3 Research methodology

George Groenewold and Laurence Lessard-Phillips

3.1 Introduction

Mainly under 'guest worker' policies, Turkish, Moroccan and former Yugoslavian immigrants in the 1960s began to constitute sizeable diasporic communities in many European cities. They formed families and had children born in the country of immigration. The so-called second generation has now come of age and is well represented among adolescents and young adults in the European Union.

The TIES project's methodological objective was to obtain statistically representative information on integration-related topics from second-generation Turks, Moroccans and former Yugoslavians in fifteen cities: Amsterdam, Rotterdam (the Netherlands); Antwerp, Brussels (Belgium); Paris, Strasbourg (France); Barcelona, Madrid (Spain); Basel, Zurich (Switzerland); Linz, Vienna (Austria); Berlin, Frankfurt (Germany); and Stockholm (Sweden). Our envisioned strategy was to survey the second generation using probability sampling, as this provides a theoretical basis from which to infer objectively to the entire populations of second-generation Turks, Moroccans and former Yugoslavians in the selected cities. This also meant facing constraints similarly encountered when sampling other minority populations (cf. Groenewold & Bilsborrow 2008). These include:

- Lack of readily available sampling frames from which to sample members of minority groups, including the second generation;
- Tendency of minority groups to concentrate in particular regions and parts of cities;
- Scarcity of members of minority groups in the general population.

This chapter discusses the TIES research methodology. The following sections address the envisioned model sampling strategy, and sampling frame availability and constraints for each participating country and city. We then present summaries of country-specific sampling designs and strategies. Our conclusion entails a discussion about problems encountered in the sampling of the second generation, our adopted solutions and repercussions for the statistical representativity of the collected data.

3.2 Model sampling strategy

The IMES-NIDI project coordination team began by agreeing on a straightforward general sampling strategy to be presented to the collaborators. The individual national institutes would then adapt the model to fit local conditions in terms of available sampling frames (see section 3.3) and financial constraints. The first step was to formulate eligibility criteria. We decided to set the age range for all study groups at eighteen to 35, while group membership was to be determined by birth in the survey country to at least one parent who was born in the country of ethnic origin (Turkey, Morocco or former Yugoslavia). Also sampled in each city would be a comparison group comprising members in the same age range who were born in the survey country and whose both parents were born in the survey country. For instance, the group of Turkish respondents included children of Kurdish immigrants born in Turkey. Conversely, the comparison group included children whose parents were born in the survey country, but whose grandparents were born abroad, e.g. in a survey country's former colony.

The second step was to formulate a general sampling strategy. In each city, members of one or two second-generation immigrant groups and of a comparison group were to be randomly sampled from a suitable sampling frame and interviewed. The target sample size of each group was set at 250 successful interviews. Within cities, respondents of all groups were to be sampled from the same spatial context (e.g. neighbourhood) to ensure parity across the broad social and economic characteristics of the context. Technically speaking, respondents were to be selected in multiple steps, by first sampling neighbourhoods with probabilities proportional to estimated numbers of second-generation Turks, Moroccans and former Yugoslavian residents and then sampling fixed numbers of respondents within sampled neighbourhoods. This decision was made a priori to ensure that a sufficient number of neighbourhoods be sampled in each city, in turn allowing for a multi-level analysis of neighbourhood effects at a later stage. Depending on whether the aim was for two or three study groups, the objective was thus to successfully interview a total of 500 or 750 persons per city, amounting to 1,000 or 1,500 respondents per country.

Setting equal target sample sizes for each study group implied that each person would be given the same selection probability in the *sample* of respondents. In real life, dealing with the general population, the situation proves quite different: study groups differ in size, so selection probabilities also differ. Because the actual population of potential comparison group members in cities is generally much larger than that of the second generation, the chance of selecting a comparison group member from its reference population is much slimmer than the chance of selecting a member of the second generation. Furthermore, non-response rates between cities, neighbourhoods and study groups are likely to differ, which has an ex post effect

on selection probabilities. To deal with these two factors, so-called compensation weights can be computed and used for all analyses. Such weights ensure that the selection probability of all respondents, irrespective of group membership and spatial location, will be the same. One subsequent effect is that the response of each interviewee is weighted to its proportionate representation in the responding population. Another effect is that outcomes of statistical significance tests are influenced (Cochran 1997; Kish 1965; Lee, Forthhofer & Lorimor 1989; Purdon & Pickering 2001). The use of weights is of particular importance for comparative analysis (ESS 2004, 2007) involving respondents from different cities and countries.

To allow for the ex post calculation of compensation weights, country teams and cooperating survey bureaus were given specific instructions. They were asked to carefully detail procedures used when sampling study group members in cities and neighbourhoods and to document who had been home-visited, who refused to participate and which replacement respondents were added during field-work. Country teams were asked to adapt this model sampling strategy to reflect differences in local sampling frames and interviewing conditions.

