Surveying Emigration II.

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

Zsuzsa Blaskó
Demographic Research Institute (DRI)
with
Natalie Jamalia
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 2nd stage of 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), **Edit S.Molnár** (External Expert), **Endre Sik** (External Expert), **Ágnes Hárs** (External Expert), **Gergely Fraller** (weighting and statistical Expert, HCSO), **Zoltán Kmetty and Dávid Simon** (External Experts on RDS methodology), **Adél Rohr** (CAWI Programming, DRI), **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 **Gordana Bjelobrk, Marija Jovanović, Gordana Jordanovski** and **Ivana Lukić**. Further valuable support was given by **Jelena Milojković, Nedjeljko Ćalasan and Dušan Vuković** (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, VesnaLukić, Vladimir Nikitović** and **Goran Penev**.

Suggested citation: Blaskó, Zsuzsa – Jamalia, Natalie (2014): *Surveying Emigration II. Report on the second 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/SEEMIGPilotReport2.pdf

Table of contents

List of acronyms and abbreviations

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	Statistical Office of the Republic of Serbia	Republika Srbija Republički zavod za statistiku
RDS	Respondent Driven Sampling	
CATI	Computer Assisted Telephone Interviewing	
CAWI	Computer Assisted Web Interviewing	

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 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 second stage of this study and records the experiences learned during the research process. Findings from the first stage can be found in 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. However a short overview of the first stage is also given in the present report and conclusions and recommendations presented in the final section also refer to the first stage.

Executive summary

The current report documents the process of and findings from the second stage of the SEEMIG pilot study on emigrants from Hungary and Serbia. In the first stage of the study, the Labour Force Survey (LFS) was utilized in both countries to produce a large, representative sample of emigrants from the respective countries. Besides collecting basic statistical information (such as age, gender, country of destination etc.) about them, an attempt was also made to get contact information (preferably telephone number and / or e-mail address) to these persons and thus to produce a sample of emigrants from the two countries. Targeted groups included current as well as former household members and also siblings of any of the household members living abroad. In terms of age we limited our focus to the age-group between 15 and 74.

Potentially high attrition rates especially at the stage of providing contact information to third persons had been acknowledged as the most challenging part of the survey design that might easily jeopardize the ultimate aim of the pilot study that is to reach out to a sufficiently large and unbiased sample of emigrants. During the research design as well as the field-working process notable attempts were made to maximize cooperativeness of the LFS respondents.

Despite all the efforts made however, respondents' level of cooperation in providing contact details to their emigrant acquaintances remained very low both in Hungary and in Serbia. The majority of the respondents in the LFS households were reluctant to provide any contact information to the emigrant person linked to their household. Compared to the total number of migrants identified during the LFS surveys in the households, any piece of contact information was given only to 27 per cent of the emigrants in Serbia and 23 per cent in Hungary – resulting in at least one piece of contact information to 546 emigrants from Hungary and 298 from Serbia. Based on previous experiences in the relevant literature as well as on our own experience on emigration being a particularly sensitive domain, it could be foreseen that the second stage was not likely to produce successful interviews with a sufficiently large number of emigrants who would also form an unbiased sample of the population they are expected to represent.

Nevertheless we have decided to carry on with the second stage of the study for several reasons. Firstly, it was SEEMIG's intention to test the full research design planned and not to stop at any stage even when a failure of fully achieving the ultimate goal can be envisaged. Indeed, being a pilot research the SEEMIG study had the mission to collect all the lessons – positive and negative ones equally – that an innovative research design can provide. Secondly, we could also foresee that valuable methodological experiences can be expected from the second stage of the survey.

Nevertheless, when faced with the low case-numbers achieved in the first stage of the study it was decided that potential alternative methods of boosting the sample gained from the LFS-SEEMIG study has to be considered and possibly tested. During intense consultations with sampling experts in Hungary[1], Respondent Driven Sampling (RDS) technique was selected – (although it was realised that it could only be carried out without meeting each and every criteria of its application). In RDS, similarly to a classic snowball-sampling, the first respondents' networks are utilized to invite further respondents into the sample. The specific conditions applied in the method ensure that RDS is a chain-referral sampling technique that produces a final sample that is independent from the initial

[1] We are grateful to Dávid Simon and Zoltán Kmetty for their invaluable recommendations on the application of RDS in the SEEMIG project.

respondents. To meet RDS's data requirements, a special set of questions were added to the end of our emigrant-questionnaires, collecting information on the number of emigrant acquaintances of the respondents, some demographic and social characteristics of them and finally, we also asked for contact details to two of the persons mentioned.

Concerning the content of the questionnaire developed, purpose of the second stage of the study was to benefit from the opportunity of contacting the emigrants directly (as opposed to the first stage of the study) and this way to collect more detailed and in-depth information about their migration history as well as their demographic and labour market characteristics than it was possible in the first stage. At the same time we also intended to cross-check data previously gathered in the home-country households.

As the majority of contact details collected were either telephone numbers or e-mail addresses we had to prepare a mixed method survey: a combination of CATI (Computer Assisted Telephone Interviewing) and CAWI (Computer Assisted Web Interviewing) was applied both in Hungary and in Serbia. While in Serbia the Statistical Office has a call centre available within the office, in Hungary a telemarketing company had to be hired to carry out the telephone interviews. To maximize response rates with a minimum cost, in both countries specific strategies were developed how to contact emigrants in the different circumstances (e.g. when both e-mail address and telephone number was available), when to modify originally chosen mode of contact and when to give up getting in touch etc. Fieldwork was carried out during June and July 2013 in Hungary and October and November in Serbia.

Finally, out of the total contact of 561 in Hungary, altogether 125 successful interviews were made: 66 on the web, and 59 via telephone. Corresponding figures in Serbia are: out of 298 persons with a contact information 98 were successfully interviewed – the majority of them (88) via telephone and only 10 had filled out the electronic questionnaire. These add up to a success rate of 22 per cent in Hungary and 33 per cent in Serbia.

As it was said before, in Hungary out of the 561 persons with contact details collected in the first stage of the study a successful interview could be made with 125 (22%). When it came to providing a contact detail in the RDS block to the emigrant acquaintance, only 31 respondents were willing to do so. Altogether they provided a contact detail to 54 further emigrant persons from Hungary. Now taking the response rates in this survey (22%) from the 54 contact details we can expect no more than 12 successful interviews in a potential second RDS round. Again, assuming a response rate similar to what we have experienced in the pilot study, we can expect to collect any contact information to 5 more emigrant persons this way. Following a similar logic for the Serbian case leads to similarly disappointing conclusions. These calculations clearly suggest that RDS would not provide any satisfying solution to the problem of small number of cases and the biased emigrant sample resulted from the first stage of the SEEMIG study.

From the results of the second stage of the SEEMIG study the following conclusions can be drawn. The low number of contact information collected in the first stage, together with the low response rates both in the CATI and especially the CAWI study method led to small sample sizes and potentially highly biased samples in Hungary as well as in Serbia. From this we can conclude that the ultimate aim of the SEEMIG pilot study, that was building and successfully interviewing large, unbiased emigrant samples on the basis of national representative studies (LFS) had failed to

succeed. The reason for this most likely lies in the lack of sufficient trust and confidence in the formal interview-situations to provide contact details to third persons – especially to emigrants, who form a sensitive target population in the first stage of the study, together with potentially bad quality contact details and a similarly low confidence of the emigrants in the second.

