

30 September 2019

---

## United Nations Economic Commission for Europe

### Conference of European Statisticians

#### Work Session on Migration Statistics

Geneva, Switzerland

29-31 October 2019

Item 4 of the provisional agenda

#### Use of longitudinal data for migration statistics

## Reconciling administrative and survey data for measuring circular migration in Hungary

Note by the Hungarian Central Statistical Office\*

### *Summary*

Circular migration is often seen as a vehicle towards triple-win solutions of migration and development. Its measurement however is not without challenges. Since appropriate survey data is scarce, most attempts in official statistics use administrative sources from a longitudinal perspective to measure circularity and report on the drawbacks and uncertainties related to these sources. The Hungarian Central Statistical Office recently developed an administrative longitudinal migration database – using social insurance data – in order to get insights into the details of this phenomenon. This paper aims at identifying the weaknesses and assessing the quality of such database through comparing it with data from the Hungarian Microcensus 2016, a unique survey data source based on an unusually large sample of Hungarian households. Significant differences were found as regards the characteristics of the previous migration trajectory, especially when focusing on the foreign-born population. These and other differences as regards age, sex and place of residence were detected and potential explanations for the incongruences were discussed.

\*Prepared by Zoltan Csanyi, Anna Sára Ligeti

## I. Introduction

1. Technological, socio-economic and political transformations of the globalised world have changed global migration dynamics substantially in terms of extension, structure, individual motivations or behavioural strategies. Currently used concepts in official migration statistics were designed to measure immigration, emigration and/or return flows and stocks of migrant populations, suggesting that migration is a once-in-a-lifetime event. To measure processes of today's rather complex dynamics of repeated migrations, return or back and forth movements, the development of new approaches in official migration statistics is needed.
2. Longitudinal approaches emerged to analyse specific forms of multiple migration. Since circular migration has become a key policy instrument for managing migration as a "triple win" situation – which brings benefits for sending and receiving communities as well as for the individuals involved in migration processes – a growing demand for its measurement is observable. Policy-based approaches often juxtapose circular migration with traditional forms of migration, which have been mired in negative new connotations. While one-off migration involving long-term residence is often paired with such concepts as brain drain, population decline, labour force shortage or various difficulties of integration, circular migration appears as the exact opposite of these notions.
3. The Hungarian Central Statistical Office (HCSO) – with an enhanced use of administrative social insurance data – created an experimental, longitudinal data set that contains consecutive migratory events of individuals that makes possible to detect circular migration in emigration and return migration flows. This data set however suffers from weaknesses related to input data quality concerns. The present paper aims at identifying such weaknesses and assessing the quality of the longitudinal data set through comparing the resulting circular migration data with that produced on the basis of the Hungarian Microcensus 2016.
4. The paper is structured as follows: first longitudinal approaches in official migration statistics and conceptualizations of circular migration will be briefly reviewed; second, the experimental, longitudinal data set – based on administrative social insurance data – and Microcensus data will be described; third, this will be followed by a comparison of the circular migration data produced using both data sources to detect weaknesses of the administrative data set. The concluding remarks will close the paper.

## II. Longitudinal approaches and circular migration in official migration statistics

5. Longitudinal migration data refer to the attributes of the same respondents – individuals or households – observed in consecutive points in time, providing information on the duration of stay or the repetition of movements. That is, longitudinal approaches offer a dynamic view of migration behaviour and reveal migration trajectory over time. While traditional migration data most often refer to migrants as units of observation (e.g. flows and stocks of migrants in a given year), in longitudinal data sets migratory events in migrants' life become central (e.g. changes in the legal status of an immigrant).

