Proceedings of Statistics Canada Symposium 2016
Growth in Statistical Information: Challenges and Benefits

Creating a longitudinal database based on linked administrative registers: An example

Philippe Wanner and Ilka Steiner[1]

Abstract

This paper describes the creation of a database developed in Switzerland to analyze migration and the structural integration of the foreign national population. The database is created from various registers (register of residents, social insurance, unemployment) and surveys, and covers 15 years (1998 to 2013). Information on migration status and socioeconomic characteristics is also available for nearly 4 million foreign nationals who lived in Switzerland between 1998 and 2013. This database is the result of a collaboration between the Federal Statistics Office and researchers from the National Center of Competence in Research (NCCR)–On the Move.

Key words: Linkage, population registers, migration, longitudinal approaches.

1. Introduction

In the last few decades, international migration has played a key role in the population growth of many industrialized countries and has sparked numerous debates on the economic and social integration of migrants. Switzerland is no exception, as almost 25% of its population is made up of foreign nationals and one-third consists of persons born abroad (Wanner 2014). Primarily due to economic activity, this migration was dominated until the early seventies by the rotation of workers from neighbouring countries (mainly Italy), who had few skills and who would arrive in Switzerland on a temporary basis to work in industries such as construction, agriculture or hotel and food and beverage services. Since then, migrants have stayed longer, the reasons for migration flows have multiplied, and the countries of origin as well as the occupational characteristics of workers have diversified. This diversification of migration has sparked a need to better understand the lifestyles and integration of populations with more varied socioeconomic and cultural profiles.

However, the Swiss statistical system was caught off guard by the change in migration flows, and existing cross-sectional data have difficulty capturing the complexity of these flows and the integration processes. Central questions such as what portion of migrants stay or eventually leave Switzerland, or what proportion of migrants become naturalized, could not be answered with traditional statistics. Furthermore, information on the integration of migrants in the workforce could only be provided as of a specific date, but the process itself could not be studied.

In this context, a major project to examine the new forms of mobility and migration, NCCR–On the Move, funded by the Swiss National Science Foundation, was launched in 2014. Part of this project involved developing statistical tools and, specifically, a longitudinal database, to document contemporary migration and the lifestyles of migrants using a longitudinal approach. A priority of the statistical portion of this project was the use of existing data. This paper describes the steps involved in constructing this database and comments on the impact of these new statistics on knowledge of contemporary migration.

[1]Philippe Wanner, NCCR–On the Move and Institut de démographie et socioéconomie, Université de Genève, Pont d'Arve 40, 1211 Geneva 4 (philippe.wanner@unige.ch). Ilka Steiner, Institut de démographie et socioéconomie, Université de Genève, Pont d'Arve 40, 1211 Geneva 4.

2. Methodology

2.1 Population studied

The purpose of the project was to create a longitudinal, sociodemographic database using existing registers and statistics to describe the stay of the "migrant" population in Switzerland. This population is defined using the criterion of nationality and not place of birth for two reasons. First, German-speaking countries have always used the criterion of passport, which gives access to certain rights (in particular, the right to vote federally), while Anglo-Saxon countries have generally used place of birth. Second, and this flows from the first reason, statistics have long focused on the criterion of nationality. Thus, from 1981 to 2010, an automated register of foreign nationals (Central Register of Foreign Nationals) was available in Switzerland, but there was no register covering the population of Swiss nationals. As for measuring the migrant population in the strict sense of the word (place of birth criterion), no information was available, except for decennial censuses: persons born abroad but who had acquired Swiss nationality were essentially not counted at all.

However, since late December 2010 and the introduction of the harmonized register of residents, Switzerland has had an individual record of its entire resident population, both Swiss and foreign, and their migration status. This register of residents provides the basis for statistics on individuals and households and contains data on each person living permanently or temporarily in Switzerland.

The longitudinal database covers the period from 1998 to 2013, which is why we decided to use nationality as the indicator (imperfect) of migration status. Persons of foreign nationality living permanently in Switzerland are included for the entire period. Persons of Swiss nationality, which represents a reference group, are included only as of December 31, 2010. Since then, it has also been possible to distinguish migrants from non-migrants (based on place of birth).

Until 2009, it was only possible to track persons on an individual basis and not to identify households. Since 2010, with the introduction of the harmonized register of residents, which includes a household number, it has also been possible to reconstruct households (defined as persons living in the same dwelling). However, it is not possible to identify the relationships between the various members of a household so that, in the case of two persons living in the same household, for example, we cannot determine if they are a couple or two individuals with no conjugal relationship.

2.2 Legal basis

The work to construct this database was carried out under the Swiss *Federal Statistics Act* that makes it possible to match data for research purposes. Section 14a of the Act stipulates that, to carry out its statistical work, the office may match data provided that the data remains anonymous. If sensitive data are matched or if the matching of the data make it possible to establish individual profiles, the matched data must be erased once the statistical work has been completed. The matched data may be provided to researchers provided that a protection agreement is in place between the Federal Statistical Office (FSO) and the researcher.

