**UNITED NATIONS STATISTICAL COMMISSION
and ECONOMIC COMMISSION FOR EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

CONFERENCE OF EUROPEAN STATISTICIANS

Joint Eurostat/UNECE Work Session on Demographic Projections
(28-30 April 2010, Lisbon, Portugal)

Item 8 – Constructing assumptions for migration: data, methods and analysis

International migration data as input for population projections

Anne Herm, Estonian Interuniversity Population Research Centre, Estonia
Michel Poulain, Université catholique de Louvain, Belgium

1.   INTRODUCTION

It is widely known that the migration component is the most problematic basic data for population projections. Most countries are facing difficulties to produce international migration data needed for reliable national population projections.

The first part of this contribution will discuss the situation of availability and reliability of data on international migration while the in the second part of the contribution we intend to discuss the possible way to face that situation and some concrete proposals will be given in order to estimate the missing figures on immigration and emigration flows for country which data are not sufficiently reliable or are missing. For conclusion, we intend to evaluate the impact of using bad data on international migration in population projections.

2.   IMPLEMENTING THE MIGRATION COMPONENT IN POPULATION PROJECTION

There are several reasons why the migration component is often handled on a simplified way in population projections, a choice that does not help to increase the reliability of the results:

- In several countries migration flow data, both immigrations and emigrations are not available while in some other countries only immigration data are produced.

- Where data are available they are often not fully reliable as can be seen when confronting immigration and emigration data related to the same migration flow between a pair of countries in a double-entry migration matrix.

- Moreover, for a given country, the level of reliability may vary largely between immigration and emigration and between different groups of migrants, like nationals, EEA citizens and non-EEA citizens.

- Time series on international migrations present brakes that are related to changes in data collection or applied administrative rules. They have often limited length, so that trends may not be easily identified for the projection purpose.

- Finally the situation can be even worse when the detailed age and gender composition of migrants is concerned.

Considering the international migration component by using the concept of "net migration" is not appropriate. The net migration is the arithmetic result of the difference between two flows that cannot be subtracted as they are composed by different persons with specific characteristics. Moreover different levels of reliability are observed. In fact immigration and emigration flows are never collected with the same level of reliability due to the fact that a country does not have the same interest to measure immigrations and emigrations for nationals, EEA citizens and non-EEA citizens. While higher level of reliability is often observed for immigrations of non-EEA citizens in the EU Member States through residence permits databases, the weakest level is related to

emigration of nationals, and that has also a consequent impact on (return) immigration data on nationals. Moreover the age and gender distribution of "net migrants" does not fit with any classical "model migration schedule". Therefore it is more appropriate to project separately the age and sex composition of immigration and emigration flows than to project the age and sex composition of 'net migrants'.

Even if the projection is related to the whole population, it is also better to project the migration flows by considering separately different groups of migrants, namely nationals, other EEA citizens, non-EEA citizens. Age and sex composition of each flow can be dependent on the main phases of the migration process, the latter being specific for each group of migrants. In addition, the impact of migration-related policies is clearly different towards immigration and emigration flows as well as to migration of national citizens and most of foreigners. Accordingly different assumptions may be done for immigration and emigrations of these groups of migrants and these will depend if the country is mainly an emigration country or an immigration country.

3.   REQUESTED MIGRATION FLOW DATA

Net migration versus immigration and emigration flows

In many population projections the migration component is reduced to the net migration. In the literature on population forecasts it is quite common to find discussion about "age- and sex-specific mortality", "age-specific fertility" and "net migration by age and sex" and not about "age and sex-specific immigration" or "age and sex-specific immigration". From the viewpoint of migration researcher the annual immigration and emigration flows are independent and consequently the net migration has no meaning when disaggregated by age and sex. *There are no 'net migrants'; there are, rather, people who are arriving at places and leaving them. Why they are doing so is central to understanding the dynamics of (...) growth and decline* (Morrison 1977).

To consider better assumptions for the future of international migration, replacing the immigration and emigration data by net migration will cause additional uncertainty because the net migration is more volatile. Moreover, the net migration is very often calculated as a residual between total population change and natural increase or decrease being not always fully consistent to the immigration and emigration flows.

The practice of using net migration in population projections and forecasts is assumedly caused by relatively bad availability of migration flows data compared to those of births and deaths, and introducing net migration as the migration component is considered being more acceptable than keeping "zero migration" assumption. Actually, the advantages and disadvantages exist in both cases - when using net migration and when separately immigration and emigration flows data (Figure 1). On one side, the positive argument for using net migration is that it could be more easily available than migration flow data as it can be easily calculated as the difference between total population change and natural change. Also, in the situation where both immigration and emigration flows are under- or over-registered the under-coverage has less impact on net migration than on migration flows as the same person moving shortly may not be counted neither as immigrant nor as emigrant. From another side, net migration is more volatile that immigration and emigration flows considered separately and therefore more difficult to predict for a longer period, even if it may be more easily linked to economic changes. When net migration is calculated as difference between immigration and emigration, the net migration may have double effect from errors or under-coverage of both flows. Net migration rates cannot be properly calculated for net migration as there is no corresponding "at risk population", and when calculated, *they hide well-established regularities in the age pattern of geographical mobility* (Rogers, 1990). In case of migration flows data, missing age distribution of immigrants and emigrants can be solved with using age models for migration schedule such standard tool does not exist for the net migration.

