Surveying Emigration I.

Report on the first stage of the SEEMIG pilot study in Hungary and Serbia

Zsuzsa Blaskó
Demographic Research Institute (DRI)
with
Jamalia Natalie
Hungarian Central Statistical Office (HCSO)

2014

This research report was developed in the framework of SEEMIG – Managing Migration and its Effects in SEE – Transnational Actions towards Evidence-based Strategies. SEEMIG is a strategic project funded by the European Union's South-East Europe Programme.
Project code: SEEMIG - SEE/C/0006/4.1/X

The report was prepared within the SEEMIG activity *Enhancing data production systems of migration and human capital in the South-East European area* coordinated by the Demographic Research Institute at the Hungarian Central Statistical Office.

The information published here reflects the authors' views and the Managing Authority is not liable for any use that may be made of the information concerned.

The pilot study in Hungary was carried out in close and intense cooperation with colleagues from the **Demographic Research Institute (DRI)** and the **Hungarian Central Statistical Office (HCSO)** and also **External Experts** were involved. We would like to say special thanks to the followings persons for their invaluable contribution to this research: **Irén Gödri** (Thematic Expert DRI), **Attila Melegh** (SEEMIG Project Manager), **Endre Sik** (External Expert), **Ágnes Hárs** (External Expert), **Gergely Fraller** (weighting and methodological support, HCSO), **Erika Csaba, Rita Váradi** and **Roland Kadlecsik** (LFS expertise, HCSO), **Zsolt Papp** (IT programming), **Ádám Dickmann** (migration statistics expertise, HCSO), **Orsolya Sármásy** (administrative support, DRI), **Ildikó Simonfalvi** and **Béla Soltész** (administrative support, HCSO).

The Serbian pilot study was administered at the **Statistical Office at the Republic of Serbia (SORS)**. The SEEMIG Project Manager at SORS was **Zoran Jančić**. Thematic expertee was provided by **Jelena Manojlović, Slavica Novaković** and **Mirjana Novaković**. Further valuable support was provided by **Mirjana Ogrizović** (weighting), **Jelena Milojković** and **Petar Jovanović** (IT programming). Administrative support was given by **Nikica Rodin** and **Andrea Hajdar**. External thematic support was also given by colleagues from the **Institute for Social Sciences (ISS)** in Serbia **Mirjana Rašević** (SEEMIG Project Manager ISS), and Thematic Experts at ISS **Jelena Predojević Despić, Vesna Lukić, Vladimir Nikitović** and **Goran Penev**.

Suggested citation: Blaskó, Zsuzsa – Jamalia, Natalie (2014): Surveying emigration I. Report on the first stage of the SEEMIG pilot study in Hungary and Serbia. Research report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

http://www.seemig.eu/downloads/outputs/SEEMIGPilotReport1.pdf

Table of contents

List of figures

List of tables

List of Acronyms

Acronym/ Abbreviation	English translation	Endonym
DRI	Demographic Research Institute	Népességtudományi Kutatóintézet (NKI)
HCSO	Hungarian Central Statistical Office	Központi Statisztikai Hivatal (KSH)
LFS	Labour Force Survey	Munkaerő-felmérés (MEF)
SORS	The Serbian Statistical Office of the Republic of Serbia	Republika Srbija Republički zavod za statistiku
GWSM	Generalized Weight Share Method	
RDS	Respondent Driven Sampling	

Purpose of the report and target audience

This report was prepared in the framework of the SEEMIG project (funded under the third call of the South-East Europe Programme, number SEEMIG/SEE/C/0006/4.1/X) as part of Work Package 4. Work Package 4 aims at data enhancement in the field of migration. The pilot study activity (Activity Nr 4.3.) is aiming at one hand 'to improve data sets on migration and related labour market and human capital processes, and on the other to comparatively evaluate different ways of 'reaching' in a statistically representative manner migrant populations.'

The main purpose of this report is to inform the funding authority (South-East Europe Programme), the national and international statistical bodies as well as the academic audience about the potentials and the limitations of an innovative method of researching emigrants from a particular country in a systematic and statistically reliable manner. Also, we believe that also local stakeholders (e.g.: local municipalities) can benefit from learning about the research methods outlined here.

The present report describes the design and the fieldwork of the first stage of this study and records the experiences learned during the research process. Findings from the second stage will be provided in a later SEEMIG report early 2014. Detailed analyses of the data gathered are being prepared along with these reports and will be presented on various academic forums as well as on upcoming SEEMIG events such as the Expert Roundtable 2014.

Executive Summary

Shortages of migration statistics are well-documented in the relevant literature and they are also acknowledged and thoroughly analysed in earlier reports produced by the SEEMIG project (Gárdos and Gödri 2013). To improve the situation, in the framework of the project an attempt was also made to test and further develop an innovative research design. The design is intended to be based on an internationally comparative, firm and standardized methodology, and also to be financially sustainable. This way it might serve as a best practice for statistical and research bodies Europe-wide to conduct surveys on emigrants from the country of origin, in a systematic and reliable manner. To achieve these goals two pilot studies were carried out (one in Hungary and one in Serbia) during 2013 to test the research method proposed.

The research design consisted of a two-stage methodology with the key idea to *derive a representative sample of emigrants from a representative national survey.* In the first stage of the study the Labour Force Survey (LFS) was utilized and international migrants were identified through the households included in the LFS household-sample. LFS is a highly standardized survey carried out regularly in each EU member countries and (in most cases) it also consists of a sufficiently large sample of households to lead to a sufficiently large number of emigrants from the country. After identifying migrant persons linked to the households, a set of basic statistical information was also collected about them. After this, an attempt was also made to record contact information (e-mail address, telephone number, etc.) to the migrants reported in the household. These contact information then served as the base of the second stage of the pilot study when migrants were directly

contacted and asked to answer a series of more in-depth questions. The second stage of the survey was carried out via telephone and the internet.

The detailed research design of the SEEMIG pilot studies was elaborated at the Demographic Research Institute with intense and valuable help from colleagues at the Hungarian Central Statistical Office (HCSO) after extensive discussions with Thematic Experts and External Experts in the SEEMIG project. Documentation of the design developed in Hungary was translated into English and sent to project partners at the Statistical Office Republic of Serbia (SORS). Since the first stage in Serbia started two months after the start in Hungary, Serbian colleagues were in the position to take over the methods previously worked out in Hungary. Moreover, it was also possible to apply some small corrections to the method – based on the Hungarian experiences.

In the pilot study any person was recorded as migrant who at the time of the survey was declared as 'currently living abroad' according to his (her) household member in the country of origin and who was not born in the country where he (she) is currently living. To 'live' abroad was defined in line with the Regulation (EC) No 862/2007: 'spends most of his (her) time abroad – resting time included – either for work or any other purposes.' People on holiday were excluded. Following the LFS age-limitations, only persons aged 15 to 74 were included in the sample. Building our survey on LFS it was also obvious that household members as defined by LFS will form part of our target group as long as they meet our criteria of migrants. Consequently, SEEMIG data was collected about (1) any LFS household member who lived abroad at the time of the survey. The circle was then further extended to (2) 'any person who left abroad from this household' – setting a time limit in 1990, i.e. recording only those who left the country in 1990 or later. This way we broadened the circle of household members by abolishing the one year limit set by the LFS household membership definition. Finally, the targeted group was further extended by collecting information about (3) migrant siblings of any household member. Our aim with this was twofold. Firstly we wanted to increase the resulting sample-size and secondly, we wanted to reach out to migrant persons who moved abroad together with all their previous household members.

Any population survey that deals with confidential data of people has to face issues of ethics as well as of data protection. This was especially so with SEEMIG, where the survey was not only aimed at collecting data about the respondents themselves, but also about 'third persons' – (former) migrant household members and siblings. Given the special aim of SEEMIG study to generate a migrant sample, data collected about these persons made it also possible to identify and to directly contact them. In SEEMIG it was a priority to handle the evolving data protection issues with much care and responsibility. In Hungary, a detailed *data protection protocol* was developed to attend the relevant stages of the study from the questionnaire design to the data analysis.

From the very beginning, collecting contact details was considered as the most challenging and potentially most risky activity in the project. It is not only an attempt that raises the issue of data protection but it was also an unfamiliar exercise for the interviewers and it was also the stage when the highest attrition rate was expected to appear. Several elements of the research design were specifically tailored to achieve maximum success despite the difficulties embedded in this activity. A careful questionnaire-design aimed at guiding the interviewer through the interview-situation with various options offered to the respondent.

Respondents were for example invited to get in touch with their migrant acquaintance and ask for his (her) permission either on the spot or later. This flexible approach proved to be successful in delivering several valuable contact details for the project. On the other hand, the so-called 'SEEMIG Research Participant Card', left in the household so that the migrant person could get in touch with the researchers via the internet has been less successful. Beside these options offered to the respondents, a specially tailored bonus-scheme for the interviewers (giving extra bonus for contact details collected); trainings for the interviewers as well as the local managers; small incentives given to the respondents in Hungary and a media campaign before the fieldwork in Serbia all aimed at enhancing response rates at the various stages of the project but especially when it came to providing contact data. Previous studies with similar design have also shown that a key challenge for studies of this kind is to achieve a sufficient migrant sample size and to keep the sample representative of emigrants from the country. A high level of attrition rate is not only problematic because it reduces the sample size but also because non-response is likely to be unevenly distributed across the various segments of the target population. If this is the case, representativeness of the sample will be jeopardized.

From all the households where a SEEMIG battery was administered, at least one emigrant was identified in 10per cent in Serbia and in 8per cent in Hungary. This ratio resulted in a household sample with at least one reported emigrant of 816 in Serbia and of 1785 in Hungary. The number of emigrants reported was above these numbers in both countries since in many cases there were more than one emigrant persons reported in one household. This way altogether we had 1079 emigrant persons reported in Serbia and 2401 in Hungary – coming from the three emigrant groups we identified. However, after admitting the existence of an abroad-living sibling or household member, quite a high proportion of respondents in the LFS sample decided not to provide any further information about them. The attrition rate at this stage of the survey was 25 per cent in Serbia and 40per cent in Hungary. These figures are not possible to compare directly because of differences between the techniques applied in the two countries and also differences between the techniques applied in the case of the various migrant groups. Nevertheless, rich data was provided by their home-staying household- and family members about an unusually big sample of migrants both in Serbia (819) and in Hungary (1430). In this group we have valuable information on their gender, age, time of emigration, destination country, etc. which are, from now on, available for further analyses.

Unfortunately – but not unexpectedly – the most significant attrition appeared in the last step of the study, when contact information to the migrants was requested. Interestingly, attrition rates in Serbia and Hungary were very similar at this stage. Compared to the number of emigrants about whom the respondents provided detailed statistical data, a contact detail was also given in 36 per cent of the cases in Serbia and in 38 per cent in Hungary, resulting in a final emigrant-sample of 546 persons in Hungary and 298 in Serbia. Although results from the data collected in the second stage of the study are not yet available, from these low case numbers the lack of success of the attempt to collect information from a large, representative sample of emigrants in a direct way can be predicted.

High attrition rates have also affected earlier stages of the study. Evidence from ex post interviews with the interviewers, as well as some direct comparisons of the estimate of the number of emigrants from Hungary based on the SEEMIG data on the one hand and on

estimates based on data from other sources on the other, suggest that respondents were concealing information and did not report all their migrant household members (or siblings) in the study. Also, as it is apparent from the significant difference between the number of migrants declared and the number of migrants with statistical information given, that many respondents withdrew after reporting the existence of such a person. Although a systematic evaluation of the attrition is still under way, it is evident that the lack of trust from the respondents' side was a key factor that reduced response rates. In Hungary, the recent and very intense media- and political attention towards emigration from the country might have contributed to this. In Serbia, where at the time of the survey no similar factors were present, the SEEMIG-estimation seems to be less biased – at least when it is compared to the only available estimate that is results from the latest Census.

At that same time SEEMIG study has proved to be successful in providing valuable methodological experiences that can promote further improvements on collecting information about emigration. Our recommendations based on the first stage of the study include specific fieldwork-techniques that can be useful in a smaller-scale study when contact details to emigrants are collected but also research tools for collecting statistical data about a representative, large sample of emigrants in an indirect way.

Very importantly, data collected during the first stage of the study has great value in improving our understanding on emigration. Respondents in the LFS provided valuable statistical information on over 800 emigrants in Serbia and more than 1 400 in Hungary – representing **the biggest, and most systematically collected set of data on emigrants in both countries**. After a systematic evaluation of the selection processes throughout the study we will be able to analyse emigration from Hungary on an exceptionally large sample of emigrants. Moreover, individual level data will be possible to link to information on the sending household, which is again exceptional in the history of migration research in this country. Estimations derived from SEEMIG are expected to prove particularly valuable on recent emigrants from the country.

