

Population size estimation is one of the most challenging problems in official statistics (Wallgren and Wallgren, 2007). Information about the number of people living in a given country or region has an impact on the distribution of European funds, resources for health care or the GDP calculation. Furthermore, Eurostat's visions for the post-2021 census emphasise the idea of moving away from a traditional, decennial census to a rolling virtual census based solely on administrative data supplemented with sample surveys. The above mentioned actions are the result of a paradigm shift in statistics (and official statistics in particular), which can be summarised as the reuse of all available data instead of creating new ones (cf. Zhang, 2012). However, the use of these non-statistical data leads to fundamental statistical and economic questions. For instance, in an experimental study conducted by Statistics Poland (2020) and involving 9 registers, the count of the foreign population in Poland was 2.1 million as at 31.12.2019. But is this figure *the actual number* of foreigners staying in Poland? How is this estimate affected by non-sampling errors and how can we correct them to provide an unbiased estimate of the size of the foreign population? What is the uncertainty associated with this estimate? How does this estimate and the associated uncertainty change our knowledge of economic indicators such as the unemployment rate or GDP components? Is the information contained in registers sufficient to meet user needs or should it be supplemented with random and non-random samples? If so, in what way?

The project seeks to answer these questions by filling methodological and knowledge gaps. On the one hand, we propose new approaches to estimating the population size along with their characteristics; on the other hand, applying them with a view to estimating the number of foreigners in Poland will provide knowledge that will be useful beyond the 2021 census.

The main goal of the project is to develop methods for estimating the size of the foreign-born population in Poland and its characteristics using multiple data sources that contain potential errors. The aim of the project is twofold. On the one hand, we focus on estimating the size of the foreign-born population based on multiple, overlapping data sources that are likely to contain errors, such as administrative data. The main motivation behind this approach is to provide census-like statistics without conducting a traditional census. Secondly, characteristics of foreign-born populations included in administrative data are scarce and can be supplemented by non-probability samples, such as mobile big data derived from smartphones. The project aims to broaden the knowledge of the foreign-born population in Poland by using administrative data and other non-statistical data sources and developing sound and novel statistical methodology.

The novelty of the project consists in: 1) proposing new estimators that take into account different sources of non-sampling errors (e.g. over-coverage) and account for the resulting uncertainties, which propagate to economic indicators, such as GDP; 2) developing methods for estimating the irregular and illegally working population and proposing new estimators that account for misclassification to provide unbiased estimates of the foreign-born population characteristics derived from mobile big data; 3) developing methods of inference for non-probability samples based on selection models and non-ignorable non-response.

We plan to closely collaborate with Statistics Poland, use anonymised unit-level data from the police and aggregated data from the Polish Border Guard and the National Labour Inspectorate. In addition, we plan to use sample surveys conducted by the National Bank of Poland and unique mobile big data about behavioural and socio-demographic characteristics derived from thousands of foreigners smartphones (aggregated data).

To make our contribution more international, we have established a collaboration with experts in this field from the University of Southampton, Utrecht University, Istat and the Sapienza University of Rome, who will work with Polish researchers.

The results of the project should also contribute to producing more reliable estimates of the contribution of immigrants to the labour force and to GDP growth in Poland. Project results will be the basis for conducting statistical analyses on the socio-economic situation of Poland's regions and for projections of their development. We also note that estimators and results of our study are not only limited to migration statistics but will also have an impact on other fields, such as macroeconomics, economic policy or sociology.

References

- Statistics Poland (2020). The foreign population in Poland during the covid-19 pandemic.
- Wallgren, A. and Wallgren, B. (2007). *Register-based statistics: administrative data for statistical purposes*, volume 553. John Wiley & Sons.
- Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, 66(1):41–63.