho). The project also involves the Universitat Pompeu Fabra (P. Baizan), the Consejo Superior de Investigaciones Científicas (A. González-Ferrer), and the Forum Internazionale ed Europeo di Ricerche sull'Immigrazione (E. Castagnone). The survey was conducted with the financial support of INED, the Agence Nationale de la Recherche, the Région Ile de France and the FSP programme 'International Migrations, territorial reorganizations and development of the countries of the South'. The MAFE project was enlarged to Ghanaian and Congolese migration in 2008. For more details, see: http://www.mafeproject.com/.

Some attempts were made to generate random samples of migrants using Heckathorn's respondent-driven sampling (RDS) method (Bilsborrow and CEPAR 2007; McKenzie and Mistiaen 2009; Friberg 2010). In principle this technique allows one to reach a representative sample if seeds are randomly selected and the surveyed individuals are separated by an average of 6 intermediaries (Heckathorn 1997; Heckathorn 2006). In practice when applied to migrants, this method led to mixed results. Likewise Bilsborrow et al. (2007), McKenzie and Mistiaen (2009) and their respective colleagues failed to reach a representative sample of migrants using this technique in Ecuador or in Brazil. Friberg (2010) seems to have succeeded but does not describe very precisely the methodology used in his project. In any case these experiments were carried out within the limits of only one country. Transnational surveys have never used the RDS methodology. Most of them applied a snowballing approach, but without any serious ambition of representativeness. Using this approach two options are possible: starting at destination, or at origin. When starting in the receiving country, migrants are interviewed first and are asked to provide contacts in their home country; the survey is then carried out at origin. This method was used, for instance, in an OECD survey on African international migration in the Senegal River Valley (Condé and Diagne 1986), in the Ghana TransNet project (Mazzucato 2008), and also for migration between Mexico and the US (Parrado, McQuiston, and Flippen 2005). The second option is the reversal of the previous one and can be labelled 'origin-based snowballing method:' the survey starts at origin, collecting contacts of migrants abroad, before implementing the survey in the destination country (more rarely in several countries).

This option has been applied in at least two surveys of Mexican migrants: the Mexican Family Life Survey (Arenas et al. 2009) and the Mexican Migration Project (Massey 1987) – the dataset used by most of the best known studies on the determinants and effects of international migration. The goal of these two surveys was not to build a representative sample of migrants at destination but rather to create transnational samples of Mexicans, starting with representative samples at origin. The Mexican Family Life Survey is a nationally representative panel survey in Mexico that aimed to follow members who left the selected households, be they in Mexico or in the United States. The MMP survey is not strictly representative at the national level, either in Mexico or in the United States. It rather aims at obtaining a representative transnational sample of each origin community included in the survey (Massey 1994). Thus for each of the 128 selected communities in Mexico, 200 households were randomly selected at origin and 20 households were sought at destination, i.e. in the US, and recently also in Canada. The snowballing chain started in Mexico, where the first contacts were recorded. It continued at destination since migrants could give contacts of other migrants originating from the same community (Massey and Zenteno 2000). The

community-based design of this survey explains why a snowballing technique is used: there is simply no other solution for building samples that are representative at the level of the transnational communities. However the samples are not strictly representative. First, the Mexican sample does not include members of the community who have migrated internally within Mexico. Second, samples in the US are concentrated in selected regions (typically in California). This concentration of migrants at destination was a way to reduce fieldwork costs and is justified on the grounds that migrants from the same community tend to cluster in the same destination areas, because of chain migration. However Mexican immigrants have tended to adopt increasingly varied destinations (Parrado, McQuiston, and Flippen 2005). Third, the extension of the MMP from rural to urban communities complicated further the objective of representativeness at the community level since community boundaries are difficult to define in urban areas. The authors of the Mexican Family Life Survey and of the MMP surveys do not claim that their samples are strictly representative of the migrant population at destination. However they provide little information on the potential biases that may be induced by the origin-based snowballing method, and do not discuss in detail their potential impact on the results obtained regarding the causes and consequences of international migration.

Generally speaking the origin-based snowballing method entails a major inherent limitation: by construction, migrants surveyed at destination tend to be those who have kept close ties with their home community, otherwise the origin households would not have been able to provide their contacts in the first place. This means that migrant households that have settled permanently at destination are likely to be underrepresented (Parrado, McQuiston, and Flippen 2005). On the other hand this selection technique also presents important advantages. Obtaining the contact details of migrants through their acquaintances and relatives makes it potentially easier (and cheaper) to locate individuals of the population of interest, to reduce the reluctance of potential interviewees to participate in the survey, and probably increases –for the same reason– the chances of including undocumented migrants in the final samples, whereas they are usually a difficult-to-reach population. For instance the MMP sample in the United States includes about 50% undocumented migrants (Massey 1994).

Beyond this general overview of the advantages and drawbacks of the origin-based snowballing method, there is so far no systematic assessment of this selection technique. From a quantitative point of view it appears to work well when used to survey Mexican migrants, since the MMP has applied this technique from 1982 up to now and the Mexican Family Life Survey was able to reach 92% of the migrants declared at origin (Arenas et al. 2009). However it has proved completely ineffective with other populations. Grasmuck and Pessar (1991), for instance, tried to record contacts in the Dominican Republic with a view to surveying Dominicans in New York,

but ended up applying the snowballing method only in destination after failing to build a sample based on contacts recorded at origin. This example suggests that acquaintances and relatives may remain quite reluctant to provide interviewers with contact details to locate their relatives in the country of destination, especially those who are undocumented.

Previous applications of the origin-based snowballing method to build samples of immigrants in destination areas raise some questions that need to be addressed so that designers of new surveys on international migration know how far it can be relied upon. To what extent does recording contact details at origin allow one to reach a sufficient number of migrants at destination? To what extent do the results obtained depend on the severity of immigration policies at destination? Does it help to select undocumented migrants, as indicated by the MMP experience? Do 'more anonymous' sources to select migrants at destination work better than origin-based snowballing or not? To what extent does this origin-based snowballing technique induce biases in the survey population? In particular to what extent does it –in fact– lead to an overrepresentation of migrants who are more closely connected to their country of origin (recent migrants, migrants who visit their home country more frequently, who send more remittances, etc.)? The objective of this paper is to answer these questions. All our analyses are based on the experience of the MAFE-Senegal project, which used various sampling techniques –including origin-based snowballing– to collect data both at origin, among households in Senegal, and at destination, among migrants in several European countries (France, Spain and Italy).[4]

3. Data and survey methodology: The MAFE-Senegal Project

3.1 Objectives and questionnaire content

Broadly defined the main goal of the MAFE-Senegal Project was to build a large longitudinal dataset to investigate the causes and consequences of migration between Senegal and Europe. More specifically we aimed at identifying: (1) the main factors driving migration from Senegal to Europe (out-migration) and from Europe to Senegal (return migration and circulation); (2) changes in the profile of Senegalese migrants and their migration strategies (routes, destinations, etc.), and the extent to which immigration policies at destination are responsible for those changes and; (3) the main consequences of international migration for both individuals and households.