When exploring household-level characteristics of sending and non-sending households in Hungary, we will be able to build on the larger sample of 2 400 identified migrants and offer a unique insight into the process of how migrant-sending households get selected. In this step of the analysis regional patterns and the impact of demographic and social composition of households will be explored. On the individual level we will then be able to provide valuable individual-level data on the composition of the most recent emigrant groups from Hungary in terms of some key demographic and labour market indicators such as age, gender, educational attainment or employment situation in the country of destination. We also have information available on their financial linkages to Hungarian households – i.e. some badly needed insight into the field of remittances will also be given. At this stage, our analysis will be built on the sample of around 1 400 migrants about whom detailed information were given by their household-members or siblings in Hungary.

1. About this report

This report was prepared in the framework of the SEEMIG project (funded under the third call of the South-East Europe Programme, number SEEMIG/SEE/C/0006/4.1/X) as part of Work Package 4. Work Package 4 aims at data enhancement in the field of migration. The pilot study activity (Activity Nr 4.3.) is aiming at one hand 'to improve data sets on migration and related labour market and human capital processes, and on the other to comparatively evaluate different ways of ''reaching'' in a statistically representative manner migrant populations.'' The pilot study also aims to contribute to 'a methodological best practice which shall be described in details and suggested for further improvements ... and in the long term it will also contribute to an improved, evidence based policy making''. 'The pilot will also facilitate the effective cooperation of data suppliers, research institutes, national, regional and local level authorities.'

In this report we focus on the first stage of the pilot study. The first stage of the study consists of a national survey, carried out in the home-countries with the aim of collecting statistical information on and contact details to emigrants from the country. The second stage then consists of an emigrant survey, carried out on the migrant-sample collected in the first stage of the study. Information on the second stage will be provided in a separate SEEMIG report, early 2014.

The structure of this report is as follows. In Chapter 2 we provide a short background to the SEEMIG study: the information-gap the SEEMIG pilot study wants to fill is described and earlier research with similar approach is introduced. After summarizing the key lessons from these, Chapter 3 explains the design of and the actual activities within the SEEMIG study in detail. In the subsequent sessions, definitions applied are described, the questionnaire gets introduced and so do the main considerations in relation to data protection in the SEEMIG study. A separate session is dedicated to the advantages of connecting the SEEMIG battery to the LFS and also the compromises deriving from this are explained. In the last session of this Chapter preparation of the fieldwork is described as well as the activities developed to enhance response rates in the survey.

In Chapter 4 some of the key results from the pilot study will critically be reviewed. In particular we will look at the level of attrition at the different stages of the survey comparing results from Serbia and Hungary. Different methods applied for gathering contact information will also be evaluated showing their success-rates in the survey situation. Individual survey-items will also be assessed on the basis of the participants' ability to provide answers to them. Finally show an estimate on the number of emigrants from Hungary on the basis of the SEEMIG study. The SEEMIG estimation will be evaluated in the light of comparable data at hand and also possible reasons for the revealed underestimation in the SEEMIG study will be discussed.

Chapter 5 concludes by listing the key lessons learned from the exercise and providing methodological recommendations for future attempts. It is important to note that on the basis of the first stage of the project only preliminary conclusions can be made. A full evaluation of the method and detailed methodological recommendations based on the findings will be included in the second report.

Throughout the report, general features of the study as well as their realization in the Hungarian pilot is explained in the main text while deviations from the general method in

Serbia and other country-specific elements of the Serbian pilot are given in additional boxes. Information on the Serbian study was written by colleagues from SORS.

2. Background to the study

2.1. Shortcomings of migrations-statistics

Shortages of migration statistics are well-documented in the relevant literature and they are also acknowledged and thoroughly analysed in the frame of the SEEMIG project (Gárdos-Gödri 2013). It is well known that administrative data on migration is often unavailable and provides a poor coverage of the relevant population. For definitional inconsistencies they also hinder international comparison even in a European context and their timeliness is also problematic. Moreover, migration data available from administrative sources also lack the necessary richness for an in-depth analysis and a sociological understanding of the social phenomenon of international migration. All these difficulties are especially significant in the case of measuring emigration.

Although survey type data collection might appear as an obvious substitute that could overcome the problems listed above, making an appropriate research design to capture important features of a representative set of the migrant population poses serious challenges to the researcher. The key problem being the lack of a sampling frame for emigrants from a given country, it is not surprising that most attempts made so far use non-random sampling methodologies (such as snowball sampling) and often concentrate on a selected set of migrants – by profession or most often by country of destination. Although it is not impossible to produce a representative sample of migrants e.g. directly on the borders of the country such methods are not only expensive but also rather time-consuming.

To improve the situation, in the framework of the SEEMIG project an attempt was made to test and further develop an innovative research design which is based on an internationally comparative, firm and standardized methodology, and which is also financially sustainable and therefore might serve as a best practice for statistical and research bodies Europe-wide to survey emigrants in a systematic and reliable manner. To achieve these goals two pilot studies were carried out – one in Hungary and one in Serbia.

2.2. The SEEMIG approach

To achieve the goals set above, SEEMIG ambitiously aimed at building a ***sufficiently large, representative***, ***unbiased*** sample of migrants having left from a specific country. It was acknowledged however, that this is a great challenge for at least three reasons – all deriving from the mere nature of the population targeted. Firstly, there is no appropriate sampling frame to use not only because many of the emigrants are unregistered in their destination country but also because of the large number of countries such a study aims to reach. Secondly, migrants typically constitute a rare population, members of which are not well concentrated in geographical locations. Finally, but most importantly, non-response is likely to become a major challenge for a series of general reasons but also for some that are especially remarkable in the countries targeted.

In our research plan explained below it will be shown how SEEMIG attempted to overcome these difficulties.

In the SEEMIG study a two-stage methodology was planned the key idea being to *derive a representative sample of migrants from a representative and large-scale national survey.* In

the first stage of the study the Labour Force Survey was utilised and international migrants were identified through the households included in the LFS household-sample. LFS was chosen to be used for the purposes of the pilot for several reasons. It is a highly standardized survey carried out regularly in all EU member countries and (in most cases) it also consists of a sufficiently large sample of households to realistically lead us to a sufficiently large number of emigrants from the country[1]. In the additional SEEMIG battery attached to the LFS questionnaire not only the emigrant persons linked to the household were identified, but also a set of basic statistical information was collected about them. After this an attempt was also made to record contact information (e-mail address, telephone number, etc.) to the migrants reported in the households. These contact information will then serve the base of the second stage of the pilot study when migrants will directly be contacted and asked to answer a series of more in-depth questions.[2] The second stage of the survey will be carried out via telephone and the internet.

The potential advantages of a research design of this kind are twofold. First, it collects information on the migrant persons irrespective of their destination country – i.e. it has the potential to represent a wide and heterogeneous group of emigrants. This is an important feature of the design, since most of the techniques applied in emigration research focus on migrants in a specific destination country. Secondly, collecting information both in the country of origin and the destination country makes it possible to link information on the migrant person to information on their originating communities. Consequently, it also becomes possible to compare households with and without migrants, that is the process of selection into emigration can be analysed.

2.3. Earlier surveys on emigrants

Surveys on emigration often compromise between representativeness and thus the capacity to provide a reliable estimation on the volume of emigration on the one hand and richness of data collected necessary for in-depth analysis on the process of migration on the other. Large-scale surveys with a complex sampling methodology usually do not have the capacity to go beyond some basic statistical data whereas rich data usually comes from more in-depth studies with non-representative samples of varying sizes.

As it was said before, in emigration research it is common that surveys are conducted without aiming at representativeness. Instead they choose to focus on providing detailed data on a relatively large sample of migrants. Origin-based surveys often use snowball techniques (e.g. Massey 1987; Arenas et.al. 2009). In these studies the number of destination countries reached was also limited to one country or to a small set of countries to ease fieldwork. Although snowballing is a useful technique to overcome several difficulties embedded in researching vulnerable groups, it does not claim to result a sample representative to the target population. An alternative to ordinary snowball sampling is Respondent Driven Sampling (RDS)[3] which is a specific form of snowball technique with very

[1] On further important features but also drawbacks of choosing LFS see later chapters.

[2] This report only covers the first stage of the survey whereas description of the second stage will be provided in a later SEEMIG report.

[3] See SEEMIG attempts to apply RDS as an extension of the original SEEMIG design in the report on the second stage of the research to be issued later.

specifically defined rules, thus allowing to reach a representative sample. As described by Beauchemin and Gonzalez-Ferrer (2011; pp.106) RDS has also been used in emigration studies concentrating on one country only, and they have not led to convincing, well-documented result so far.

Surveys aiming at representativeness and thus providing reliable estimate for the extent of emigration also tend to be started from the origin country (community). A possible approach is to sample travellers on the country borders. Such method is applied for example in the UK (Jensen et.al. 2012) and – among the SEEMIG countries – in Bulgaria. (Kostova and Yakimova 2013) Limitation of this method is the high costs implied and the restricted depth of data that can be gathered. Also, as it is conducted at the time of leaving the country it has no capacity to collect information on migration-experiences.

Collecting information on emigrants in the country of origin through their relatives of household-members is an approach that has potentials not only for being the base of a snowball sampling, but also for providing data on a representative set of emigrants[4]. The aim of such surveys can be twofold. Once, sample-members in a national representative survey can provide information on their emigrant relatives or household-members. When all details are appropriately designed, the sample of emigrants reported in the survey can properly represent the emigrant population. This way, not only a reliable estimate of emigration can be given but it will also be possible to provide estimates on distributions of this population based on the responses to the survey-questions given by the household-members (relatives) in the origin country. A second ambition of this approach is to use the sample of emigrants collected in the origin-country for the purposes of an emigrant-survey when emigrants are directly contacted using the contact-information collected in the households. A success of this second ambition can lead to unexceptionally rich data on a representative sample of emigrants.

The first of these two sets of goals, i.e. relying on an indirect, origin-based data-collection in household-surveys to estimate emigration, has been used in several earlier attempts. (An early critical review of them is given e.g. in Zaba 1987). In these studies indirect estimation methods are used to estimate the number and the composition of emigrants on the basis of the number of abroad-living siblings, children or previous household members of the respondents in the national survey (see Jensen et.al. 2012). As the surveys collect information on third persons, special statistical techniques are needed for data weighting as well as for deriving reliable estimates.

A recent study in Nepal has set both of the possible goals explained above: to collect information on a representative sample of emigrants through a household-survey in the origin community and to carry out a direct emigrant survey based on the first data collection (Ghimire, D.J. et.al. 2012). The survey was built on a well-established panel study, the

[4] A less frequent method is to use information sources on the community level. In the so called ''community censuses'' in Romania questionnaires regarding emigrants from the local community were sent by post to the local (communal) police offices. The questionnaire were completed by so called key-informers who, in a proportion of 60 per cent, were employees of the major's office, but we could find teachers or other representatives of local intelligentsia among them. Although not free from validity problems, results from the survey has been widely used for estimating emigration from Romania (see SEEMIG country report *Analysis of existing migratory data production systems and major data sources in Romania*: http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportRomania.pdf)

Chitwan Valley Family Study in Nepal, and identified migrants from the originating community to the Gulf Cooperation Council countries. In the first stage, personal interviews were carried out by interviewers well acquainted with the members of the households and in fact also with the wider neighbourhood.

The Nepal study was extremely successful in collecting contact information as well as in eventually finding and interviewing members of their target population. In 92 per cent of the cases when a migrant person was identified, contact information was also provided to the interviewer. In the second stage 87 per cent of their target respondents were successfully interviewed in 6 months and 95 per cent in 26 months.

A key factor in achieving such a high level of response rates in both stages of the survey was the intense fieldwork applied. Very importantly, the survey was administered on a long-running, well-established sample in Nepal with experienced and well-trained fieldworkers being in ongoing contact with the interviewees. Throughout the fieldwork a flexible and also personal approach was taken. Households were revisited when the first person was unable to provide a contact; the wider social network at the place of origin and also at the destination was utilized to generate the necessary contact information when it was needed; interviewees were provided a mobile phone to ask for permission of the migrant declared, etc. Another personal element that was likely to further enhance cooperation was fieldworkers offering to deliver messages between household and migrant.

Of course the social environment in which the Nepal study was taking place was also markedly different from the (South-East) European one. Nepal is a low-income agricultural country, with a massive increase in emigration during the past decades. The society is a traditional one, concerning both its way of living and its value system with small, closed local communities with strong ties and famylistic values.

Although the measures from this research are impressive, earlier attempts suggest that special features of the Nepal study have played a crucial role in achieving them. Earlier research attempting to collect contact details to migrants in their former household led to varying and sometimes very low success rates – with only 5 per cent of declared migrants interviewed successfully in the MAFE project for example[5] (see e.g. Beauchemin and Gonzalez-Ferrer 2011). Lessons learnt from earlier studies

To sum up the mixed lessons from previous research in the field, the proposed method was chosen as the potentially most effective design albeit with well-known risks embedded. In the SEEMIG research we decided to apply an origin-based method when a household survey is used (1) to estimate emigration and provide estimates on the distribution of the emigrant population and (2) to build and use representative sample of emigrants in a subsequent survey.