3.3 Sampling frames

Ideally, we expected up-to-date municipal population registers to be available and accessible to researchers in each city. We hoped such a register would offer a database of all documented city residents' personal records consisting of: name, sex, current address, date of birth, birthplace, father's birthplace and mother's birthplace. We found, however, that this optimal situation – one that would permit a sampling of the second generation directly via register details – existed in only a couple survey countries (the Netherlands, Sweden). In others (Belgium, Germany, Switzerland, Austria), population registers were available though did not offer sufficient access. This was either due to rules and regulations concerning privacy (e.g. Belgium) or lacking enough details to unambiguously identify the second generation, notably parents' birth countries (Germany, Switzerland, Austria). In France, it proved almost impossible to access or, for that matter, identify suitable sampling frames. Keeping records of a person's ethnic affiliation is subject to strict rules of privacy and even contradicts the state philosophy that all legal residents are considered first and foremost French. A similar situation exists in Spain, where municipal records do not permit the identification of persons belonging to the second generation.

Similarly, other potential sampling frames, such as telephone directories, electricity company customer files and national labour force surveys (LFS), do not offer sufficient and/or relevant details permitting identification of the second generation (Klevmarken, Swensson & Husselius 2005).

Telephone directories only reflect a select part of the population; certain groups are often underrepresented due to some persons exclusively owning a mobile telephone with an unlisted number, not owning a telephone whatsoever or not consenting to having a listed number. Although large-scale nationally representative panel surveys such as the EC/ESF European Social Survey (ESS) and EC/UNECA's Gender and Generation Project (GGP) permit identification of the second generation, the actual numbers of second-generation respondents is too small for meaningful statistical analysis.

The TIES research teams in Austria, France, Germany and Switzerland decided to join forces, collectively developing a strategy in which existing registers and/or directories (see section 3.4 for details per country) served as the basis for onomastic analysis on persons' forenames and surnames. These lists of names were thus used to determine persons' probable ethnic origin and subsequently derive city- and ethnic group-specific sampling frames. The quality of the frames was then examined by taking name samples from the list and screening the persons through a short interview.

This strategy proved to have highly accurate results for Turkish forenames and surnames, though less so for Yugoslavian-sounding names. Meanwhile, in France it was inappropriate for establishing a sampling frame for the Moroccan second generation because this group could not be singled out from the similarly named Tunisians and Algerians who live in their communities; this study group was therefore excluded. We acknowledge that the onomastic approach has some shortcomings. For instance, it remains uncertain which percentage of the entire second-generation Turks, Moroccans and former Yugoslavians are actually covered in such synthetically derived lists of names and address. Another problem is that Turkish and Moroccan women who marry a person from another ethnic group and take on his surname may go undetected (Humpert & Schneiderheinze 2009).

3.4 Country-specific sample designs and implementation

Between April 2006 and December 2008, the TIES teams developed and implemented country-specific sampling strategies. They led to 6,145 successful interviews with second-generation respondents and 3,626 comparison group respondents – a total of 9,771. The ensuing section describes the principal characteristics of country-specific surveys and sampling strategies. Main survey statistics are presented in table 3.1. Our data presentation is based on the sequence in which the surveys were designed and implemented: first in the Netherlands and somewhat later in Belgium, Sweden, Austria, Switzerland, Germany, France and Spain.

The Netherlands

For the Netherlands, second-generation Turks and Moroccans and members of a native comparison group were sampled in Amsterdam and Rotterdam. Comparison group members were sampled in the same neighbourhoods as the second-generation groups. Minimum total effective sample size was set at 1,500 persons. Independent random samples of equal size were taken in cities (750) and in study groups (250). The municipal population registers served as sampling frames because these personal records comprise information permitting the identification and classification of residents according to age, sex, birthplace and parental birthplaces (BPR 2006).

What follows is a summary of the sampling method. As a first step, in each city, neighbourhoods were sampled with probabilities proportional to the sum of second-generation Turkish and Moroccan neighbourhood residents. To determine how many neighbourhoods were to be sampled, we decided a priori to set cluster size to 30 respondents (3 groups x 10 persons). This number was a compromise between our desire to secure a fair number of respondents from each study group in the neighbourhoods and to secure a sufficient number of neighbourhoods for sampling. Neighbourhood sampling was guided by the systematic selection method (Kish 1965), whereby a neighbourhood could be sampled more than once, depending on the number of second-generation Turkish and Moroccan residents. Effectively, cluster size was increased to a factor four (4 x 30 = 120 respondents) because research into non-response rates in comparable studies suggested that high non-response rates could be expected in the field (e.g. Stoop 2005). Initially, 6,000 addresses (= 4 x 1,500) were thus sampled from the municipal population registers. By the end of the fieldwork, non-response among second-generation respondents appeared higher than expected, so an additional 271 records were sampled from the registers. Of the total 6,271 addresses actually sampled from the registers, 4,999 proved valid. Discrepancy was mostly due to more than one eligible person living at the same address or so much time having elapsed between the sampling and the interview that the eligible person had moved. In the former scenario, we selected the eldest eligible household member (Kish 1965).