The FSO enforces the *Statistics Act* in three ways. First, data matching is carried out in the FSO's offices. A work station has been made available to carry out these matchings onsite. Second, researchers do not have access to non-anonymous data or to the unique personal identifier (social security number) that makes certain matchings possible. A pseudo-identifier is created by the FSO from the social security number and made available to researchers. Lastly, each researcher or research group is able to access only the variables specifically required for their research, with a pseudo-identifier specific to each research project.

2.3 Registers and statistics included

The registers included were selected based on various criteria, one being the exhaustive nature of the data (with one exception). A second criterion involves the information provided by the registers, which must relate to the research topic. At this stage, the focus is on migration behaviours in particular, as determined by changes in status (arrival, departure, change of residence permit, naturalization) and changes in place of residence. As well, a variety of work has been done in the area of structural integration, measured in particular by position in the labour market (including the relationship between the occupation exercised and the occupation learned). However, there are plans to include other dimensions (for example, related to education or population behaviours) that would lead to other registers or statistics being included.

Table 2.3-1 presents the registers included. Three of them (central register of foreign nationals, register of asylum seekers and statistics on individuals and households, which are based essentially on the harmonized register of residents) provide data on the population living in Switzerland, and the other four on economic and occupational characteristics. The registers of individuals are structured as two annual files, one covering the population at the end of the year and the other covering the recorded movements of the population during a year (for example, an arrival or death). No one source is exhaustive (structural report) and its inclusion in the database is linked to the fact that this annual survey, which replaces the traditional exhaustive census that took place until 2000, provides data not found in the other registers; notably on occupation learned, occupation exercised, number of hours worked weekly, type of household, language spoken and primary language, religion or even commuting and use of public transit or private transportation.

Table 2.3-1
Registers included in the population database

	Characteristics	
	Population and period	Description
Central register of foreign nationals	Persons of foreign nationality, 1998 to 2010	Exhaustive record of foreign nationals with residence permits (as well as cross-border workers) at the end of the year and changes over a one-year period (migrations, births, deaths, status changes, etc.). This register was integrated in 2010 in the statistics on individuals and households.
Statistics on individuals and households	All resident persons, 2010 to 2013	Exhaustive record of persons living in Switzerland and movements (migrations, change of community, naturalization, etc.).
Register of asylum seekers	Asylum seekers and provisionally admitted persons, 1998 to 2010	Exhaustive record of persons in the asylum field at the end of the year and changes over a one-year period (migrations, births, deaths, status changes, etc.). This register was integrated in 2010 in the statistics of individuals and households.
Structural report	Sample of the population aged 15 and older living in Switzerland for more than 12 months, 2010 to 2013	Survey of family, economic, linguistic and mobility characteristics for a sample of more than 200,000 persons per year (Swiss and foreign nationals).
Individual accounts	Persons performing paid work in Switzerland, 1998 to 2013	Exhaustive record of work income subject to old age contributions. This exhaustive register provides the type of contribution (salaried, self-employed) and the amount of the wage.
Unemployment register	Persons registered as unemployed, 1998 to 2013	Exhaustive record of persons registering to seek work.
Social assistance statistics	Recipients of social benefits subject to resource conditions	Exhaustive record of social assistance recipients[2]

[2] At present, social assistance statistics are used only for a subsample of the population (asylum seekers, recognized refugees and provisionally admitted persons in Switzerland).

2.3 Matching work

The various registers were compared by matching using two distinct approaches.

First, the availability of a personal identifier in the different registers facilitates the linkage of several registers. This 13-character identifier number (known as the AVS13) is completely anonymous and was introduced at the end of the first decade of the 21st century. It replaces the former social security number (AVS) that was calculated using a simple algorithm based on the family name, person's sex and date of birth, and thus did not guarantee anonymity. The new identifier is available for anyone present in Switzerland in 2010 and after. Every foreign national arriving in Switzerland (immigration or by birth) is assigned a number.

The matching work also involved validation of this identifier, which may be problematic in some cases. Some individuals had mistakenly been assigned two AVS13 numbers and the changes to these numbers had to be taken into account. For other individuals, the FSO had no AVS13 number at the time the harmonized register of residents was validated and thus assigned a temporary number, which then had to be replaced later by the real number.

A different approach was needed to match the register of asylum seekers (AUPER) with other registers. This register contains persons currently going through the asylum process or provisionally admitted to Switzerland for security reasons. Once a residence permit is issued (following recognition of their refugee status, but sometimes also as a result of marriage), these individuals are removed from the AUPER register and reappear in the central register of foreign nationals. Consequently, it is crucial to have information on the asylum process to link the two registers.