It was clear that in the SEEMIG context it is not possible to build our survey on a well-established research that involves intense qualitative elements and establishes close links between the interviewers and the respondents. As we believe that these elements were crucial factors that led to the great success of the Nepal study we acknowledged that the SEEMIG attrition rates would be more likely to resemble those from the MAFE study. This is even more so, since we also assumed that in the South-East European (SEE) social context

[5] Beauchemin, C. and Gonzalez-Ferrer A. (2011); pp.106.

emigration might be a more sensitive domain, especially in Hungary, where rapidly increasing volume of emigration is a new phenomenon.

Nevertheless, we still decided to carry out the proposed pilot studies. This decision was not only made because the method – if carefully applied – still offered itself as the best available choice for improving the current situation of emigration statistics on the SEE region. It was also obvious that even if the ultimate aim of producing a large representative sample of migrants to be contacted directly would fail, the project is still likely to provide us with a range of useful outcomes. Size of LFS ensures that emigration will be measured on a sample bigger than what has ever been used for estimating size and composition of emigration in Hungary before.

Furthermore, testing a research method in a (South-East) European environment that has only been tested in very different settings (African countries) before can be a valuable contribution to the common knowledge base in emigration research. Conducting the survey provides an excellent opportunity to test and understand the possibilities and limitations of surveying emigration in the SEE region with a relatively small budget. Based on our experiences it was expected that lessons would be learned that help us further improve the methodology and possibly also to better adjust it to the specific European environment.

Furthermore, (after the second stage) the pilot study is expected to help us identifying a set of attributes on which the migrants' relatives in the home-country can reliably report. From a systematic comparison of the answers provided by the migrants themselves and their household members in the home-country, we can identify those migrant-attributes that can validly be investigated in a representative household-survey in the origin-country. This can serve as a validation of survey questions which can then be included in upcoming surveys on attributes of emigrants.

Even if we expect the representative sample to get reduced and also biased through the subsequent stages of the project, we can still expect to collect information about an exceptionally big sample of migrants from their household members and relatives. This data will provide us (at least in Hungary) with detailed information about migrants in a greater number than any former surveys did. Even if formal tests of representativeness of the resulting sample would fail, describing most common patterns of emigration still remains possible.

The process will also provide an opportunity to test further alternative methods (e.g. applying Respondent Driven Sampling) at later stages of the project.

3. The first stage of the pilot study in Hungary and Serbia

3.1. Introduction

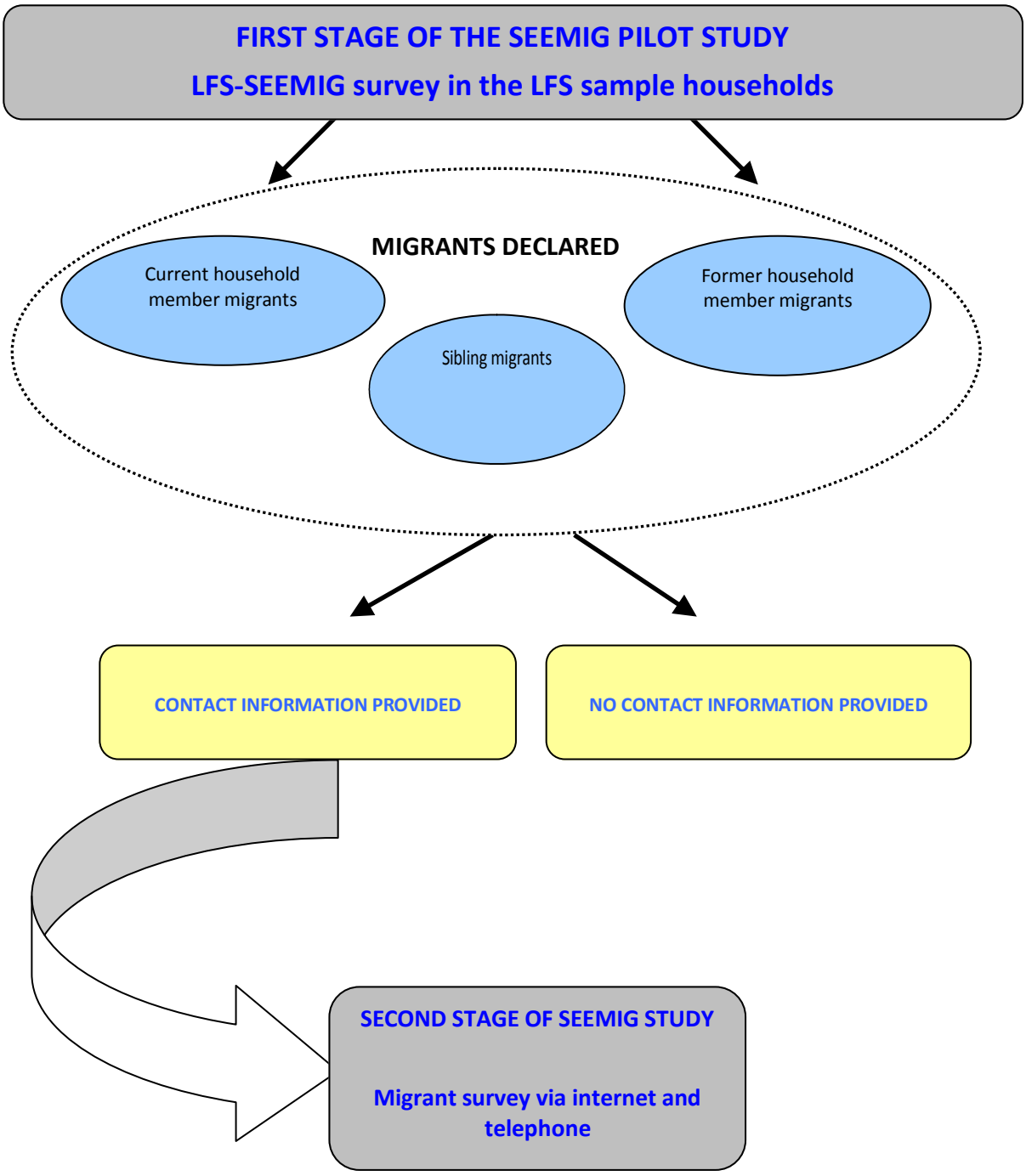
Taking all the above-mentioned experiences and considerations into account, Thematic Experts worked together with External Experts in the SEEMIG project to develop an efficient research design to achieve the goals set and maximize results despite the obstacles embedded. After the international SEEMIG discussions in Bratislava September 2012, the final research design was elaborated at the Demographic Research Institute with intense and valuable help from colleagues at the Hungarian Central Statistical Office (HCSO). A close cooperation with the LFS expert team at the HCSO was especially important since it was them who could efficiently represent the viewpoint of the LFS, making sure that the implementation of the SEEMIG battery will not in any way harm the LFS sample and data collection.

Documentation of the design developed in Hungary was translated into English and sent to colleagues at the Statistical Office Republic of Serbia (SORS). Since the first stage in Serbia started two months after the start in Hungary, Serbian colleagues were in the position to adapt the methods designed in Hungary. Moreover, it was also possible to apply some small corrections to the method – based on the Hungarian experiences.

3.2. The SEEMIG research design

The SEEMIG research design constitutes of a two-stage research plan as it is shown on Figure 1.

Figure 1 The SEEMIG research design



In the first stage a large number of households – all members of the LFS sample – in the originating country are contacted and asked whether there is any migrant person linked[6] to their household. About migrant persons identified this way, a small set of questions is asked to collect basic data on their education, employment and migration history. The quantity and type of data gathered about migrants enables us to describe some basic characteristics of the migrant population (e.g. distribution by country of destination, age, gender, activity abroad, etc.) directly after the LFS-SEEMIG data collection. At the very end of the LFS-SEEMIG survey interviewed household members were asked to provide contact information (e-mail address, phone number) to the migrant person identified.

In the second stage of the study, contact information collected in the first stage will be utilized and an attempt be made to contact emigrants (either in their destination country or during their visit at home) and to administer a short questionnaire with them. This second stage was planned to be multi-method surveying, applying telephone calls as well as electronic questionnaires. The questionnaire of the second stage was planned to cover a more detailed migration history and also some additional information on demographic and social characteristics. Very importantly, the second questionnaire should also make it possible to test the validity of information gathered from household members in the country of origin.

3.3. Definition of migrants

In Hungary any person was recorded as migrant who at the time of the survey was declared as 'currently living abroad'' according to his (her) household member in the country of origin and who was not born in the country where he (she) is currently living. To 'live'' abroad was defined in line with the Regulation (EC) No 862/2007: 'spends most of his (her) time abroad – resting time included – either for work or any other purposes.'' People on holiday were excluded.

This additional explanation was especially important because of the (potentially) large number of commuters who might follow individual regular or irregular patterns in their time-use. According to our definition, daily commuters did not form part of our sample but for example weekly commuters or those who commute on an irregular basis in an intense manner (e.g. two weeks of work abroad, one week stay at home) did. Additional questions regarding the frequency and length of home-visits included in the questionnaire made it possible to distinguish between classic migrants and commuters as described above[7].

As can be seen, when defining 'migrants'', we decided not to impose any limitations regarding the time spent abroad[8]. Instead, anyone living abroad was recorded and an

[6] For how being ''linked'' is defined see later in this chapter.

[7] In fact the extra explanation ''*Please regard a person as a migrant if this person spends most of his/her time abroad (including rest time) because of a job or something else. Even someone who visits home on a weekly basis counts as a migrant.' was added to Questions 1, 10 and 19 after one month of fieldwork . Although this definition had formed part of the interviewers' guidelines from the beginning it was felt that a more explicit declaration is needed during the interview to help clarifying the status of some abroad-working acquaintances.*

[8] Except that in Hungary we defined ''former household migrants'' as persons who left the household in 1990 or later. This was, however, not to exclude early migrants, it rather was an attempt to control the notion of ''household'' by not extending the category for an unrestricted time period in retrospect. This approach was, however, not followed in Serbia.

additional question on the date of emigration made it possible to categorise migrants according to their length of stay ex post.

Also, every person who currently lives abroad but previously lived in Hungary is included irrespective of his (or her) nationality and citizenship. Following the age-restrictions applied by LFS only migrants aged between 15 and 74 were included in the sample.

As the data collected in the questionnaire allows us to distinguish between migrants by their length of stay abroad, the frequency of their visits home as well as by their country of birth and citizenship, the broad definition applied gives us flexibility to look at different groups of migrants depending on the specific analytical needs.

Definition of migrants in Serbia[9]

A similar migrant definition was used in Serbia except that persons who were born abroad but had lived in Serbia for at least one year were also considered as migrants.

3.4. Defining 'persons linked to the household''

Building our survey on the LFS, it was obvious that household members as defined by the LFS will form part of our target group as long as they meet our criteria of migrants. Consequently, SEEMIG data was collected about (1) any LFS household member who lived abroad at the time of the survey. This was, however expected not only to be a too small group of migrants but also one that is defined too strictly for our purposes – in the case of Hungary this includes only those who 'live abroad for no more than one year' and who also 'share their income with the household'. Therefore, we extended the circle defined by the LFS by also enquiring about (2) 'any person who left abroad from this household'', setting a time limit in 1990, i.e. recording only those who left the country in 1990 or later. This way we broadened the circle of household members by abolishing the one year limit.

Finally, the targeted group was extended by collecting information about (3) migrant siblings of any household member. Our aim with this was twofold. Firstly, we wanted to increase the resulting sample-size and secondly, we wanted to reach out to migrant persons who moved abroad together with all their previous household members. This was a crucial step, since data collections that gather information about missing household members only (censuses for example), will by definition omit this significant group of migrants from their target group.

By including migrants in our resulting migrant-sample who are not (or who have never been) members of the households included in the LFS sample, we are in fact applying *indirect sampling methodology* and we are bound to use the consequent weighting process afterwards[10]. (Deville and Levallee 2006)

[9] For a historical overview of the migration processes in Hungary and in Serbia see the SEEMIG WP3 Country Reports on Serbia and Hungary. Upcoming.

[10] Note that this aim also motivated the inclusion of some specific questions in the questionnaire (e.g.: Questions on the siblings of household members or questions on income-transfers in the case of migrant siblings.) This also implies that it is very important not to omit any questions from the battery since it might jeopardize the usability of the final sample.