---

[4] For the sake of simplicity in writing and reading, the term "Europe" refers in the rest of the text only to the three European countries involved in the MAFE-Senegal project (France, Spain, Italy).

In order to achieve these three main goals, it was clear that two parallel surveys had to be carried out in both origin and destination areas. Moreover due to the diversity of the Senegalese destinations, it appeared necessary to select more than one destination country[5] and, as far as possible, different types of receiving areas in those countries. France was selected as the historical target of Senegalese migrants, and Italy and Spain were added in the survey to represent new European destinations. All in all these three countries accounted for 45 percent of the international Senegalese migrants declared in the 2002 Senegal Census.

Two different questionnaires were designed: a household and an individual questionnaire.[6] The household questionnaire was used in Senegal: it contains sociodemographic variables on all the current members of the household, but also on international migrants declared by the household respondent. In this questionnaire the following persons living abroad at the time of the survey could be declared as "household migrants:" the head's children, his/her spouse(s), and also other relatives of the head or of his/her current spouse with whom the household had had regular contacts within the last 12 months. The questionnaire included a specific module aimed at obtaining contact information for each of the declared migrants. In addition, it contained specific modules on short migration histories, on relationships between each migrant and the household, and on housing conditions and owned assets.

The individual questionnaire was used both in Senegal and in the European countries to collect life histories of migrants, non migrants and return migrants. It contains retrospective information on the following topics: dwelling, family, work, international migration of the interviewee (including attempts to migrate, return trips to Senegal, transit migration and legal status in foreign countries), migration history of the migrant's relatives (list of their stays abroad, including dates and country names), goods and assets, and remittances and contributions to associations in the country of origin.

---

[5] Despite the multiple recommendations for progress in this direction (Bilsborrow et al. 1997), previous examples of parallel sampling that also included several destinations remain rare. The Pull-Push Project, for instance, aimed to permit comparison of migrants from a specific country of origin in different countries of destination, but it ended up interviewing migrants from specific origins in just one destination country (i.e., Moroccans and Senegalese in Spain, and Ghanaians and Egyptians in Italy). A truly representative sample of international migrants should include not only 'final' destinations but also the so-called 'transit countries,' since migrants who end up staying for long periods in places initially conceived just as 'transit' places might significantly differ from those who reach their final destinations. However budgetary limitations discouraged us from adding such an innovative dimension to the research design of the MAFE project.

[6] The questionnaires are available in English, French, Spanish, and Italian on the MAFE project website: http://www.mafeproject.com/

3.2 Sample schemes

The MAFE-Senegal survey was conducted in 2008 both in Africa and in Europe. Its sample at origin was designed to overrepresent households declaring migrants and to be representative of the population living in the region of Dakar,[7] which is home to about a quarter of the national population and which is an area of high prevalence of international migration. In the end 1,141 households and 1,067 individuals were interviewed in Senegal. In Europe, the objective was to select 200 migrants in each country. For many reasons, and as usual in surveys on migrants, no straightforward sampling strategy could be applied (Bilsborrow et al. 1997; McKenzie and Mistiaen 2009). The final choice was to combine different sampling methods with the aim of choosing the best available option in each country and diversifying the sources and directions of potential biases associated with each sampling strategy. Some features are shared by all countries, however. In all of them two types of sources were used to build the samples:

1. First, contact details of migrants recorded in Senegal, obtained from households previously surveyed in the region of Dakar. Most of these contact details concern migrants declared in the household questionnaires (labelled "household migrants" in this text). Others are additional contact details of migrants who are close friends, acquaintances, etc. of the respondents but who do not "belong" to the household. They are thus not included in the household data. Obtaining these contact details at origin required special efforts that are described in Appendix 1.
2. Second, an additional sample constituted in each country to complement the number of interviews ultimately obtained from the contact details of migrants collected in Senegal, until the final target of 200 completed questionnaires per country was achieved. The strategy applied to obtain these complementary national samples in the destination countries consisted of:
    2.1. Probability sampling method in Spain, which used the Municipal Population Register (Padrón) as a sampling frame to draw a random sample of people born in Senegal and living in Spain at the time of the survey. The register presents the unique advantage of including undocumented migrants as well as documented ones (see http://www.mafeproject.com/ for more information on this source).
    2.2. Quota method in France and Italy. This sampling method - which is not probabilistic - requires: (a) having auxiliary data that can be used to set quotas of respondents according to different characteristics (gender, age, region of residence, etc.) and (b) that respondents be recruited through a variety of

---

[7] For more details on the sampling methodology, see: http://www.mafeproject.com/

channels to limit biases (this implies strict monitoring of the fieldwork). Various sources were thus combined: Senegalese migrants' associations, public places, and snowballing techniques.

In sum, we ended up using multiple sources to obtain contact information on potential interviewees for the fieldwork in European countries. Selected public places and snowballing techniques were used in both Italy and France, but not in Spain. Contact information supplied by migrants' associations was used only in France but not in Italy or Spain. And, obviously, the Municipal Population Register could only be used in Spain. In fact households previously surveyed in Dakar were the only source of contact information available for all the countries.

Whatever the variety of the selection sources, the same eligibility criteria applied in all countries to homogenise the type of people we would interview. Interviewees had to be individuals: (a) born in Senegal, (b) with Senegalese nationality at some point in his/her life, (c) who had migrated to Europe for the first time at age 18 or older and, (d) aged between 25 and 70 at the time of the survey. In all countries the samples were stratified by sex (half men and half women) and age (half aged 25-40 and the other half aged 41-70). For budgetary reasons we were also obliged to select specific regions within each destination country, instead of carrying out the surveys over their whole territory. In Spain, the twelve selected provinces included approximately three quarters of the total Senegalese population living in the country as of January 1st, 2008 (INE 2009). In France, the selected areas –Ile de France, Rhône-Alpes and Provence-Alpes-Côte d'Azur– included approximately 64% of the total population born in Senegal living in France at the time of the survey (INSEE 1999). Finally the regions selected in Italy were Campania, Emilia-Romagna, Lombardia and Toscana, where approximately 64% of the Senegalese population lived in 2006 (ISTAT 2006). In Italy and Spain, where Senegalese immigration is more recent and partly oriented towards rural areas, special efforts were made to reach Senegalese migrants living in places of lesser concentration, on the grounds that migrants who live in areas where many other conationals reside might differ in a substantial manner from those who reside in more isolated areas. This is the rationale for the inclusion of Campania in Italy, a region that accounts for only 2.6% of Senegalese migrants (compared to 39.3% in Lombardia). And in Spain a third quota was imposed and respected: the proportion of interviewed migrants living in areas with a large concentration of Senegalese residents had to be equivalent to the real proportion of Senegalese migrants living in those areas in the selected regions.

3.3 Datasets

In this article we use three types of data from the MAFE-Senegal survey: (A) data on "household migrants," collected in Senegal (household questionnaire), (B) data on the migrants surveyed in Europe (individual questionnaire), and (C) fieldwork data.