This linkage was not possible for persons who obtained a residence or settlement permit prior to 2010 except through a match based on the availability in both registers of non-modifiable (i.e., date of birth, sex, nationality at time of arrival in Switzerland) or rarely modifiable (place of residence, civil status) variables. Various attempts were made to optimize the match by applying algorithms in the literature (e.g., Fellegi and Sunter 1969). Ultimately, given the special nature of the data and the sometimes mediocre quality of the data in the AUPER register (place of residence is often missing, nationalities are regularly corrected, etc.), a specific algorithm was programmed. This algorithm takes into account two characteristics in particular: national territories change and a given individual may see his nationality change over time (for example, a Yugoslavian takes on Serbian-Montenegrin nationality, then Serbian nationality). The next step was to consider specific matching rules. Certain records may be missing (e.g., a person who received refugee status in 2005 may appear late–in 2008–in the central register of foreign nationals): this situation requires matching attempts that allow for a time lag. The matching procedures used the SAS program and are documented elsewhere (Wanner et al. 2016).

3. Results

3.1 Trajectories

The results appear in the form of more than a hundred different annual files with a single pseudo-identifier for all registers. The files may cover the status of the population (e.g., 15 annual files from the register of foreign nationals, as of December 31 for the years 1997 to 2010) or movements (13 movement files from the register of foreign nationals for the years 2008 to 2010).

Reconstructing life trajectories becomes simply a matter of combining these different files and the different variables of interest to the researcher. Certain trajectories may involve a short-term migration (for example, an arrival and departure in a short period of time), while others may cover the entire period examined.

Nearly 10.64 million individuals were tracked for a period between one year and 15 years. This number is greater than the size of the Swiss population (8.2 million permanent residents in 2013) because migration involves the regular replacement of the migrant population, with certain migrants leaving and others arriving.

In total, 3,431,527 persons of foreign nationality were included at least once in the "status" file of the register of foreign nationals: 672,432 of them (19.6%) were tracked across the entire 1998-2013 period and thus recorded 17 times at year end between December 31, 1997 and December 31, 2013. In contrast, 415,700 persons (12.1%) were recorded at year end only once. Furthermore, some 7.2 million individuals of Swiss nationality or those who arrived in Switzerland after 2010 appear in the harmonized register of residents: some 84.6% of them were present during the entire 2010 to 2013 period. Given the number of reconstructed trajectories, a detailed analysis of the data is warranted.

3.2 Validation

While the sources used are considered of good quality, they still contain quality issues inherent to registers and the data must be validated before it is used. Special attention was given to the various sources of error mentioned by Roos and Nicol (1999), particularly duplicates or missing records (e.g., disappearance of individuals), contradictory information (e.g., an individual with a birth date later than the immigration date), late updates of elements that may change over time (e.g., civil status), and a lack of coherence in the various data sources (e.g., an individual of male sex in one source and female sex in another). Instances of these different sources of error were identified and corrected using specific criteria.

The availability of information on both the status of the population (at the end of each year) and movements (births, deaths, immigration, emigration, naturalizations, change of status, etc.) also made it possible to validate the reconstructed trajectories and to correct cases of undocumented appearances or disappearances.

Specific, systematic and documented correction rules were used to deal with identified errors (see Steiner and Wanner 2015).

3.3 Longitudinal analyses initiated

This population database has already been used in a number of ways, some of which have been published in specialized journals. These analyses adopt a longitudinal approach since the database allows tracking of migration trajectories.

One area of research focused on the factors impacting emigration of the migrant population in Switzerland, emigration that may be synonymous with a return to the country of origin or a departure to a third country (the two cases can be differentiated because of the information available in the harmonized register of residents). The results showed that factors related to poor structural integration (such as low employment income or not employed in a sector reflective of the skills acquired through training) increase the risk of return emigration, while factors related to success in the workforce (such as higher income than anticipated based on acquired training) increase the likelihood of leaving for a third country.

Another area of research focused on the pace of integration of the population in the asylum category (asylum seekers and refugees) from a longitudinal perspective over a period of observation of up to 15 years (Wanner et al. 2016). The matches made in this context used the variables measuring unemployment, use of social assistance and income levels. The analyses documented extremely variable levels of integration among the communities affected by asylum and also revealed the lengthy period of time required for the relative improvement of the socio-professional status of the asylum population.

In conclusion, the matches were carried out and the preparation of this population database resulted in a better understanding of different aspects of migration.

References

Fellegi, I. P. Sunter, A.P, (1969). « A Theory for Record Linkage. » *Journal of the American Statistical Association*, 64(328): p. 1183-1210.

Roos, L., Nicol, P., Cageorge S. (1987). « Using administrative data for longitudinal research: comparisons with primary data collection », *Journal of Chronic Diseases*, 40(1), p. 41-49.

Steiner, I., Wanner, P. (2015), Towards a new data set for the analysis of migration and integration in Switzerland. Working Paper no 1. Neuchâtel: NCCR On the Move,

Wanner, P. (2014), *Une Suisse à dix millions d'habitants. Enjeux et débats*, Lausanne: PPUR.

Wanner, P., Bertrand, A.L., Steiner, I. (2016), *Intégration structurelle et déqualification de la population réfugiée en Suisse*, rapport non publié, Genève: Université de Genève.