Figure 2 Overlapping circles of migrants

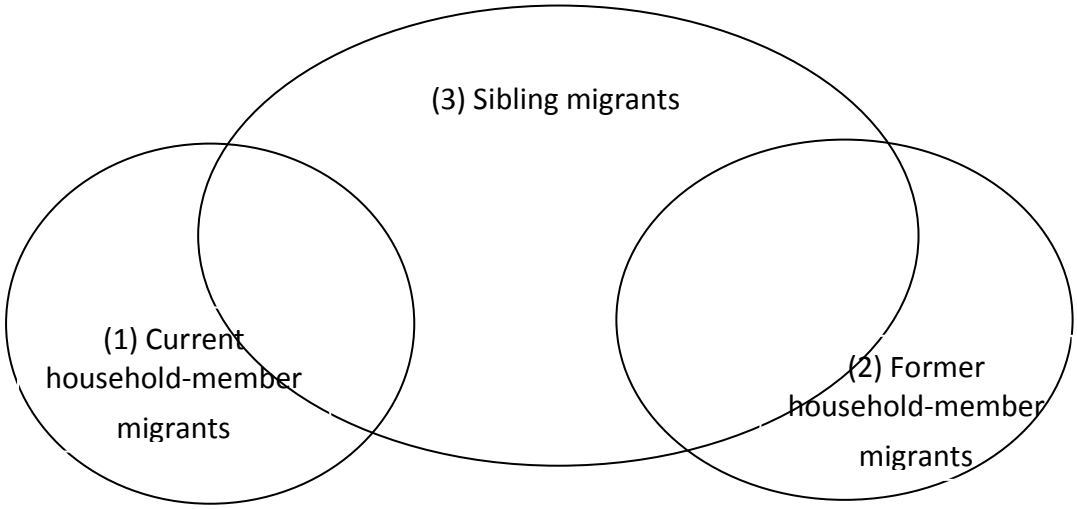


Figure 2 provides a representation of the three groups of migrants the SEEMIG study covers. As can be seen, the three circles are overlapping suggesting that a person who is a sibling of one (or more) member of the household can also be a (former) household member. This possibility had to be dealt with in the questionnaire-design (to avoid double-reporting) but it also affected weighting as it will be shown later.

From the previous sections it follows that the target population of the SEEMIG pilot study constitutes of the following group:

- Hungarian citizens and persons born in Hungary who live abroad and are between 15 and 74 years old

  AND

- they are either a current or former member of a Hungarian household and they moved abroad either in 1990 or thereafter

  OR

- they have a sibling aged between 15 and 74 living in Hungary[11].

3.5. Data protection

Any population survey that deals with confidential data of people has to face issues of ethics as well as of data protection. This was especially so with SEEMIG, where the survey was not only aimed at collecting data about the respondents themselves, but also about 'third

[11] Persons born in the country where they are currently living (mostly Hungarian nationals in neighboring countries) were also excluded.

persons' – (former) migrant household members and siblings. Given the special aim of the SEEMIG study to generate a migrant sample, data collected about these persons also made it possible to identify and to directly contact them. In SEEMIG it was a priority to handle the evolving data protection issues with much care and responsibility.

In Hungary therefore we asked for legal a statement from the Data Protection Committee of the Hungarian Central Statistical Office. The Committee raised the objection that asking personal information about a third person which also makes this person identifiable in a survey might go against the EC directive Nr 95/46. The Committee therefore decided to forward our request to the Hungarian National Authority for Data Protection and Freedom of Information. The Authority provided us with a detailed statement which, however, was not fully conclusive and warned us to handle the issue with care[12].

Following these steps, a detailed ***data protection protocol*** was developed which attends to all the relevant stages of the study from questionnaire design to the data analysis. The protocol includes the following elements:

1. All staff participating in the process of collecting and handling data sign a confidentiality statement.
2. Respondents (household members) will be informed about the objectives of the research in much detail. This is partially done orally by the interviewer, partially via written documents such as a special edition of the SEEMIG project newsletter and the Data protection statement (See next point).
3. Every time the respondent is requested to provide a contact detail, the interviewer leaves a document called *Data protection statement*[13] in the household – no matter whether a contact detail is eventually provided or not. This document describes the aims of the research and ensures the respondents that all data protection rules are followed. The statement is signed by the main researchers who provide their phone numbers so that the respondents could contact them with any arising questions or problems.
4. Respondents (household members) get informed that information collected within the survey will only be used for research purposes, and no direct link will be made between household members (or their relatives) and the statistical information collected about them.
5. During the interview situation respondents are offered the possibility to contact their migrant household members (or relative) and ask for permission to give their contact details to a third party. (See the detailed description of this procedure in the next section.) Interviewers do their best to ensure that respondents take this opportunity of getting consent from their migrant household members and to arrange a next meeting or telephone contact with the respondent, once the consent is given.
6. No matter whether or not the consent of the migrant person was already provided in the first stage, the consent will directly be asked in the second stage of the research, when migrants themselves complete the online or telephone questionnaire.

[12] For the statements see Annex II.

[13] See Annex II.

7. *Data* are only *used for the purpose of the research.* All the collected contact details are handled with special care, complying with all the legal rules and regulations regarding data protection.
    - Data suitable for identifying persons (contact details and names) are *stored separately* from personal data (data gained from the LFS survey in the first stage of the research) and they are used exclusively to get in touch with the migrant.
    - As soon as the contact with the migrant has been established, all the information related to the contact details will be *destroyed*.
    - Personal identity data are *not connected* in any way to personal data collected directly from migrants.

Serbia

Data protection in the Republic of Serbia is regulated by the Official Statistics Law. The Law is relevant for all statistical surveys under the jurisdiction of the Statistical Office of the Republic of Serbia. Survey respondents were informed about the respective legislation and the data protection procedure during the interview and – as in Hungary – they were offered to ask for the permission from their migrant relatives before providing their contact details.

3.6. Design of the questionnaire

The questionnaire was designed in Hungary then translated into English. After a detailed discussion between the partners in Vienna February 2013, Serbian colleagues translated the questionnaire into Serbian and also made the necessary changes to better adjust it to their specific needs.

3.6.1. Main considerations

The design, the logic, wording of the questionnaire, etc. had to be in harmony with LFS standards. This requirement has affected the placing of our battery at the end of the LFS block so that it did not interfere with the usual flow of the LFS interview (even when it would have produced a more sound sequence of questions from other aspects). Basic educational and employment characteristics that are usually collected about each LFS household members, including migrants were not collected again in the SEEMIG battery, and the same information was collected in the same way (the 'LFS way') about other groups of migrants (former household members and siblings).

It is important in any population-survey that the questionnaire is short and manageable in a short period of time so that it does not overload the respondents. With the SEEMIG battery however it was even more vital than usual, since respondents were faced with the battery after an already time-consuming LFS questionnaire. Also, we had to make sure that we do not deter them from remaining on the LFS panel for the upcoming waves.

The specific weighting method, the Generalized Weight Share Method (GWSM, see later) we had planned to use made some questions necessary to be included in the questionnaire. In particular, questions number 10, 37 and 60 were added to the questionnaire because the information they provide are a prerequisite for conducting a proper GWSM weighting on the sample.

As described in Chapter 3.5, data protection was of special concern in this study. Aspects of data protection were reflected on mostly in the section where contact information was collected from the respondents. A detailed explanation on the purpose of the study, the data protection letter provided at this stage of the interview, the opportunity offered to contact the migrant person before providing his or her contact details are all elements of the questionnaire that had been motivated (also) by the data protection requirements.

As the core version of the questionnaire was prepared in Hungary, it was designed for Computer Assisted Personal Interviewing (CAPI).

3.6.2. The structure of the questionnaire

The SEEMIG battery comes directly after the general questions of the LFS, with the following structure[14]. In the main SEEMIG battery we start by asking about the three reference circles of migrants as described above. First about the household members themselves, then an extended group of household members (those who left the household for more than a year), and finally the siblings of the household members are covered. We follow the same procedure with all the three groups: first we ask if there is a person living abroad in this circle, then we ask for their first names to ease identification during the interview process and finally go through a series of personal questions, filling in the so called *emigrant data sheet*. These data sheets vary to some extent, depending on the reference circle. This is partly due to the fact that questions in this block overlap with the base-questionnaire, which makes it unnecessary to include them in the block relating to current household members and partly because certain complementary questions are required by the weighting process only for certain groups of migrants. All in all, the emigrant data sheet makes the following set of information available about each migrant in the survey.

- *Year of birth*
- *Gender*
- *Country of residence*
- *Time of emigration (year and month)*
- *Employment status*
- *Family status*
- *Educational attainment*
- *Citizenship*
- *Number of visits to Hungary during the past 12 months*
- *Total amount of time spent in Hungary during the past 12 months*
- *Financial linkages to home-country household (whether or not financial support is sent home / received from home)*

At the very end of the questionnaire we ask the respondent for the contact detail of the person living abroad. The majority of the questions are simple closed questions.

[14] In the first section there is also a short series of questions about household members who live in Hungary but work abroad on a more or less regular basis. This block was also extended by the statistical office with some migration potential questions – these are not strictly necessary for SEEMIG purposes.

3.6.3. Collecting contact details

Successfully gathering contact details to abroad-living persons from our respondents in the LFS-SEEMIG survey is crucial and also the most sensitive part of the interview. It is clear that collecting individual data which does not only make the persons identifiable but also makes them approachable for the data gathering body is a very sensitive issue that requires a great amount of trust between the interviewer and the interviewee. As we could see in the Nepal case, the necessary amount of trust had been built up through a long and intense panel-maintaining process, which is typically not the case with the LFS surveys in Europe. Also, it was not possible for us to apply costly fieldwork techniques (e.g. offering mobile phones to the respondents, etc.) that are not standard part of the LFS procedure. Thus we had to aim at gaining the necessary amount of trust and maximising the cooperation with measures that are easy to standardise and to attach to the LFS protocol.

A carefully designed process of gathering contact information was therefore applied. At the end of each interview in which a migrant person (no matter whether a household member or a sibling) was identified, the interviewer briefly explained the importance of getting in touch with the migrant person directly and also described the data protection protocol applied in the study. At the same time, the data protection letter (a declaration signed by the main researchers of the project) as well as a SEEMIG project newsletter and a small incentive (a textile bag with SEEMIG logo) were handed over to the respondent.

After this, respondents were offered to get in touch with the migrant person they declared – either on the spot via (their own) phone, or at a later time. Those who decided not to take this option but provided the requested details were asked to give at least two of the following information: e-mail address; skype contact; mobile phone number; other phone number; date of next visit home, together with a contact information at home. Those respondents who chose to contact their migrant acquaintance right away and received permission followed the same procedure. When a later communication with the migrant person was chosen, the interviewer fixed the time and the mode (face to face or telephone) of another appointment with the respondent. This way we successfully introduced some element of flexibility into the otherwise highly standardized process of data collection.

If at any stage of the interview-process the cooperation was denied by the respondent (but NOT when the migrant person him- or herself denied the cooperation via the telephone) a SEEMIG Research Participant Card[15] was left in the household. The Card included a personal identification code and a link to the project website with the electronic version of the questionnaire prepared for the second stage of the study. Household members were then requested to give (or send) this card to their migrant acquaintance[16].

[15] See Annex I.

[16] After the first month of the fieldwork we decided to provide the card even in cases when the interview gets terminated by the respondent at a stage before the data-gathering section, but after a migrant person is identified. After this change, when the household member declared that there was a former household member or a sibling living abroad but refused to answer the questions about this person, they were directed to the contact details part of the questionnaire. This way we increased our chances to receive the contact detail (not very likely) or to leave a SEEMIG respondent card in the household.

Difficulties in adapting the SEEMIG questionnaire – Serbia

Since the SEEMIG emigration pilot study in Serbia was administered in the paper form, adjusting the Hungarian questionnaire that was originally designed for an electronic application was not always straightforward. As can be seen in the attached questionnaire (See Annex I), in the paper-form of the Serbian questionnaire, questions were inserted at the top of the page, horizontally. Down space was left for enrolling the answers for the household members – there were eight rows for this purpose. The Serbian questionnaire had all questions the Hungarian version did, except questions related to the future plan of living abroad of the household members. The structure of the questionnaire very closely followed the Hungarian one.

Serbian colleagues had problems with the identifying of those household members or siblings living abroad who were not mentioned in the questionnaire before. This problem was overcome by adding new empty columns which have specified the sequence numbers for different groups of the migrants in advance. One person could only be mentioned once in the questionnaire, this way there was no overlapping.

3.7. LFS and SEEMIG

As it was said before, LFS was chosen as the national representative survey to which the SEEMIG battery should be added. It is not new to utilize LFS for emigration research purposes. As it is discussed in the SEEMIG summary report *Analysis of existing migratory data production systems and major data sources in eight South-East European countries* (Gárdos and Gödri, forthcoming), the LFS has been used for analyses on labour-migration in Romania and also in Hungary and we are also aware of similar attempts in Moldavia. The target population, however, is rarely extended to a population beyond the household members.

Advantages of the LFS in emigration studies include the large sample size; the standardized methodology applied across Europe; the regular data collection sessions and the rich set of data collected which are also relevant for the analysis of international emigration. Also, linking the SEEMIG battery to a panel survey (rather than to a single cross-sectional one) offers the advantages of relying on ongoing contacts between the interviewers and the respondents and building on already established, potentially positive attitudes towards the survey.