Sample design weights were derived and corrected for differential non-response rates across neighbourhoods and study groups. Furthermore, selection bias was examined by comparing age, sex and marital status characteristics of non-respondents with those of respondents. This was possible because the personal records of all sampled persons (non-respondents and respondents both) offer such information. Our finding was that non-response bias seems slight in terms of the compared characteristics (Groenewold 2008).

Ultimately, 738 persons, a total slightly below our target, were successfully interviewed in 23 of Amsterdam's 90 neighbourhoods. In 24 of Rotterdam's 77 neighbourhoods, 767 persons were successfully interviewed, slightly above our target of 750. Overall, the response rates were low: 30.1 per cent in Amsterdam and 29.2 per cent in Rotterdam (see table 3.1 for variation across study groups).

Belgium

In Belgium, second-generation Turks and Moroccans and members of a comparison group were sampled in the same neighbourhoods. Our aimed minimum effective sample size was set at 1,650 – that is, 900 successfully interviewed respondents in Antwerp's ten districts (3 groups x 300 respondents) and 750 respondents in Brussels' nineteen communes (3 groups x 250 respondents). Like the Netherlands, Belgium maintains a national population register. However, in 2005, a modification to privacy regulations made the register effectively inaccessible to researchers.

As such, our objective was to derive a sample of respondents of comparable ages and levels of education from all study groups, whereby the probability of respondent selection would be proportional to the presence of second-generation target groups in a particular neighbourhood. Because the availability of sampling frames differed in Antwerp and Brussels, somewhat variant sampling strategies had to be pursued in the two cities.

In the case of Antwerp, access was obtained from personal records in the population registers of the city's ten districts. Anticipating some non-response, net target numbers of respondents for each target group were scaled-up. The scaled-up target samples of second-generation Turks (667) and Moroccans (668) were sampled with probabilities proportionate to their distribution over the statistical sectors with each district, thus essentially reflecting their actual geographical distribution over the city. Allocation of the scaled-up target sample (701) of comparison group members to districts and statistical sectors called for another tactic. In a first step, the target sample was subdivided over the ten districts according to the prevalence of combined totals of second-generation Turks and Moroccans residing in districts. In a second step, district allocations were allocated to statistical sectors according to the prevalence of second-generation Turks and Moroccans residing in these sectors. A main difference from the designs for Amsterdam and Rotterdam is that the Dutch primary sampling units (PSU) were neighbourhoods, while in Antwerp they were statistical sectors sampled within each of the city's ten districts.

A suitable sampling frame of names and addresses was absent in Brussels, but information on numbers and spatial distribution of members from the three study groups was available. The first step here was to develop a sampling frame of Brussels' street segments. In a second step, street

segments were independently sampled for each second-generation study group, with selection probabilities proportionate to the respective numbers of second-generation members living in them. Sampling and interviewing comparison group members was guided by the same philosophy as our research in Antwerp. Street segments were sampled with selection probabilities in proportion to the combined numbers of second-generation Moroccans and Turks residing there. The research team gained access to a listing of addresses from a commercial database supplier (who uses the information for direct marketing purposes) with information on residents (age, nationality and name, from which we inferred whether the individual might be of Turkish or Moroccan background). On the basis of this information, we could identify the second generation and members of the comparison group as well as their addresses. Based on estimates of expected non-response and misclassifications in addresses, the number of target addresses for interviewing was scaled-up for second-generation Turks (250 to 1,110), Moroccans (250 to 1,114) and comparison group members (250 to 952). The scaled-up number of respondents' addresses was then sampled, the eligibility of persons living there was screened and, if appropriate, they were then interviewed. If more than one eligible person was living at the same address, we selected the eldest eligible household member. Sampling these respondents in the selected street segments was done by the simple random sampling without replacement (SRSWOR) method. Due to the field-work's slow progress – a result of the mediocre quality of our available address list – we decided to switch to a semi-quota sampling approach. Now interviewers were allowed to search for the first eligible potential respondent they could find in the same or adjacent street segment, thus distorting the initial probability sampling strategy.

These distortions occurred in both Antwerp and Brussels, prompting a decision not to derive probability weights (corrected for non-response rate variation), but rather to resort to the computation and application of compensatory weights. The weights were derived by comparing age and sex distribution of the respondents in the three study groups to that of comparable persons in Brussels and Antwerp, as covered by the Belgian censuses in 1991 and 2001 (Swyngedouw, Phalet, Baysu, Vandezande & Fleischmann 2008).

Sweden

Sweden's study groups comprised second-generation Turks and members of a native comparison group. Due to financial constraints, sampling goals were set at successfully interviewing just 250 persons in each group who live Stockholm County. The Swedish TIES team subcontracted sampling issues and field-work to the Swedish Central Bureau of Statistics (Statistiska Centralbyrån, SCB). This cooperation proved advantageous in

that the project gained direct access to one of the best-maintained population registers (Wallgren & Wallgren 2007; Swedish Tax Agency 2007) for sampling purposes and to teams of seasoned sampling experts and interviewers. The Swedish population register comprises the name, address, date of birth, sex, birthplace and parental birthplace of all Swedish citizens. One disadvantage of twinning was that the project proposal had to be screened and approved by various government agencies.