6. As regards the data sources to be used for producing longitudinal migration data, panel surveys and administrative sources should be mentioned. When using the former, repeated data collection from the same respondents is carried out in different points of time. However panel surveys provide most detailed quantitative data allowing for deeper socio-economic analyses, they require disproportionate resources, while also having their own limitations (e.g. high respondent burden, panel attrition).
7. Longitudinal data production using administrative sources is less costly. Three main strategies of using administrative data in longitudinal approaches might be distinguished (UNECE 2018b). First, a single source is often used when the administrative data source in question allows for the association of migrants with each of their migratory events. Second, integrated administrative data – that requires the combination of at least two different data sources into a single dataset (UNECE 2017a) – might be needed when consecutive migratory events registered in different administrative sources can be linked to individual life cycles. Third, in some countries, population registers – that usually involves the integration of multiple sources – have become crucial for longitudinal approaches in official statistics.
8. While migration researchers offered several tentative definitions for circular migration, a working group, established by the Conference of European Statisticians (CES), adopted unified definitions of circular migration in official statistics for the first time in 2016 (UNECE 2016):
  - *General statistical definition*: A circular migrant is a person who has crossed the national borders of the reporting country at least 3 times over a 10-year period, each time with a duration of stay (abroad or in the country) of at least 12 months;
  - *Extended statistical definition*: A circular migrant is a person who has crossed the national borders of the reporting country at least 3 times over a 10-year period, each time with a duration of stay (abroad or in the country) of at least 90 days.
9. For measuring how the patterns of circular migration change over time, the CES working group recommended the following categories for circular migration flows applying either to immigration or emigration:
  1. Non-Circular immigration/emigration
    - 1.1. First immigration/emigration
    - 1.2. At least one previous immigration/emigration, but none in the last 10 years
  2. Circular immigration/emigration
    - 2.1. One previous immigration/emigration in the last 10 years (+any number earlier)
    - 2.2. Two or more immigrations/emigrations in the last 10 years (+any number earlier).

### III. Data sources

#### A. Longitudinal data set based on administrative data

10. A single-source, experimental longitudinal data set on the migration trajectory of Hungarian citizens was created by the HCSO using administrative social insurance data owned by the National Health Insurance Fund (NHIF). Please note, that this data set is appropriate for producing only migration flows data (its current use in official statistics is producing migration flows data of Hungarian citizens). For this reason, our intention was to detect circular migrants among the flows of Hungarian citizens leaving or entering/returning to Hungary. The following subgroups were used to identify circular migration flows:

- a. Hungarian citizens who suspended (that is temporarily invalidated) their social insurance,<sup>1</sup> or permanently settled down abroad (deregistered their Hungarian addresses)<sup>2</sup>;
  - b. Hungarian citizens who reactivated their social insurance after a period of being insured abroad.
11. Once these migrant subgroups were established in the input data source, the data set was re-structured and cleaned as follows. In constructing the longitudinal structure, migratory events and their attributes (starting/closing dates and destinations) were directly connected to individual IDs. Table 1 and Table 2 illustrate the original and the re-established, longitudinal data structures.

Table 1  
Original structure of the data set (illustration)

ID	Case	Starting date	Closing date	Destination	Migratory event
56	1	04.09.2012	11.10.2013	Austria	Return migration
56	2	01.02.2015		Germany	Emigration

Table 2  
Longitudinal structure of the data set (illustration)

ID	1. event Starting date	1. event Closing date	1. event Destination	2. event Starting date	2. event Closing date	2. event Destination	Status (last event)
56	04.09.2012	11.10.2013	Austria	01.02.2015	-	Germany	Emigrant

12. It is well-known that emigration/return migration flows data based on administrative sources underestimates real migration flows, since migrants often do not report on their leave/return to the authorities. The lack of (de)registrations at the social insurance register is one of the most important quality concerns of using this data set. Another part of the quality concerns related to the input data might be explained with the technical difficulties of data collection (e.g. lack of drop-down menus in data entry or missing validation rules), and that the administrative processes related to social insurances are not standardized across different migratory events (e.g. outmigration is often reported only when returning). In general, the administrative procedures are often too complex and related information is hard to access, further decreasing individuals' will and ability to correctly fulfil administrative obligations.
13. During the data cleaning process, first, duplicates were identified and deleted. Second, missing/contradictory values were imputed or adjusted. Third, migratory events were contrasted against definitions of circular migration, in order to correctly select the individuals under examination. First, a total of 136,663 events (21% of all) were detected as duplicates, including simple (that is repeated records – 96,916 cases) and assumed duplicates (when multiple, contradictory information was associated with the same events – 39,747 cases). In the latter cases, selection rules were specified: e.g. when in different data entries, different closing dates are associated to the same starting date of an individual's foreign

<sup>1</sup> In accordance with Article 80. § (5) of Act LXXXIII/1997 on Health Insurance the deregistration of emigrants is a legal obligation

<sup>2</sup> Data from the population and address register of the Ministry of Interior

insurance, we considered the earlier closing date as valid (and vice versa); or when in different data entries, different destination countries were associated to the same migratory event, the earlier filed case was deleted assuming that the later one was a correction of the earlier, incorrect data entry; etc.