**Figure 1 - Advantages and disadvantages related to concept used for the migration component**

NET MIGRATION			MIGRATION FLOWS
Can be more easily available (census results, calculated as difference of population and natural change)	+	-	Ad hoc data collection system is needed
More reliable even if flows are under or over counted with the same error or if the two censuses are fully comparable.	(+)	-	Difficult to ensure the complete coverage of all groups of migrants
More volatile and more difficult to predict as it is a difference of two flows	-	+	Trends can be identified even if fluctuation takes often place
Rates cannot be properly calculated (no at risk population)	-	+	Target and at risk population groups can be identified
No models are available for age distribution of net migrants	-	+	Age distribution of immigrants or emigrants can be modelled
Such data are not possible to collect but can only be obtained through calculation	-	+	Data can be collected by statistical or administrative systems

In order to understand the possible future developments in the international migration, the data on both immigration and emigration flows should be analysed distinctly and not just net migration. Therefore, where any data exist on migration flows, these should be used (and not net migration) even if they cover only partly the migration pattern in the country. All possibilities to assess and improve the available immigration and emigration flows data should be checked and introduced in order to project future migration. Still, there are reasons why some countries may want to use net migration. As the possibilities to improve the reliability of net migration figures and forecast its trend are limited, we do not suggest using net migration in population projection and will no longer consider this possibility in this contribution.

Age- and sex-specific immigration and emigration flows

As for the population data generally, age and sex distribution is necessary in case of migration data. Even if the age distribution of immigrants and emigrants has the same general character, the concrete immigration and emigration flows of a country represent a specific age and sex distribution that depends on several factors, including such as if country is predominantly immigration or emigration country. The following example proves that flows of male and female migrants can be significantly different (Figure 2).

**Figure 2 - Immigration of male and female foreign EU citizens in Germany, 2006
(Eurostat database, April 2010)**
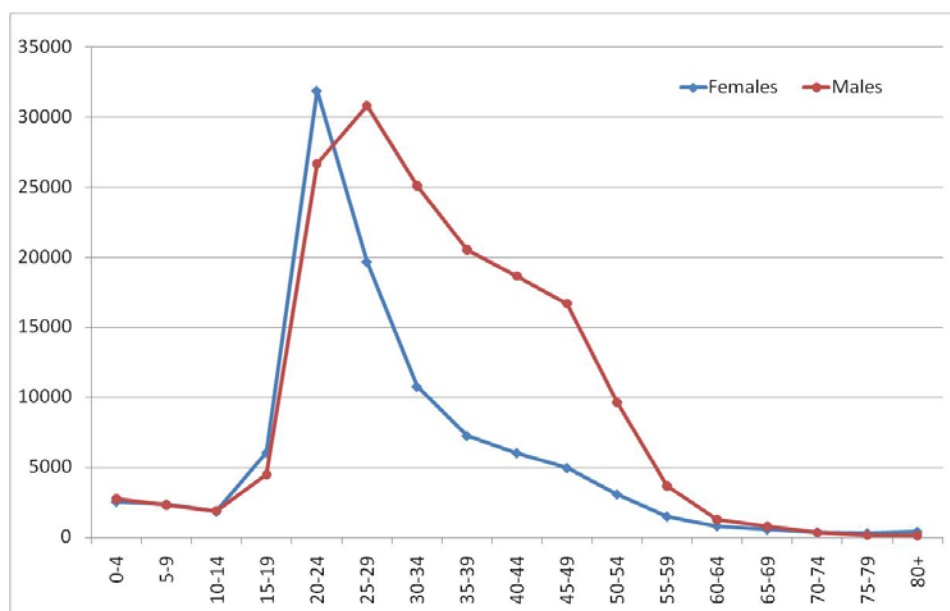
Disaggregation by sex is compulsory in all demographic data collection, including migration data, as well in population projections. Where data exist on migration flows, these are generally disaggregated by sex. Indeed, according to the information available in Eurostat database, all countries in UNECE region having data on international migration flows have these data separately on males and females.

Disaggregation by age is also compulsory in all demographic data, including migration. The question is on which level disaggregation by age the data are most appropriate for producing projections. One may argue if by single year of age is necessary or five year groups would be satisfactory. We believe that even if the projection is done by single year of age up to 110+, such detailed level of disaggregation is not needed and even not recommended for international migration flows. If data are available by single year age, it may include age peaks and random variations, and therefore would need to be smoothed, while data by 5 years age could be disaggregated, in order to be introduced in the population projection by single year of age, by using models.

Concerning the highest age group, the distribution up to age 110 years is considered as useless. Ideally, the age distribution up to age 100 years would be the best, but in fact, there are only a limited number of migrants in old ages. Thus, the most appropriate age distribution for migration flows would be 5-years groups up to the age-group "80 years and older" as these data are usually more reliable and less volatile compared to more detailed age distribution. Moreover, if needed, the single year age distribution up to 110+ can be calculated on this base by using migration schedule models.

Citizenship or origin and destination of migrants

In addition to age and sex, which other characteristics of migrants should be considered as essential? Clearly only total numbers of immigrants and emigrants are not sufficient for predicting future trends of the international migration to be included in population projections. Data by country of citizenship are needed as country of citizenship is the most policy-relevant characteristic of international migration. Despite there is a general pattern of age distribution of international migrants, the total flow consists from several groups of migrants that have different age structures. These groups of migrants are accepted in country under very different conditions and administrative registration rules. Concretely, the conditions for immigrating or emigrating depend on the citizenship of migrants and the most important difference exists between migrants with free or with restricted movement rights. More precisely there are fully different conditions for immigration of nationals and foreigners to the most of countries. Even among foreigners, in EEA countries, rules are different when migrant is EEA citizen, and when he or she does not have citizenship of an EEA country. Finally, the differences were introduced between EU citizens as the result of enlargement – the majority of "old" 15 Member States established restrictions for free movement for citizens of new Member States for a transition period of 3 or 5 years.

Thus, in order to project the total migration flows we recommend to disaggregate the available migration data into citizenship groups. As most of countries have important flows of migrants of some particular citizenship (most often the neighbour countries), data on these citizenships would be useful to analyse separately.