A. *Household survey in Senegal*. For the purpose of this paper, the survey was used to generate a dataset in which each "household migrant" living in France, Spain, or Italy represents an observation (N=783). Descriptive variables are available at two levels: (1) migrant characteristics, i.e., sociodemographic variables and several variables describing his/her migration experience and his/her relationships with the surveyed household; (2) household characteristics, i.e., sociodemographic characteristics of the heads of household and interview conditions.

B. *Individual survey in Europe.* The original dataset contains complete life histories of the migrants interviewed in Europe. In total it includes 602 Senegalese migrants living in France (199), Italy (203), and Spain (200). For this paper all individuals were kept in the sample, no matter how they were selected, and life histories were used to generate synthetic variables on topics of special interest: legal status of the migrants both at arrival and at the time of the survey; and several variables on the migrants' links with their home country (number of return trips since arrival in Europe, duration of stay in case of long return to Senegal, contribution to a collective remittance system through a Senegalese migrant association, and number of years remitting to people living in Senegal).

C. *Fieldwork data.* This dataset contains information on the selection method of the migrants interviewed or to be interviewed in Europe. Two types of individuals are included in the file: (1) the 602 migrants who were actually interviewed in Europe, whatever their source of recruitment; and (2) people who were not interviewed but for whom we had recorded contact details in Dakar or, in Spain only, through the Padrón. For the migrants who actually took part in the survey, the data indicate the source of recruitment (public place, association, contact details through the Padrón, contact details obtained in Dakar, and snowballing at destination). For all the migrants for whom we had contact information before starting the fieldwork, additional information is available: number of attempts to reach the individual and final output of the attempt(s) (interview done, ineligible individual, impossible to reach/not found, refusal).

These three data files can be used jointly since they contain common identifiers. First merging the household (A) and the fieldwork (C) data, it becomes possible to assess the efficiency of contact information recording at origin. Second merging the individual dataset (B) with the fieldwork data (C) offers the opportunity to test for the selectivity of the origin-based snowballing method not only in the phase of contact information recording, but ultimately in the sample of individuals actually interviewed.

4. Results

In practice in the field, the origin-based snowballing method relies on a step process, each step giving rise to sample attrition and selection. The first step consists in the recording of contact details in the home country: households agree or refuse to give the phone number, address or e-mail of "their" migrants living abroad; and their agreement or refusal is probably not random. The last step consists in conducting the survey with migrants at destination. At this stage, some may refuse either explicitly (they say they don't want to take part) or implicitly (some argue they are not available, others simply do not answer the phone calls). In the meantime some contacts are lost because migrants are ultimately ineligible, or because the phone numbers appear to be wrong, etc. The first part of this section details attrition issues, examining the quantity of contact information recorded at origin and also the reasons why some contact details could not ultimately be used in Europe. The second part is devoted to selection biases at origin: we provide results on the factors explaining the probability of obtaining migrants' contact details when doing the household survey. The final part of this section gives insight into the characteristics of the migrants surveyed through the origin-based snowballing method, compared to the migrants recruited through other channels. It focuses especially on the migrants' legal status (documented or not) and on the links maintained with the home country.

4.1 Assessing the quantitative efficiency of the origin-based snowballing method

As can be seen in Table 1 the origin-based snowballing technique applied in MAFE-Senegal led to poor quantitative results. The number of migrants who were finally interviewed in Europe thanks to contact information recorded in Dakar is very small: only 36 individuals, i.e., 6% of the total European sample. From the very beginning, it was clear that the number of contacts recorded at origin would not cover a very large share of the sample: only 364 contacts were recorded for a final target of 600 interviewed migrants. Furthermore the loss between the potential number of migrants that could have been interviewed through origin-based snowballing and the number of interviews actually carried out in Europe is huge. A high level of attrition was recorded at three steps:

1. Obtaining contact details from households in Dakar was quite difficult: in total, respondents gave us phone numbers for only a third of the total number of household migrants declared as living in France, Spain, or Italy at the time of the survey (248 phone numbers out of 783 declared migrants). This ratio does not vary

greatly across countries (30% in France, 33% in Spain and Italy) and is also very similar to the ratio we achieved during the pilot survey in November-December 2006 (8 contacts were obtained out of 25 migrants declared to be living in France by the households surveyed in Dakar). Note that 116 additional contacts were recorded in Dakar for migrants who did not belong to the surveyed household.

2. The contact information obtained was not always accurate. Only 78% of all recorded phone numbers were actually correct (i.e., someone answered our call, regardless of the eligibility of the individual and his/her willingness to participate in the survey). Why were 22% actually wrong? First hypothesis: the phone number was not properly recorded. This is quite plausible given the complex structure of the phone numbers in Europe, different in each country. Second hypothesis: the household respondent was reluctant to give us the real number but did not want to overtly refuse and thus provided us with a wrong one. However this is unlikely since the percentage of inaccurate numbers for non household migrants equals the that of household migrants.

3. Finally accurate contact details did not guarantee that migrants would participate in the survey: only 17% of the correct contact details resulted in interviews in Europe (Table 1, line 7). The loss was even greater for the non household migrants (only 9% were interviewed) than for the household migrants (21%). This difference is probably due to a lesser proximity (relationships less intense or frequent) between the non household migrants and the household respondents. In both cases, the huge loss is due to the fact that some migrants were, in fact, not eligible (18% of the contacts, Table 2, line 4) or refused to be interviewed (19%, Table 2, line 2).

In sum there was a dramatic difference between the number of contacts recorded and the actual number of interviewed migrants in Europe: only 5% of the declared household migrants were finally interviewed in Europe.

Table 1: **Efficiency of the origin-based snowballing method in the MAFE-Senegal survey**

	Household migrants	Non Household migrants	Total
(1) All household migrants declared in the household questionnaires	783	-	-
(2) Migrants with a *given* contact	248	116	364
(3) Migrants with a *correct* contact*	193	91	284
(4) *Interviewed* migrants in Europe	36	12	48
Ratios			
(5) = (2) / (1)	32%	-	-
(6) = (3) / (2)	78%	78%	78%
(7) = (4) / (3)	21%	9%	17%
(8) = (4) / (1)	5%	-	-

* A correct contact is a contact that could be used to reach a person (someone answered the phone call).
Sources: MAFE-Senegal, Household survey in Senegal, Fieldwork data