14. Second, besides the duplicates, the most important quality concerns were 1) those related to missing data, 2) the inconsistencies of dates and 3) the ill-defined cases and categories. Based on these concerns 49,365 cases (9.8%) were adjusted. Determining the current migration status on the basis of the information available was also often difficult. While conclusions on the latest migratory event – that might be either emigration or return – can be drawn using the variable “current status of social insurance”, this variable and its categories are not defined precisely. We assume that current status of emigrants’ social insurance is a) Hungarian citizen with social insurance abroad; b) Cancellation or suspension of the insurance; or c) No longer insured. Based on this assumption, additional 65,432 return dates were imputed and 3401 return dates were deleted.
15. Third, further adjustments were needed in accordance with the definitions of circular migration as regards the duration of migratory events, since the NHIF system does not require minimum lengths of stay for migratory events. For the sake of simplicity (see also: UNECE 2018a) here we explain these data cleaning steps only as regards the general statistical definition, thus, how stays shorter than one year (abroad or in Hungary) were removed from the data set. Two main situations were considered (it should be noted that a similar process of data cleaning was also carried out applying the extended statistical definition):
  - c. If the only duration of staying abroad was less than one year: these individuals were not considered as migrants (since none of their migration experiences were considered as migratory events).
  - b. If the duration of staying in Hungary between two migratory events (in case there were more than one) was less than one year: the period of staying in Hungary was deleted and the durations of staying abroad were summed up (since the stay in Hungary was not considered as a return migration). Then summed up durations of stay were checked.
16. In contrast, adjustments related to the 10-year timeframe in the definitions of circular migration was not possible: the NHIF data set is relatively new, migration flows data can be reliably produced using this data source since only 2012. For this reason, a 5-year timeframe was applied in the exercise.

## **B. The Microcensus 2016 data set**

17. In 2016 October, with the aim of tracking social trends between full-scope censuses, the HCSO carried out Microcensus, a population survey based on an unusually large sample covering 10 per cent of the Hungarian households (for more details, see HCSO, 2018). Apart from the basic questionnaires on dwellings (N= 406.023) and personal information (N= 815.521), selected households were asked to fill in complementary questionnaires on specific topics, among which the International migration complementary data set (N=41.367) might be of interest.
18. Information on the actual place of residence as well as on a maximum of eight previous – short or long term – migration experiences abroad, together with the durations of staying abroad, is available in this data set. Nevertheless, since answering the questions of complementary questionnaires was optional, in some cases imputing the missing values was

necessary: e.g. missing data on the durations of staying abroad was estimated as the minimum duration of stay in a given category (i.e. 3 months or 12 months, in the case of short and long term migration experiences, respectively), or as the average values of respondents.

19. Despite – as it was explained previously in this paper – usually panel surveys are considered as the appropriate survey sources for the longitudinal approach, the international migration complementary Microcensus data set provides a unique opportunity to identify circular migrants. It does so – we should add – in a way that results would be comparable with the flows data produced using the longitudinal data set, described in the previous section.
20. Two specific features of the Microcensus data set should be taken into account however, a) that only the flows of immigrants / return migrants can be calculated reliably using this data source; and b) that Microcensus data was collected in October, 2016, that is, migration flows calculated for the year 2016 will be necessarily biased, since it only counts with the migrations occurring the first three quarters of that year.