The additional effort made by the SEEMIG team to compensate for the low case numbers by Respondent Driven Sampling technique has also failed to deliver the expected results. Again, lack of willingness of the respondents (in this case: the emigrants) to provide contact information to further emigrant persons had made it clear that further attempts to follow the collected links would be purposeless. As an earlier study in Hungary researching labour-migration to Austria has successfully applied RDS technique (Hárs 2009) our failure seems to suggest that it was more the lack of personal contact with the interviewer (CATI and CAWI) together with some limitations in our application of the method than the sensitivity of the research-topic that has doomed the SEEMIG RDS attempt to fail both in Hungary and in Serbia.

Besides demonstrating the inappropriateness of certain techniques to collect information on emigrants, even the second stage of the SEEMIG pilot study has produced valuable data from a couple of aspects. Firstly, with detailed information about 125 emigrants from Hungary and 98 in Serbia invaluable qualitative analyses can be carried out, aiming more at describing different patterns explored than at generalizing to a wider population. Secondly, a comparison of answers provided to the same questions by household members in the home country and their emigrant acquaintance on a one by one basis will make it possible to test the reliability of some data collected in the first stage of the SEEMIG study.

Finally, we suggest that failure of the second stage of the SEEMIG attempt has only demonstrated that it is not realistic to build a large representative sample of emigrants through a *large, highly formalized, national survey*. We recommend that the methodology is still worth being tested on smaller scale, local surveys in localities with high emigration, applying a more flexible approach in field-working and also building on the local knowledge of interviewers. Useful guidance to how such a local survey should be developed can be found in both reports on the SEEMIG pilot study.

Also, the failure of the second stage should not hide the inevitable positive results from the first stage of the SEEMIG study. As explained in the first report, the first stage has not only proved to be successful in providing valuable results methodologically but also in terms of improving our understanding of emigration. It has in fact provided us with a rich set of data on an exceptionally large set of emigrants, even though the representativeness of this data requires further investigation.

Nevertheless, after a systematic evaluation of the selection processes throughout the study, we will be in the position to analyse emigration from Hungary and also from Serbia on a large sample of emigrants. Moreover, individual level data will be possible to be linked to information on the sending households, which is again exceptional in the history of emigration research in these countries. On the individual level we will be able to provide valuable data on the composition of the most recent emigrant groups both from Hungary and from Serbia in terms of some key demographic and labour market indicators such as age, gender, educational attainment or employment situation in the country of destination. SEEMIG has also provided us with information on the financial linkages of emigrants to their home-country households – i.e. some badly needed insight into the field of remittances will also be given.

Finally, our experiences has also proved that although large, formal, nationally representative sample-research are not likely to sufficiently support direct surveys with emigrants, they can still be utilised as a basis for developing reliable estimates for the number of international emigrants from a country. Based on the lessons learnt from the pilot study a research tool can be developed, that can be utilized by members of the research community to produce reliable estimates on the number of emigrants as well as to estimate their distributions by some basic characteristics such as age, gender or country of destination.

1. About this report

This Report is the follow-up of the SEEMIG report on the first stage of the pilot study on emigration (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). The two-phase pilot study was prepared in the framework of the SEEMIG project (funded under the third call of the South-East Europe Program, 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 the first report the first stage of the study was described in detail together with a thorough introduction of experiences and lessons learned in that exercise. In the first stage households in the sending countries were surveyed and information about the emigrants was collected in an indirect way. In the second stage then an attempt was made to approach emigrants directly and conduct a survey among them. In the present report methodology and methodological lessons learned from this second stage will be provided.

The structure of this report is as follows. In Chapter 2 we briefly describe the overall research design then provide a brief summary of our findings from the first stage of the study. This is especially important since results of the first stage already forecast the outcomes of the second stage. Chapter 3 will then provide a detailed description of the methodology followed in the second stage both in Hungary and Serbia. Design and content of the questionnaire will be described, the survey strategies (CATI and CAWI – Computer Assisted Telephone Interviewing and Computer Assisted Web Interviewing) introduced, data protection protocol given. In this chapter we will also explain our motivations for experimenting with Respondent Driven Sampling (RDS) methodology in the SEEMIG study and describe the RDS block in our survey.

Chapter 4 will then introduce the response rates in our surveys – both in the general parts and in the RDS section. Main lessons learned from the exercise will be briefly explained and conclusions regarding the feasibility of an emigrant survey started from the sending country will be drawn.

Chapter 5 is of key importance in the SEEMIG pilot study exercise since it will provide conclusions gained from the two subsequent stages in Hungary and Serbia. The original research design will be evaluated and limitations and potentials of the methodology will be summarized. On the basis of these some, methodological recommendations will also be provided.

Finally, it must be noted that the present report will be followed by further project outputs related to the SEEMIG pilot study. A Working Paper will describe the full pilot study in a condensed way, focusing on the research methodology and on how it fits into the international stream of the

empirical studies on emigration. Besides, further academic appearances are planned based on analyses to be carried out on the results from both phases.

2. Background to the study: the first phase of the SEEMIG pilot study

2.1. The research design

As it was discussed in the first report, shortages of migration statistics are clearly suggesting a strong need for innovative methods not only in surveying emigrants in more in-depth ways but also in merely estimating their flows and stocks. It was acknowledged that survey type data collection can be a fruitful approach to overcome existing data-problems, even though challenges involved in producing a representative sample of the migrant population and also in successfully approaching them are numerous.

Not losing sight of the methodological difficulties, 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. It was expected to 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 two-stage pilot studies were carried out – one in Hungary and one in Serbia.

In designing the research methodology the aim of SEEMIG was to *derive a representative sample of emigrants from a representative 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. In the additional SEEMIG battery attached to the LFS questionnaire not only the emigrant persons linked to these households were identified, but also a set of basic statistical data was collected about them[2]. These covered information on age, gender, education, employment and migration history. The quantity and type of this data enables us to describe some basic characteristics of the migrant population on the basis of the LFS-SEEMIG data collection, that is the first stage of the study. Among emigrants 'linked' to the household we took account of (1) current household members living abroad, (2) former household members who left the country after 1989 living abroad and (3) any brothers or sisters of any household members living abroad. Living 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'.

