

Distr.  
GENERAL

Working paper 9  
3 March 2008

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and EUROPEAN COMMISSION  
ECONOMIC COMMISSION FOR EUROPE STATISTICAL OFFICE OF THE  
CONFERENCE OF EUROPEAN STATISTICIANS EUROPEAN COMMUNITIES  
(EUROSTAT)**

Joint UNECE/Eurostat Work Session on Migration Statistics  
Geneva, Switzerland, 3-5 March 2008

Item 2 of the provisional agenda

#### SELECTED METHODS TO IMPROVE EMIGRATION ESTIMATES

### **TOWARDS AN IMPROVEMENT OF CURRENT CURRENT MIGRATION ESTIMATES FOR ITALY\***

Submitted by Istat, Italy

#### I. INTRODUCTION

1. This paper deals with two main topics, both related to possible strategies to increase the quality of statistics on migration, especially emigration flows. Firstly, attention is focused on the current surveys based on administrative data, as it is recognized that some improvements in the process of data collection and a more extensive use of estimation methods can turn into major improvements in the quality of data, i.e. timeliness and accuracy of information provided. Secondly, the possibility to use alternative sources of data is explored; in particular household surveys currently carried out by Istat for scopes different from the study of migration.

2. This analysis aims to use all available information (population census, current administrative surveys, sample surveys) in order to get a clear indication on the actual size and composition of migration flows. Special attention is devoted to an evaluation of emigration

---

\* This paper has been prepared by Domenico Gabrielli and Maria Pia Sorvillo at the invitation of the secretariat.

numbers, in order to determine if they are actually underestimated in current surveys and how large is the undercount and to look for possible solutions.

## II. AVAILABLE SOURCES FOR MIGRATION STATISTICS

3. In current statistics, migration flows are measured by using the Municipal Population Registers and therefore concern the resident population. Being based on usual residence, the registrations and cancellations from and for abroad may be assimilated to long-term migration, since it may be hypothesized that all residence transfers are due to migratory movements of more than 12 months.

4. Summary and individual data on migratory flows are gathered through current surveys, requested from the municipal registry offices with a statistical form.

5. Summary data are collected as part of the demographic balance, including all the relevant natural and migratory flows and population stocks. The demographic balance also contains the initial stock of the population, as well as that calculated at the end by adding and subtracting the demographic flows, as the main aim of the survey is actually to update the amount of resident population.

6. Data collection of individual data on migratory flows is based on the survey of residency transfers (registrations and cancellations in population registers). This survey annually supplies data at an individual level on the migration flows between Italian municipalities and abroad. Transfers that take place within the same municipality are excluded. Unlike the collection of summary data this survey allows to gather data on individual transfers pertaining to the characteristic of migration and of migrants (at the time of migration<sup>1</sup>). The data collection on demographic balance therefore verifies in advance the total of the migratory flows, while the data collection on residency transfers supplies detailed information on each migratory movement, allowing also analyses on the origin and destination of migration. It was actually this survey which revealed the changes that had taken place during the 1980s, when Italy began to be a destination country for migratory flows of foreign citizens.

7. There is a systematic undercount of the outward flows in the data collected with the survey on individual change of residence, compared to the survey on municipal demographic balances (see table 1) In fact, collected data can be affected by three kind of problems. First of all it is possible that a limited number of modules are not delivered to Istat or are discarded because of the amount of errors in it, causing a slight underestimation of the number of recorded registrations and cancellations. Secondly, the whole process still relies hardly on paper,

---

<sup>1</sup> The data collection model (APR/4) has included a question on nationality with codes for a limited number of countries since 1980. A full list of nationalities and countries was adopted from 1995 onwards. The information collected also includes the country of birth and the country of origin/destination (previous or future residence), as well as demographic characteristics (sex, date of birth, marital status, etc), qualifications, professional status and sector of activity. (more information is reported in section 4). There is no specified reason for the move.

difficult to move and heavy to handle, and is therefore quite long and time consuming, causing a notable delay in data availability. But the main reason of the difference lies probably in the administrative process, as Municipalities don't always fill in the individual form when residents move abroad, even though they count the migration in the summary form.