Recommended groups of country of citizenship:

NATIONALS

FOREIGNERS

of which :

- EU or EEA
- Non-EU or non-EEA
- most important partner countries (e.g. neighbouring countries).

Such grouping of countries may be useful also because data may be more easily available on some specific groups (i.e. non-EU or non-EEA foreigners who need residence permit to live in country). Figures hereunder show how the trends by groups of citizenship may be different (Figures 3, 4 and 5).

Figure 3 - Immigration by groups of country of citizenship for Lithuania (1999-2008)
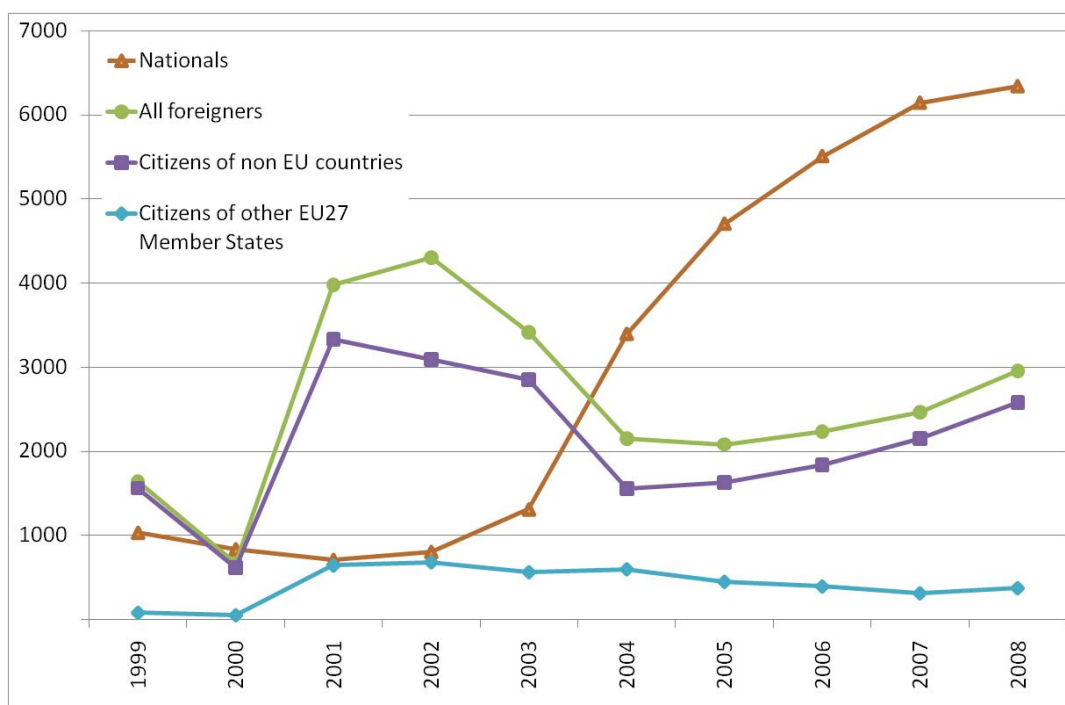**(Eurostat database, April 2010)**

Figure 4 - Immigration by groups of country of citizenship for the Netherlands (1999-2008)
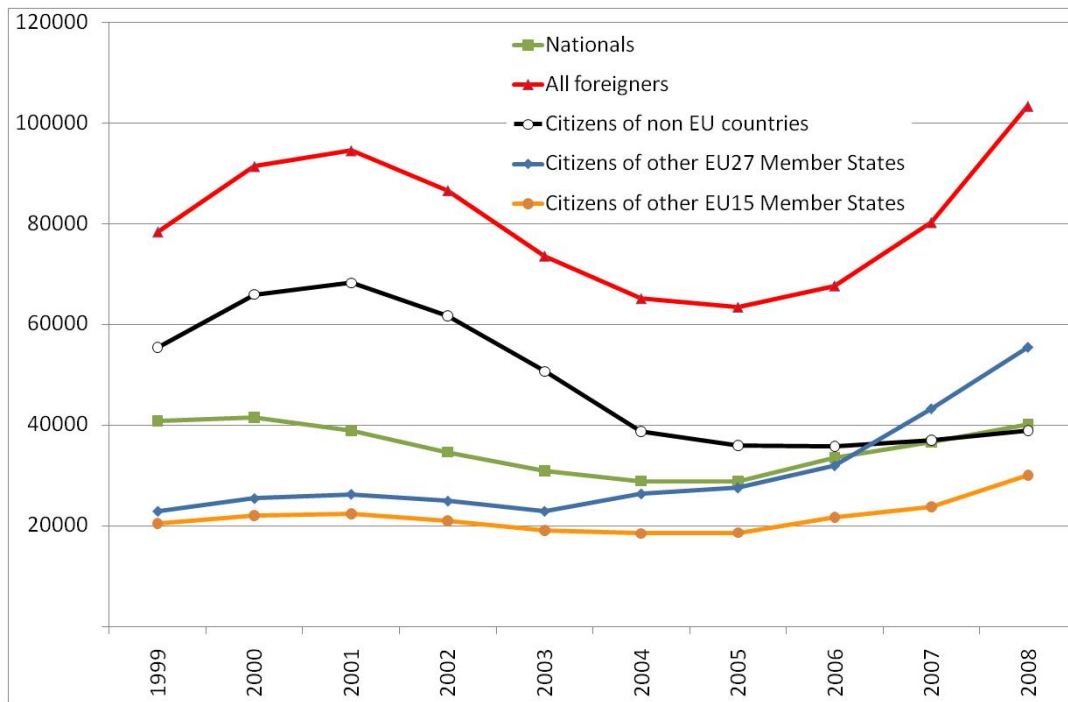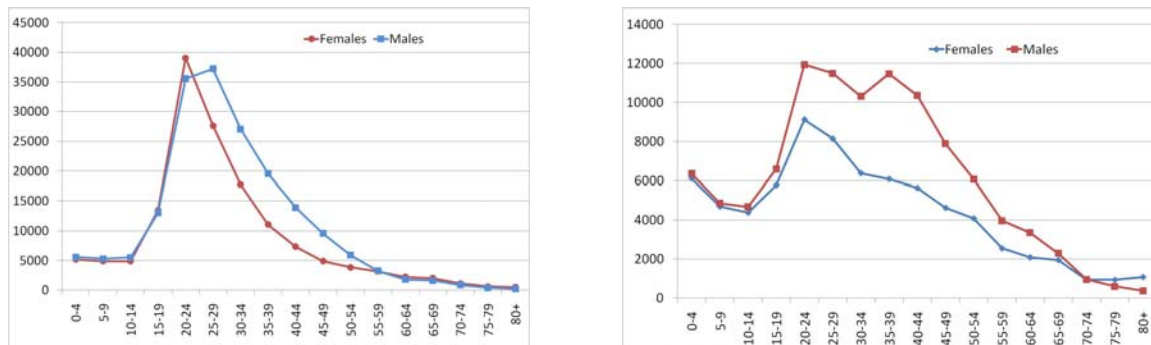**(Eurostat database, April 2010)**