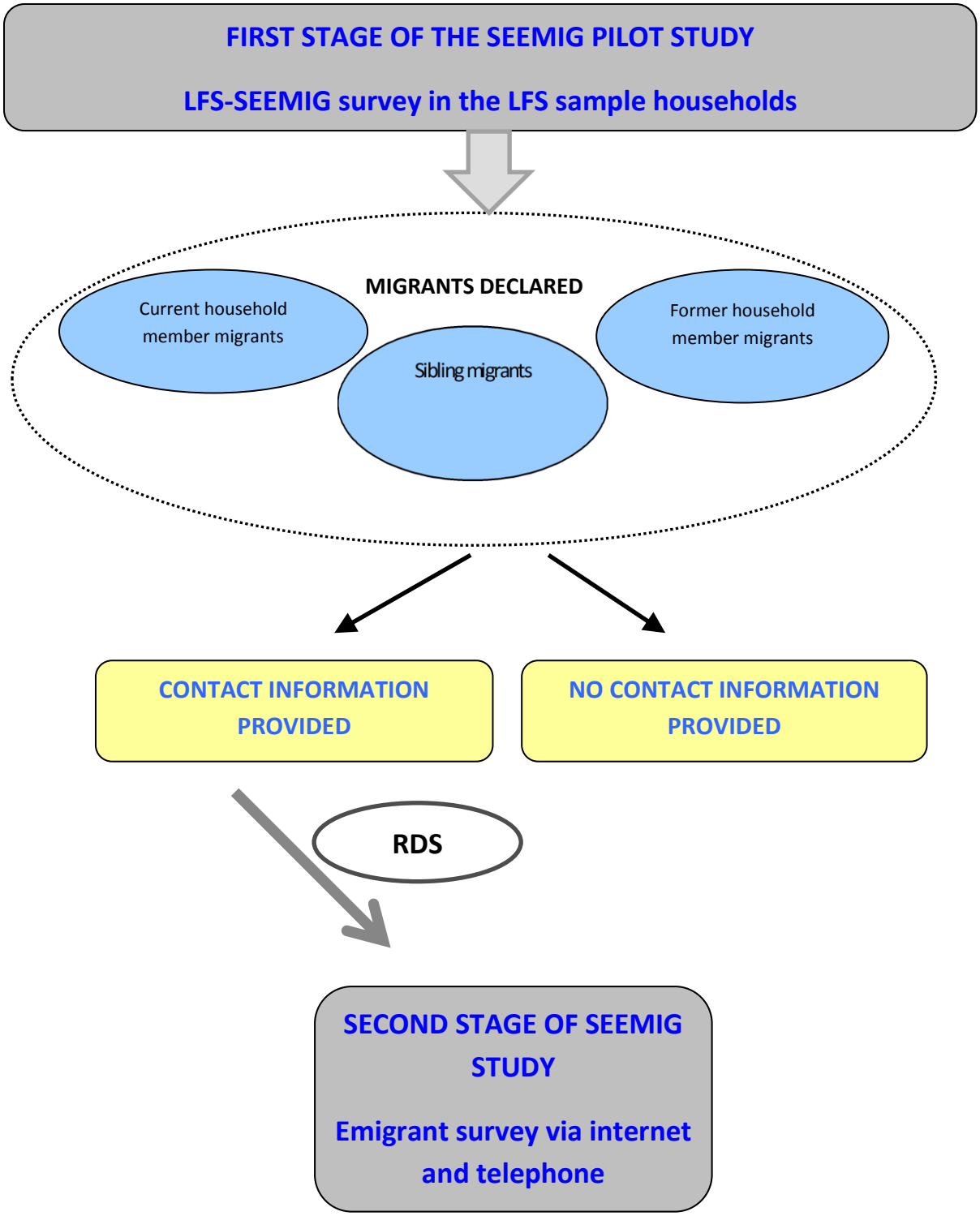
In the very last phase of the LFS interview, respondents in households where an emigrant was recorded were asked to give at least two of the following information concerning the emigrant person: e-mail address; skype contact; mobile phone number; other phone number; date of next visit home (together with a contact information at home). Contact information collected this way in the first stage was meant to form the basis of the second stage of the study when a short questionnaire was planned to be administered through the links collected. The second stage was planned to be

[2] Age; Gender; Highest level of education; Year of emigration; Country of residence; Citizenship; Country of birth; Frequency of home visits; Current employment status; (Assumed) plans for returning the origin country; Occurrence of financial support received from a household in the origin country; Occurrence of financial support provided to a household in the origin country.

multi-method surveying, applying telephone administered interviews (CATI) as well as electronic questionnaires (CAWI). The questionnaire of the second stage was planned to cover a more detailed migration history and also some additional information on demographic and social circumstances. Very importantly, the second questionnaire should also make it possible to test the reliability of information gathered from household members in the sending countries.

The SEEMIG research design is shown on Figure 1.

Figure 1 The SEEMIG research design


FIRST STAGE OF THE SEEMIG PILOT STUDY
LFS-SEEMIG survey in the LFS sample households
MIGRANTS DECLARED
Current household member migrants
Sibling migrants
Former household member migrants
CONTACT INFORMATION PROVIDED
NO CONTACT INFORMATION PROVIDED
RDS
SECOND STAGE OF SEEMIG STUDY
Emigrant survey via internet and telephone

2.2. Results and response rates in the first stage

Successfully gathering contact details to abroad-living persons from our respondents in the LFS-SEEMIG survey was a crucial act and also the most sensitive part of the interview. It was 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 partners involved. 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 rather formalised LFS protocol.

In the first stage of the study a carefully designed process of gathering contact information including the following elements was therefore applied.

Before the fieldwork:

- Although hard deadlines limited the opportunities (especially in Hungary), interviewers were provided with a special training focusing on the SEEMIG-LFS battery. In this training much attention was given to the importance of collecting contact information and also to the challenges embedded in this unusual task;
- Bonus-schemes for the interviewers were designed both in Hungary and in Serbia to provide a significantly increased level of incentive for any contact detail collected;
- In Serbia an information letter was sent to each LFS household before conducting the survey;
- Also in Serbia a media campaign was organised to familiarize the general public with the aims and methods of the survey.

During the fieldwork:

- at the end of each interview in which a migrant person 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;
- a 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 (Hungary only);
- 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 – and ask for their approval;
- 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.
- If at any stage of the interview-process the cooperation was denied by the respondent (but NOT when the migrant person himself denied the cooperation via the telephone) a SEEMIG Research Participant Card was left in the household. The Card included a personal identification code and a link to the project website where an email address was requested from those registering. Household members were then requested to give (or send) this card to their migrant acquaintance (Hungary only).

Despite all the efforts described above, respondents' level of cooperation in providing contact details to their emigrant acquaintances remained very low both in Hungary and in Serbia. Table 1 shows that the majority of the respondents were reluctant to provide us with any contact information. Compared to the total number of migrants identified during the LFS surveys any piece of contact information was given only to 27 per cent of the emigrants in Serbia and 23 per cent in Hungary.

Table 1 - Number of migrants to whom a contact detail was collected in the first stage of the SEEMIG study in Hungary and in Serbia

	Serbia	**Hungary**
Households successfully interviewed – SEEMIG	7 986	23 393
number of households reporting migrants	816	1 785
migrants – identified	1 090	2 401[3]
migrants – statistical details provided	819	1 659
migrants – contact information provided	298	546
migrants total - contact provided %	**27%**	**23%**

It is interesting to note that both the ratio of contacts provided and the ratio of migrants with statistical information given were somewhat lower in Hungary (69%) than in Serbia (75%). Even more importantly, we have good reasons to assume that there was a higher level of concealment of the existence of any emigrant acquaintance in Hungary than in Serbia. We assume this, since comparison to external data sources suggest that SEEMIG has underestimated the number of emigrants in Hungary but not in Serbia (for details see Blaskó and Jamalia 2014 pp 41.). Differences of this kind between the two countries are likely to be due to various forms of non-sampling errors. One of them can be the respondents' attitude towards surveying emigration. Indeed, feedback given after the survey by the participating interviewers has pointed on respondents' mistrust in both countries, but more so in Hungary than in Serbia.

From the aspect of the second stage of the study the final response rates achieved in the first stage were clearly disappointing and already suggesting the failure of the ultimate aim of the research, that is to build a large, representative sample of emigrants both from Hungary and Serbia. Obviously, a starting sample size of 298 (Serbia) and 546 (Hungary) did not seem to be likely to produce a large and unbiased final sample of successfully interviewed migrants. From the relevant literature it is clear that interview techniques available to us – CATI and CAWI – tend to produce rather poor response rates especially in sensitive groups. Even with an optimistic scenario, expecting a response rate of 40%, we would not realistically reach at more than 120 / 220 emigrants in Serbia and Hungary respectively. This means that even in optimistic case we would have failed to produce an emigrant sample large enough for detailed statistical analyses.