**Table 1** - Immigration and emigration flows, individual an summary data. Years 1995-2004

Years	Summary data		Individual data		Differences		Differences %	
	immigrants	emigrants	immigrants	emigrants	immigrants	emigrants	immigrants	emigrants
1995	100.788	50.120	92.820	43.937	7.968	6.183	7,9	12,3
1996	178.464	51.756	171.980	47.517	6.484	4.239	3,6	8,2
1997	172.743	53.408	162.857	46.273	9.886	7.135	5,7	13,4
1998	165.696	56.707	156.885	45.889	8.811	10.818	5,3	19,1
1999	189.876	76.483	185.052	64.873	4.824	11.610	2,5	15,2
2000	236.292	66.821	226.968	56.601	9.324	10.220	3,9	15,3
2001	214.776	67.125	208.252	56.077	6.524	11.048	3,0	16,5
2002	222.801	49.383	213.202	41.756	9.599	7.627	4,3	15,4
2003	470.491	62.970	440.301	48.706	30.190	14.264	6,4	22,7
2004	444.566	64.849	414.880	49.910	29.686	14.939	6,7	23,0

### III. IMPROVEMENTS FOR THE STATISTICS ON MIGRATIONS

8. Major improvements in the data collection from Municipal Population Registers will be brought about by the completion of the process of computerization of all Registers. At the moment only 50% of data collected from Municipalities are sent to Istat by means of computerized systems (summary stock and flow data and individual flow data). An all-round computerization of data collection will certainly make data available more promptly and improve their quality. It will be achieved when all municipalities will use the web systems of data collection Istat has launched in 1999, giving the opportunity to send all requested files via Internet. A relevant increase in timeliness could be obtained also using estimate methods based on those municipalities that send their data using the Internet, and therefore are well-timed and more accurate.

9. Also centralization could help in improving statistics on migrations. Ministry of Internal Affairs launched a project, named INA-SAIA, which should have great impact on the efficiency of Population Registers and, therefore, on the derived statistics. In the next future it should be fully operating in its two components: INA, a centralized index of resident population; SAIA, a network for the communication of all variations in the position of resident population (birth, death, marriage, migration, etc...) among all municipalities and between them and various public administration offices.

10. From a statistical point of view, this system should be able to grant some major advantages. Firstly, it will help a full computerization of all municipal Registers, therefore allowing to replace paper forms, still used for the transmission of data to Istat, with web data collection. The improvements in timeliness and accuracy of data are clear and straightforward. In addition to that, the centralization of all registers will help to check the internal consistency of data, especially avoiding time gaps in transmission of information on migration. In the long term,

present surveys could be complemented with data derived by SAIA, increasing the amount of information available without any additional burden for respondents, in line with the principles of efficiency of public administration.

#### IV. WHY DO WE THINK EMIGRATION IS UNDERESTIMATED BY CURRENT SURVEYS?

11. We have seen in the previous paragraphs that the main source of data about emigration in Italy is the deregistration from population register. This way of counting emigration presents advantages and disadvantages. Among the former, the availability of detailed data, yearly updated and giving some insight about main characteristics of emigrants, such as age, sex, civil status, citizenship and country of destination. In addition to that, data obtained from this source are consistent with other data used to evaluate the population balance and, finally, the total amount of resident population. But there are disadvantages, as well. First of all, the statistical data collection relies on an administrative database, or better on more than 8.000 databases, one for each municipality. It is quite difficult to control for some relevant aspects, for example the full compliance to statistical definition of emigration and the timeliness of registration. Even more relevant, at least in the framework of the issue we are examining in this paper, are the effects of administrative rules and everyday practice on the risk of undercounting. The rules – namely the Population Register law - require that every person declares to the population register every change of his/her position, including the emigration. In the practice, there is no fine nor sanction for those who do not fulfil this obligation. Therefore, it is likely accomplished only if there is a personal utility in doing it. For example, Italians residing abroad can have an advantage in declaring it because they have the right to vote for Italian elections in the country of new residence; foreigners going back to their country, on the other hand, could not see any advantage in declaring it, and therefore just ignore the rules.