Obviously, building the SEEMIG survey on an ongoing survey, rather than establishing a new data collection can save financial resources especially considering the sample size needed to reach a sufficient number of emigrants. At the same time, however, the close links to a well-established large-scale international survey also imply compromises. Conditions of the SEEMIG survey were to a large extent determined by the standard, largely inflexible procedures applied in LFS.

As mentioned above, format of the questionnaire, communication style applied in the wording of the questions, basic definitions applied were all pre-set according to the LFS standard. Similarly, the interviewers were originally employed for LFS and SEEMIG had only very limited possibilities to direct or control their work. Interviewers had to work according to LFS regulations, and there was no room for much flexibility that could possibly enhance cooperation of the respondents (other than the possible return or recall after the respondent has collected the migrant's permission).

Naturally, a key priority of the LFS team was to avoid any chance of jeopardizing the successful LFS data collection. Collecting contact details, however, appeared as a non-standard activity which does not only require a complicated set of questions to be included in the questionnaire, but which also puts extra burden on the interviewer (and the interviewee). Of course the SEEMIG questionnaire had to be designed to minimise the risk of evoking distrust in the respondents and does not in any way endanger the further cooperation of the LFS panel members.

The interviews were administered electronically in Hungary and the software (BLAISE) designed for LFS purposes had to be used. Special features involved in the SEEMIG battery (such as the need to avoid double-coding of migrants who are both current household members and siblings or former household members and siblings at the same time) posed extra challenges for the LFS informatics personnel.

Finally, timing of the survey was also strongly affected by factors related to LFS-administration. In Hungary the length of other EU batteries already settled to be included in certain waves of the LFS had to be taken into account. In the case of Hungary this meant that SEEMIG data collection had to be done between January and April 2013, leading to rather tight deadlines throughout the design and implementation stages.

Key features of the LFS in Hungary and in Serbia

Hungary	**Serbia**
The emigration pilot study was conducted with core LFS in the first wave of 2013. **Responsible institution (for LFS):** Hungarian Central Statistical Office (HCSO) **Frequency of surveys:** monthly **Sample size of the first quarter year**: 35 835 households. **Sampling:** The Labour Force Survey is based on a multi-stage stratified sample design. The sample design strata were defined in terms of geographical units, size categories of settlements and area types such as city centres, outskirts, etc. **Base used for the sample (sampling frame):** 2001 Population and Housing Census. **Last update of the sampling frame:** 2011 **Primary sampling unit (PSU):** In case of self-representing settlements dwellings are PSUs and in the other part of the sample settlements are PSUs. **Final sampling unit (FSU):** Dwellings are FSUs. **First (and intermediate) stage sampling method:** Sampling method: non-self-representing localities are selected with Probability Proportional to Size method.	The emigration pilot study was conducted with core LFS in the first wave of the 2013 year. Cooperation between LFS team and colleagues from SEEMIG project was without any difficulties. **Responsible institution:** Statistical Office of the Republic of Serbia (SORS) **Frequency of surveys:** semi-annual. From 2008 until 2012 Labour Force Survey was conducted twice a year. In 2013 we conduct LFS three times, and the plan is that from 2014quarterly surveys will take place. **Sample size:** 10 000 households (as from 2013) **Sampling***:* panel survey*,* two-stage stratified random approach, six rotation groups in each wave, 50per cent overlap with previous wave, four rotation groups CATI, two CADI. Two-stage, stratified sample. The first stage units are enumeration areas and the second stage units are households. Enumeration areas, as primary units, are stratified by the type of settlement (urban and other) and by territory covered by administrative districts. Frame for sample selection for the first stage is the list of enumeration areas with 20 or more households, and for the second stage the frame

Housing units are selected in sampled localities with systematic random sampling. The quarterly sample is made up of three monthly sub-samples. The monthly sub-samples have no overlap. **Final stage sampling method:** In sampled localities with systematic random sampling. **Overall theoretical yearly sampling rate (i.e. including non-response):** In the different strata of the LFS sample different sampling rates are used. Overall sampling fraction: f = 3.68 per cent **Target population:** all persons aged 15 and above, living at the national territory; persons from institutional households, diplomatic and consular personnel were excluded. Basic set includes all households and persons who work or reside on the territory of Hungary for one year and more. Excluded are persons who live abroad for more than one year and persons in institutional households (students' homes, homes for children and young people with developmental disability, homes for socially imperilled children, old / retired people homes, homes for adults with disability, monasteries, nunneries, etc.). **Method of interviews:** Computer Assisted Personal Interviewing, CAPI. **Fieldwork:** 331 interviewers worked on the field in the first quarter year.	is made of the lists of households for the selected enumeration areas. The sample was reduced by 1per cent in relation to the basic set. Rotation scheme was prepared for the purpose of quarterly survey and then modified for biannual survey. The rotation scheme established this way allows for possible rotation of households in case of transferring to quarterly survey. For biannual survey, six different rotation groups are taken, three of which are repeated in relation to the previous six-month period, i.e. 50per cent of the households appearing in the sample in April are interviewed again in October. **Target population:** all persons aged 15 and above, living at the national territory; persons from institutional households, diplomatic and consular personal were excluded. Basic set includes all households and persons who work or reside on the territory of the Republic of Serbia for one year and more. Excluded are persons who live abroad for more than one year and persons in institutional households (students' homes, homes for children and young people with developmental disability, homes for socially imperilled children, old / retired people homes, homes for adults with disability, monasteries, nunneries, etc.). **Method of interviews:** There are two interview methods: on the field with the questionnaire in the paper form are interviewed those households who are for the first time in the sample and households which were interviewed before but did not give phone number. By phone are interviewed households which were interviewed before at the field. **Fieldwork:** 170 interviewers worked on the field in the whole territory of Serbia, and there were about 50 supervisors from 15 regional districts. Interviewers and supervisors were well trained with LFS methodology and most of them had been engaged in LFS survey for several years. The core LFS questionnaire contains 136 questions.

3.8. Preparing the SEEMIG fieldwork

Tight deadlines in Hungary together with the starting date of the fieldwork being shortly after the Christmas-New Year holiday period (7 January) restricted the possibilities of an extended training session to be held for the participating colleagues. We consider that in an ideal case a full-day-long session would be necessary also for trained and experienced interviewers to transfer and to practice the extra skills needed for the SEEMIG survey. Although the SEEMIG battery represented a technically complex set of questions with several skips and repetitions of the same blocks etc., the survey software could efficiently guide the interviewers through these difficulties. The real challenge in this case lay in gaining the trust of the respondent not only to provide valid information in a sensitive topic but also to help us to get in touch with further persons. To successfully complete these tasks, the interviewers needed extra communication and other personal skills – part of which could possibly have been provided them on a well-focused training.

Unfortunately in our case a centralized training session held by the leaders of the study could only be organised for (A) the regional managers responsible for the interviewers' work in their region (B) the interviewers in the capital city Budapest and the central region. The sessions lasted one hour and a half and consisted mainly of frontal presentations on the study. Regional managers then either held a special session for their interviewers or provided them instructions on a one by one base.

As the SEEMIG survey was administered together with the LFS, the usual LFS personnel were employed. They all had already had their usual LFS training and the majority of them were very experienced in administering standard population surveys. However, it was acknowledged that SEEMIG included unusual elements and needed special preparation. Although it was obvious that it was not possible to make up for a sufficient participatory training it was necessary to introduce some other forms of preparation. Therefore:

- Detailed Manuals were prepared to help individual preparation for the work.
- Interviewers were instructed to fill in two SEEMIG questionnaires with very specific instructions provided. Two imaginary households with complicated links to migrant persons were described in these, and the questionnaires had to correctly represent the situations described.
- A test with ten questions regarding possible difficult situations during the fieldwork as well as the mechanisms of identifying migrant acquaintances had to be filled in by each interviewer before starting their work.

Interviewers were obliged to present their test-questionnaires as well as the filled in tests before starting their fieldwork. Regional managers provided extra assistance when the tasks presented difficulties. On the other hand, an additional sum was paid (above the interviewers' fee) when over 70 per cent of the questions were found to be correctly answered.

Two weeks after the start of the fieldwork, completed questionnaires were controlled in the central office. This control did not detect any problems that could directly be linked to the interviewers' work. Obviously, the electronic questionnaire made it possible to avoid skip logic or other technical type mistakes. We had no tool, however, to control the communications between the respondents and the interviewers. Therefore, we were not in the position to tell how well the interviewers did in convincing the participants about the importance and the safety of their cooperation with us.

Training of the local managers and interviewers in Serbia

Several trainings were organized in March 2013 before beginning the survey. First, training for the executives and statisticians from the Serbian statistical regional districts (15 regional districts, one executive and statistician from each district) was held in Belgrade (19 March 2013) where all details of the questionnaire were explained. Later on, executives from each regional district held trainings for the interviewers from their regional offices. Interviewers got appropriate knowledge for the fieldwork. Possible ways to approach the household members were also explained as well as the goal of the interview. There were about 170 interviewers in total and all interviewers had previous experience in LFS and other surveys. At the trainings they were introduced with a questionnaire paper form and all methodological explanations regarding the SEEMIG emigration battery and they received all necessary material for the fieldwork: blank questionnaires for their census district, list of the households and all necessary office material. Training practice was based on complicated examples for the interview-situations in order to present and practice all possible situations regarding different group of migrants. During the training, interviewers were invited to fill in the SEEMIG emigration battery in pairs, and they had plenty of the questions based on which various difficulties were highlighted and solved. Although filling the questionnaire was not easy either for the supervisors or for the interviewers, later on it was very helpful during the fieldwork. Participation at the training also encouraged fieldwork interviewers to contact us every time they had any kind of dilemma.

The SEEMIG fieldwork in Serbia

About half of the sample size (around 5 000 households) were interviewed at the field, face to face (new households of the sample and households that did not provide their phone numbers), while the other half (around 5 000 households) were interviewed by phone with paper form questionnaire. In each regional district, statisticians from the regional statistical offices supervised and controlled the work of the interviewers and gave them instructions and advice when it was necessary. In the regional offices (Belgrade, Zajecar and Smederevo) additional supervisors were engaged due to the higher number of migrants from these regions. Additional personnel was engaged to control the accuracy of the completed paper questionnaires and some of them also participated in data entering via computer BLAISE application.

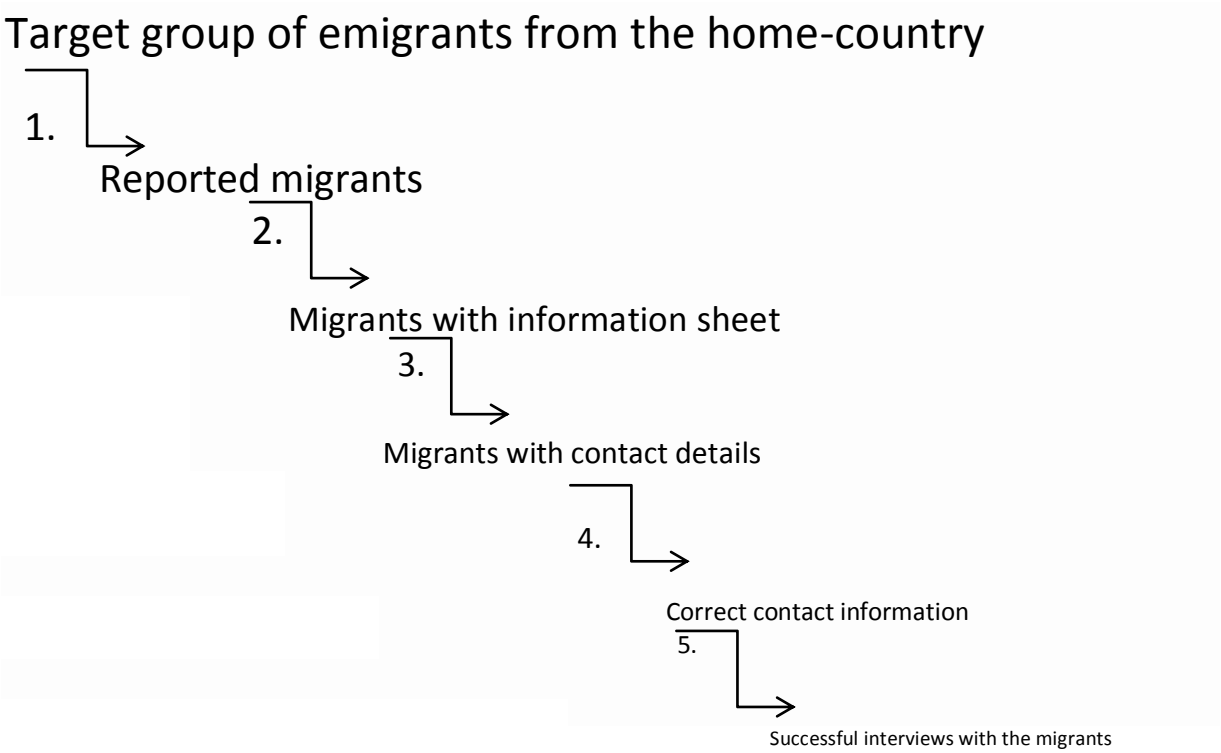
3.9. Enhancing response rates

As it was also clear from previous studies, a key challenge for the SEEMIG pilot was to achieve a sufficient migrant sample size and to keep the sample representative of emigrants from the country. A high level of attrition rate is not only problematic because it reduces the sample size but also because non-response is likely to be unevenly distributed across the various segments of the target population. If this is the case then representativeness of the sample will be jeopardized.