Figure 5 - Immigration of citizens of non-EU countries (left) and nationals (right) in Germany, 2006
(Eurostat database, April 2010)

It may be argued that country of birth is a better characteristic than country of citizenship, particularly because the latter can be changed while the first is fixed. We consider that in the situation where people are more and more moving between countries, country of birth is not any more the indicator that identifies the most important bound between a person and a country. Moreover, in some situations it is difficult to establish the country of birth for time series (as well as for different countries) in a comparable way because the borders between several countries have changed during the last century. Nevertheless, analysing stocks by country of birth collected by census may help to validate the existing migration flows data. In addition when migrants have specific origin or destination country, it would be helpful to consider these flows separately in details.

4.   AVAILABILITY OF PAST IMMIGRATION AND EMIGRATION FLOWS

Eurostat website shows that almost all EEA countries and EU candidate countries have some data on international migration flows. Information given hereunder represents the situation with electronically available data at the beginning of April 2010 (Figure 6). At first, data were extracted from Eurostat database and thereafter checked and completed by adding information that was available on national websites, mainly at national statistical institutions. Availability of data on Eurostat database depends largely on the programme of the data collection by the Eurostat and that has been linked to the stages of the EU enlargement. One may assume that more data have been produced and published by countries during last decades but probably these are available only in publications that are not available by electronic means.

At the beginning of April 2010 the most complete data as far as totals on immigration flows are concerned were available for the year 2007 (Figure 6). From 34 observed European countries only 3 did not have these data on Eurostat or national websites. However with more and more countries become available, there are also more gaps in their data series. Migration flow data are not necessary more available on census year but seems that census results may have some influence on both, for starting publishing international migration data or in some cases, for stopping to publish.

Figure 6 - Availability of data on total number of international immigrations and emigrations
(Eurostat database and NSI websites, April 2010)

Note: Black cells indicate that total number of migrants is available for this year.
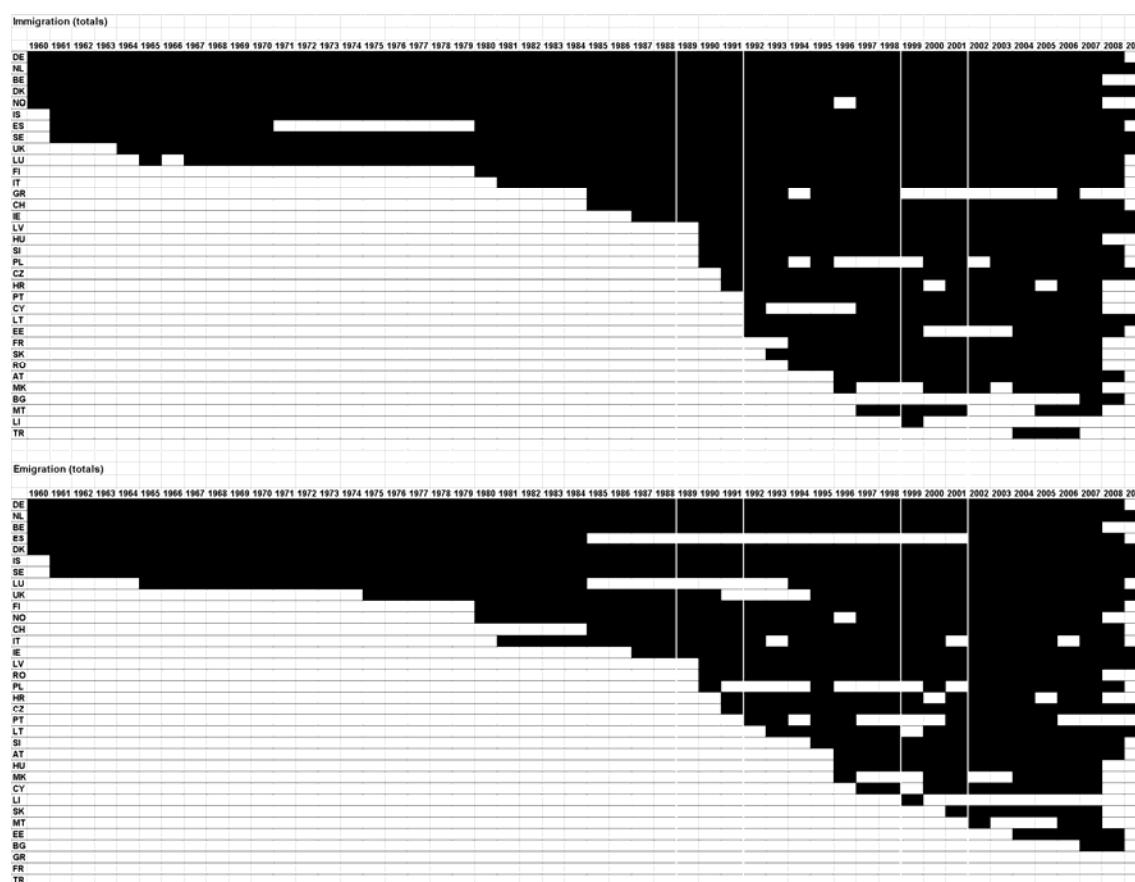
How long should be time series in order to identify a correct trend that can be extrapolated? As seen in Figure 6 migration data of several EU Member States (Belgium, Denmark, Germany, Netherlands, Norway, Spain and Sweden) are available in the Eurostat database from the 1960's. Assumedly these countries may have even longer data series in their own databases or paper publications and such series exist also in many other countries but not in Eurostat database.

