

How to get better Data on Emigrants? Lessons learned from the SEEMIG¹ pilot emigrant survey in Hungary and Serbia

(Extended abstract)

Zsuzsa BLASKÓ

HCSO Demographic Research Institute, Budapest

E-mail: blasko@demografia.hu

Irén GÖDRI

HCSO Demographic Research Institute, Budapest

E-mail: godri@demografia.hu

Shortages of migration statistics and in particular problems of *emigration* data are well-documented in the relevant literature. Although survey type data collection might appear as an obvious substitute that could overcome the shortcomings of administrative data in the field, making an appropriate research design to capture important features on a representative set of the emigrant population poses serious challenges to the researcher. The key problems are (1) lack of an appropriate sampling frame, (2) the fact, that emigrants most often constitute a rare population in the destination countries and that (3) migration is a sensitive domain often biased by low response rates². Because of difficulties (1) and (2), survey studies on emigrants tend to use non-random sampling methods (very often snowball sampling) not aiming at representativity. Respondent Driven Sampling, although an attractive alternative to these has led to no convincing, well-documented results in the field of migration as yet³. Beside being non-representative, emigrant surveys usually are also of limited scope in concentrating on a selected group of migrants e.g. by profession or most often by receiving country.

¹SEEMIG, Managing Migration and its Effects in South-East Europe. (www.seemig.eu) Funded under the third call of the South-East Europe Programme, number SEEMIG/SEE/C/0006/4.1/X

²Beuchemin, 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

³Beauchemin, C. and Gonzalez-Ferrer A. (2011); pp.106.

An innovative approach offering the potential to produce a representative sample of emigrants from a community of origin (potentially even from an entire country) was first introduced in the Nepal study by Ghimire D.J. and his colleagues.⁴ The Nepal research project started off from the *Chitwan Valley Family Study* by identifying family-member migrants to the Gulf Cooperation Council countries in the sample-member households. During the interview, contact information was also collected to the migrants making it possible to directly contact them in the second phase of the survey. The Nepal study was extremely successful in collecting contact details as well as in eventually finding and interviewing members of their target population. The carefully designed and administered research process resulted a sample of 460 migrants – 95% of the total number of identified migrants⁵. As this sample was based on a random sample of households at origin and as the links between the household and the out-migrant were systematically defined, the Nepal study successfully produced a large and representative sample of emigrants providing. Thus it not only provides the opportunity make in-depth analyses on the basis of the rich information collected from the migrants themselves but it can claim to serve as a basis for estimating the extent of outmigration from the Chitvan Valley.

Although the authors of the report suggest that their research strategy “*will have applicability elsewhere*”, they also acknowledge that risks and difficulties inherent in the method are significant. The major challenge lies in gaining the trust of the respondents in the origin- as well as in the destination country. Earlier research attempting to collect contact details to migrants in their former household lead to varying and sometimes very low success rates – with only 5% of declared migrants interviewed successfully in the MAFE project for example⁶. This also suggests that success of the approach applied in the Nepal study depends largely on the actual details of the research design as well as on the specific cultural and social context.

After reviewing the experiences from the Nepal study, as well as carefully reviewing the situation of migration statistics in the South-East European region, in the frame of the SEEMIG

⁴ Ghimire, D. J. Williams, N., Thornton A., Young-DeMarco L., Bhandari P.B. (2013): "Innovation in the Study of International Migrants." Presented at the Population Association of America Annual Meeting, April 11-13, New Orleans, LA.

⁵ In 92% of the cases when a migrant person was identified, contact information was also provided to the interviewer. In the second phase, 87% of their target respondents were successfully interviewed in 6 months and 95% in 26 months.

⁶Beauchemin, C. and Gonzalez-Ferrer A. (2011); pp.106.

project a decision was made to implement a survey approach similar to the one in Nepal⁷ - albeit with a greater territorial scope. Realising also the notable risks inherent in the method, it was chosen as the potentially most effective design to improve the situation in researching emigrants from the various countries of the region. In spring 2013 two pilot studies (one in Hungary and one in Serbia) were carried out to further improve and test the methodology in a (South-East) European environment and – based on the experiences gained – to provide guidance for future research on emigration.

The SEEMIG research design consists of two-phases the key idea being to *derive a representative sample of emigrants from a representative national survey*. Due to its relatively large sample-size and high level of standardisation across countries, in the first phase of the study the Labour Force Survey was utilized and out-migrants were identified through the households included in the LFS sample. In addition to the standard questions a SEEMIG battery was attached to the LFS. The battery not only identified out-migrants linked to the households, but also collected a set of basic statistical information about them. In the most sensitive, final part of the interview an attempt was then made to record contact information (email address, telephone number...) to the migrants reported in the household. These pieces of contact information serve as the basis of the second phase of the pilot survey when migrants are directly contacted and asked to answer a series of more in-depth questions. The second phase of the survey has been carried out via telephone and the internet during the summer and autumn 2013.

Although hindered by the low level of cooperation of the respondents, the SEEMIG pilot study has provided us with unique results. First, it has the capacity to improve our understanding of out-migration from the two countries involved in the pilot and secondly, it provides valuable lessons for the improvement of emigration surveys in the future. The proposed paper will focus on this latter set of new insights – the methodological lessons built mainly on the Hungarian experiences.

The paper will introduce the research design in detail. A special attention will be given to the innovative elements of the SEEMIG study – as for example including migrant siblings among

⁷ In fact, authors of the SEEMIG project proposal could not be aware of the existence of this study at the time of writing the proposal. The Nepal papers first became available to a wider audience in Spring 2012 when the SEEMIG project was already about to get started. Nevertheless, we could make very good use of their findings and experiences at the final state of the research design.

members of the targeted population, and also some new methods introduced in the LFS fieldwork. The consequences of these innovations on the outcomes of the research will also be discussed.

Estimating the stock of out-migrants from Hungary on the basis of the SEEMIG data has certainly been one of the main ambitions of the study. Comparing this estimate to estimations based on other data-sources is also an important test of the SEEMIG study. Although (following from the very nature of the problem) firm reference points are hard to find, SEEMIG seems to somewhat underestimate the number of emigrants from Hungary. The extent of this bias will be evaluated and potential reasons to this systematically analysed. We will look at the attrition rate in the subsequent phases of the data collection and (where it is possible) analyse the selection into the upcoming phases. For example we will systematically compare the households where a migrant person was declared but no more information was provided about him (her) with those households that also gave statistical data on the migrant person. Similarly, households providing statistical data as well as contact information will be compared to households providing statistical data only. This will help us to reveal potential systematic biases in the sample.

As underestimation is probably related to the low level of cooperation from the respondents' side in the survey, the paper will also cover this issue. In doing so, we will discuss new findings from an external test to SEEMIG data. This test⁸ aims to control for potential LFS-effects in the SEEMIG-findings. When discussing the general but also the context-specific reasons for non-cooperation we will also rely on the survey carried out among the interviewers after the Hungarian pilot study.

The paper will conclude by formulating key lessons and recommendations for future emigrant surveys. In this we will also rely on findings from the second phase of the study. In particular we will use the unique opportunity to validate the survey items by comparing data provided by the migrant to data provided by his or her Hungarian household-member.

⁸ The test is being carried out on two consecutive Omnibus surveys in September and October 2013 by Hungarian research institute TÁRKI.