Interestingly there were large cross-country differences in the final outcome of the contact details collected in Dakar (Table 2). The results were not especially poor in Spain, despite the delay between contact information recording in Dakar (March) and the Spanish survey (May-July). In fact Italy is the country with the worst indicators. Firstly the proportion of interviews conducted was much lower there (only 5% of the contacts obtained in Dakar resulted in a completed interview, Table 2) than in France and Spain (17% and 18%, respectively). Secondly the refusal rate was much higher in Italy (86%) than in the two other countries (42% and 57%, respectively). Thirdly the percentage of migrants who could not be found was also very high (56%) compared to France (30%) and Spain (33%).[8] Why did the origin-based snowballing method worked so poorly in Italy? Could it be due to the fact that the survey agencies that we hired in France and Spain made bigger and better efforts to find people for which we had contact details from Dakar than in Italy? This is unlikely. First all agency managers received exactly the same instructions on how to proceed using the phone contacts during a joint training session held in Paris. Besides the average number of phone calls made to each correct number was quite similar in the three countries: 4.1 in Spain, 3.7 in France, and 3.6 in Italy. Could it be due to a higher number of undocumented

---

[8] There are two types of "not found" migrants: those who never answered the phone calls and those whose numbers were wrong. The proportion of "not found" contacts due to wrong numbers was not higher in Italy (44%) than in France (67%) or Spain (54%). In other words the overall quality of the contacts seemed quite similar across countries but the proportion of people who never answered our phone calls (to apparently correct numbers), for some reason, was higher in Italy.

migrants (probably more reluctant to pick up the phone if the caller is unknown) in Italy? In fact this is not the case: according to our data, the proportion of undocumented migrants is the same in Italy and Spain (19% in both countries, in contrast to 5% of the migrants for whom we obtained contact information in France). The poor results of the origin-based snowballing method in Italy may actually relate to the particular political context in this country during the fieldwork period. Indeed public debates and policy measures against migrants were especially tough in Italy right at the time of the survey: in May 2008, Silvio Berlusconi, the new prime minister, presented a forceful anticrime plan including extremely restrictive measures against immigrants.[9] In this troubled context for migrants, mistrust probably explains the proportion of migrants that refused to participate in the survey either explicitly or implicitly (by saying they were not available or by not answering unknown callers).

Table 2:   Outcome of contacts collected in Dakar by country, and from the Padrón in Spain

Outcome	Contacts collected in Dakar (Household and non household migrants)								Padrón	
	All countries		France		Italy		Spain		Spain	
	N	%	N	%	N	%	N	%	N	%
(1) Done	48	13	30	17	6	5	12	18	188	40
(2) Refusal[1]	70	19	21	12	29	23	20	30	30	6
(3) Not found[2]	144	40	52	30	70	56	22	33	213	45
(4) Ineligible	65	18	43	25	12	10	10	15	16	3
(5) Other[3]	37	10	26	15	9	7	2	3	26	5
Total	364	100	172	100	126	100	66	100	473[4]	100
Refusal Rate (2)/(1+2)	58%		42%		86%		57%		14%	

[1]   People who answered the phone but said won't be available for the interview ever during the whole period of the fieldwork were classified as "implicit refusals." In some cases it was not the person him/herself who answered but one who claimed to be a friend or a relative and denied the possibility of reaching the cell phone's owner during the fieldwork period. The figures for implicit refusals were 14 cases in Italy, 4 in France and 2 in Spain.

[2]   "Not found" means that we were not able to actually locate the migrants either because the phone number was wrong (no connection, as for 54% of the not found migrants) or because nobody ever answered the call (46% of the not found).

[3]   This category includes people who were not contacted because they were too clustered with other migrants (spouses, brothers, sisters, children).

[4]   Out of the 600 contacts obtained from the Padrón and provided to the survey firm, only 473 were actually used to reach the objective of 200 surveyed migrants.

Source: MAFE-Senegal, Fieldwork data

[9] See, for instance: http://www.nytimes.com/2008/05/21/world/europe/21iht-italy.4.13105375.html

As mentioned in the previous section, one of the main reasons for choosing the origin-based snowballing method over other alternatives is the belief that contacting the migrant through a person he/she knows (and probably trusts) will reduce the difficulty of locating the person and the overall non response, which in turn will substantially lower the survey cost. Our results so far suggest that this technique applied to Senegalese migrants, suffers from severe attrition problems. Our data allow us to go further in the assessment of this technique by comparing its outcomes with those of a more random and anonymous source of selection, namely the Padrón. Although the number of contacts obtained in Dakar for the survey in Spain is limited (66), the comparison should be restricted to this country. If we focus only on those contacts who were correct and eligible, differences in refusal rates are huge: 57% for contacts obtained in Dakar *versus* only 14% for contacts obtained via the Padrón. In total 18% of the total contacts obtained in Dakar finally resulted in interviews in Europe (12 out of 66). The ratio was much higher for contacts obtained through the Padrón (40%). This anonymous source seems thus substantially more effective for recruiting migrants. However part of the difference stems from differences in the data collection process between the two sources: for contact information obtained in Dakar, we used phone calls instead of home visits, as was the case with the Padrón, which may have increased the migrants' reluctance to participate and, at the same time, made it easier for them to refuse. In any case contrary to common expectations, our results show that an anonymous source of recruitment, such as the population register in Spain, can lead to better results in terms of response rate than a transnational snowballing strategy.

4.2 First step: selection biases in recording of contact details at origin

As shown earlier, the origin-based snowballing method is subject to high attrition at different steps. This attrition is very probably not random. For instance when collecting contact details among households at origin, it might be more difficult to obtain phone numbers of undocumented migrants from households with a poor understanding of the survey objectives. This section is precisely aimed at analyzing these potential biases at the first step of the origin-based snowballing method (recording of contact details at origin). Basically we distinguished three types of potential biases: those related to the migrants' characteristics, (sociodemographic variables and migration experience); those related to the households (sociodemographic characteristics and understanding of the survey); and those related to the links between the migrants and their origin household. We also examined (and controlled for) the influence of the interviewers themselves on the recording of contact details. All the analyses were performed using the data on all household migrants who were declared as living in France, Spain, or Italy at the time of

the survey (N=783). We thus excluded from the analyses the migrants for whom we obtained contact information in Dakar but who are not recorded in the household dataset, and for whom we have no information on the origin household characteristics, personal migration experience, etc. (additional contacts of non household migrants, N=116). For each of the 783 declared household migrants, it was possible to know from the fieldwork data whether the survey team was able to obtain contact information (N=248), regardless of the actual quality of the contact details (as we already mentioned, some of them were finally not usable). Two types of analyses were conducted. Descriptive statistics show to what extent migrants and households' characteristics vary when comparing the whole population of declared migrants living in Europe and the selected population of migrants for whom we could obtain contact information. In addition we ran a multivariate logit model to see whether the probability of obtaining contact information is significantly dependent on the variables cited above.

Overall the results suggest that the probability of obtaining contact information depends much more on the household than on the migrants' variables (Table 3). Odds ratios for the latter do not vary greatly and differences with the reference categories are never significant. It is neither more nor less easy to obtain contact information when the migrant is young vs. old, male vs. female, or less educated vs. highly educated. Migration experience does not make any difference either. In the model as in the descriptive results, there are some differences by destination country, but they are not significant. The duration since last departure has no impact either. This could result from counter-balanced effects: on the one hand a long absence may weaken the relationships with the origin household; but on the other a long sojourn abroad means that the migrant is more settled and thus might also be easier to reach. It seems that the illegal status of the migrants at destination reduces the odds of obtaining a contact in the model, but the result is not significant. Intuitively it would not have been surprising to observe that households are more reluctant to give the contact details of their migrants when they are vulnerable from a legal point of view. Finally, the non significance of the difference between undocumented and documented migrants indicates that the survey design and all the efforts done in the field to obtain contacts were profitable.