In the first sampling step, a subset of the population register database was created by selecting records only of persons who, as of January 2007, were between eighteen and 35 years old, were currently living in Stockholm County, were born in Sweden and whose both parents were either born in Turkey or in Sweden. In a second step, this subset database was sorted according to parental country of birth in order to create two strata. Stratum 1 thus comprised persons whose both parents were born in Turkey, while stratum 2 comprised persons whose both parents were born in Sweden. From each stratum, a simple random sample of 250 personal records was drawn and the persons were subsequently home-visited and interviewed. Thus, unlike the sampling strategy in other countries (see section 3.2), this selection method did *not* aim to sample comparison group members coming from the same neighbourhoods as our second-generation Turkish respondents.

During the field-work, refusal and non-response rates were high. As it turned out, second-generation Turkish women refused to be interviewed by male interviewers. In general, second-generation Turkish men were reluctant to participate: they often refused to be interviewed at home or, if an interview was fixed at a neutral location, they would not show up to the appointment. This unexpectedly low overall response rate of 42 per cent made it necessary to draw a second sample from the register. The second sample anticipated high non-response rates by drawing a sufficiently large sample: 2,250 names of second-generation Turks and 750 names of members of the comparison group, whereby the variation in group sample sizes reflected the different expected non-response rates. Once we achieved the targeted 250 interviews, the field-work was terminated.

The sampling approach and a careful documentation of how field-work proceeded permitted a derivation of sample design weights. These were calibrated to account for differences in the characteristics of respondents vis-à-vis non-respondents. Similar to the situation in the Netherlands, basic socio-economic characteristics (i.e. age, sex, educational attainment, income group, marital status, group size within the general population) were available in the population register, thus allowing the two groups to be compared. Differences were analysed, leading to the derivation of a so-called calibration weight for all respondents. This weight was subsequently combined with initial sample design weights, resulting in one single calibrated sample design weight for each respondent. The sum of the

calibrated design weights by group reflects both study groups' relative size within the general population of Stockholm County.

Austria

The study groups for our sampling in Austria were second-generation respondents from the country's two largest immigrant groups, Turks and former Yugoslavians, as well as a comparison group of respondents of native-born parentage. We selected two rather contrasting cities. Vienna, with a recorded 1.7 million inhabitants in 2007, is much larger than Linz, with its population of 190,000. Although a person's migrant status is recorded in Austrian municipal records, parental birthplace is not. The status of migrant also changes once full Austrian citizenship has been acquired so that children of immigrants holding an Austrian passport do not continue to be classified as persons with a migration background. Existing administrative records were therefore unsuitable as a sampling frame for the second generation. Moreover, Austrian privacy protection laws generally prevent social scientists from accessing administrative records.

Fortunately, the Austrian team managed to secure cooperation from both cities. The municipal administrations provided forenames and surnames of *all* inhabitants in the age range of eighteen to 35. These names were then screened by a survey bureau specialised in onomastics, permitting the derivation of an ethnic classification and identification of second-generation study groups. However, different types of 'frame pollution' appeared to be present. For instance, on screening respondents at their doorsteps in Linz, interviewers found that 13 per cent of persons with names identified as 'Yugoslavian' and 8 per cent with names identified as 'Turkish' did not belong to their presumed ethnic groups.

Once an appropriate sampling frame was constructed, the objective was set to the general TIES model sampling strategy of successfully interviewing a total of 1,500 respondents, i.e. 250 per city and per study group. As existing figures on non-response were unavailable for these study groups, a buffer of names and addresses four times the above-mentioned target numbers was created. In the case of Linz, with its much smaller numbers of second-generation residents, this factually boiled down to home-visiting *all* persons with a seemingly Yugoslavian (835) or Turkish (315) name in the database.

While the second generation in both cities was approached by means of a simple random sample (without replacement) straight from the list of names and addresses, the sampling of comparison group members occurred in a different manner. The address of each successfully interviewed second-generation respondent was taken, literally, as the starting point for identifying a comparison group member living nearby. Using the random route method (Kish 1965), the fifth street address following the address of

the interviewed second-generation respondent was screened for the presence of an eligible comparison group member. If present, the person was interviewed. If more than one eligible person was present, eldest household member (Kish 1965) was selected and interviewed. If absent or if the person refused to be interviewed, a new random route was pursued until an eligible comparison group member could be located.

Response rates do not differ much between groups and cities, hovering around 40 per cent, except for Turkish respondents in Linz, where the response rate was about 70 per cent. An onomastic respondent selection strategy does not permit the derivation of conventional sample design weights. To ensure resemblance of the TIES survey population with that of the representative reference population, so-called post-stratification weights were derived. The nationally representative 2008 LFS was used to derive these weights by comparing the distribution of TIES respondents and LFS respondents according to city of residence, ethnic group, age, sex and educational attainment. Thus, the TIES survey population was modelled to resemble the LFS population in terms of the aforementioned characteristics.