[3] In Hungary, respective figures also include those siblings of household members, who live abroad not because they have migrated from Hungary, but because they are Hungarian nationals in a neighbouring country. Without this group in Hungary we have 1430 migrants with relevant statistical data in our sample.

Moreover, several signs were suggesting that the optimistic scenario is not the one we are likely to achieve. The high level of distrust experienced during the first stage of study suggested that emigrants themselves might also be reluctant to provide information about their circumstances and therefore they might produce a below-average response rate. Also, fieldwork experiences in the first stage evoked doubts that the contact details provided by the household-member would always be correct – either for lack of appropriate information or for other reasons.

As we know from a similar study in Nepal (Ghimire et.al. 2012), it is not entirely impossible to achieve very good success rates in contacting emigrants in a social survey. In that study however not only the social context was markedly different from the South-East European one, but also the emigrants were geographically more concentrated and – most importantly – resources (both time and money) were more readily available than in our case. In the SEEMIG project an eminent aim was to develop a methodological best practice that is also financially sustainable. Clearly, very long interviewing-period in the second stage of the study or an intense fieldwork that also includes revisits to the original households when the contact information fails to work were not realistic options in our case.

Besides the high likelihood of achieving a sample too small for appropriate analyses, obviously the fear from the final sample being statistically biased was also present. In the first place we could expect the emigrant sub-population with valid contact information in our sample to be systematically different from the overall sample of emigrants identified. Possible forms of bias at this stage can be closer links to the Hungarian households and a higher likelihood of working legally in the destination country. Similarly, we could expect that emigrants positively responding to our survey would be different from the overall population in several aspects.

Nevertheless we have decided to carry on with the second stage of the study for several reasons. Firstly, it was SEEMIG's intention to test the full research design planned and not to stop at any stage even when a failure of fully achieving the ultimate goals can be envisaged. Indeed, being a pilot research the SEEMIG study had the mission to make available all the lessons – positive and negative ones equally – that testing an innovative research design can offer. Secondly, we could also see that valuable methodological experiences can be expected from the second stage of the survey. As mentioned before, it was planned that wherever possible, information provided by the emigrant person will be cross-checked with the relevant information given about the same person by his or her acquaintances in the sending country. Finally we also acknowledged that information collected even on a small and not fully representative sample can be an object of important qualitative analyses, providing valuable insight into the process of emigration from our countries.

3. The second stage of the survey in Hungary and Serbia

Despite the low number of emigrants with contact information collected in the first stage of the study, the second stage was designed with much effort maximizing potential benefits from the research. As in the first stage, design of the questionnaire as well as the general planning of the study was done in Hungary by members of the SEEMIG team at the DRI and the HCSO together with external experts. Documentations of the design were then translated into English and sent to colleagues at the SORS in Serbia who then adapted the material for their circumstances.

3.1. Definition of migrants, target population

Definition of migrants and consequently the target population was directly following from the design of the first stage of the study. Accordingly, 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 home country and who was not born in the country where he (she) is currently living. To 'live' abroad was defined in the following way: 'spends most of his (her) time abroad – resting time included – either for work or any other purposes.' People on holiday were excluded. We included commuters on a regular but not very frequent base – consequently daily commuters were excluded. We also excluded people who were born in the country where they currently stay. Following the age-restrictions applied by the LFS only migrants aged between 15 and 74 were included in the sample.

Following from the definitions above and from our method of identifying migrants (as described before), members of the target population are:

- Hungarian citizens, who live abroad and are between 15 and 74 years old

AND

- they were either born in Hungary or they live in a country different from their place of birth

AND

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

OR

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

3.2. An attempt to boost the initial sample: Introducing Respondent Driven Sampling

When faced with the low case-numbers achieved in the first stage of the study it was decided that potential alternative methods of boosting the sample should be considered and possibly tested. During intense consultations with sampling experts Dávid Simon and Zoltán Kmetty, the following methods available for sampling rare and hidden populations were considered: Disproportionate Stratified Random Sampling, Random Sampling with Screening, Multiple-Frame Sampling, Space-Time Sampling, Adaptive Cluster Sampling and Respondent Driven Sampling[4]. After a careful review of the statistical prerequisites as well as of the inherent costs of each methods Respondent Driven Sampling was chosen as the most appropriate approach that would also benefit from the qualities of the emigrant samples collected through the LFS-SEEMIG exercise. Obviously, using the sample in the first stage of the pilot study was an asset we wanted to make use of. In RDS sampling (first introduced in Heckathorn 1997.), our contacts from the first stage could be applied as seeds. RDS applies a special mathematical model to compensate for the biases arising from the fact that the starting sample (the set of seeds) was not randomly selected. Similarly to a classic snowball-sampling, the seeds' networks are then utilized to invite further respondents into the sample. The specific conditions applied in the methodology (e.g. the way the referred persons are selected and also certain characteristics of the population studied etc.) ensure that RDS is a chain-referral

[4] For a detailed description of the selection process see Kmetty and Simon 2013a.

sampling technique that produces a final sample that is independent from the initial respondents from which the sampling process begins (Simon 2012).

When considering the applicability of RDS in the SEEMIG case it was established that the population studied and the starting sample meet all the criteria necessary for the application of the method. Doubts were only raised regarding the assumption that members of the target population are all linked to a single component in the network. Based on an earlier research on Hungarian emigrants to Austria applying RDS methodology (Hárs 2009) it was assumed that even this assumption is likely to hold in the target population. Also, fulfilment of this assumption was planned to be tested a posterior after the sampling is completed.

Moreover, it was also acknowledged that the emigrant-sample derived from the first stage of the pilot study has better qualities than RDS would in fact require. Since the SEEMIG emigrant sample is an indirect sample taken from a nationally representative household-sample it can be expected to proportionally represent the target population. This quality of the starting sample cannot only be capitalized during the process of verifying the validity of the prerequisites for the RDS, but might also promote a faster convergence of the RDS sample (Kmetty and Simon 2013a).

Despite these advantages, we could still not fully meet the prescriptions of RDS methodology. Most importantly – due to applying CATI and CAWI – we could not ensure full anonymity for the respondents. Instead of requesting them to directly connect the researchers with their peers, (without giving out their contact details) we asked them to identify their emigrant peers and provide us with contact information so that we could contact them later. In the lack of the necessary means to build up a special infrastructure (either to buy or to develop specialized software) that would enable us to make these connections without handling the contact data, we were aware that a restricted version of RDS was applied.

RDS methodology prescribes the information to be collected from the respondents rather specifically. Following these prescriptions[5], the following questions were inserted into our questionnaire.

- How many friends / relatives / colleagues of Hungarian (Serbian) citizenship do you have who currently lives abroad? Please only consider those with whom you had been in contact with during the last month.
- Please provide some information about each of these persons. (Gender, age, country where they live, type of relationship – type of relation: friends, family members, etc.)
- Please provide us with some contact details (phone number and / or e-mail address) to the *first* and the *last* person on your list above.