12. The theory and practice of administrative process behind data on emigration justify some doubts about the completeness of estimates based on this source. It is necessary to verify this hypothesis more carefully, and to try to quantify it. It is not a straightforward task, as the sources available for such an elaboration are not promptly available, and therefore we have to rely heavily on a number of estimates. The main source of information about this topic is the population census, that enables us to make a comparison between the population as resulting from the registers and the numbers counted by the census. If the census were perfect, the amount of population registered but not counted by the census could be considered as “shadow emigration”, i.e. persons emigrated during the last ten years but still present in the population register. Actually, the situation is quite more complicated, as we have to take into account the census errors, namely the under-coverage (persons resident but not counted by the census) and the over-coverage (persons included in the census as resident, but not having the requirements). In addition to that, it exists a quota of Anagrafi undercoverage, due to foreigners who do not register as they do not any advantage in doing it, even though they would have all necessary requirements.

13. A survey conducted right after the census in order to evaluate census errors gives an estimates of the coverage rate, as reported in table 2<sup>2</sup>. On the basis of the deriving undercoverage rate and considering the over coverage of the census, as can be evaluated looking at subsequent operations of realignment, we can estimated the error due to mireregistrations happened in Anagrafi in the period 1991-2001, which is around 166 thousand. If it were entirely due to “black emigration” it would be equivalent to an average rate of 27% during the period 1991-2001<sup>3</sup>. This value has to be seen as a estimation of the upper limit, as in this exercise, with a rough approximation, we disregarded Anagrafi undercounting, focusing on the other causes of the difference between registered population and census. This analysis, even if not suitable to give accurate estimates of emigration flows, indicates clearly enough that it could be useful to look for different sources of information to complement the administrative surveys currently available.

**Table 2** - Population register performance (estimated indicators)

Difference Register-census (2001)	961,949
Coverage rate (2001 census)	98.55
Persons not recorded by the census	833,417
Persons incorrectly recorded by the census	38,000
Persons incorrectly registered in Anagrafe (net 1991-2001)	166,532
Total number of emigrants (administrative data) <sup>3)</sup>	608,164
Maximum “Shadow emigration” rate (average 1991-2001)	27.4

<sup>2</sup> "La progettazione dei censimenti generali 2010 – 2011” in “ Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento” Documenti, n.10 anno 2007 in [www.istat.it](http://www.istat.it).

<sup>3</sup> The total amount of emigrants, as resulting from administrative data, is obtained adding the number of persons who declare to Anagrafi they moved abroad to the number of foreigners who are cancelled as, during periodical checks, they turn out to be no longer residing in their municipality, nor emigrated in another Italian municipality.

## V. HOUSEHOLD SURVEYS AS COMPLEMENTARY SOURCES OF INFORMATION ON EMIGRATION

14. Istat carries out a number of sample surveys on households. Generally, the sample frame is the list of resident population as drawn from Anagrafi, excluding those living in an institution. In order to examine the possibility to use results from these surveys to disentangle the problem of “black emigration” we need to rely on panel or pseudo-panel surveys. Through them it should be possible, in principle, to obtain estimates on the number and characteristics of those who left the country, actually disappearing from the sample after the first interview.

15. The two surveys whose features are consistent with our needs are: EUSILC (European statistics on income and living conditions) and LFS (Labour force survey). The first one has sample based on a panel of about 25,000 households and is carried out every year since 2004. The second one is a continuous survey, and has a much larger sample (more than 270,000 households every year) where the households are re-interviewed following the scheme 3-12-15 months. It means in the 12 months sample, for example, about 50% of the original sample is re-interviewed; referring to a single quarter it is about 37,000 households that are asked the question about changes in households composition during a period of one year.

16. The project we are working on foresees, as first step, an analysis of the structure of the questionnaire of both Eusilc and LFS, looking for the questions suitable to answer to two points: how many persons emigrated during a one year period, leaving some member of the household in Italy? How many households left all together?

Of course, the two questions are to be treated separately, as the two different phenomena cannot be investigated in the same way in the framework of a panel (or pseudo-panel) survey. In principle, it should be much more probable to obtain accurate answers to the first question, when it is possible to ask it to some member of the household still in Italy.