Switzerland

Second-generation Turks and former Yugoslavians and a comparison group of native parentage constituted our study groups in the agglomerations of Zurich and Basel, also the main settlement areas for the two immigrant groups. In Switzerland, the number of residents with a migration background is difficult to determine because administrative records do not record parental birthplace or whether Swiss nationality has been gained by birth or naturalisation. Population figures (see appendix 3) were therefore estimated on the basis of the 2000 Swiss census and the times series of recorded numbers of immigrants by origin in the central aliens register.

Similar to their Austrian colleagues' approach, the Swiss team thus built sampling frames for each of the three study groups using available municipal registers. Municipal registers in Basel and Zurich consist of a system of commune-level personal registers interlinked across cities. A survey bureau was subcontracted to develop a sample design in consultation with the Swiss team. This incorporated use of the same computer software that was successfully used in Austria to derive sampling frames (Humpert & Schneiderheinze 2009). The onomastic method was applied to commune-level population registers to identify all persons in the age range eighteen to 35 with forenames or surnames linguistically akin to Turkish and Yugoslavian names.

As a preparatory activity for designing the sampling strategy, we analysed the 2000 census in order to determine the spatial distribution of the three study groups within the boundaries of the two city agglomerations. An important finding was that the two immigrant groups appeared to live

in specific areas within Basel and Zurich. This implied that the TIES model sampling strategy had to be adapted. In other words, though feasible in other countries, the sampling design being developed could not a priori guarantee that members of all three study groups in Switzerland would be sampled in the same area (e.g. communes or neighbourhoods).

In a first step, two strata consequently had to be defined in each city. The Turkish stratum and the Yugoslavian stratum thus each comprised all communes in a city in which at least twenty members of the respective study groups resided. Communes with smaller numbers were excluded. The objective was to sample 250 respondents in each stratum. This was realised by first drawing a sample of communes and then sampling a fixed number of study group members within each sampled commune. If a particular commune was selected in both strata, a fixed number of Turkish and former Yugoslavian respondents was sampled and interviewed in that commune.

In a second step, we determined the numbers of communes and respondents of each study group. The aforementioned analysis of the spread and prevalence of study group members over the city and communes at the time of the 2000 census led to our conclusion that the optimal situation would be five Turkish respondents in each of 50 clusters (5 x 50 = 250) and six former Yugoslavian respondents in each of 42 clusters (6 x 42 = 252). This strategy implied that the target sample of 250 comparison group members was to be redistributed over the two strata according to the ratio of clusters to be sampled from the two strata (i.e. 136 in the Turkish stratum, 114 in the Yugoslavian stratum).

Our third step was to sample communes and respondents. This was achieved by applying the systematic selection method (Kish 1965; Cochran 1977). It provided a convenient way to allocate the 50 and 42 clusters to a cumulative list of communes in each stratum. Application of the method leads to self-weighing samples in each stratum and each city. Communes with high numbers of target group members have higher probabilities of being selected than those with smaller numbers; communes with the highest numbers could be selected more than once through multiple clusters (i.e. multiple batches of five or six respondents). Once subset communes were sampled by this method and the number of persons to be interviewed was known, commune authorities were requested to provide access to their commune registers, allowing researchers to identify and sample names and addresses of potential respondents from each study group.

The success of the Swiss sampling strategy depended heavily on cooperation from commune authorities. As it turned out, not all were cooperative. Some sampled communes thus had to be replaced by ones with a similar proportion of relevant second-generation residents. Furthermore, as non-response turned out to be high (see appendix 3), similar to the situation in other countries, it was necessary to repeatedly sample from the same name register in sampled communes. In some communes, the list of names was

eventually depleted and names had to be selected from other communes, thus distorting the sample design. These sources of errors, among others, complicated the derivation of sample design weights and, to an unknown extent, jeopardised statistical representativity of the data in Basel and Zurich.

Germany

Second-generation Turks and former Yugoslavians as well as members of a comparison group of native parentage were also the subjects of the TIES surveys in Berlin and Frankfurt. Similar to the situation in Austria and Switzerland, no readily available representative sampling frames were feasible via population registers in Germany. Although coverage and quality of German municipal population registers are good, personal records do not offer all information required for identifying the second generation.

In cooperation with the Austrian and Swiss TIES teams, the German team also pursued an onomastic approach (Humpert & Schneiderheinze 2009) to develop appropriate sampling frames for the three study groups. Once the municipalities of both cities had cleared the project and expressed their support, the details of all native-born eighteen to 35 year olds in the population registers could be obtained. These lists comprised 725,040 persons in Berlin and 121,374 persons in Frankfurt. Apart from name, address, sex and age, each personal record included the person's place of birth (though not of the parents) and citizenship status. In the case of Berlin, the onomastic software classified 5 and 1 per cent of the records as being second-generation Turks and former Yugoslavians, respectively. For Frankfurt, this turned out to be 7 and 4 per cent, respectively. Little over 50 per cent of the records in both cities were classified into the stratum of the comparison group of native-born parentage. This constituted the population universe from which statistically representative random samples of names were taken in each city separately. Contrary to the stratified multistage sample designs for cities in Austria and Switzerland, simple random samples of 250 names were taken straight from the deduced sampling frames of names of each study group. This deviates from the general TIES strategy of insofar as possible sampling all study groups residing in the same neighbourhood.