For example it is very likely that household members will be less willing to report on illegal migrants than on legal ones, which in itself is a source of distortion to the sample. At the same time it is possible that the above-average non-response in these groups will also lead to the underrepresentation of emigrants to certain countries and / or of emigrants with certain qualifications in the sample. For these reasons it was crucial to keep the attrition rates at a minimum level.

Figure 3 shows the main threats to achieving a large and undistorted sample in the SEEMIG study.

Figure 3 Chain of attrition in the SEEMIG pilot study



As it is represented by the decreasing font-sizes on the figure, on each step on this 'chain' we can expect that some part of our target group will be lost for our sample – how big and this part is, and what characteristics it has we cannot always tell.

From the originally targeted group of migrants we can expect that (1) not each person will be declared in the surveyed households – due to non-cooperativeness or maybe to the ignorance of the responding member of household. It is crucial to try to mitigate this type of non-response during the fieldwork, especially because there will be no information available neither on the extent nor on the nature of this loss. Besides providing sufficient training for the interviewers to get the respondents cooperate with them, we had also tried to define our target group so that LFS household members should have the necessary information on their whereabouts.

Even if a migrant person gets reported it is still possible that further information about this person will be denied by the respondent. (Step 2) In such cases, the mere information on the emigrant person's existence will help us to improve our estimate on the total number of emigrants and household level data collected with the LFS battery will also be available for household-level analysis, but no individual data will be produced.

In step 3 then we can expect that not all respondents who are willing to declare their migrant acquaintance and also to give some basic statistical data will at the same time be willing to provide a contact detail. It is also possible that the respondent does not have the necessary information we need. As it was described before, asking for a contact detail is also a sensitive step in our study from data-protection points of view. Being attentive to data

protection and implementing the related measures described above is expected to gain the confidence of our respondents at this stage of the study, too. Also, offering them to provide the contact details at a later time was expected to help them when lack of information was a problem to them. All in all it was envisaged that a substantial loss in sample size is likely to occur at this stage of the project which can be a major obstacle to carry out a successful survey among the migrants themselves. At the same time, however, it is important that selectivity at this stage will be possible to measure from the resulting data by comparing the declared migrant population to those to whom a contact detail is provided.

As opposed to the above mentioned one, sample-loss in step 5 is not possible to estimate. Although it is very likely that a segment of the contact information we gather will not be accurate (out-dated, incorrectly reported / coded, incorrectly reported by purpose, etc.), it will not always be possible to tell why a migrant is not responding to an e-mail or a telephone call. Incorrect contact details will not always be possible to separate from other forms of migrant non-response. To minimize the number of incorrect contact details, interviewers were trained to pay much attention to correct coding and some automatic control was also built into the interview-software.

Finally, it is also expected that not all the technically correct contact information will lead to a successful migrant-interview. To increase response rates among the migrants themselves is, however, a problem to be solved for the second stage of the SEEMIG pilot study.

As can be seen, maximizing the LFS respondents' trust and cooperation during the SEEMIG interview is crucial from the point of view of the SEEMIG study's success. In Hungary, respondents were given some small gifts: a SEEMIG project newsletter in Hungarian as well a textile bag with a SEEMIG logo. Interviewers were instructed to hand over these items to each respondent who had declared a migrant acquaintance preferably before starting to ask them about the contact details of this persons. However, interviewers were free to provide the gift at a different time-point if they found that more apt to the situation.

Besides trying to motivate respondents with these small gifts, it was also evident that incentives for the interviewers can also play an important role in boosting the number of successful interviews. The following bonus-scheme was therefore created in Hungary. Above the usual fees paid for the LFS interview, further bonuses were offered:

- for successfully administering questions Nr 1 to 10, for each person covered: 220 HUF
- for each migrant person identified and successfully recorded in the migrant-information sheet: 240 HUF
- for each contact information collected: 1000 HUF
- if more than one contact information was collected to the same person: 200HUF extra
- if a SEEMIG respondent card was left in the household and the card gets used to access to the web-based registration page for the online questionnaire: 1000 HUF

The scheme was intentionally designed so that it rewards successful contact-detail collection to a disproportional extent. As these fees represent relatively high amounts as compared to the usual fees paid to the interviewers, it was expected that they play a significant role in motivating the interviewers to make a great effort in order to get the respondents cooperate with us.

Interviewers' bonus-scheme in Serbia

Interviewers were paid for each five parts of the completed questionnaire: for the first four parts they got equal sum (approximately 2 EUR), and for the completed fifth part with contact details they got double sum so they have additional incentive. Compared to to the standard LFS survey, the completed SEEMIG emigration battery was relatively well-paid. No gift was provided for the respondents.

Information letter sent before the interview in Serbia

Sending an information letter to each LFS household before conducting the survey is a regular practice in Serbia, so the same procedure was followed before conduction LFS with the SEEMIG emigration battery survey. This way households are informed about the forthcoming survey: by whom it is organized and what it's the purpose is, the timing of survey (when they can expect an interviewer) and a statement regarding data security and confidentiality. The information letter had a huge impact on success/response-rated of the SEEMIG pilot study – household members were more willing to provide the requested information.

Media campaign to promote SEEMIG in Serbia

The Serbian LFS-SEEMIG Team decided to present the SEEMIG project at a national press briefing on 26 December 2012. The idea was to make an attempt to reach out to the local population through local media. The SORS SEEMIG Team prepared a promotion material with a short introduction of the SEEMIG project and pilot migrant survey and the campaign started one week ahead of the fieldwork on local and national media (TV and radio), newspapers of the Regional Offices. The press release and the list of media appearances can be found in Annex III.

4. Results and main lessons learned

4.1. Sample attrition

Table 1 presents the key measures from the SEEMIG pilot study in Hungary and in Serbia. The figures show that response rates in the LFS were around three quarters in both countries. Household-level denial to the SEEMIG battery was non-existent in Serbia and was very low in Hungary (1 per cent).

From all the households where a SEEMIG battery was administered, at least one migrant was identified in 10 per cent of the cases in Serbia and in 8 per cent in Hungary. This ratio resulted 816 households with at least one reported migrant in Serbia and 1785 in Hungary. The number of migrants reported was above these numbers in both countries since in many cases there were more than one migrant persons reported in one household. This way altogether we had 1079 migrant persons reported in Serbia and 2401 in Hungary – coming from the three migrant groups we identified[17]. (We will refer to these groups as migrants reported / identified.)

After reporting the existence of an abroad-living sibling or household member, quite a high proportion of respondents in the LFS sample decided not to provide any further information about them. The attrition rate at this stage of the survey was 25 per cent in Serbia and 40 per cent in Hungary. These figures, however, are not possible to compare because of differences between the techniques applied in the two countries and also differences between the techniques applied in the case of the various migrant groups.

In Hungary for example, a specific method was used for identifying migrant siblings. In question Nr 10 respondents reported about the number of their (live) siblings living in Hungary. From this figure the number of their migrant siblings was automatically calculated by the interview software. In other words, respondents in Hungary did not explicitly report about their migrant siblings, therefore, they had no opportunity to conceal this information. From the disproportionally high attrition rate at this stage of the study in Hungary (from 1383 migrant siblings identified this way, further information was provided about 759, representing only 55 per cent) we can assume that more migrants would have been identified if we had used a similar, non-direct way of getting information about migrant (former) household members as well. In other words, it is suspected, that a significant level of non-cooperation of the respondents remained hidden when they simply chose not to report their migrant relatives in our survey. (See step 1 type sample-loss on Figure 3.)

Nevertheless, rich data was provided by their home-staying household and family members about an unusually large sample of emigrants both in Serbia (819) and in Hungary (1430). In this group we have information on their gender, age, time of emigration, destination country, etc. available for further analyses.

Unfortunately – but not quite unexpectedly – the most significant attrition appeared in the last step of the study, i.e. when contact information to the migrants was requested.

[17] It is important to note that the aggregate of the numbers of migrants in the three individual categories (household members / former household members / siblings) exceed 2041 in Hungary, because one person could be recorded in more than one category – e.g. both as a former household member and a sibling of one of the household members.

Interestingly, attrition rates in Serbia and Hungary were very similar at this stage. Compared to the number of migrants about whom the respondents provided detailed statistical data, a contact detail was also given in 36per cent of the cases in Serbia and in 38per cent in Hungary.

Finally, results clearly justify the decision to include sibling migrants in the sample, since this has proved to be a significant group increasing the size of the final migrant sample to a notable extent.

Table 1 - Response rates and number of migrants recorded in the SEEMIG study in Hungary and in Serbia

	Serbia	Hungary
Households (HH) in the LFS sample	10 294	35 835
Successful LFS HH interviews	7 986	26 936
Successful LFS HH interviews %	78%	75%
From this: part of the SEEMIG sample	7 986	23 749
Households successfully interviewed – SEEMIG	7 986	23 393
HHs successfully interviewed %	100%	99%
number of HHs reporting migrants	816	1 785
migrant current HH member – identified	75	439
migrant former HH member – identified	657	579
migrant - sibling – identified	358	1 383
migrants total – identified	1 090	2 401
migrants total – details provided	819	1 430
migrants total - details provided %	75%	60%
migrants total - contact provided	298	546
migrants total - contact provided %	27%	23%
contact provided in relation to info provided	36%	38%

4.2. Evaluation of the different methods of collecting contact details

Although an emigrant information sheet was completed in only 1430 cases in Hungary, an attempt was made to get contact details to 1531 migrants. This is because we decided to make such an attempt also when a migrant was declared but statistical information were denied (migrant information sheet not filled in) whenever it seemed possible in the interview situation. From these attempts, one or more pieces of contact information (typically e-mail addresses and / or telephone numbers) were successfully received in 561 cases, that is, 37 per cent.

The vast majority of the contact information was provided by the respondent during the interview, without the specific approval of the migrant (446 cases). Another 37 pieces of contact information was given by the respondent after a successful call to the migrant. In these cases the migrant person readily gave his/her permission to this act. In another 50 cases a second visit or an additional call to the LFS respondent was needed for getting the necessary information – proving the usefulness of this flexible approach in our fieldwork.

Finally, in 10 cases the migrant could be contacted via the same channels as another migrant person linked to the same household.

Out of the 969 cases when we could not get a contact detail, the most typical case was that the requested information was denied immediately by the respondent (790 cases). In these cases the LFS respondents did not even make an attempt to contact their migrant acquaintances. An immediate telephone-call to the migrant resulted a refusal in 20 cases: in 16 cases the migrant explicitly gave no permission to the respondent, while in 4 cases the respondent did not manage to get in touch with his (her) migrant acquaintance and decided to refuse cooperation thereafter. In quite a large number of cases (159) the respondent made no straight refusal during the interview but asked for a second visit (phone call) from the interviewer and by this second contact he (she) decided not to give any information. We cannot tell whether these refusals were preceded by a discussion with the migrant or not.

Table 2 - Results of the different methods of collecting contact information during the SEEMIG study

Migrants to whom we attempted to get contact information	**1531**
From this: a successful attempt	**561**[18]
household gave contact detail without asking the migrant	446
the migrant's contact detail is the same as another migrant's	10
contact detail provided on the spot after receiving permission from the migrant on the phone	37
contact detail provided at a later interviewer visit or telephone call	58
migrant got in touch using the SEEMIG Research Participant card	10
Unsuccessful attempt	**969**
contact detail denied on the sport without asking for the permission of the migrant	790
contact detail denied on the spot after an unsuccessful attempt to contact the migrant	**4**
the migrant refused on the phone	16
contact detail denied at a later visit / call by the interviewer	159

4.3. Evaluation of the questionnaire items

When collecting statistical information about third parties (i.e. others than the respondents themselves), especially when the subject of the questions is geographically far away and / or he (she) is not very closely related to the respondent, the reliability of the answers given needs further investigation. A full test on this can only be carried out when (after the second stage of the study) information on the same subject will also be available from the migrants themselves. At this stage, however, we can already investigate whether respondents in the

[18] In 15 cases a contact detail was given to the migrant without a completed information sheet.

LFS sample could at all provide any answers to the questions in the SEEMIG battery. In other words, we can look at whether the research design was realistic in the sense that respondents were in the position to provide information about the circles of migrants we had defined. The following table shows the ratio of 'do not know' answers to each question in the SEEMIG emigrant data sheets for each group of migrants separately.