As collecting contact information about a third person is a sensitive attempt, this block was inserted to the very end of the questionnaires to avoid losing cooperation at an earlier stage. Also, sensitivity and consequent data protection issues were considered and dealt with (See Chapter 3.7)

Considering the limitations of approach (lack of full anonymity) and also the pilot nature of our study, we decided to take a risk-averse approach with the application of RDS. The plan was to do the first

[5] See e.g. http://www.respondentdrivensampling.org

round of data collection of the SEEMIG pilot second stage with the necessary RDS questions included in the questionnaire and to decide about the continuation of the chain-referral method only after reviewing response rates achieved in this first round.

3.3. Content and structure of the questionnaire

Purpose of the second stage of the study was to benefit from the opportunity of contacting the emigrants directly (as opposed to the first stage of the study) and this way to collect more detailed and in-depth information about their migration history as well as their demographic and labour market characteristics than it was possible in the first stage. At the same time we also intended to cross-check data gathered in the home-country households. The questionnaire was developed in the Hungarian Demographic Research Institute, translated to English and then adapted to the Serbian case by experts at the Statistical Office at the Republic of Serbia.

Considering that similar sample-surveys of the emigrant population are practically without precedent in Hungary (as well as in other countries in the region), a wide range of topics and research-questions had offered themselves as important and relevant to our study[6]. Finally, migrants' motivation, plans for the future and labour force situation abroad were chosen as main focal points of the questionnaire. Besides, we had added a small series of question items (Question nr 147) relating to Developmental Idealism Theory (Thornton et.al. 2012).

The structure of the final questionnaire was the following[7]:

- Circumstances of migration
- Purposes and motivation of migration
- Circumstances abroad
- Education, occupation and employment
- Contact with relatives and friends in Hungary
- Plans for the future
- Developmental idealism
- Respondent Driven Sampling – RDS

3.4. Design of the questionnaire

When designing the questionnaire a range of practical considerations had to be kept in mind. These included the specific requirements of a mixed-mode survey, the necessity of minimizing the burden on the respondents and the absence of the interviewer during the interview.

As we collected primarily e-mail addresses and telephone numbers to the migrants we had to design the questionnaire to CATI and CAWI as well. For CATI and CAWI the questions were grammatically a little bit different but the content was the same in the two methods. As we conducted mixed mode

[6] As the SEEMIG pilot study had originally been planned as a methodological experiment, no specific focus of the study had been defined previously in the project plan either.

[7] For the full questionnaire please consult the Appendix.

survey we had to program the web and the telephone questionnaires as uniform as possible to reduce non-sampling bias.

A further difficulty had arisen from the definitions of migrants applied. As described before, 'living abroad' was defined as 'spending most of the time abroad'. Therefore also frequent (e.g. weekly, monthly – but not daily) commuters were included in the sample. To meet their needs, most of the questions in the questionnaire needed to be phrased in two different ways – although with the same content. Question Nr 10 in the questionnaire served the purpose of dividing the sample in two along these lines in the original questionnaire. Cognitive and also technical testing was carried out via Skype (Hungary), telephone and also in person (Serbia) to test the length of the interview and also other aspects of the questionnaire.

In Hungary the web questionnaire was programmed with Formsite which is an online form builder[8]. As neither HCSO nor DRI has a call-centre available within the institution, a telemarketing company (Telequest) had to be hired to conduct the telephone interviews. The CATI questionnaire programming was made by the company. Before the fieldwork we trained some CATI operators in the company with a presentation and wrote them training material. After the training they immediately started the fieldwork and we had the chance to check out the first interviews.

Serbia

In Serbia, the design of the questionnaire was the same for both kind of interviewing. The Serbian questionnaire slightly differed from the original Hungarian one: a few questions were formulated differently in a way that some answers provided to the respondent were deleted as too sensitive to ask and some news ones were also added (answers on relation with friends, relations with broader family etc).

SORS IT department programmed the web and the telephone questionnaire application with no differences. Respondents were allowed to choose their preferred way of interviewing.

The SEEMIG web questionnaire was developed in VB.NET (programming language) using Microsoft Visual Studio 2010. For the purpose of the web questionnaire an SQL database was developed.

In Serbia a short training for the interviewers was also held to inform them about the importance of the questionnaire and with its contents.

Because the interviewers worked in the SORS building (SORS has its own call center for conducting the telephone surveys), we were able at any moment to follow the telephone interviewing and review fulfilled questionnaires (quantity and quality of the telephone survey).

3.5. Fieldwork and data collection

In the first stage the following types of contact-details were collected:

- E-mail address
- Telephone numbers (cell phone or landline phone numbers)
- Skype name
- The date when the migrant visit Hungary with home address or telephone number

[8] We are grateful for Adél Rohr for the great job she did in this activity.

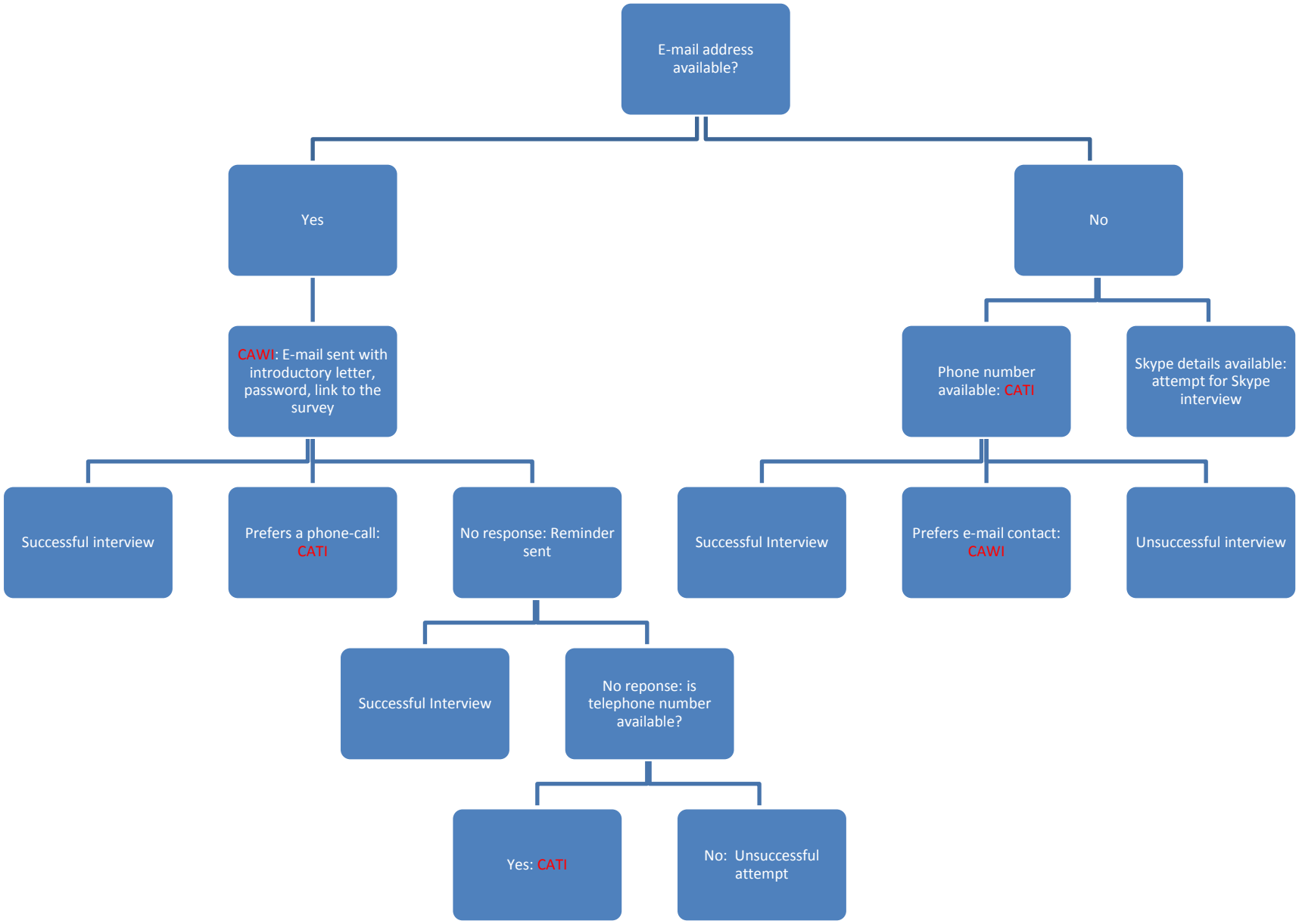
Table 2 - Nature and number of contact details collected during the first stage of the study