#### IV. Comparing results

21. Having in mind the limitations of both data sources, in the exercise we focused on circular migration flows among returners in 2016 with migratory events between 2012 and 2016. It should be underlined that only Hungarian citizens were considered in the exercise, and – contrary to the proposed definitions of circular migration, a 5-year timeframe was applied in which previous migratory events were taken into account.
22. Results of comparing the circular migration data produced using the administrative longitudinal data set on the one hand and the Microcensus data set on the other, were expected to confirm the following presumptions:
  - a. Since administrative migration data contains information exclusively on the migratory events declared to the authorities, it is expected that circular migrations will be underestimated using the administrative longitudinal data set compared to Microcensus.
  - b. Despite incongruences across the data sets in absolute terms, we expect that the shares by socio-demographic variables of circular migrants would be similar.
  - c. It is further expected that comparing the socio-demographic characteristics (age, sex), as well as the migratory trajectories (destinations, durations of stay) of circular migrants, would reveal the migrant subgroups which are systematically underestimated the administrative longitudinal source.
23. In accordance with the main results (see Table 3 and Table 4), when using the general definition, the total number of returners were similar in absolute terms across both data sets (approximately 26 thousand and 23 thousand), while the number of circular migrants in the NHIF data set was significantly lower than in Microcensus (439 and 1,383 respectively). Their shares of the total return flows were also different across the two data sets: 1.7% compared to 6.1%. When using the extended definition, the results are quite different regarding both the total of return flows (approximately 35 thousand and 45 thousand) and the shares of circular migrants (7.3% and 26%).

Table 3  
Comparing results on return and circular migration flows in 2016 based on administrative longitudinal and Microcensus data – General definition

General definition	Administrative longitudinal		Microcensus	
	Non-Circular (return)	25 947	98.3%	21 381
Circular	439	1.7%	1 383	6.1%
Total	26 386	100%	22 764	100%

Table 4  
Comparing results on return and circular migration flows in 2016 based on administrative longitudinal and Microcensus data – Extended definition

Extended definition	Administrative longitudinal		Microcensus	
	Non-Circular (return)	32 393	92.7%	33 484
Circular	2 534	7.3%	11 781	26.0%
Total	34 927	100%	45 265	100%

24. Further details can be observed in Table 5 and Table 6: first immigrations (categories 1.1.1. and 1.1.2.) are more accurately measured, while the rest of the categories are underestimated by the administrative source. This might be explained by the fact a) that immigration data is generally of higher quality in administrative sources than emigration or return migration data; b) that the number of immigrants are often underestimated in surveys.
25. On the other hand, circular migrants with two or more immigrations in the last 5 years (category 2.2.) practically do not appear in the administrative longitudinal data set when using the general definition. Even combining both categories of circular migrants (categories 2.1. and 2.2.) their share is much lower than in Microcensus. In contrast, when using the extended definition, the category 2.2 becomes measurable also in the administrative data set, however their number still remains lower compared to Microcensus data.

Table 5  
Immigration flows to Hungary and circular migrants in 2016 – General Definition

Circularity (low-detail)	Circularity (high-detail)	Hungarian citizens			
		Born in Hungary		Foreign born	
		Administrative longitudinal	Microcensus	Administrative longitudinal	Microcensus
1. Non-Circular immigrations	1.1.1 First immigration to Hungary without emigration prior	-	-	13 600	3 462
	1.1.2 First immigration to Hungary, following a previous emigration	11 134	16 381	-	-
	1.2 At least one previous immigration, but none in the last 5 years		1 346	1 213	192
2. Circular immigrations	2.1 One previous immigration in the last 5 years (+any number earlier)	416	963	20	233

	2.2 Two or more immigrations in the last 5 years (+any number earlier)	3	187	0	0
TOTAL		11 553	18 877	14 833	3 887

<sup>a</sup>: Data calculated using the population and address register of the Ministry of Interior and the National Directorate-General for Aliens Policing

Table 6  
Immigration flows to Hungary and circular migrants in 2016 – Extended Definition

Circularity (low-detail)	Circularity (high-detail)	Hungarian citizens			
		Born in Hungary		Foreign born	
		Administrative longitudinal	Microcensus	Administrative longitudinal	Microcensus
1. Non-Circular immigrations	1.1.1 First immigration to Hungary without emigration prior	-	-	13600 <sup>a</sup>	2291
	1.1.2 First immigration to Hungary, following a previous emigration	17047	25690	-	-
	1.2 At least one previous immigration, but none in the last 5 years		4193	1746	1310
2. Circular immigrations	2.1 One previous immigration in the last 5 years (+any number earlier)	1819	7337	69	823
	2.2 Two or more immigrations in the last 5 years (+any number earlier)	630	3506	16	115
TOTAL		19496	40726	15431	4539