As it can be seen, the ratios of such answers are typically not very high: they rarely exceed 4 per cent. Not surprisingly, the migrants' intentions regarding the length of their staying abroad was the most difficult question to answer (with 24 to 27 per cent of do not know answers). On the other hand, key "basic information' such as country of stay, family status, employment status, length of stay, year and place of birth, citizenship seemingly caused no particular difficulties for respondents to provide. Even information on more complicated issues relating to the special situation of the migrant (financial links to Hungarian households, visits to Hungary) were duly provided in the majority of cases.

The amount of information household members in the originating have on their emigrated (former) household members does not seem to be affected by the amount of time this person has spent abroad. A valid answer was given by the respondent in a similar ratio of cases no matter whether the emigrant person belonged to the group of the present or the former household—members. It is different, however, in the case of sibling migrants: relating this group the ratio of non-responses varies between 0 and 25 per cent, depending on the question asked. Although some of these ratios are not negligible, they are still moderate and do no question the value of collecting important information on migrants in such an indirect way. Of course (as it was said before) a full evaluation of the validity of these answers can only be tested after data from the migrants gets collected in the second stage of the research.

Table 3 - Ratio of 'do not know' answers to the different questionnaire items relating to the different groups of migrants

	CURRENT HH MEMBER		FORMER HH MEMBER		SIBLING MIGRANT	
	% 'do not know'	N	% 'do not know'	N	% 'do not know'	N
Country of current residence	0	439	0	579	0,5%	1383
Intentions to stay abroad	24%	430	-	-	27%	539
Financial support provided to household in Hungary?	0,7%	430	0,43%	461	9%	539
Financial support provided by household in Hungary?	0,93%	430	0,2%	461	7%	539
Year of birth	0	439	0	579	0	496
How long has been living abroad?	1,6%	434	4,3%	466	5,6%	759
Employment status	0,9%	439	0,9%	461	4,8%	539
Marital status	0	439	0,7%	461	3,7%	539
Relation to the head of the household	1,13%	439	0	460		-
Highest level of education	0	439	0,9%	461	7,9%	539
Country of birth	0	439	0	461	0,18%	539
Citizenship	0	439	1,5%	461	7,8%	539
Number of visits to Hungary	12,9%	430	3,7%	461	9%	539
Time spent in Hungary during the visits	0	256	11%	227	25%	258

4.4. Estimation of the number of emigrants from the SEEMIG data

If successfully conducted, the SEEMIG pilot should have the potential to provide a reliable estimate on the number of emigrants from the country of origin where the survey is carried out. In this chapter the Hungarian pilot study will be evaluated from this aspect. In Annex IV a detailed description is given on the weighting process applied in the Hungarian (and the Serbian) pilot study. In the followings, estimates based on these weighting schemes in Hungary will be introduced and discussed.

4.4.1. Estimating the number of emigrants from the SEEMIG data - Hungary

Based on the weights introduced in Annex IV, we can calculate that the number of emigrants from Hungary was around 195 500[19]. The estimation is based on the total number of migrants reported in the SEEMIG study, i.e. on 2 401 cases. According to the definitions applied in this study, this figure reflects the number of migrants who

- are living abroad, or spending most of their time (resting time included) abroad either for employment or other purposes;
- are aged between 15 and 74;
- were not born in the country where they currently live (i.e. people who returned to their country of birth or who – in the case of siblings – may not even have lived in Hungary are excluded);
- are considered as (former) members of a Hungarian household or have at least one live sibling aged between 15 and 74 living in Hungary.

Although we have no fully reliable reference point to evaluate this figure, we still have reasons to believe that it significantly underestimates the number of Hungarian emigrants. Partially comparable data available in this respect are (1) data from the 2011 Census, (2) data based on mirror statistics and (3) estimates from an earlier survey carried out by the Demographic Research Institute (DRI). Although all these measures refer to somewhat different target populations, the SEEMIG estimate falls short of either of these.

Table 4 - Comparing estimates on emigration from Hungary

Definition	Data source	Figure
(1a) Hungarian citizens abroad for more than a year on 1 October 2011	Census 2011	143 000
(1b) Hungarian citizens abroad for less than a year on 1 October 2011[20]	Census 2011	70 059
(2) Hungarian citizens living in EEA countries in 2012.	Eurostat (2013), supplemented by data from Statistik Austria (Austria) and Annual Population Survey (2011) (Gödri)	239 000
(3) Hungarian citizens abroad with permanent residency in Hungary – age group 18-49	DRI 2013	335 000
Hungarian citizens and Hungarian born-population abroad, age group 15-74	**SEEMIG 2013**	**195 000**

[19] This figure is based on the total number of migrants identified – those who live in the country (outside Hungary) where they were born excluded. However, the information on the place of birth is only available on those 1 659 migrants who the respondents did not deny to give further data about. Consequently, in 742 cases we cannot tell, whether the person should have been excluded for the reason explained above. Assuming that the ratio of those living in their country of birth is similar in this group of 742 as it is in the larger group, we would exclude a further 102 persons from our sample. In this case the estimate on the number of emigrants will be reduced to 187 222.

[20] KSH 2013

In the case of the most recent population census, the figure to be compared to SEEMIG data is 143 000 plus 70 059, that is 200 059. Although the actual value is not unlike the SEEMIG estimate, it is not reassuring, since the Census is also expected to underestimate the number of emigrants. This is because it only partially included those who emigrated together with all of their household-members and their residence in Hungary is empty; and fully excluded those whose Hungarian property is either rented out, or sold to new owners. Since in the SEEMIG study we expected to reach entire emigrant households in our sibling-subsample irrespective of the current state of their property in Hungary, and therefore members of entire emigrant households are not excluded from the SEEMIG sample, it is not clear why SEEMIG has not resulted in a figure above the Census data.

The incompletion is also evident if we take the mirror-statistics as a reference point. Since in this figure of 239 000 only Hungarian emigrants in the EGT countries are included, we would again expect the SEEMIG figure to exceed this one.

Finally, we get the most striking difference if we take a recent estimate produced at the Demographic Research Institute. In this case, the number of emigrant Hungarian citizens with permanent official residency in Hungary was calculated on the basis of a representative survey. Although the estimate is restricted to the age group 18 -49 only, it has produced a figure far above our estimates from the SEEMIG study. Since we have no reason to believe that the value of 335 000 would overestimate the actual size of the population targeted, moreover, it refers to a target population from a narrower age-group than SEEMIG does, it again suggests that SEEMIG provides an underestimation of the number of emigrants.

Notwithstanding the differences in stock data explained above, it also has to be noted that emigrant flow as estimated from the SEEMIG data for the past few years exceeds the current, official emigrant flow estimates of the Hungarian Central Statistical Office (HCSO) to a notable extent. At the same time, however, SEEMIG flow estimates are still significantly lower than estimates derived from the mirror-statistics, despite the fact that 'mirror statistics' estimate includes only those emigrating to EU countries. On the other hand, mirror statistics also include migrants who have possibly returned since their departure, whereas in SEEMIG we only have data on those still living abroad. Looking at the trends between 2010 and 2011 it is also evident that SEEMIG data reflects a similar (although slower) trend of increase in the period as do data from the other data sources.

Table 5 - Estimates for yearly emigration flow, Hungary

Definition	Data source	Figure
Hungarian citizens immigrating into European countries - 2010	'Mirror' statistics	43 005
Hungarian citizens emigrating from Hungary – 2010	Social Security data, HCSO	7 318
Hungarian citizens emigrating from Hungary - 2010	**SEEMIG**	**16 245**
Hungarian citizens immigrating into European countries – 2011	"Mirror' statistics	58 861
Hungarian citizens emigrating from Hungary – 2011	Social Security data, HCSO	12 413
Hungarian citizens emigrating from Hungary - 2011	**SEEMIG**	**19 673**

As it was said before, we expect SEEMIG to provide more valid data on recent emigration – therefore, it is very likely that information on migrants from the last few years will prove to

be useful even for estimation purposes. This would be especially welcome since shortcomings of the currently used HCSO method (data derived from social insurance statistics) have for long been evident, but so far no reliable alternative has become available. Now the SEEMIG method will also be evaluated from the point of its potential to replace current HCSO methods to provide such data.

Evaluating the SEEMIG estimate on the number of emigrants from Serbia

In Serbia the only reference point for evaluating the figures derived from the SEEMIG study is respective data from the latest Census in 2011. In the Census emigrants from Serbia were defined in the following way: ''Serbian citizens living abroad for a one year or longer and those who are abroad for less than one year but intend to stay longer than 1 year''. From the methodology of the census data-collection it follows that full emigrated households are partially included. Acknowledging this shortcoming of the Census we can conclude that on the basis of the latest Census in Serbia the number of emigrants was 285 116 (Persons aged 15 and older). The Serbian estimate derived from the SEEMIG study is significantly higher than this figure.

In Serbia weighting method different from the Hungarian one was applied. Weighting scheme in Serbia SEEMIG study coresponded to the estimates of two-stage stratified sampling design (PPSWR at first stage and SRSWOR at second stage) within six rotation groups as a representative sub-samples. Initial weight for each household was a product of inverse inclusion probability at the each stage with correction of household non-response. Thus, initial weight was LFS final cross-sectional weight for household. Then data was corrected for non-response at the household level for households that have migrants and refused to give any data about migrants. We can estimate that the total number of emigrants from Serbia was 386.884 according to the SEEMIG study in first half of 2013. Having in mind that SEEMIG encompassed some households (''sibling channel'') which are not covered by Census, we believe that the results of the total number of emigrants received by SEEMIG study are better. However, given the sample size, it is difficult to give a realistic assessment of emigrants.

On the basis of these considerations we conclude that the SEEMIG study has provided a fairly good estimate for the total number of emigrants from Serbia. The difference calculated between the two estimations together with our understanding that neither of the two approached can sufficiently cover household-migration we evaluate that SEEMIG has produced a remarkable result to replace Census-estimate in the intervening years between consecutive censuses for a reasonable cost. This is so despite that further breakdowns of the sample (eg. distribution by age, education etc.) indicate more marked, statistically significant differences between the estimates calculated from the two different data-collections. With the increasing number of surveys to 4 times a year, it would be increased the coverage and quality of the data of emigrants at a higher level.

Also, in the Serbian case we can see less evidence that non-cooperation of the respondents have affected the number of migrants identified during the LFS study to any significant extent. Interviewers have reported very small number of such experiences and general experiences from SORS surveys suggest that the general public are supportive towards data-collection carried out by the official statistical body of Serbia.

4.4.2. Why does SEEMIG underestimate emigration?

To understand why SEEMIG might underestimate the number of emigrants – especially in Hungary – two possible reasons need to be considered. First we have to look at the design of the sampling and how our target population was constructed. Secondly, we need to investigate the possibility that it is not the restriction embedded in the design, but rather a significant level of 'data-concealment' that leads to the low estimate in our study.

Looking at first on the actual target population of SEEMIG, it can be seen that the following groups are excluded:

- only children and persons with no alive brothers or sisters and without any link to a Hungarian (Serbian) household (that is being neither a current nor a former member of);
- emigrants with all of their (alive) brothers and sisters living abroad and without any link to a Hungarian (Serbian) household;
- emigrants whose (alive) brothers or sisters in Hungary (Serbia) are outside the 15-79 age group and without any link to a Hungarian (Serbian) household;
- emigrants either below 15 or over 74.

Unfortunately, the size of these age groups is not easy to estimate, therefore, it is not easy to compare them to other groups not covered by other data sources. These deficiencies of the sample-design are present both in Hungary and in Serbia.

On the other hand, factors linked to survey-nonresponse and mistrust from the respondents' side have most likely to have a greater affect in Hungary then in Serbia. In Hungary, another obvious reason why SEEMIG might underestimate the level of emigration is that LFS respondents in our survey concealed the fact that they had any household members or siblings who lived abroad. They could easily do so, either by saying 'no' to the related questions or simply not responding to them.

Indeed, a subsequent survey we did among our interviewers in Hungary showed that they had experienced several signs of the respondents' concealing this information. Essentially all of our interviewers reported that they had observed signs of serious mistrust during the fieldwork. In many cases the respondent informally told the interviewer that they did have such a relative but did not want to report it formally in the interview. It also happened that after reporting a migrant person in the first stage of the interview, the respondent changed her mind at a later stage and then she requested the interviewer to remove even the information provided previously. Most often this happened when they reached the section when contact-details were requested. As foreseen, asking for contact information was indeed the most sensitive part of the interview and often it was at this stage that uncomfortable feelings were expressed.

Below is given a list of the most typical respondent-attitudes the interviewers reported after the fieldwork.

- Respondents *do not believe that data is needed for statistical purposes only;*
- The respondents did not understand why the Hungarian Central Statistical Office is interested in these kinds of data.