17. The questions we identified for further analysis are:

in EUSILC: Presence in the household – item: person not longer living in the household because moved abroad; in LFS: The person “name” is still part of the household? If not: Can you tell me why he/she is no longer part of the household? – item: He/she moved abroad.

A first remark refers to the different nature of the two surveys: the first is a panel survey, therefore it is a nodal point to assure the possibility to follow at the best households, considering every single member and the household as a whole. Therefore, great effort is devoted to minimize the risk to lose the household, tracking its movements on the territory, and to keep track of any change in composition, such as the departure of one of the household’s member. LFS, on the other hand, is only partially longitudinal, and a household missing the following interviews is not such a crucial problem.

18. Results from the 2006 wave of EUSILC – the last currently available – give an estimate of about 75.000 persons emigrated during the year 2006, considering only individuals. Taking into account households too, the numbers are shown in table 2. It has to be noted that the standard error is quite high, the event of emigration being rare in the population and the sample being not so large (sampling rate is around 1 percent). Furthermore, the size of an household moving abroad has been arbitrarily fixed at 2, vs. an average for all households of 2.5. The

underlying hypothesis is that emigration is a feature specific of younger and therefore smaller households, with a strong incidence of one-person households. The resulting number of emigrants could vary from 114 to 209 thousand.

**Table 3** - Emigration as resulting from EUSILC 2006

	Estimates	Sampling error (%)	Confidence Interval ( $\alpha=0.05$ )	
			Lower Limit	Upper Limit
Households	42,827	16.7	28,828	56,826
Household members (A)	85,654		57,656	113,652
Individuals (B)	75,612	13.2	56,031	95,193
Total (A+B)	161,266		113,687	208,845

19. A comparison with data for the same year deriving from current surveys shows some interesting convergence. The present numbers on emigration are equal to 113 thousand (see note 3), which practically matches with the lower limit of the confidence interval for survey estimates. Considering the undercount rate observed in the period 1991-2001, it should add up to 144 thousand, that is anyway internal to the confidence interval.

20. First elaborations carried out on LFS data are not as interesting as the previous ones. Even though this survey is potentially more informative than EUSILC, due to a much larger sample, actual results show severe problems in the accuracy of this information, as collected by interviewers. Less than 10,000 persons are detected as emigrated during the period, and even taking into account the fact that this number does not include households movements, it is clear that the question is not properly registered (as asked by interviewers or answered by interviewed persons). A further exploration of this source is surely relevant, as it is the main household survey carried out by Istat, and a mine of information about socio-economic condition of population, including those who emigrate. But first results shows clearly that the use of this source requires more elaborations, first of all in order to consider a larger sample, taking into account different quarters, and secondly working on an improvement in the process of collection of this kind of data, as already undertaken for the 2008 wave of the survey increasing the interviewers' awareness of the relevance of an accurate tracking.

## VI. SOME PRELIMINARY CONCLUSIONS

21. The process of amelioration of statistics on emigration has to be seen as a long path, including different activities. First of all, it is necessary to monitor the changes in legislation and in the management of Population registers, with a particular attention to their impact on statistics. This task requires a strict cooperation with the Ministry of Internal Affairs and with associations of public officers in charge of Registers, making them aware of the value of their work in a statistical perspective.

22. Second, it has already been started the work to realize improvements in the production of data from current surveys, in terms of timeliness and accuracy. Italy is in a rather lucky situation, compared with other European countries, having an administrative system able to give detailed information on resident population stocks and flows. Even though not always perfectly in line with statistical requirements, data already collected are very valuable, and they have to be exploited at their best. A systematic use of estimates could help to increase this value for statistical purposes.

23. Last point, preliminary results seem to be encouraging as far as the possibility to use sample surveys to complement existing data on emigration, especially referring to EUSILC. This can be a source of additional information on migrants socio-demographic characteristics, and can be used as a benchmark for data coming from administrative sources. Of course, further work is needed to fully exploit informative potentialities and to ameliorate estimates, as those presented in this paper are quite rough and rely on a number of hypothesis to be further verified. In addition to that, it will necessary to evaluate the consistency of results with administrative data, that could be done both at individual and aggregate level. It could help to better understand point of strength and weaknesses of the two sources and, finally, to ameliorate migration statistics for Italy.

\*\*\*\*\*