In contrast to migrant characteristics, household characteristics play a major role in explaining why contact details are or are not obtained at origin. First of all our results indicate that migrants for whom contact details are given are more likely to belong to modest households, as can be deduced from odds ratios related to the housing status: heads who own their home are 30% less likely to give contact details than those who rent it.[10] This result is consistent with the result related to the proportion of origin households equipped with latrines (the less comfortable type of amenities), which is

---

[10] The variable head's education was highly correlated with the variable housing status, and we preferred to keep this one since it seems to be more informative of the living conditions of the household as a whole.

16% for the whole sample of declared migrants living in Spain, France, or Italy, but is higher for the population of migrants with correct contact information (Appendix 2). Both results are in line with interviewers' observations during the fieldwork: those assigned to wealthy areas reported much more difficulty in finding the households and persuading them to take part in the survey. In fact refusal rates are globally higher in rich census districts. All in all the origin-based snowballing method tends to overrepresent the migrants linked to the less wealthy households. The chances of obtaining contact details also depend on the characteristics of the head himself: while there is no gender effect, older heads tend to provide contact information more easily than younger ones. This age effect could be explained in terms of respect for social norms: the elders do not fear that they will be blamed by migrants who are younger than them if they give their contact details without first asking permission. Interestingly even when controlling for other characteristics of the head, his/her migration experience is another important factor in explaining the likelihood for the survey team of obtaining contact information to reach migrants in Europe: the team was 40% more likely to get contact information when the head had migrated out of Senegal for at least one year in the past. The sample of migrants contacted at destination through the origin-based snowballing method thus tends to be biased towards people who originate from households with previous migration experience.

The variables that inform us about the influence of the migrant-household links tend to confirm the idea that when households have strong relationships with their migrants, it is much easier to obtain contacts. The odds of obtaining a phone number are twice as high when the migrant contributes to a large share of the household budget through his/her remittances, as compared to the migrants who send nothing. Moreover this result is in line with other descriptive statistics showing that migrants with a contact address are also those who more frequently send both goods and money. However it is worth noting that the migrants' visits to the household or the fact that they received help from the household to migrate do not make any difference in this regard (Appendix 2). In the end it seems that the origin-based snowballing method tends to oversample migrants who remit more or more frequently than others.

Finally the general understanding of the survey appears to be a determinant of contact information retrieval. Clearly the better the respondent understands the survey (as reported by the interviewers), the higher the odds of obtaining contact details. We also hypothesized that interviewers might be unequally successful in obtaining contact details. Descriptive statistics confirm this assumption. On average contact details for 13 migrants were obtained per interviewer, but this number varied from 1 to 37 (Appendix 2). These differences could be explained by a large range of factors: the socioeconomic level of the census districts to which interviewers were assigned (see above), the prevalence of migration in the district (few migrants in Europe explains few contacts),

the willingness of the interviewers (some of them were afraid of being blamed by the households if a migrant was deported after the survey), and their skill in establishing an atmosphere of trust during the interview. In the end after controlling for all the migrants' and households' characteristics (that captured some of the factors just mentioned), interviewers themselves play no role in explaining the successful recording of a contact.

Table 3: **Factors explaining the collection of migrants' contacts in origin households**

		All household migrants in France, Spain, Italy		Migrants with a given contact		Logit Model (odds ratios)
		N	**%**	**N**	**%**	
Migrants' characteristics						
Age	0-24	57	7	18	7	1.00
	25-34	225	29	83	34	0.74
	35-44	245	31	76	31	0.69
	45-54	141	18	42	17	0.91
	55 & +	74	9	18	7	0.58
	Don't know – Missing	41	5	11	4	0.47
Gender	Homme	534	68	158	64	1.00
	Femme	249	32	90	36	1.12
Education	No education	53	7	13	5	1.00
	Primary school	169	22	49	20	0.77
	Secondary School	217	28	76	31	0.89
	University level	151	19	52	21	0.83
	Koranic School	69	9	23	9	0.86
	Don't know – Missing	124	16	35	14	0.98
Migrants' Migratory Experience						
Destination	France	394	50	118	48	1.00
country	Italy	260	33	87	35	1.31
	Spain	129	16	43	17	1.63
Duration since	21 years or more	88	11	25	10	1.00
last departure	11 to 20 years	117	15	43	17	0.98
	10 years or less	374	48	127	51	1.07
	Don't know – Missing	204	26	53	21	0.62

Table 3: (Continued)

		All household migrants in France, Spain, Italy		Migrants with a given contact		Logit Model (odds ratios)
		N	%	N	%	
Legal status	Documented	603	77	196	79	1.00
	Undocumented	103	13	31	13	0.86
	Don't know – Missing	77	10	21	9	1.24
Head's characteristics						
Housing Status	Tenant	207	26	71	29	1.00
	Owner	454	58	143	58	0.69**
	Other	108	14	30	12	0.50*
	Don't know – Missing	14	2	4	2	0.00
Head's gender	Male	483	62	148	60	1.00
	Female	290	37	97	39	0.88
	Don't know – Missing	10	1	3	1	61200000
Head's age	under 35	80	10	16	7	1.00
	35-44	137	17	39	16	1.35
	45-54	181	23	60	24	1.47*
	55 & +	357	46	127	51	1.76**
	Don't know – Missing	28	4	6	2	0.97
Has the head ever migrated outside Senegal?	Never or less than a year	622	79	183	74	1.00
	Yes, for at least a year	151	19	62	25	1.39*
	Don't know – Missing	10	1	3	1	
Migrants' Links with the Household						
Partner within the HH?	Yes	120	15	43	17	1.00
	No	456	58	136	55	0.78
	No Partner	197	25	66	27	0.61
	Don't know – Missing	10	1	3	1	0.26
Ever lived with the HH head?	Yes	600	77	198	80	1.00
	No	180	23	48	19	0.86
	Don't know – Missing	3	0	2	1	45.18
Contribution of the transfers to the household budget	No transfer	263	34	58	23	1.00
	Small share	117	15	37	15	1.40*
	Moderate share	200	26	78	32	1.99***
	Large share	178	23	65	26	1.96**
	Don't know – Missing	25	3	10	4	1.30

Table 3:   (Continued)

		All household migrants in France, Spain, Italy		Migrants with a given contact		Logit Model (odds ratios)
		N	%	N	%	
Interview conditions						
Understanding of	Excellent	297	38	118	48	1.00
the questions by	Good	428	55	119	48	0.62**
the respondent	Correct	39	5	7	3	0.57**
	Don't know – Missing	19	2	4	2	0.71
Interviewer's ID						1.00
Number of		783	100	248	100	
observations						783.00
LR chi2						80.26
Prob > chi2						0.00
Log likelihood						-448.76
Pseudo R2						0.08