- *Fear that their migrant relative will suffer from some administrative consequences:*
    - loss of home-country social benefits;
    - illegal work gets found out;
    - they will be forced to come home;
    - they get double-taxed – or taxed when they are avoiding tax-paying
    - etc.
- Some respondents blamed (typically parents of emigrant youngsters) the state for the emigration so they felt that it was not fair from a public institute to collect information about their leaving.

Besides the explicit expressions of distrust, explanations related to the sensitivity of providing information about a third party have also come up – although in a more limited number of cases. Examples include fear that the migrant person would not want to be reported on, or the household member not possessing the information needed.

All these concerns became most explicit when it came to providing contact information. Interviewer-questionnaires have clearly justified our anticipation that asking for contact details would be the most sensitive part of our study and this was so both in Hungary and in Serbia. Of course, the level of attrition at this stage of the study can easily be measured by comparing the number of migrants to whom contact information was provided (546 in Hungary and 298 in Serbia) to the number of migrants reported and also an information sheet provided on (1430 in Hungary and 819 in Serbia). The ratio of cooperating respondents at this stage was very similar in the two countries: 38 per cent and 36% respectively. Alternatively, we can calculate the ratio of contact information in relation to the number of migrants reported (but a data sheet not necessarily completed - 2401). As the estimation of the total number of migrants is based on the number of migrants reported, this is a more relevant calculation, resulting a ratio of 23 per cent in Hungary and a somewhat higher 27% in Serbia. (See Table 1)

It is more difficult to estimate the ratio of information lost because of non-cooperation at the first stage of the interview. In the case of Hungary a reference point to such estimation could possibly be taken from the number of sibling-migrants unintentionally reported in Question 10 compared to the number of sibling-emigrants a data sheet was completed about. One can speculate that the ratio of concealed information in the case of siblings this way was similar to the ratio of concealed information in the case of former household members.

In the case of former household members, an explicit question was asked whether or not a former household member has emigrated. The number of admitted migrants to this question (579) exceeds the number of former household member migrants with a data sheet by ‘only’ 23 per cent, (579 compared to 469) It is significantly lower than the ratio of concealing information in the case of sibling migrants. From this one can suspect that a significant number of respondents concealed the information about their former household members by (a) not responding to the relevant question, (b) explicitly denying the existence of such an acquaintance. This assumption is also supported by the fact that the number of non-responses to Question 19 (230) was unusually high.

An external test to some of our results in SEEMIG will be available from a survey carried out by the independent research institute TÁRKI in Hungary. The small battery of questions was designed to test the SEEMIG questions regarding the number of former household members. Relevant questions from the SEEMIG survey have been repeated in a TÁRKI omnibus survey in September and October 2013 with the aim of testing a possible 'LFS-effect' on the number of migrants successfully identified in the LFS-SEEMIG study. Results from this test are expected to come out by early 2014.

To assess the nature as well as the extent of the sample-bias at the subsequent steps of the data-gathering process, a systematic evaluation of the selection is needed. Some attributes of the data-loss can however be hypothesised on the basis of the relevant literature and understanding the nature of emigration and characteristics of sample surveys. Thus we expect that illegal migrants are underrepresented in our survey – this is quite understandable, especially given the fear of administrative consequences as expressed by the respondents. In contrast to this, recent migrants and also migrants with close links to the home-community are likely to be overrepresented, since they are more likely to be (a) considered as a member of the household and (b) recalled by their relatives in the interview situation. For similar reasons we also expect that emigrants to nearby countries are overrepresented as they are more likely to pay frequent visits to their country of origin. A preliminary overview of our data seems to support these assumptions.

5. Summary and evaluation of the method

In the frame of the SEEMIG project an innovative method was tested to improve knowledge and in-depth understanding of emigration. The SEEMIG study has proved to be successful in providing valuable results methodologically but also in terms of improving our understanding of emigration. It has resulted valuable experiences that can promote further improvements in collecting information about emigration. Furthermore, it has also provided us with a rich set of data on an exceptionally big set of emigrants, even though the representativeness of this data requires further investigation.

A series of more systematic tests of the methodology are already on their way. An external test of the surveying method has been carried out in September-October this year by an independent research institute in Hungary. Findings from this survey are expected to shed light on the reasons of the low number of migrant household members declared in the SEEMIG study. Results from this supplementary study will also be included in a paper to be prepared by Hungarian experts[21]. The paper will focus on possible sources of differences apparent in the various estimations of emigration from Hungary.

Using the only available reference point (estimate from the 2011 Census) in Serbia however we find that the SEEMIG method has produced a higher and presumably more reliable estimate for the number of emigrants. This is most likely due to the fact that SEEMIG was more successful in covering entire migrant households than the census was. Also, general experiences in Serbia suggest that SORS enjoys an especially high level of reputation and trust in the general public and respondents in their surveys are usually happy to cooperate. Moreover, in Serbia the SEEMIG study took place in a period of time when emigration was not so much in the focus of the public communication as it was in Hungary. It has to be

[21] Prepared by Endre Sik, Irén Gödri, Zsuzsa Blaskó.

noted however, that SORS followed a weighting method not so much specifically tailored to the indirect nature of the SEEMIG data-collection method as DRI and the HCSO did. It is not possible to tell, to what extent this could have affected the actual estimate in Serbia.

Although definite conclusions and a full set of recommendations can only be provided after these analyses get implemented and also the second stage of the study is carried out, some preliminary conclusions can be made at this stage already.

First it can be seen that the number of contact details provided and especially the low ratio of reported migrants to whom a contact detail was also given is not sufficiently high to realistically result in a successful survey in the second stage. We attribute this failure to the lack of trust from the respondents and the limitations of the flexibility in the LFS fieldwork both in Hungary and in Serbia. Although LFS was an obvious choice for accommodating the SEEMIG battery, the fact that the survey was carried out by a public institute might also have had a special impact on the readiness to provide information on a sensitive domain. These assumptions suggest that a similar study less closely linked to a formal, state-administered survey might be more successful in collecting contact details than SEEMIG was. Taking the last point further, a local study applying a similar method might again be in a better position to reach out to a representative sample of migrants. Emigrants from a smaller community – e.g. a town or a couple of settlements – could be researched with a similar technique potentially building also on the social embeddedness of the interviewers. A smaller-scale approach with more intense fieldwork (as seen in the Nepal study) without giving up the ambition of producing a representative migrant sample is likely to succeed. SEEMIG-findings can support such a research design with the outcomes of our various data-gathering techniques as described earlier. We have shown for example that the opportunity to contact the migrant relative on the spot as well as to make a similar call later have both proved to be efficient in enhancing the cooperation of the respondents.

Unfortunately, mistrust has also affected the outcomes of the first stage our study – certainly so in Hungary, but presumably also in Serbia although to a much lesser extent. Beside the well-documented, general sensitivity of the domain of emigration, the specific cultural and political context might also have a considerable impact on the success of an emigration survey. Clearly, the increased political and media-attention to the issue of emigration in Hungary has created a very special climate to our research. Although in the beginning it was not clear how this would affect our results, in retrospect we can conclude that it has very likely had an unfavourable impact. This can be said not only on the basis of the low response rates but also on the feedback received from the interviewers.

Although it is hard to provide a reliable estimate on the proportion of those who concealed the existence of a migrant sibling or household member, we can assume that SEEMIG has somewhat underestimated the level of emigration also for such reasons. However, we believe that the level of concealment would decrease if no attempt would be made to collect contact information in the survey. In other words, the already notable results of the SEEMIG approach could be further increased if the ambition to collect contact details would be dropped and LFS would solely be utilised to collect information on emigrants indirectly. Beside limiting the goals of the study to indirect data-collection, the level of cooperation could be further increased by providing more efficient, well-focused trainings to the

interviewers and possibly launching a media campaign to support the survey should also be considered.

From the SEEMIG study we have learned that home-country household members are in a good position to provide valuable information on emigrant acquaintances – siblings of household members included. This latter finding is especially important, since including siblings in the study can at least partially resolve the problem of including members of entire emigrant households in the sample and can also notably increase the number of migrants available through such channels. Further possibilities to increase the number of migrants reached would be to cover children / partner of household members, too.

Further findings from SEEMIG that can be utilised in designing any future emigrant survey aiming at indirect data-collection through a national household-survey also include the followings. Non-explicit questioning as applied in the case of sibling of migrants (i.e. when the number of emigrant siblings was calculated from a question relating to siblings in the country) has proved a useful method to find out about a higher number of migrants. An obvious shortcoming of the Hungarian pilot was the lack of sufficient training provided to the interviewers. A well-targeted trainings session with interactive elements could certainly help the interviewers to overcome the inevitable difficulties during this special sort of fieldwork (as this was also a possible advantage of the Serbian study as compared to the Hungarian one).

From the second stage of the study, further valuable lessons can be expected. Beside learning about the accuracy of the contact information gathered as well as the emigrants' willingness to participate in a survey we also expect to test the method of Respondent Driven Sampling in the field of emigration research. In particular, we will find out whether it is possible to boost our emigrant sample with RDS method. Furthermore, after comparing the actual answers given by the emigrants to the information collected from their home-country household members (siblings), we will be in an exceptional position to test the validity of data home-country respondents can provide about their migrant relatives.

Concerning the lessons we can learn not about methodology but about emigration from the SEEMIG study, unique findings can also be expected. After a systematic evaluation of the selection processes throughout the study, we will be in the position to analyse emigration from Hungary on an exceptionally large sample of emigrants. Moreover, individual level data will be possible to link to information on the sending household, which is again unexceptional in the history of emigration research in this country. As it was said before, estimation derived from SEEMIG is expected to prove particularly valuable on recent emigrants from the country.

When exploring household-level characteristics of sending and non-sending households, we will be able to build on the larger sample of 2 400 identified migrants and offer a unique insight into the process of how migrant-sending households get selected. In this step of the analysis regional patterns and the impact of demographic and social composition of households will be explored.

On the individual level we will then be able to provide valuable individual-level data on the composition of the most recent emigrant groups from Hungary in terms of some key demographic and labour market indicators such as age, gender, educational attainment or employment situation in the country of destination. We also have information available on

their financial linkages to Hungarian households – i.e. some badly needed insight into the field of remittances will also be given. At this stage, our analysis will be built on the sample of around 1 400 migrants about whom detailed information were given by their household-members or siblings in Hungary.

References

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

Beuchemin, C. and Gonzalez-Ferrer A. (2011): Sampling international migrants with origin-based snowballing method: New evidence on biases as limitations. Demographic Research, Vol.25:103-134

Deville J. and Levallee P (2006): Indirect Sampling: The Foundations of the Generalized Weight Share Method. Survey Methodology, December 2006, Vol.32. No.2, pp.165-176, http://www.statcan.gc.ca/ads-annonces/12-001-x/9551-eng.pdf

Gárdos, É. – Gödri, I. (2013): Analysis of existing migratory data production systems and major data sources in eight South-East European countries. Research paper developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

Ghimire, D. J. Williams, N., Thornton A., Young-DeMarco L., Bhandari P.B. (2012): 'Innovation in the Study of International Migrants.' http://paa2013.princeton.edu/papers/130951

Jensen et.al. 2012 A Review of Methods for Estimating Emigration. Report of the Suitland Working Group.
http://www.unece.org/fileadmin/DAM/stats/groups/suitland/SWG_emigration_Jensen_UNECE.pdf

Központi Statisztikai Hivatal 2013: 2011. Évi Népszámlálás. Országos adatok. 2013 Budapest, KSH. https://www.ksh.hu/docs/hun/xftp/idoszaki/nepsz2011/nepsz_orsz_2011.pdf
(Hungarian Central Statistical Office, 2013: Census 2011. National data.)

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

Producing migration data using household surveys. Note by the National Bureau of Statistics, Republic of Moldova. paper presented on the Conference of European Statisticians. Work Session on Migration Statistics. Geneva 16-17 October 2012.
http://www.unece.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/DAM/stats/documents/ece/ces/ge.10/2012/WP_10_MOL_01.pdf&t=1384947060&hash=1eaf15c55c8e7034bda3f5e338f37627c51cd124

Zaba, B. (1987): The Indirect Estimation of Migration: A Critical Review. International Migration Review 21:1 395-445.

Kostova, M. – Yakimova, E. (2013): Analysis of existing migratory data production systems and major data sources in Bulgaria. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.
http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportBulgaria.pdf

Annexes

I. Questionnaires
    a. Questionnaire in Hungary – English translation
    b. Questionnaire in Hungary – original Hungarian version
    c. Questionnaire in Serbia – English translation
    d. Questionnaire in Serbia – original Serbian version
    e. SEEMIG respondent card
II. Documents of Data Protection
    a. Data protection letter in Hungary – English translation
    b. Official Statistics Law of the Republic of Serbia – English translation
III. Media Campaign in Serbia
    a. Press release issued by the SORS – English translation
    b. List of media appearances
IV. Description of the weighting process
V. Preliminary tables from the Hungarian pilot study