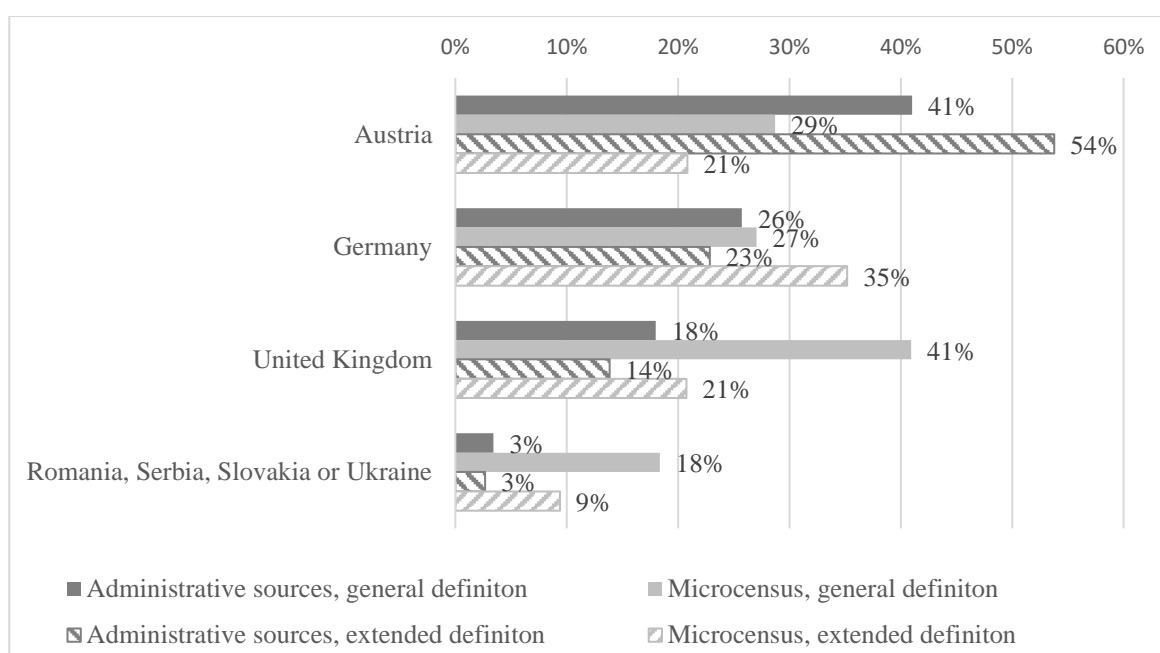
<sup>a</sup>: Data calculated using the population and address register of the Ministry of Interior and the National Directorate-General for Aliens Policing

26. Since data on circular migrants in both data sources showed significant differences, in our attempt to identify the weaknesses of the administrative longitudinal source, results were compared also along the main socio-demographic variables and characteristics of the migration trajectory.
27. However, the most common destination countries in circular migrants' trajectory were the same – Austria, Germany, the U.K. – their relative weights diverge across the data sources (Figure 1). Using the general definition, while in accordance with the administrative longitudinal data source, most circular migrants returned from Austria (41% of all circular migrants lived in this country for at least one year), the survey data suggests that the U.K. was the most popular destination country from which the Hungarians returned (41%). At the same time, if using the extended definition, Austria appears to be the most popular destination in the administrative longitudinal data set (54%), while it is Germany in the Microcensus (35%).
28. As it can be seen, the administrative source compared to Microcensus – using any of the two definitions, but especially in case of the extended definition – significantly overestimates the proportion of circular migrants returning from Austria. An explanation might be the differences in destination countries administrative systems potentially affecting also the (de)registrations in the Hungarian social security system. Previous research showed (Dickmann – Ligeti 2018) that Hungarians choosing these three destination countries differ by socio-demographic characteristics: the U.K. is chosen mostly in younger age groups