For the projection purposes, however, such long trends are not necessary and even using these for projections is not recommended. Indeed, a very long time series (e.g. 30 years) may appear useless because it includes too

many historical changes influencing trends that will not probably occur any more in the future (Figure 7). Longer period means that more changes may have occurred in the administrative rules or statistical methods of the data collection. Moreover, it is usually not easy distinguishing in the migration trends the administrative or methodological changes and the actual changes that occurred long time ago. Series longer than 20 years will not necessary give additional value and help to do more clear assumption for future trend.

Figure 7 - Time series of immigration flows in selected countries, 1960 – 2009
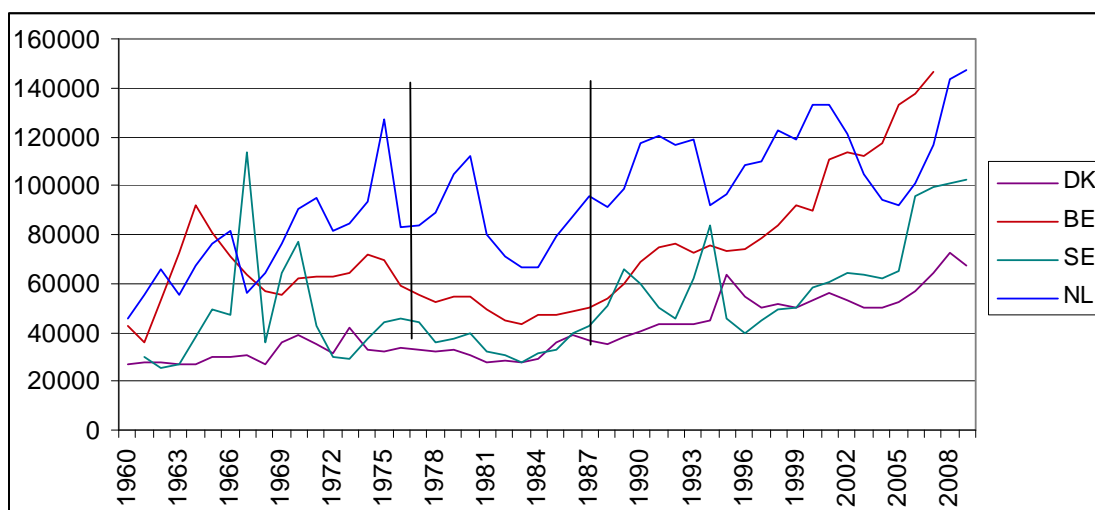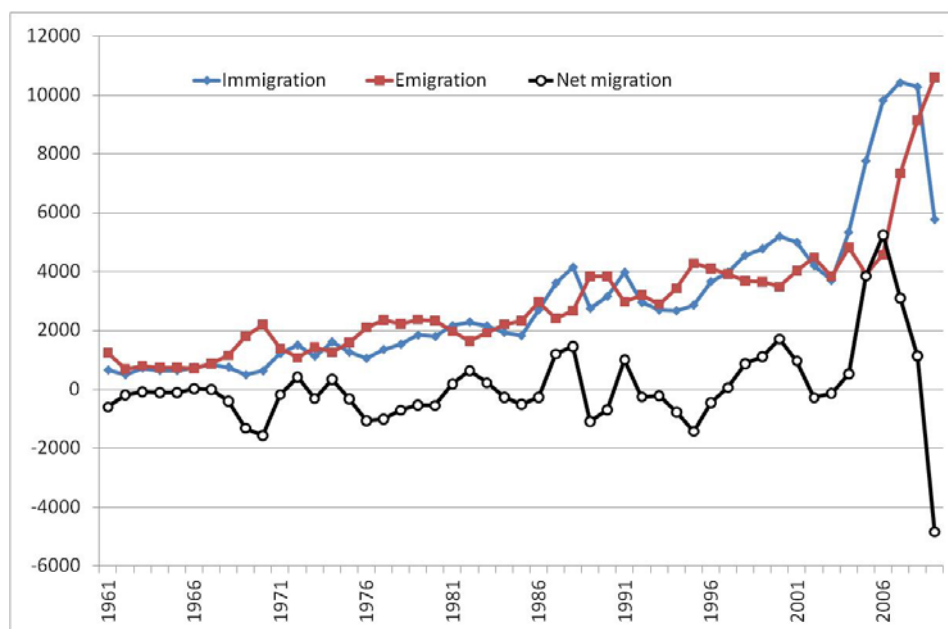**(Eurostat and NSIs websites, April 2010)**

Figure 8 - Total immigration flow, total emigration flow and net migration for Iceland
(Iceland Statistics and Eurostat database, April 2010)

Also, a too short time series (e.g. less than 5 years) is not appropriate as this reflects only recent temporary fluctuations in the levels while global trends may be hidden (Figure 8). It seems that a period covering 10 to 15 years would be the most suitable. However, the length of series of data to be used may vary somewhat depending on the country specific situation and migration history of its population. Looking on the availability of data by countries in the figure 6, one can see that majority of EEA countries may satisfy this requirement at least on the level on total flows. Some small gaps in series (lasting one or two years) would not be problem because both level and trend can be indentified based on available data.