Similar to experience in other countries, adolescents and young adults in the three study groups proved difficult to contact and to convince to participate in an interview. For each city, two sampling waves involving the names of 750 and 1,000 persons, respectively, were required to draw a sufficient number of names and addresses to achieve the targeted number of 250 successful interviews in each study group.

France

Identifying second-generation Turks in Paris and Strasbourg was difficult due to lacking access to suitable sampling. France's most recent census and municipal population registers have no record of parents' country of birth. Similar to the approach followed in the German-speaking countries, the French team therefore pursued the onomastic identification procedure to build a sampling frame of names and addresses of the second generation.

The sampling frames in both cities were based on telephone directories. In total, 10,658 names were identified as Turkish: 7,823 in Paris and 2,745 in Strasbourg. Shortcomings of the frame used in France are similar to those experienced in the German-speaking countries, though were compounded by the aforementioned exclusionary tendency of telephone directories.

The sampling frame was established on the basis of postcode areas in each city with probabilities proportional to the number of Turkish names listed as residents in the area. The first stage consisted of a telephone screening among a sample of respondents from the target group. This screening, which consisted of a few basic questions (age, sex, individual and parental country of birth), was intended to quickly determine whether sampled persons did indeed belong to the intended target group and, if so, whether other target group family members were living within or apart from the contacted household. If all criteria were met, the names and addresses were included in a list of potential respondents for the main survey. The sampling frame was updated to include family members as potential respondents, i.e. via the snowball method, frequently used to identify and interview respondents (Kish 1965).

Including a screening stage was advantageous in that it permitted our collecting basic socio-demographic information, including that of persons who would later refuse the main interview or could not be contacted for follow-up. Background characteristics of both respondents and non-respondents were later used to inflate compensation weights for variation in non-response rates between postcode areas and study groups.

A list of names and addresses of members of the comparison group in the age range was compiled during the screening of sampled second-generation Turks in the selected postcode areas. The comparison group was subsequently sampled by postcode area, as was done for second-generation Turks.

The first stage of our field-work yielded response rates (here calculated as the number of successful interviews out of overall eligible individuals) of 25 per cent for second-generation Turks and 37 per cent for the comparison group. In the first stage of the sampling, the number of respondents was too low. A second stage was thus implemented, mainly through re-contacting individuals who had agreed to participate but were unavailable at the time of field-work, re-contacting initial refusals and using snowball sampling.

Spain

In Spain, our groups of interest were second-generation Moroccans in Madrid and Barcelona and a comparison group of native-born parentage in both cities. The lack of recent suitable sampling frames meant having to pursue three separate identification and selection strategies. The first was a sample provided by the Spanish National Statistical Institute (INE), though this one only included a few fitting cases.

The second strategy involved identifying districts within Madrid and Barcelona according to their share of resident Moroccans, regardless of their generational status, and allocating a set number of interviews to each district (proportional to the size of the Moroccan population in the district). Interviewees were then sampled from streets surrounding the centre of a given district. Respondents from the comparison group were selected in the same districts and in the same proportion. A drawback to this method is that during the initial identification stage, it was not possible to distinguish naturalised Moroccans from the comparison group. This might have underestimated districts with a high proportion of naturalised Moroccans (and hence the second generation).

A third method, used in Barcelona towards the end of the field-work period, involved asking Moroccan immigrant organisations for the names of potential interviewees.

In the case of Spain, an appropriate sampling frame could thus not be established. Potential respondents had to be identified in the field by interviewers who went in search of them, i.e. via the snowball method (Kish 1965). Data collected via this method for both study groups in Madrid in Barcelona, however, cannot claim statistical representativity.

Concluding this section, table 3.1 presents some basic statistics of the TIES surveys conducted in the eight participating countries.

3.5 Conclusions

The TIES project used the survey instrument to collect data on various dimensions of integration from second-generation Turks and former Yugoslavians and comparison group members of native parentage in fifteen cities in eight European countries. Our objective was to collect statistically representative data for these target groups, all the while acknowledging that certain constraints had to be overcome. First, we had to identify our target group members; from there, create a sample; next, contact all the persons; and, last but not least, secure their collaboration.

To support the coordinators of the TIES country teams, notably in the initial phase of the project, meetings were held to discuss potential survey sampling approaches and how to overcome technical, logistical and

Table 3.1 *Size estimates of reference populations, numbers of successfully interviewed persons and response rates, by country, city and study group*

	The Netherlands		*Belgium*		*Sweden*	*Austria*	
	Amsterdam	*Rotterdam*	*Brussels*	*Antwerp*	*Stockholm*	*Vienna*	*Linz*
Population							
Turkish	5,088	6,941	18,575	1,480	5,723	13,125	5,432
Moroccan	8,649	4,117	61,155	4,506			
Yugoslavian						26,269	3,817
Comparison	102,491	71,288			275,505	217,623	60,845
Sample							
Turkish	237	263	244	358	251	252	206
Moroccan	242	251	246	311			
Yugoslavian						253	242
Comparison	259	253	257	301	250	250	234
Response rate (%)							
Turkish	29.9%	30.5%	31.5%	63.5%	32.0%	40.0%	70.0%
Moroccan	25.9%	24.2%	30.6%	55.9%			
Yugoslavian						38.0%	38.0%
Comparison	40.1%	34.8%	31.1%	55.8%	54.0%	43.0%	42.0%
Total	31.1%	29.2%	31.0%	58.4%	42.0%		