Legend: * p<0.05; ** p<0.01; *** p<0.001
Source: MAFE-Senegal, Household data, Fieldwork data

4.3 Final step: Does origin-based snowballing lead to selection of specific migrant profiles at destination?

As mentioned earlier the number of interviews finally conducted in Europe thanks to the contact information obtained in Dakar is unfortunately too small (N=48, including contact details of both household and non household migrants) to make a proper analysis of the entire selection process involved in the origin-based snowballing method. However by comparing the characteristics of the migrants included in the European sample by their method of selection (origin-based snowballing versus others), it is possible to check whether the snowballing method leads ultimately to individuals with specific traits. In this section we use data from the individual survey in Europe to look at whether the snowballing method was more effective than more anonymous sources in including undocumented migrants in our survey, as commonly expected; and whether migrants interviewed through the contact details obtained in Dakar are indeed more strongly connected to their origin country (and household) than migrants contacted through other selection methods.

We analyzed the legal status of the interviewees at two different points in time: at the time of their first arrival in Europe, and at the time of the survey. In principle the

advantages of the snowballing method are expected to be more visible with regard to the legal status of individuals at the time of the survey, which is what makes them vulnerable and distrustful, rather than at the time of arrival.

As can be seen in Table 4 the proportion of interviewed migrants who lacked proper documents to reside in the country where they were surveyed in 2008 (time of the survey) was substantially smaller among migrants whose contact information was obtained in Dakar compared to other interviewees (2% versus 15%, respectively). Even though numbers are small, this difference is significant at 1%. Clearly this result challenges the idea that contact details obtained at the country of origin substantially increase researchers' chances of including undocumented migrants in surveys at countries of destination, in comparison to other methods. Moreover this difference remains after going back to the migrants' legal situation at the time of their first arrival in Europe. While 46% of the migrants interviewed through other sampling methods (i.e., Padrón, public places, etc.) reported having entered Europe without proper documents, only 25% of those interviewed thanks to the contact information obtained in Dakar did so. This difference is also significant at 99%.[11]

Table 4: **Legal status among interviewed migrants in Europe:**
Differences according to the sample source

Legal status		Interviews from origin-based snowballing contacts		Interviews from other sampling sources	
		N	%	N	%
At the time of the survey	Documented	46	96	447	80
	Undocumented	1	2	85	15**
	Missing	1	2	23	5
	Total	48	100	555	100
At the time of arrival	Documented	36	75	275	50
	Undocumented	12	25	257	46**
	Missing	0	0	23	4
	Total	48	100	555	100

Source: MAFE-Senegal, Individual Survey in Europe
** p<0.01

---

[11] It is important to note that the direction and robustness of these results does not change even if interviewees who were selected through the Padrón in Spain are excluded from the comparison.

Thus, as summarized in Table 5, interviewees who were selected through the origin-based snowballing method had returned more often to Senegal during the time they spent abroad, and had also remitted for longer periods than interviewees who were contacted through alternative methods. Besides these results are not due to longer length of stay among the former group in Europe, since migrants recruited through this channel were not significantly different from others in this regard (results not shown here). Therefore our results in Europe confirm the idea that the utilisation of transnational linked samples in immigrants' surveys tends to overrepresent individuals more strongly connected with their home country and, in addition, seriously challenges the idea that origin-based snowballing methods are better than others in providing us with reliable samples of undocumented migrants at destination.

Table 5:   Relationships with the home country:
Differences according to the sample source

	Interviews from the origin-based snowballing	Interviews from other sampling sources
Number of Return Trips	0.27*	0.10
	(.01)	(.08)
Years of stay during return trips to Senegal	4.4	6.4
	(1.0)	(0.5)
Number of years remitting to Senegal	11.5*	8.9
	(8.5)	(9.0)
Ever contributed to associations in Senegal	0.29	0.22
	(.06)	(.02)
N	48	554

Source: MAFE-Senegal, Individual Survey in Europe
* $p < 0.05$

5. Discussion

The aim of this paper was to assess the reliability of the origin-based snowballing method for researchers wanting to plan a transnational survey of migrants. Our analysis is limited to a survey conducted in the region of Dakar, in Senegal, and in selected European countries. Obviously our results cannot be generalized to the whole world. However some useful lessons can be drawn from the MAFE-Senegal experience in terms both of sample sizes and selection biases. Let us now come back to our initial questions and to the answers our analyses provide.

To what extent does collecting migrants' contacts in the country of origin allows one to reach a sufficient number of migrants at the destination? One of our most striking results is that the origin-based snowballing technique, applied to Senegalese migrants, led to the collection of a very small number of reliable contacts. This echoes the poor results reported by Grasmuck and Pessar (1991) in their survey on Dominicans. But it also contrasts with the results of other similar surveys carried out on Mexican migration. In the Mexican Family Life Survey, for instance, 91% of the migrants who left their origin household between 2002 and 2005 to go to the US were successfully tracked and interviewed (Arenas et al. 2009). The high level of this ratio might be partly explained by the fact that only recent migrants were targeted in this survey. But it still contrasts hugely with our own result, whereby only 5% of the migrants declared in the sending households could finally be interviewed in Europe (section 4.1), knowing besides that the duration since last departure has no effect on the probability of obtaining a contact in origin households (section 4.2). Thus our results suggest that the origin-based snowballing method is terribly uncertain. It definitely depends on the type of population observed. Mexican migrants, for instance, might be easier to track because they maintain stronger relationships with their origin households than African migrants, thanks to a shorter distance between origin and destination that allows for more frequent visits. And it also depends on the context at the very precise moment of the survey.

Is the final result in terms of numbers of migrants actually interviewed very much dependent on the particular political context concerning immigration issues at destination? Our experience suggests that the information that origin households and migrants themselves receive on how migrants (documented or not) are treated in destination countries is indeed a crucial issue. And it appears that tiny pieces of information can have tremendous and immediate effects on the collection of migrants' contact details. For instance, one of the three tests of the MAFE-Senegal survey took place in 2005 just after the Senegalese national TV broadcasted a report on violent expulsions of Senegalese migrants from Spain. This time the interviewers obtained almost no contact information from the households of migrants living in Senegal. The third test took place in 2007 just after Nicolas Sarkozy, the French President, made a speech in Dakar that was widely condemned in Africa.[12] This time again we obtained very poor results, especially regarding migrants living in France. By chance the second test was not affected by any unexpected event and we ended up with better results. During all these tests, Italy appeared as a country where contact details were easier to obtain than in the other countries. And as fully discussed above, this is the country where we obtained the worst results during the real survey, probably because we did the

---

[12] On the reception of this speech, see, for instance: http://www.guardian.co.uk/world/2007/aug/27/southafrica.france

survey at the wrong moment in terms of policy context in that country. In sum, the MAFE-Senegal experience suggests that the origin-based snowballing method is a very risky one in terms of quantitative success. In fact, for this reason, we abandoned it when the MAFE project was extended to Congolese and Ghanaian migration.