	Hungary	Serbia
migrants' identified in the first stage	2401	1090
migrants with contact(s)	561	298
migrants with one contact	390	290
Out of this:		
migrant with only e-mail address	134	31
migrant with only telephone number	245	245
migrant with only Skype	11	14
migrants with two contacts	146	20
migrants with three contacts	8	
Out of this:		
e-mail & telephone	120	3
e-mail & Skype	15	2
telephone & Skype	11	3

Contact details collected required some preliminary cleaning to eliminate or (preferably) to correct obviously faulty pieces of information. Accentuated letters were filtered, obvious typos such as 'co' instead of 'com' were corrected. We revised the telephone numbers to callable forms and finally checked the prefix numbers whether they match the country of residence.

3.6. Fieldwork, data collection

To maximize the number of successful interviews in a cost-effective way in Hungary the following set of contact-making strategies was developed. (Only most typical scenarios are included.)


E-mail address available?
Yes
No
CAWI: E-mail sent with introductory letter, password, link to the survey
Phone number available: CATI
Skype details available: attempt for Skype interview
Successful interview
Prefers a phone-call: CATI
No response: Reminder sent
Successful Interview
Prefers e-mail contact: CAWI
Unsuccessful interview
Successful Interview
No reponse: is telephone number available?
Yes: CATI
No: Unsuccessful attempt

CAWI being the most cost-effective way of surveying, much effort was made to reach as many emigrants through this channel as possible. Therefore whenever an e-mail address was available – no matter whether a phone number was also at our disposal or not – the first trial for contact-making was via web. An e-mail was sent altogether in 277 cases in Hungary.

The introductory e-mail sent to the migrants included a user name and a password enabling respondents to enter the web questionnaire. We also attached a letter to explain the purpose of our study and also from where we got their e-mail addresses. The first e-mails were sent on a Monday 10 June 2013. In cases when no response had arrived until Friday, a reminder e-mail was sent (218 cases). After waiting for another week for their response, the telephone number of the emigrant person (when available) was sent to the telemarketing company. When the e-mail address was all the information we had, the attempt was given up at this stage. The option to provide a telephone number and carry out a telephone interview was also offered in the e-mails but no one took advantage of this.

Telephone numbers to emigrants without a valid e-mail address were given to the telemarketing company in Hungary. The interview-period started on 24 June. Beside the numbers, some further attributes of the emigrants were also provided to the company in order to ease their work (the migrants' name, gender, where they live, etc.). Applying an automatic call device, the company called the phone numbers 5.2 times on average – sometimes as many as 10 or 11 times. When choosing the time of call, time zone of the destination country was taken into account. Timing of the calls were also arranged to cover various time-periods of the day / week hoping to find the most convenient time for the sample-member. If the migrant denied cooperation, the operator offered him/her to answer the questionnaire on the web.

The attempt for reaching the emigrant was eventually given up by the telemarketing company when one of the following situations occurred:

- on average over five unsuccessful calls were made,
- the phone number did not operate,
- a fax or answering machine answered the call on all the attempts made.

In 11 cases we only had Skype name to the migrant. In these cases we sent a message on Skype which was seen when the migrant signed in. In this message we informed them how we got their contact, what we wanted and why. We asked them to respond to the request we could send them the link to the questionnaire, the user name and the password. We also offered that they could respond on Skype.

Serbia

In Serbia the SORS collected primarily telephone numbers (83%), e-mail addresses (12%) and Skype contacts (5%).

To provide better results, we had to hire two interviewers with experience in similar research. The first ten days interviewers worked only in the morning and in the rest of the period they worked in the afternoon, because of the time difference in some countries and considering when people come home from work. To reduce phone costs, we signed a contract with the telecommunications company (Telekom) for two phone lines at a certain time.

To migrants whose contacts were e-mail address we sent one e-mail with the log in data (user name, password and link) and a cover letter explaining the importance and purpose of the survey. Migrants who did not respond to the web survey a week after the first e-mail received another e-mail with the same content but at the weekend, assuming they have more free time then.

In order to get an answer from respondents on the web, as a cheaper way of interviewing, we had two waves of informing sample-members via e-mail addresses. After two unsuccessful attempts we chose to call the respondent. After checking the telephone numbers provided in the first stage of the study, we found 15% non-operating numbers. Fax or answering machines were represented in a few contacts as the reason for unsuccessful establishment of the connection with the respondent. There were no cases when somebody else answered instead of the sample-member. Serbian telephone numbers were represented in a small number of total telephone numbers provided. In order to avoid high costs of calls paid by the respondent (cell phones mostly), these numbers were not used in the survey. We concluded this could not significantly affect the results. In the beginning interviewers only called landline phone numbers. For mobile phone numbers we created the option of sending an SMS which contained a briefer version of the information included in the e-mails. For sending text messages to migrants we used both our application (generating an invitation message with username, password and link for the web questionnaire) and a public android application *Mighty Text* (to send text message from computer, sync'd with Android phone). We expected this way of informing migrants to be successful, but since it was not, the interviewers had to call these numbers in the last ten days. Five per cent of the total collected contacts were Skype name of the migrant. Several attempts were made, but this kind of contact was not established.

A major and unfortunately unforeseen problem was that in as many as 230 cases we only had either a Hungarian household's telephone number or a Hungarian cell phone number at our disposal. In fact we had a Hungarian phone number to more emigrants than to whom we had a foreign one (120 cases). It was obvious, that the likelihood of successfully administering an interview using a home-country based phone number is rather low in the case of an emigrant person. When the number belonged to a household in the country of origin, an attempt was made either to speak to the migrant in this household or to get a direct contact detail to him / her. However, these attempts were rarely successful. Having a Hungarian mobile number only was also problematic because we were afraid that the migrants would not answer any call with an unknown number from Hungary for financial reasons. As we could find no way around these situations, this kind of problem had remained an important lesson to remember in future studies of this kind.

Interestingly, the problem was much less prevalent in Serbia, where a home-country number was given in only 13 cases whereas a foreign-country number in 232 cases. This might at least partially be

due to the differences in the nature of emigration in the two countries. With relatively more recent migrants in Hungary often in the neighbouring countries with more intense links in to the home-country households it is possible that they are more likely to keep their home-country mobile phones. Maybe some emigrants only use their Hungarian mobiles when they (temporary) stay within the home country. It is also possible that providing a Hungarian mobile phone number to the emigrant household member (or sibling) was a hidden way to refuse cooperation from the LFS respondents' side.