5.   RELIABLILITY OF LEVELS AND TRENDS OF INTERNATIONAL MIGRATION FLOWS

The availability of data on time series is not all what is needed for doing assumptions concerning international migration flows for population projections. In order to introduce correct assumptions, one has also to ensure that observed trends and levels of immigration and emigration flows are reliable, meaning that they reflect the real situation in the migration considering all relevant population groups.

Levels of migration flows observed and reported in statistics depend from one hand on both statistical methods of data collection and availability of data from administrative sources, and from another hand, on the strength of administrative systems that provide these data. Because of the first two, some groups of migrants can be excluded from statistical data. For example, some countries do not include in the immigration flows data students who enter country for studies and in emigration flows who leave for studies. Some countries are not able to count national migrants. In such cases, estimations on these specific groups would be needed. In another case, where the data on migration flows are based on voluntary registration and deregistration of the place of residence in the country, the total migration flows may be undercounted even if all relevant groups are covered.

Generally countries are publishing migration data as they are observed, without adjusting for eliminating the under-coverage due to incomplete registration etc. Therefore, the observed level of flows can be underestimated compared to the actual migration flows. In opposite situation, where country legislation requires registration of all arrivals, including those for short-term stay, the level of migration flows may be theoretically overestimated. In order to assess the reliability of available data on international migration flows it would be useful to analyse census data (stocks by citizenship and country of birth, place of residence 1 or 5 years before census) in order to ensure that the levels are correct as well as the age and sex composition of migrants.

6.   COMPARABILITY OVER TIME , IMPACT OF EU REGULATION ON TRENDS

Data on migration trends may be biased because of (i) administrative or legal measures against migrants like changes in the registration and regularisation of illegal migrants, or (ii) the changes in the statistical methodology, i.e. using new data sources, definitions etc. In order to introduce correct assumptions on future migration in the projection, the real changes in the trends must be identified in the available statistics on immigration flows. The artificial changes or breaks in series can easily be taken by error as the change in migration trend. Examples shown in the Figure 9 and 10 represent changes in trends due to above mentioned reasons: Sweden introduced asylum seekers in the migration flows statistics in 2006 while Italy had increased immigration figures in 2003 and 2004 due to regularisation of illegal migrants. Spain have done several change in their methodology during recent years, including adding new categories of data in both immigration and emigration statistics (Table 1 and 2).

Figure 9 - Sweden: total immigration, total emigration and net migration, 1998 – 2009 (Eurostat and Statistics Sweden, April 2010)
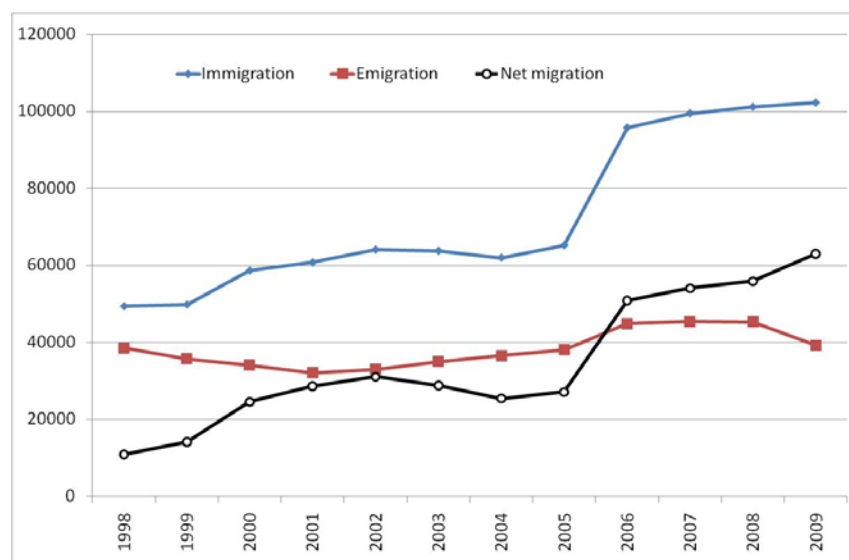
**Figure 10 - Italy: total immigration, total emigration and net migration, 1998 – 2005
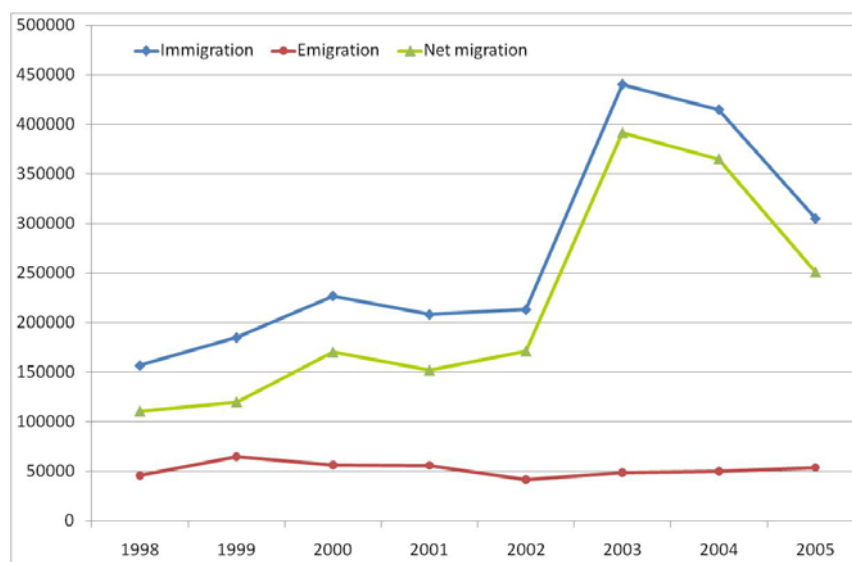(Eurostat database, April 2010)**