Does the origin-based snowballing method help to select undocumented migrants in comparison to other 'more anonymous' methods of selection? Regarding the widely-held belief that snowballing starting at origin may present an advantage for selecting undocumented migrants, our results are mixed. On the one hand it is not more difficult to collect contact information for undocumented migrants than for the documented ones: there is simply no effect of the legal status on the probability of obtaining a phone number in origin households (section 4.2). However, in the end, when looking at the migrants who were actually interviewed, contact information collected in Dakar provided us with substantially smaller proportions of undocumented migrants than other alternative methods, no matter whether we focus on their current legal status or their legal status at first entry into the destination country (section 4.3). This last result indicates that more anonymous sources of selection do not exclude irregular migrants from surveys, compared to more "personal" sources of information.

To what extent does this origin-based snowballing technique induce biases in the surveyed population? According to our results, migrants selected through transnational snowballing starting at origin tend to be more connected to their origin household (section 4.2) or, more generally, to their origin country (section 4.3). As a consequence surveys that rely exclusively on this kind of sample probably overestimate the frequency and size of remittances, as well as circulation practices (return trips). They may also overrate poverty in households at origin since it is more difficult to obtain migrants' contact details from the better-off households (section 4.2). Additionally this selection technique could blur the understanding of the migration process. As shown earlier, households in which the head used to live abroad are more likely to provide migrant contact information (section 4.2). This result means that the origin-based snowballing method tends to overestimate the influence of previous migration experiences of social networks to explain out-migration. When demonstrated with this kind of sample, the idea that migration is a self-feeding process within households should be qualified. All these selection biases, in addition to the attrition issues mentioned above, should dissuade survey designers from using the origin-based snowballing method exclusively to select migrants in destination areas. In the absence of a sampling frame, there is no perfect method to select migrants: so far, there is no straightforward alternative to building fully probabilistic samples of migrants in receiving countries (McKenzie and Mistiaen 2009). Respondent-driven sampling and intercept point methods do not work perfectly and they are not usable for small samples (less than 400 or 500 migrants per country). Each technique entails its own biases. One

way to limit them is to mix the methods of selection. In any case, survey designers should always make efforts to collect the data needed to properly assess sampling biases. This is a real necessity in order to improve collection of data on international migration, a field of research where surveys are often better than worse, but not fully satisfying.

6. Acknowledgements

References

Arenas, E., Teruel, G.M., Rubalcaba, L., and Herrera, C. (2009). Tracking beyond borders: Experience of the Mexican Family Life Survey. Paper presented at *Population Association of America Annual Meeting*, Detroit.

Bilsborrow, R.E. and CEPAR (Centro de Estudios sobre Población y Desarrollo Social). (2007). *The Living Conditions of Refugees, Asylum-seekers and Other Colombians in Ecuador: Millennium Development Indicators and Coping Behaviour.* Geneva, Switzerland: United Nations High Commissioner for Refugees.

Bilsborrow, R.E., Hugo, G., Oberai, A., and Zlotnik, H. (1997). *International Migration Statistics: Guidelines for Improving Data Collection Systems*. Geneva: International Labour Office.

Condé, J. and Diagne, P.S. (1986). *Les migrations internationales Sud-Nord: une étude de cas: les migrants maliens, mauritaniens, sénégalais de la vallée du Fleuve Sénégal en France*. Paris: OCDE.

Fawcett, J.T. and Arnold, F. (1987). The role of surveys in the study of international migration: An appraisal. *International Migration Review* 21(4): 1523-1540. doi:10.2307/2546523.

Friberg, J.H. (2010). Working Conditions for Polish Construction Workers and Domestic Cleaners in Oslo: Segmentation, Inclusion and the Role of Policy. In: Black, R., Engbersen, G., Okólski, M., and Pantîru, C. (eds.). *A Continent Moving West? EU Enlargement and Labour Migration from Central and Eastern Europe.* Amsterdam: Amsterdam University Press: 23-50.

Grasmuck, S. and Pessar, P.R. (1991). *Between two islands: Dominican international migration*. Berkeley: University of California Press.

Groenewold, G. and Bilsborrow, R.E. (2008). Design of samples for international migration surveys: Methodological considerations and lessons learned from a multi-country study in Africa and Europe. In: Bonifazi, C., Okólski, M., Schoorl, J., and Simon, P. (eds.). *International migration in Europe; new trends and new methods of analysis*. Amsterdam: Amsterdam University Press: 293-312.

Heckathorn, D.D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems* 44(2): 174-199. doi:10.1525/sp.1997.44.2.03x0221m.

Heckathorn, D.D. (2006). Studying second-generation immigrants: Methodological challenges and innovative solutions. (Migration Information Source). [http://www.migrationinformation.org/usfocus/display.cfm?ID=441#top (June 30, 2010)].

Massey, D.S. (1987). The Ethnosurvey in theory and practice. *International Migration Review* 21(4): 1498-1522. doi:10.2307/2546522.

Massey, D.S. (1994). The Methodology of an Ethnosurvey. In: Bogue, D.J. (ed.). *Readings in the Methodology of Population Research*. New York: UNFPA.

Massey, D.S. and Zenteno, R. (2000). A validation of the Ethnosurvey: The case of Mexico-US Migration. *International Migration Review* 34(3): 766-793. doi:10.2307/2675944.

Mazzucato, V. (2008). Simultaneity and Networks in Transnational Migration: Lessons Learned from a Simultaneous-Matched Sample Methodology. In: DeWind, J. and Holdaway, J. (eds.). *Migration and Development Within and Across Borders: Research and Policy Perspectives on Internal and International Migration.* Geneva: International Organization for Migration: 69-100.

McKenzie, D.J. and Mistiaen, J. (2009). Surveying migrant households: A comparison of census-based, Snowball and Intercept Point Surveys. *Journal of the Royal Statistical Society* 172(2): 339-360. doi:10.1111/j.1467-985X.2009.00584.x.

Parrado, E.A., McQuiston, C., and Flippen, C.A. (2005). Participatory Survey Research: Integrating community collaboration and quantitative methods for the study of gender and HIV risks among Hispanic migrants. *Sociological Methods and Research* 34(2): 204-239. doi:10.1177/0049124105280202.

Rallu, J.L. (2008). One-way or both-ways migration surveys. In: Bonifazi, C., Okolski, M., Schoorl, J., and Simon, P. (eds.). *International migration in Europe: New trends and new methods of analysis*. Amsterdam: Amsterdam University Press: 273-292.

Stillman, S., McKenzie, D.J., and Gibson, J. (2007). Migration and mental health: Evidence from a natural experiment. World Bank Policy Research Working Paper (nº 4138).