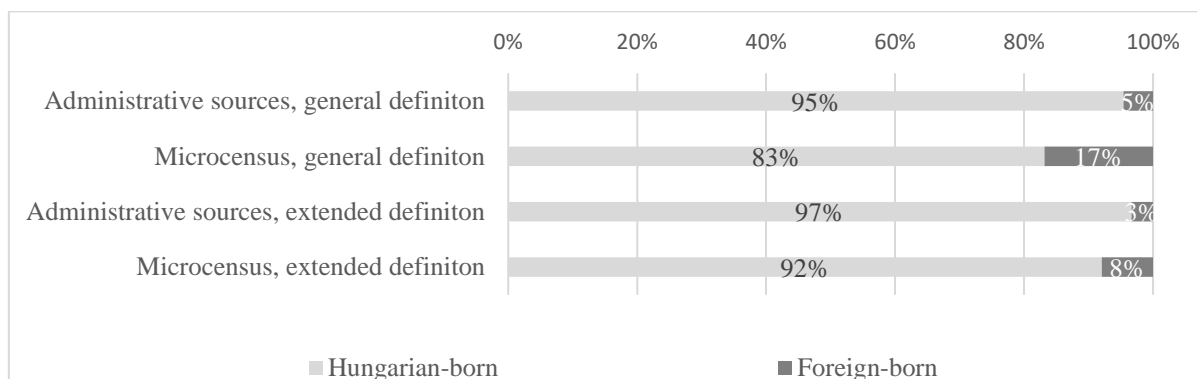
among those with at least secondary education who generally work in the services sector; while among those who migrate to Austria and Germany are typically men with lower education, working in industries and construction. These differences might be also related to emigrants' and returners' habits as for (de)registering their migratory events at Hungarian authorities.

Figure 1  
Circular immigrations to Hungary in 2016 by previous country of residence



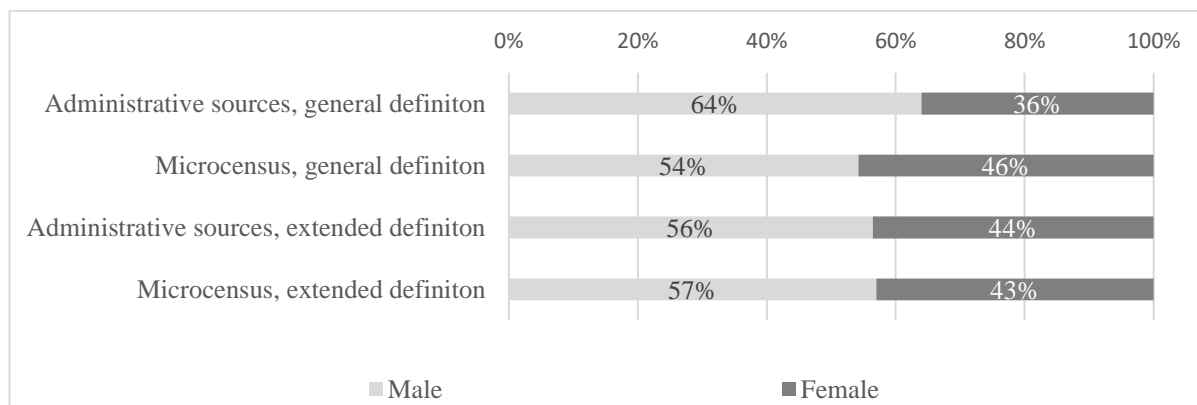
29. A significant difference is observable between the admin and survey data sources when focusing on the countries of birth of the foreign-born population (Figure 2). The foreign-born population in Hungary is typically ethnic Hungarian and the vast majority are born in one of the following four neighbouring countries: Romania, Serbia, Slovakia or Ukraine. In accordance with Microcensus data, the foreign-born population (about 5-6% of the total population) is overrepresented among circular migrants, especially when using the general definition (in this case, the share of circular migrants born abroad is 17%). This can be explained by the fact that circularity of foreign-born Hungarian citizens returning to the country requires a first immigration to Hungary, an outmigration, and a second immigration to Hungary (that is, a total of three migratory events are needed); while circularity among the Hungarian-born returners requires at least two outmigrations and two return migrations (that is, a total of four migratory events). In addition, this population easily meets the general definition's condition of staying at least 12 months abroad, since they might had lived in their country of birth for several years after their birth before moving first to Hungary.
30. On the other hand, administrative sources apparently do not capture foreign-born population properly. Using any of the two definitions, the share of circular migrants who lived in one of the four neighbouring countries is 3-3%, while the share of those born in those countries is 5% and 3% compared to the total population. This discrepancy is entirely due to shortcomings in administrative sources, and more precisely to that, in the case of foreign-born people, it is difficult to determine the date of their first immigration from social insurance data.