**Table 1 - Immigration flow in Spain by different groups of immigrants
that are included or not in statistics, 2000-2006 (INE, Spain)**

Years	National Immigrations	Foreign Immigrations with well known country of origin	Foreign Immigrations with unknown country of origin	Total number of Immigrations as available in official Statistics
2000	31 587	330 881	Not Included	362 468
2001	20 724	394 048	Not Included	414 772
2002	40 175	443 085	Not Included	483 260
2003	40 486	429 524	Not Included	470 010
2004	38 717	520 152	125 692	684 561
2005	36 573	578 736	103 975	719 284
2006	37 873	679 840	123 131	840 844
2007	37 732	807 908	112 626	958 266
2008	33 781	550 260	141 968	726 009

**Table 2 - Spain: Emigration flow for Spain by different groups of emigrants
that are included or not in statistics, 2002-2006 (INE, Spain)**

Years	National Emigrations	Foreign Emigrations with well known country of destination	Foreign Emigrations with unknown country of destination	Foreign Emigrations with unknown country of destination due to the 'expiry procedure'	Total number of Emigrations as available in official Statistics
2002	29 674	6 931	Not Included	Not Included	36 605
2003	15 990	9 969	Not Included	Not Included	25 959
2004	13 156	13 589	28 347	Not Included	55 092
2005	19 290	17 756	30 965	Not Included	68 011
2006	22 042	23 223	40 429	56 602	142 296
2007	28 091	29 630	34 659	134 685	227 065
2008	34 453	37 627	52 065	142 315	266 460

These changes were implemented because of the need to improve national statistics due to recently introduced Council Regulation on migration statistics. Accordingly, several EU Member States may show rapid changes in the trends because of need to apply definitions as requested by this regulation. Particularly, it concerns those countries which national definition on migrants was not in accordance with that internationally recommended (that is considering 12 months duration of stay for defining a migration event). These changes in data series are expected starting from 2009 data, while in 2010 all countries must produce the data according to harmonised definitions. Of course countries may continue to produce another set of migration data according to their "old" national definitions. Nevertheless, the user of data must make sure to choose comparable data over time.

Figure 11 - Change of definition of immigrant (adding temporary migrants) in Poland in 2008 (Eurostat database, April 2010)
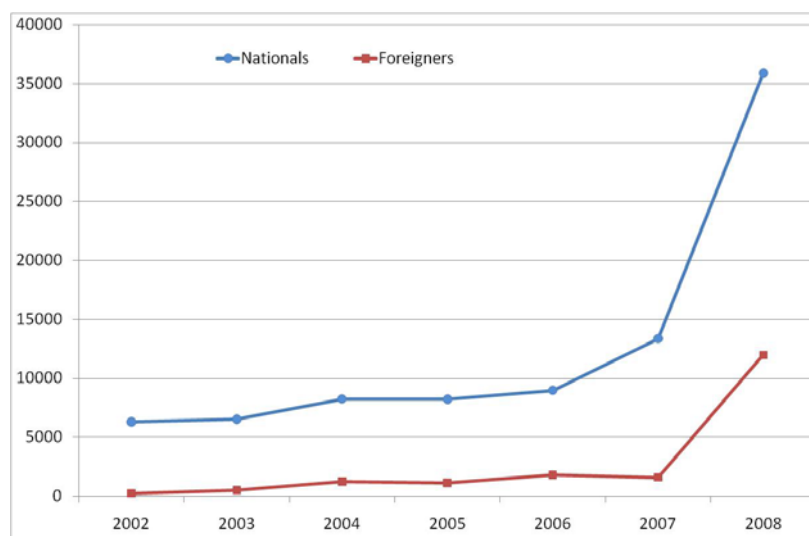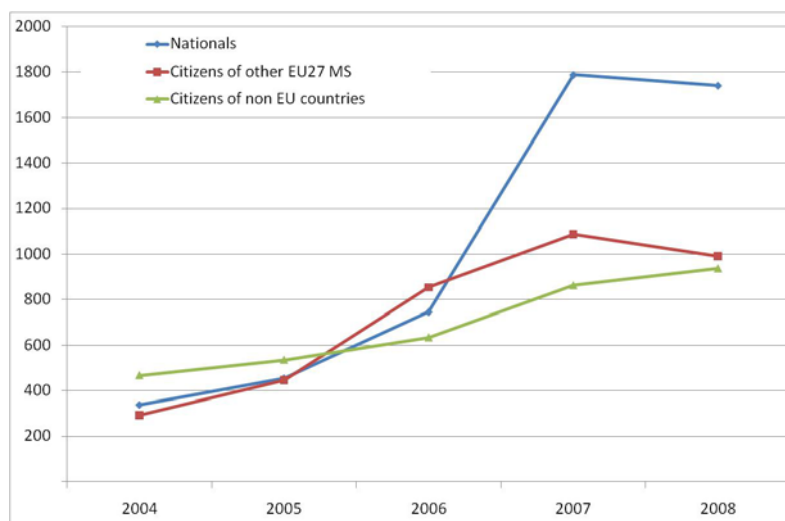

Figure 12 - Example of on-going improvements linked to the implementation of the EU Regulation : immigration in Estonia (2004-2008) (Eurostat database, April 2010

In many countries on-going improvements linked to the implementation of the EU Regulation have impact to the migration data already during several years and because of this it may be impossible to fix concrete year of the break in series. Here some examples are Poland where temporary international migrants were introduced in the migration flows statistics (Figure 11), and Estonia where the improvement of the population registration system, including the introduction the obligation to register the place of residence in the Population Register Act has been partly forced by the above mentioned regulation (Figure 12). The best example presenting the impact of the regulation is Denmark where two sets of data can be found for year 2009, one based on national definition of migrant and another based on definitions requested by the regulation (Figure 13).