3.7. Data protection

Any social survey has to face issues of ethics as well as of data protection. This was especially so with the SEEMIG study, that not only aimed at collecting data about the respondents themselves, but also about 'third persons' – migrant friends, acquaintances – both in the first stage and (in the RDS block) – also in the second one. Throughout the project it was a key priority to handle the evolving data protection issues with much care and responsibility.[9]

Elements of the original SEEMIG data protection protocol that also hold relevance in the second stage were applied and also extended with further elements meeting the specific nature of the mixed-method technique.

1. No matter whether or not the consent of the migrant person was already provided in the first stage, their informed consent was directly asked in the second stage of the research, when migrants themselves completed the online or telephone questionnaire.

2. Data were only used for the purpose of the research. All the collected contact details were handled with special care, complying with all the legal rules and regulations regarding data protection.

3. Data suitable for identifying persons (contact details and names) were stored separately from personal data (data gained from the LFS survey in the first stage of the research) and were used exclusively to get in touch with the migrant for the purposes of the study.

4. As soon as the contact with the migrant had been established, all the information related to the contact details were destroyed.

5. All staff members participating in the process of collecting and handling data signed a confidentiality statement.

6. Principles of data protection were described in the e-mails sent to the respondents and also explained to the respondents contacted via telephone.

[9] For a detailed description of the process of developing a protocol for the first stage of the study see Blaskó and Jamalia 2014 Chapter 3.5.

Serbia – Data Protection

In the course of the research complete protection of personal data were provided in accordance with the Official Statistical Law, similarly to all other statistical surveys. Potential respondents were informed (via the official statistical website, e-mail addresses and SMS information) that data obtained in this study would be used for statistical purposes only. In that way special codes were provided for each respondent, which was recoded by the respondent when entering the web questionnaire. Help and additional information were offered to respondents with the possibility to write an e-mail to the project team or call the leader of the project at the SORS.

4. Results and main lessons learned from the second phase

From the total contact of 561 in Hungary, altogether 125 successful interviews were made: 66 on the web and 59 via telephone. Corresponding figures in Serbia are: from 298 persons with a contact information 98 were successfully interviewed – the majority of them (88) via telephone and only 10 had filled out the electronic questionnaire. These add up to a success rate of 22% in Hungary and 33% in Serbia.

A detailed list of outcomes of the various contact-making attempts in Hungary and in Serbia is provided in Table 3. As figures show, telephone contacts resulted in significantly higher response rates than did e-mail approaches in both countries – although the difference was much more notable in Serbia than in Hungary. On the other hand, e-mail contacts generated better results in Hungary than they did in Serbia.

Table 3 - Response rates in the second stage of the pilot study in Hungary and in Serbia

	Hungary	**Serbia**
Total number of migrants with a contact detail from the first stage	**561**	**298**
CAWI		
number of e-mails sent to migrants	277[10]	71
non-working e-mail addresses	23	7
number of people who answered after our first e-mail	35	9
number of people receiving a reminder e-mail	212	28
number of people who answered after our reminder e-mail	31	4
partially completed questionnaire received	10	2
no response from an operating e-mail address	178	49
total number of successful interviews via CAWI	**66**	***10***
Proportion of successful interviews with CAWI	24%	14%
CATI		
number of telephone numbers attempted to call	357	245
unsuccessful calls (not answered / answering machine / fax / answered by someone else etc.)	177	55

[10] Also includes those whose e-mail address was received during a telephone-interview attempt (18 cases).

refusal (by the person targeted)	92	89
prefers to answer via e-mail	29	7
interrupted interviews	0	6
successful interviews via CATI	***59***	***88***
Proportion of successful interviews with CATI (Ratio of successful telephone interviews to the number of persons approached via telephone)	17%	36%
Successful interviews total (CAWI + CATI)	**125**	**98**
Proportion of successful interviews related to the number of emigrants targeted	**22%**	**33%**

The relatively low response rates together with the low number of successful interviews achieved in either countries are considered neither sufficient for the final sample to be statistically unbiased of emigrants nor large enough for providing any reliable estimates. However, the collected data are still considered to be appropriate for valuable qualitative analysis of looking at various types of migrants appearing (such a work is already underway in both countries) and also for a small-scale validity test of the data provided by the home-country household members about the emigrant persons.

4.1. Response rates in the Respondent Driven Sampling block

Considering the success rates to the RDS block in Hungary the following evaluation can be made[11]. Out of the 561 persons with contact details collected in the first stage of the study a successful interview could be made with 125 (22%). From the 125 respondents of the second stage valid response was given to the first RDS question (how many emigrant persons he / she knows) by 100 (81%). 89 out of these said that they had at least one emigrated person in their personal network (77%). (The average number of acquaintances mentioned was 5.4). When it came to providing a contact detail to the emigrant acquaintance, only 31 respondents were willing to cooperate. Altogether they provided a contact detail to 54 further emigrant persons from Hungary (see Table 4.).

From these ratios the following scenario can be foreseen. Taking the response rates in this survey (22%) we can expect no more than 12 successful interviews in a potential second round of RDS. Again, assuming a response rate similar to what we have experienced in the pilot study (contact information provided by 25 per cent, to 1.74 emigrant persons on average) we can expect to collect contact information to five more emigrant persons in the next round.

Following a similar logic for the Serbian case it can be seen that out of the 298 emigrant persons 98 were successfully interviewed. From these cases further contact information was given in only 13 cases – the total number of contact information provided was 17. A prediction based on these figures would tell us that no more than one additional contact information can realistically be expected in a second round. (In Serbia instead of the 'Response denied' option the 'does not know any migrants' seems to be systematically chosen by the respondents.)

[11] The upcoming chapter is based on Simon-Kmetty 2013b.

These calculations clearly suggest that RDS in this form would not provide any satisfying solution to the problem of small number of cases and the biased emigrant sample resulted by the first stage of the SEEMIG study and it is not worth continuing the exercise.

Table 4 – Answers provided to the RDS question in Hungary and Serbia

		Hungary	Serbia
CAWI	the respondent knows at least one migrant **and** gave some statistical information **and** gave contact	16	2
	the respondent knows at least one migrant **and** gave some statistical information but **did not give** contact information	15	0
	the respondent knows at least one migrant but he/she **did not** gave statistical information about her (so **did not** give contact)	20	4
	does not know any migrants	11	4
	response denied	4	4
CATI	the respondent knows at least one migrant **and** gave statistical information **and** gave contact information	15	11
	knows at least one migrant **and** gave statistical information **but** did not give contact information	23	0
	the respondent knows at least one migrant but he/she **did not** gave statistical information about him / her (so **did not** give contact)	0	34
	does not know any migrants	0	43
	response denied	21	43
Together	knows but did not give any statistical information or contact	20	38
	gave some statistical information (but no contact)	38	0
	gave contact information	31	13
	does not know any migrants	11	47
	response denied	25	47
	Total	125	98
	Responded to the RDS block	**100**	**13**
	Number of contacts collected	**54**	**17**

5. Summary and overall evaluation of the SEEMIG pilot study

The low number of contact information collected in the first stage of the study, together with the (not unprecedentedly) low response rates both in the CATI and especially the CAWI study method in the second stage led to small sample sizes and potentially highly biased samples in Hungary as well as in Serbia. From this we can conclude that the ultimate aim of the SEEMIG pilot study, that was building and successfully interviewing large, unbiased emigrant samples on the basis of national representative studies (LFS) had failed to succeed.