	Switzerland		*Germany*		*France*		*Spain*	
	Zurich	*Basel*	*Berlin*	*Frankfurt*	*Paris*	*Strasbourg*	*Madrid*	*Barcelona*
Population								
Turkish	4,967	4,706	35,363	8,456	n.a.	n.a.		
Moroccan							n.a.	n.a.
Yugoslavian	14,737	4,827	6,477	4,477				
Comparison	709,290	321,104	388,343	61,725	n.a.	n.a.	n.a.	n.a.
Sample								
Turkish	206	248	253	250	248	252		
Moroccan							250	250
Yugoslavian	235	191	202	204				
Comparison	202	266	250	253	174	177	250	250
Response rate (%)								
Turkish	38.2%	46.7%	31.2%	24.8%	n.a.	n.a.		
Moroccan							n.a.	n.a.
Yugoslavian	29.7%	45.9%	22.1%	22.9%				
Comparison	41.2%	48.4%	25.7%	24.3%	n.a.	n.a.	n.a.	n.a.
Total	36.0%	47.5%	26.4%	24.0%				

Notes: The Netherlands: Population estimates derived from municipal population register (d.d. 1 April 2006). Belgium: Technical report; no information on size of comparison group; response rates in Brussels based on first sampling wave. Sweden: Response rates based on first sampling wave. Austria: Austrian Labour Force Survey weighted estimates; estimates for Linz unavailable and pertain to Voralberg region as a whole.
Source: TIES survey 2007-2008

financial constraints when surveying the second generation. For one, a model sampling strategy was offered to help country teams start developing a strategy that would take their specific local conditions into consideration. This model design assumed the availability of suitable sampling frames.

As it turned out, truly suitable sampling frames were only available for Amsterdam, Rotterdam, Antwerp and Stockholm. These cities offered the luxury of up-to-date population registers containing personal records necessary to define the second generation. Moreover, these databases, albeit with certain restrictions, were accessible to the research community. A notable advantage here was being able to know the size of the actual second generation reference population for which the TIES survey results aim to be representative. That all study groups could be directly sampled from the registers was also beneficial.

Circumstances were less favourable in other cities, where even an estimated size of the second-generation reference population could not be derived (see table 3.1). The implication was that the a priori probability of selection of a respondent could not be determined, which is a necessary condition for probability samples (Cochran 1977). Therefore, creative and innovative, albeit partial, solutions to this problem had to be developed. In the case of Brussels, for example, area sampling was implemented by sampling street segments from the main residential areas, followed by the screening of street addresses, and subsequently sampling and interviewing eligible respondents. In Austria, Switzerland, Germany and France, the national TIES teams made concerted efforts to develop the required sampling frames. Their innovative approach allowed the names and addresses of eighteen to 35 year olds in each city to be collected and compiled. From there, onomastic software was used to analyse the lists and, in so doing, derive sampling frames for each TIES study group. From these frames, probability samples of group members were taken, leading to survey data that were statistically representative for the reference population in the name lists. In the absence of a readily available and/or up-to-date population register, sampling frames derived in this way are probably the next best option. The approach's main – and inevitable – drawback is that the study populations are likely to deviate from the actual reference populations, especially if the deviation is caused by the systematic omission of persons with certain characteristics. Consequently, survey results can only claim statistical representativeness for persons included in such type of sampling frames.

Response rates were generally low in all countries, ranging from 22 per cent among second-generation former Yugoslavians in Berlin to 70 per cent among second-generation Turks in Linz. The majority of the TIES survey target audience proved difficult to contact in the first place and difficult to pin down for an interview. Low response rates raise doubts about

whether responding persons can represent non-respondents in terms of personal characteristics and measured attitudes and opinions. For Amsterdam, Rotterdam and Stockholm, this could be examined more closely because basic information on non-respondents was available in population registers. This information revealed that age, sex and marital status differences between responding and non-responding persons proved to be slight, lending support to claim statistical representativity of the collected data in these cities.

In the French cities, basic personal characteristics were collected during a large-scale eligibility screening of persons included in the constructed sampling frames. After the survey, non-respondents and respondents were compared with respect to age, sex and educational attainment. The conclusion was that in both study groups – second-generation Turks and comparison group members – non-respondents were more likely to be men and had a lower level of educational attainment than respondents. To preserve statistical representativity, compensation weights were thus derived for dealing with this bias, giving higher weight to lower-educated and male respondents.