Figure 2  
Circular immigrations to Hungary in 2016 by country of birth



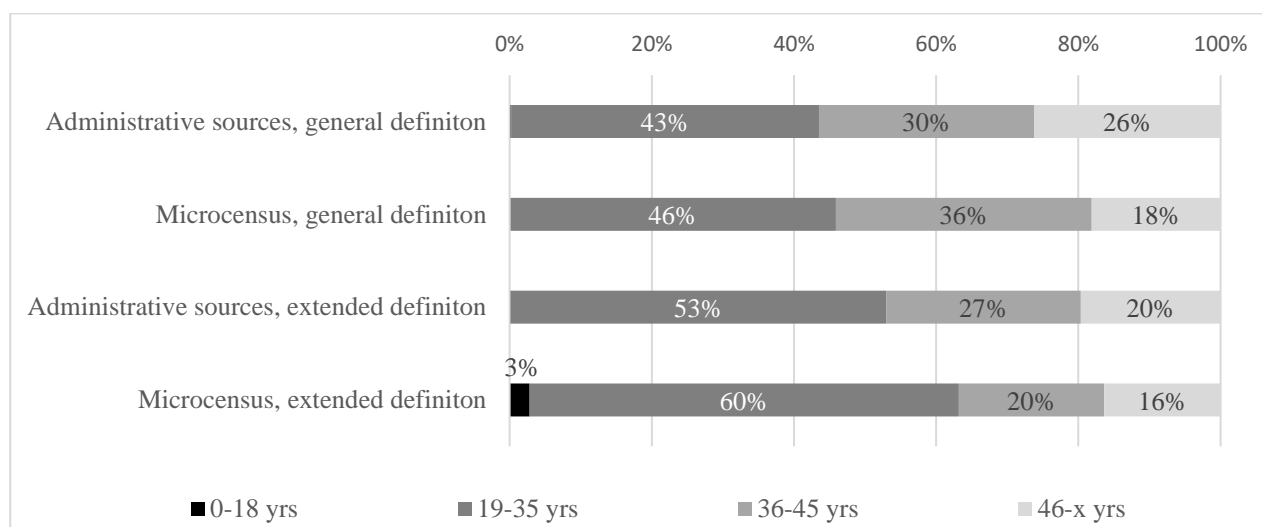
31. A great part of the differences as regards sex and age (Figure 3 and 4) is also due to the measurement problems mentioned above. Using the extended definition, the share of women and men is similar in both the administrative longitudinal data set and in Microcensus. However, the share of men is much higher when using the general definition. As the sex ratio is generally balanced in the foreign-born population, the absence of this group may be the reason for this imbalance. When using the extended definition however, the male surplus can be easily explained, since the short-term circular migrants are most likely to work abroad in the industries sector and constructions.

Figure 3  
Circular immigrations to Hungary in 2016 by sex



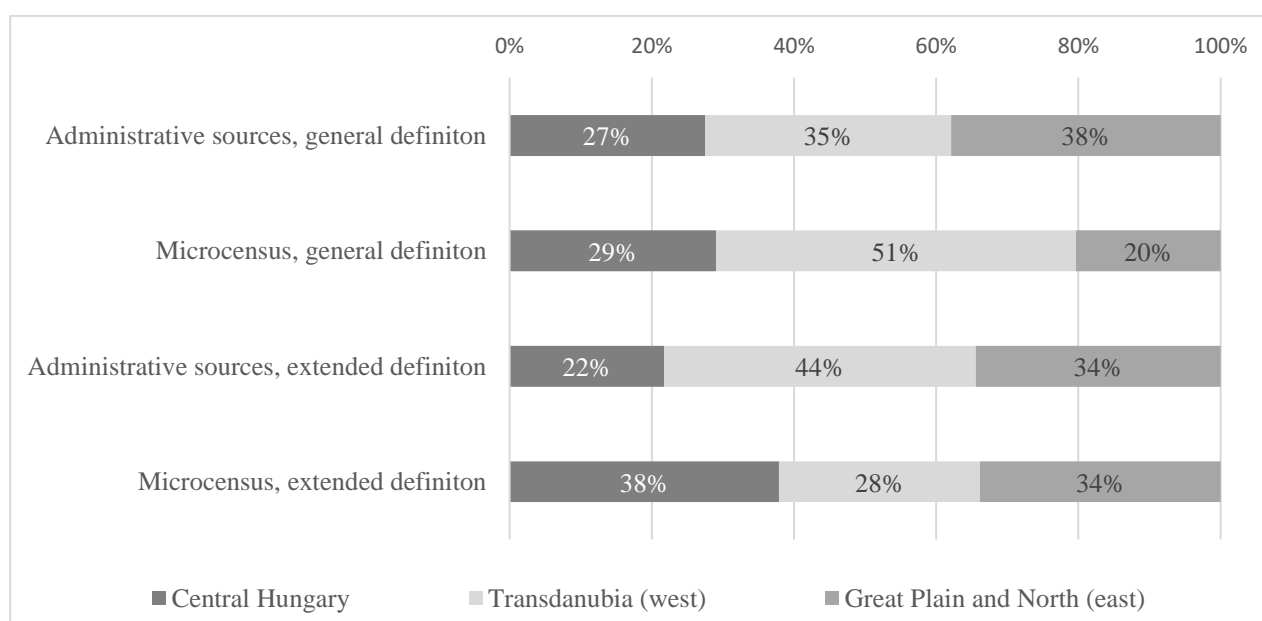
32. Along the age groups, children practically do not appear in the administrative source, only in Microcensus when using the extended definition. In contrast, the oldest age group – of those older than 45 – appear to be overrepresented in the administrative source compared to Microcensus data, probably due to that this group has a higher registration willingness.