The reason for this most likely lies in the lack of sufficient trust and confidence in the formal interview-situations to provide contact details to third persons – especially to emigrants, who form a sensitive target population in the first stage of the study, together with potentially bad quality contact details and a similarly low confidence of the emigrants in the second.

Although response rates during the first stage of the study were generally higher in Serbia than in Hungary, the reluctance of providing contact information was still not sufficiently high in either country – in fact it was of very similar magnitude in the two countries and response rates in the second stage were also not unlike each other.

An additional effort made by the SEEMIG team to compensate for the low case numbers by Respondent Driven Sampling technique has also failed to deliver the expected results. Again, lack of willingness of the respondents (in this case: the emigrants) to provide contact information to further emigrant persons had made it clear that further attempts to follow the collected links would be purposeless. As an earlier study in Hungary researching labour-migration to Austria has successfully applied RDS technique (Hárs 2009) our failure seems to suggest that it was more the lack of personal contact with the interviewer (working with CATI and CAWI) possibly together with the lack of full anonymity than the sensitivity of the research-topic that has doomed the SEEMIG RDS attempt to fail both in Hungary and in Serbia.

Besides demonstrating the inappropriateness of certain techniques to collect information on emigrants even the second stage of the SEEMIG pilot study has produced valuable data from a couple of aspects. Firstly, with detailed information about 125 emigrants from Hungary and 98 from Serbia we can carry out invaluable qualitative analyses, identifying typical patterns of migration without generalizing to a wider population. Secondly, a comparison of answers provided to the same questions by household members in the home country and their emigrant acquaintances on a one by one basis will make it possible to test the reliability of some of the results of data collected in the first stage of the SEEMIG study.

Finally, we suggest that the failure of the second stage of the SEEMIG attempt has only demonstrated that it is very difficult to reach a large representative sample of emigrants through a *large, highly formalized, national survey*. Based on the very positive experiences of the Nepal study (Ghimire et.al. 2013) as well as our experiences during the first stage of the study we recommend that the methodology is worth being tested in smaller scale, local surveys in settlements with a large number of emigrants. As we argued in our first report (Blaskó and Jamalia 2014) applying a flexible approach and building on the local knowledge of the interviewers building better trust with contacts in a smaller settlement is likely to lead to much higher response rates and a higher number of contact

information collected in the study. Successful and less successful attempts to tailor the fieldwork to the need of respondents are given both in Ghimire et.al. 2012 and the first SEEMIG report. Further pieces of important experience include that might be useful in such a local study and that was gained in the second stage of the study was that it should be avoided in every possible way that respondents provide local telephone numbers as contact information to emigrant persons.

Lack of full success of the overall research plan should not hide the inevitable positive results from the first stage of the SEEMIG study. As explained in the first report, the first stage has not only proved to be successful in offering valuable results methodologically but also in terms of improving our understanding of emigration. It has in fact provided us with a rich set of data on an exceptionally large set of emigrants, even though the representativeness of this data requires further investigation. This is more so in Hungary, where the significant underestimation of the emigrant population suggests that respondents in the LFS sample tended to conceal information. Fewer doubts were raised in Serbia, where the number of emigrants estimated on the basis of the SEEMIG study has significantly exceeded the respective estimate derived from the latest Census data.

Nevertheless, after a systematic evaluation of the selection processes throughout the study, we will be in the position to analyse emigration from Hungary and also from Serbia on an exceptionally large sample of emigrants. Moreover, individual level data will be possible to be linked to information on the sending household, which is again exceptional in the history of emigration research in these countries. This opportunity offers 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 could be explored.

On the individual level we will be able to provide valuable data on the composition of the most recent emigrant groups both from Hungary and from Serbia 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 some information available on their financial linkages to their households in their country of origin – i.e. some badly needed insight into the field of remittances can also be given.

Finally, our experiences have also proved that although large, formal, nationally representative sample-research are not likely to sufficiently support direct surveys with emigrants they can still be utilised as a basis for developing reliable estimates for the number of international emigrants from a country. Based on the lessons learnt from the pilot study a detailed set of recommendations can be developed that can be utilized by members of the research community to produce more reliable estimates on the number of emigrants as well as to estimate their distributions by some basic characteristics such as age, gender or country of destination[12].

[12] A short summary of this recommendation will be based on has been included in the SEEMIG National Action Plan in Hungary and is also given in the Annex II of this report.

References

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

Gárdos, Éva - Gödri, Irén (2014): *Analysis of existing migratory data production systems and major data sources in eight South-East European countries*. SEEMIG Working Papers, No.2, Hungarian Demographic Research Institute, Budapest.

Ghimire, Dirgha J. – Williams, Nathalie - Thornton, Arland – Young-DeMarco, Linda – Bhandari Prem B. (2013): Innovation in the Study of International Migrants. Presented at the Population Association of America Annual Meeting, April 11-13, New Orleans, LA.

Hárs, Ágnes (2009): Magyarok az osztrák munkaerőpiacon. Ingázók, bevándorlók, munkaerő-migránsok? Kutatási Zárójelentés. (Hungarians in the Austrian Labour Market. Research Report.) https://onedrive.live.com/view.aspx?resid=20226F10B70B2C25!311&cid=20226f10b70b2c25&app=WordPdf&wdo=2&authkey=!AEZMjxUfjQcQMng

Heckathorn, Douglas D. (1997): Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44(2).

Kmetty, Zoltán – Simon, Dávid (2013a): A válaszadó vezérelt mintavétel módszerének alkalmazási lehetősége a SEEMIG kutatás kapcsán. Szakértői összefoglaló. Kézirat (Applying Respondent Sampling Technique in the SEEMIG pilot study. Manuscript.)

Kmetty, Zoltán – Simon, Dávid (2013b): Az RDS továbbvitelének lehetőségei a SEEMIG projektben. Kézirat. (Feasibility of Continuing with RDS Methodology in the SEEMIG Project. Manuscript.)

Simon, Dávid (2012): Válaszadó-vezérelt mintavétel: ritka és rejtett csoportok kvantitatív vizsgálata (Respondent driven sampling: quantitative research of rare and hidden groups). In: *Statisztikai Szemle,* 90/4 http://www.ksh.hu/statszemle_archive/2012/2012_04/2012_04_249_1.pdf

Arland Thornton, Arland – Binstock, Georgina – Jalal Abbasi-Shavazi, Mohammad – Ghimire, Dirgha – Gjonca, Arjan – Melegh, Attila –Mitchell, Colter – Moaddel, Mansoor – Xie, Yu – Yang, Li-sshou – Young-DeMarco, Linda – Yount, Kathryn M. (2012): Knowledge and beliefs about national development and developmental hierarchies: The viewpoints of ordinary people in thirteen countries. *Social Science Research*, 41 (5) 1053–1068.

Annexes

I. Questionnaires
   a. Online questionnaire in Hungary – English translation
   b. Online questionnaire in Hungary – original Hungarian version
   c. Telephone questionnaire in Hungary – original Hungarian version
   d. Questionnaire in Serbia – English translation
   e. Questionnaire in Serbia – original Serbian version
II. Possibilities of using SEEMIG pilot methods (formulating new, constant LFS questions concerning outmigration)
III. Social and demographic distributions of emigrants reached in the first and the second phase of the pilot study