For the German, Swiss and Austrian cities, this kind of sampling frame information was unavailable, but the survey contained a question for the interviewers to answer after each successful interview concerning how difficult it was to get in contact with the respondent. The continuum of resistance model (Lin & Schaeffer 1995; Stoop 2005) asserts that late respondents can be considered as proxies for unobserved non-respondents. Comparison of age, sex and educational attainment of easy-to-reach with difficult-to-reach respondents in these cities revealed that: 1) in German cities, difficult-to-reach respondents of Turkish origin have slightly lower levels of education; 2) in Austrian cities, difficult-to-reach respondents of both study groups are more often males and have a lower level of education; 3) in Swiss cities, the two types of respondents did not appear to differ in terms of age, sex or educational attainment. This analysis was also applied to the responses of study groups in the Dutch and French cities. Results confirmed earlier conclusions that non-respondents and respondents do not seem to differ much according to age, sex or educational attainment profile. In the case of Sweden, the survey did not collect information required for this analysis.

From these reflections, we conclude that appropriate procedures were followed insofar as was possible and feasible. We found indications that, despite fairly high non-response rates, non-response bias in most cities may not be too problematic. In light of constraints encountered in the field, the data collected and compiled by the TIES project probably reflects the best one might expect to retrieve from second-generation study groups.

References

BPR (2006), *The Personal Records Database: For the Authorities and for You*. The Hague: Ministry of the Interior and Kingdom Relations. http://www.bprbzk.nl/Zoekresultaten?query=The+Municipal+Personal+Records+Database; accessed 15 November 2011.

Cochran, W. (1977), *Sampling Techniques*, 3rd ed. New York: Wiley and Sons.

ESS (2007), 'Weighting European social survey data'. http://ess.nsd.uib.no/ess/doc/Weighting_ESS_cumulative_data.pdf; accessed 15 November 2011.

ESS (2004), 'ESS2-2004 summary and deviations'. http://ess.nsd.uib.no/ess/round2/deviations.html; accessed 15 November 2011.

European Commission (2010), *Migrant Integration Policy Index* (MIPEX). http://www.mipex.eu; accessed 15 November 2011.

Groenewold, W.G.F. (2008), 'Appendix 1. Sample design', in M. Crul & L. Heering (eds.), *The Position of the Turkish and Moroccan Second Generation in Amsterdam and Rotterdam; The TIES Study in the Netherlands*, 169-179. IMISCOE Research Series. Amsterdam: Amsterdam University Press.

Groenewold, G. & R. Bilsborrow (2008), 'Design of samples for international migration surveys: Methodological considerations and lessons learned from a multi-country study in Africa and Europe', in C. Bonifazi, M. Okólski, J. Schoorl & P. Simon (eds.), *International Migration in Europe: New Trends and New Methods of Analysis*, 293-312. IMISCOE Research Series. Amsterdam: Amsterdam University Press.

Humpert, A. & K. Schneiderheinze (2011), 'Sprachliche Analyse von Personennamen mit dem "Onomastik-Verfahren" '. Duisburg: Sozial- und Umfrageforschung. http://www.stichproben.de/herunterladen/Onomastikverfahren_HS_GbR.pdf; accessed 15 November 2011.

Kish, L. (1965), *Survey Sampling*. New York: John Wiley & Sons, Inc.

Klevmarken, A., B. Swensson & P. Husselius (2005), '5 The SHARE sampling procedures and calibrated design weights', in A. Börsch-Supan & H. Jürges (eds.), *The Survey of Health, Aging, and Retirement in Europe: Methodology.* Mannheim: Mannheim Research Institute for the Economics of Aging (MEA). http://www.share-project.org/t3/share/index.php?id=452; accessed 15 November 2011.

Lee, E.S., R. Forthhofer & R. Lorimor (1989), *Analyzing Complex Survey Data*. Series Quantitative Applications in the Social Sciences: Sage University Paper 07-071. Newbury Park: Sage.

Lin, I.F. & N.C. Schaeffer (1995), 'Using survey participants to estimate the impact of nonparticipation', *Public Opinion Quarterly* 59: 236-258.

Purdon, S. & K. Pickering (2001), *The Use of Sampling Weights in the Analysis of the 1998 Workplace Employee Relations Survey.* London: National Centre for Social Research.

Stoop, I.A.L. (2005), *The Hunt for the Last Respondent: Non Response in Sample Surveys*. SCP report 2005/8. The Hague: SCP.

Swedish Tax Agency (2007), 'Population registration in Sweden', SKV 717B 3rd ed. http://www.skatteverket.se/download/18.70ac421612e2a997f858000102733/717b04.pdf; accessed 15 November 2011.

Swyngedouw, M., P. Phalet, G. Baysu, V. Vandezande & F. Fleischmann (2008), *Trajectories and Experiences of Turkish, Moroccan and Native Belgians in Antwerp and Brussels: Codebook and Technical Report of the TIES Surveys 2007-2008 Belgium*. Leuven: ISPO & CSCP.

Wallgren, A. & B. Wallgren (2007), 'Frame populations and register populations', Seminar on registers in statistics: Methodology and quality, 21-23 May 2007, Helsinki. http://www.stat.fi/registerseminar/session4_wallgren_pres.pdf; accessed 15 November 2011.