**Figure 13 - Impact of the change of definition of immigrant in Denmark in 2008
(Eurostat website for the EU definition and Denmark Statistics for the national definition, April 2010)**
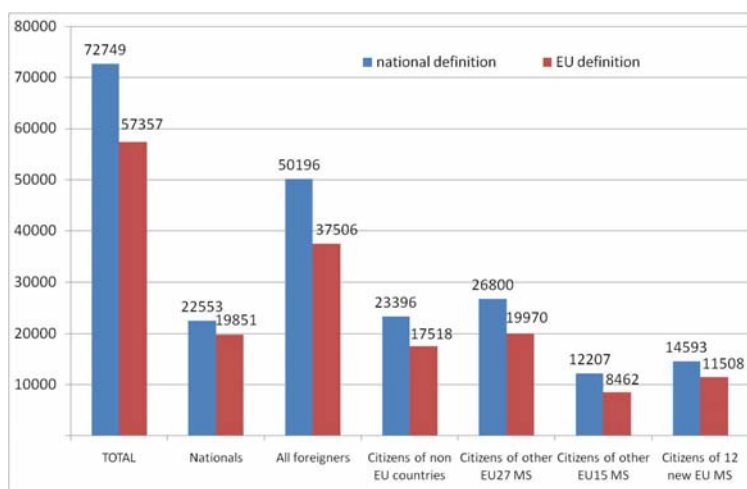
7.   COMPARABILITY BETWEEN COUNTRIES FOR MULTINATIONAL PROJECTION

International migration is a phenomenon that can be measured in two countries, in the country of origin (as emigration) and in the country of destination (as immigration). Therefore, as a referee, the data from other countries can be used for checking the reliability of levels and trends produced from national data sources for international migration data. It is particularly useful in case of intra-European (or intra-EEA) migrations where the conditions of free movement do not support the complete registration of migrants.

For assessing the reliability of the levels of migration observed in a country it would be useful to compare the flows data with the same flows observed by another (partner) country. The so-called "double entry migration matrices" have already a long history and have been used in order to assess the comparability of migration data on international level. Recently, adjustment factors were produced for immigration and emigration flows between EU Member States by an EU funded research project MIMOSA (http://mimosa.gedap.be/). These adjustment factors could be the best tool for correcting the difference in the level and overall underestimation of migration flows for EU Member States (Table 3). However these factors cannot be applied for non-EU citizens or non-EU origin and destination.

**Table 3 - Estimates of adjustment factors for immigration and emigration, selected EU Member States,
2002/2007 (MIMOSA,http://mimosa.gedap.be/Documents/Mimosa_2009b.pdf)**

Country	Immigration	Emigration
AT	1.06	1.74
IS	0.57	0.74
PL	17.85	10.64
CY	1.06	5.29
IT	1.42	2.92
SE	1.00	1.21
CZ	2.14	3.33
LT	2.33	2.45
SI	5.18	2.71
DE	1.03	0.69
LU	5.65	2.43
SK	18.90	43.69
DK	0.74	0.80
LV	2.92	6.22
UK	1.21	1.18
ES	0.82	4.90
NL	0.97	1.25
FI	1.26	1.22
NO	0.84	1.19

Time series may show trends that reflect either real trends or are the consequence of changes in data collection. In order to check if a specific trend is a real one or not, it is possible to compare this trend with the trend that is presented for the same flow by the partner country. Because the different methodology introduced in the partner country may cause difference in the level of migration, absolute figures can be not compared, but the relative figures on trends would still give a good indication if the trend is real or not. Whenever these two trends are different, it would be needed to check if there is no change in the methodology of the data collection or in administrative rules.

8.   CONCLUSION

International migration is an important component to be taken into consideration in population projections and this is more specifically the case for short-term projections. Imputing international migration figures only through 'net migration' is not appropriate for several reasons largely detailed above. The separate use of both immigration and emigration flows is recommended. Sex and age distribution is needed and the distribution by 5 years groups up to age 80+ is the most suitable option. Model migration schedules may be used to estimate single year of age and figures for older ages. Considering separately migrants by group of country of citizenship is also profitable as levels and trends of migration flows for nationals, EEA citizens and non-EEA citizens may be largely different. Time series of 10 – 15 years should be the best compromise between too short and too long series but considering specific situation and migration history of the country is helpful. Assessing real levels and trends for international migration flows is also compulsory considering the weak quality of migration data. Therefore using data from other countries and some adjustment factors as those proposed by MIMOSA could be helpful. Fortunately, the Council Regulation on Community statistics of migration and international protection (2007) is expected to have a positive impact on the reliability and the harmonisation of international migration data and a substantial improvement is expected for data related to the year 2010.

9.   REFERENCES

Morrison P (1977), The functions and Dynamics of the migration process. In Internal migration: A Comparative perspective, edited by A. A. Brown and E. Neuberger, pp. 66-72. New York: Academic Press.

Rogers Andrei (1990), Requiem for the net migrant. In Geographical Analysis, Vol. 22, No. 4 (October 1990), pp. 283 – 300. Ohio State University Press.