Appendix 1

Protocol to maximize the collection and use of migrants' contact information

The three pilot tests conducted to prepare the MAFE-Senegal survey showed us that obtaining contact information in Senegal was not a straightforward process. The households were always very suspicious, expressing their fear that their migrants could be deported and they even sometimes threatened the interviewers, who in turn became reluctant to ask for contact details. Building a climate of confidence within the survey team and winning the interviewees' trust was a big challenge. Major efforts were thus made to find the most appropriate way to inform the target population.

Content of the information

A lot of details were given to the interviewees on:
- the objectives of the project, including some examples of specific hypotheses that we intended to test and that challenged conventional ideas among policy makers (e.g., the fact that return migration is quite substantial);
- the institutional aspects of the project, highlighting the fact that it was funded only with research money, that it was carried out by academics on the basis of a true Afro-European collaboration, in relationship with an NGO (ENDA Tiers-Monde, which is a well-known Senegalese organisation). In short we emphasised the fact that the research was completely independent from Ministries involved in migration management. Basically the idea was to convince the interviewees of the benevolent neutrality of the MAFE project ;
- the interviewees' rights, especially in terms of confidentiality and the right to refuse.

Media used to disseminate the information

- **In the field**. The information was given at different steps in each survey area. (1) The information was first given to local political and administrative authorities by the supervisors. (2) All households living in the selected areas received information during the population count conducted to construct the sampling frame that was used to draw the selected households. (3) Selected households received detailed information at the time of the survey. Supervisors visited all the households that were still reluctant after the interviewers' explanations. They

could mediate when households were reluctant to take part in the survey, but also more specifically when they were reluctant to give contact details. Supervisors and interviewers were specially trained to present the survey in a way that would not raise suspicion (during the training, half a day was dedicated to this aspect of the survey, including question/answer exercises).

- **Outside the field**. Interviewers and supervisors could make reference to other sources of information: a leaflet (distributed to all potential participants), a website specially designed for the interviewees, and phone numbers of the research teams and of the NGO. Note that the same material was used in Europe, where it proved to be very useful. In particular the NGO received several phone calls from undocumented migrants residing in Spain who had been selected through the Padrón. They were worried and wondered how the interviewers were able to obtain their postal address and their names.

- **Lessons from the pilots**. One of the pilots was carried out in a semi-rural area of the Senegal River Valley. In this context other media proved to be extremely useful: town criers and local representatives of migrants' associations. There was no equivalent in the region of Dakar. As an alternative, during a subsequent test, we organised public meetings in two areas where the survey test was planned. This activity was carried out by the NGO. It had very contrasting results. In one of the areas, the meeting worked as a focus group and enabled us to broach a large range of interesting questions related to international migration in the neighbourhood. It was also a good opportunity to present the objectives of the MAFE project in detail. But, in the other area, it proved to be totally counter-productive. Discussions were overwhelmed with questions on irregular migration and deportations and people gained a distorted view of the MAFE project, so that interviewers were not welcome when they came to do the survey. As a result considering that is was virtually impossible to control the content of discussions at this kind of public meeting, we decided not to continue this approach for the real survey.

Tools used to record contact information

- **A specific module**. The questionnaire included a two-page module for the recording of contact information. It was the fourth module out of five. During the pilots, interviewers insisted it was placed far enough into the interview to ensure that they could build a climate of trust and be sure that the respondent would not be perturbed by this request for contact details (which, in turn, could have lowered the quality of the answers in the rest of the questionnaire). In this module contact

information was requested for each household migrant living in Europe. At a first stage the contact information consisted only of the migrant's first name and a phone number to reach him/her. In cases where a phone number could not be provided, the respondent was asked whether he/she could provide the interviewer with the contact details of an intermediary person or association that could help us to reach the migrant. If this information was provided, additional information about the migrant (sex, age, country and town of residence) and the respondent (first name, gender, relationship to the migrant) was also recorded. To gain the migrants' trust, it was important that interviewers in Europe could say: "I am contacting you because [respondent] gave me your name…" Note that the interviewers' training course in Dakar contained a section on the structure of phone numbers in Europe (unfortunately different in each country) in order to maximize the quality of the contact information recorded. The module can be viewed at: http://www.mafeproject.com/

- **Prepaid phone cards**: the first question asked in the contact module was: *"Would you agree to ask [migrant] whether (s)he would agree to take part in our survey? You could call him/her with my supervisor's cell phone."* This was added because during the pilots some respondents were reluctant to give migrants' contact details without their consent. Prepaid phone cards were thus provided to the supervisors so that they could give a call, with the interviewees, to the household migrants.

- **A software** application was designed to enter the data of this module. Since we wanted to keep the contacts fresh (for maximum reliability), the information collected during the household survey in Dakar was immediately recorded and transmitted to the support teams in Europe, using a secure internet connection. We intended to use these contacts immediately in Europe. This could be done in France and Italy, but not in Spain.

Despite all these efforts, it was clear that we would obtain contacts for only a small share of the migrants declared in the household questionnaires. We thus decided to also collect contact information for migrants not cited in the questionnaire but for whom respondents could nonetheless provide contact details. After going through the list of household migrants, the respondent was simply asked whether (s)he knew other migrants (other relatives or acquaintances) living in Europe who could be contacted for our survey.

Appendix 2

Table A1: Outcome of the request for contact details by characteristics of the households, the migrants and the interviewers

	(1) All Migrants		(2) Migrants for whom contact details are given		(3) % of migrants for whom contact details are given
	N	%	N	%	Ratio (2)/(1)
Household amenities					
Flush toilet, linked to sewerage	379	48	104	42	27
Flush toilet with pit	243	31	73	29	30
Latrines	129	16	61	25	47
Other	12	2	2	1	17
Don't know	20	3	8	3	40
Respondent's Status within the HH					
Head	536	68	168	68	31
Head's spouse	108	14	35	14	32
Head's children	75	10	32	13	43
Other	64	8	13	5	20
Interviewers					
Average / int.	41		13		32
Minimum nb	16		1		2
Maximum nb	88		37		71
Migrant's occupation					
No occupation	212	27	63	25	30
Employee	186	24	76	31	41
Other	151	19	50	20	33
Don't know	234	30	59	24	25
Migrant's matrimonial status					
Monogamous	469	60	144	58	31
Polygamous	108	14	36	15	33
Alone	197	25	66	27	34
Don't know	9	1	2	1	22
Transfers of goods during the last 12 months?					
Yes	234	30	95	38	41
No	529	68	144	58	27
Don't know	20	3	9	4	45
Remittances during the 12 last months?					
Yes	453	58	162	65	36
No	317	40	79	32	25
Don't know	13	2	7	3	54

Table A1:   (Continued)

	(1) All Migrants		(2) Migrants for whom contact details are given		(3) % of migrants for whom contact details are given
	N	%	N	%	Ratio (2)/(1)
Frequency of the remittances					
Monthly	184	23	81	33	44
Regularly but less often	100	13	31	13	31
Occasionally	163	21	48	19	29
No remittance	317	40	79	32	25
Don't know	19	2	9	4	47
Total	783	100	248	100	32

Source: MAFE-Senegal, Household Survey in Senegal, Fieldwork data