Figure 4  
Circular immigrations to Hungary in 2016 by age groups



33. Finally, as regards the differences in regional distributions (Figure 5), these are most probably related to the geographical distances between Hungarian migrant-sending regions and destination countries. Migration to Austria is most common in the most Western, Transdanubian region. As a result, Western regions are overrepresented in the administrative longitudinal data set compared to Microcensus. While in case of the general definition, this Western predominance is at the Eastern regions expense, when using the extended definition, the weight of the Central Hungary is lower in the administrative source than in Microcensus.

Figure 5  
Circular immigrations to Hungary in 2016 by region



## V. Conclusions

34. Despite circular migration is being seen as a key policy instrument for triple-win solutions in migration and development, its measurement with statistical tools is not without challenges. The experimental longitudinal data set created by the HCSO on the basis of administrative social insurance data suffers from typical data quality concerns own of the administrative, input data set. Some of these quality issues were treated while creating the longitudinal data set.
35. In this paper, we compared circular migration data based on this experimental longitudinal data set on the one hand, and using Microcensus on the other, with the aim of detecting the weaknesses of the former. Having in mind the limitations of both data sources, in the exercise we focused on circular migration flows among returners in 2016 with previous migratory events between 2012 and 2016. When using the general definition, we found that the total number of returners were similar in absolute terms across both data sets, while the number of circular migrants in the administrative data set was significantly lower than in Microcensus. When using the extended definition, the results are quite different regarding both the total of return flows and the shares of circular migrants. Further, circular migrants from both sources were compared by the characteristics of the migration trajectory and by the main socio-demographic variables. Results show that while the most common destination countries in circular migrants' trajectory were the same – Austria, Germany, the U.K. – their relative weights diverge across the data sources. The administrative data set significantly overestimates the proportion of circular migrants returning from Austria.
36. Significant differences were observed between the data sets when focusing on the foreign-born population, especially in the analysis of the countries of birth. In accordance with Microcensus data, the foreign-born population is overrepresented among circular migrants, especially when using the general definition. On the other hand, administrative sources apparently do not capture foreign-born population properly.
37. A great part of the differences as regards sex, age, a place of residence is also due to the measurement problems mentioned above.

## VI. References

- Dickmann, Á. – Ligeti, A. S. [2018]: Mikrocensus 2016 – 10. Nemzetközi vándorlás. Központi Statisztikai Hivatal. Budapest.
- UNECE [2016]: Defining and Measuring Circular Migration. New York, Geneva.
- UNECE [2017a]: A Guide to Data Integration for Official Statistics. Geneva.
- UNECE [2017b]: A Set of Tables for Circular Migration. Working Paper. Geneva.
- UNECE [2018a]: Circular Migration: Outcomes of the Test on the Provisional Tables. Working Paper. Geneva.
- UNECE [2018b]: UNECE Task Force: Longitudinal Data for Migration Statistics. UNECE Work Session on Migration Statistics, Geneva, 24-26 October 2018.
-