VI Kongres Statystyki Polskiej – Abstrakt

Blok tematyczny: Badania statystyczne – metodologia i zastosowania

Tytuł (PL)

Erozja operatu PESEL. Strukturalna nielosowość niedostępności i jej konsekwencje dla wnioskowania o populacji Polski (2010–2025)

Autor / autorzy

Dominik Batorski, Uniwersytet Warszawski *(autor korespondencyjny)*

Zbigniew Marczewski, CBOS

CEL

Operat PESEL stanowi fundament probabilistycznych badań społecznych w Polsce, w tym kluczowych projektów statystyki publicznej oraz badań akademickich i międzynarodowych. Celem analizy jest empiryczna ocena dwóch warstw degradacji jego komponentu adresowego – braku adresu w rejestrze oraz nieaktualności adresu wśród rekordów formalnie kompletnych – a także odpowiedź na pytanie, w jakim stopniu wynikająca z nich niedostępność dorosłej populacji jest strukturalnie nielosowa i czy może być skutecznie korygowana standardowym ważeniem poststratyfikacyjnym.

METODY

Wykorzystano dane z dwunastu comiesięcznych badań omnibusowych CBOS realizowanych corocznie w latach 2010–2025. Próby losowane są w schemacie probabilistycznym bezpośrednio z rejestru PESEL i obejmują dorosłych mieszkańców Polski. Łączna liczebność prób brutto w analizowanym okresie wynosi około 540 tys. wylosowanych osób (średnio ok. 36 tys. rocznie). Wyniki uogólniane są na populację poprzez ważenie poststratyfikacyjne wg wieku, płci, wielkości miejscowości i regionu, w oparciu o bilanse ludności GUS oraz NSP 2021.

Niedostępność operacjonalizowano dwuwarstwowo: (1) brak adresu w wykazie udostępnianym do losowania – traktowany jako błąd pokrycia operatu; (2) niedostępność adresowa potwierdzona w realizacji terenowej – definiowana jako przypadki, w których ankieter ustalił brak rezydencji osoby wylosowanej pod adresem PESEL, a zaproszenie nie zostało skutecznie przekazane także w kanałach alternatywnych (CAWI, CATI, PAPI). Kategoria ta odróżnia się od chwilowego noncontactu i obejmuje przypadki strukturalnej niemożliwości dotarcia. Strukturę niedostępności analizowano w przekrojach wieku, płci, regionu i wielkości miejscowości; skalę zestawiono z liczbą ludności rezydującej publikowaną przez GUS. Oszacowano przedziały ufności dla proporcji oraz zbadano efektywność standardowej kalibracji w usuwaniu obciążenia szacunków.

WYNIKI

Łączna niedostępność dorosłej populacji w operacie PESEL wynosi obecnie ok. 20% i obejmuje ok. 6% rekordów bez adresu oraz ok. 14–15% rekordów z adresem, lecz nieprowadzącym do skutecznego kontaktu nawet w mix-mode. We wcześniejszych latach, przed wprowadzeniem kanałów CAWI i CATI w schematach mieszanych, odsetek całkowicie nieosiągalnych był wyższy – mix-mode częściowo, lecz nie w pełni rozwiązuje problem. Niedostępność jest silnie nielosowa: dotyczy częściej osób młodszych, mieszkańców dużych ośrodków oraz regionów o wysokiej mobilności. Standardowe ważenie poststratyfikacyjne tylko częściowo redukuje obciążenie szacunków.

WNIOSKI

Niedostępność w operacie PESEL przestała być problemem operacyjnym pojedynczych projektów badawczych – jest problemem systemowym statystyki publicznej oraz nauk społecznych w Polsce. Mix-mode redukuje jej wpływ, lecz nie eliminuje twardego rdzenia osób strukturalnie nieosiągalnych. Standardowa kalibracja ma wyczerpujące się granice, ponieważ niedostępni różnią się od dostępnych również wewnątrz strat ważenia. Kierunkiem koniecznym staje się rozwój operatów hybrydowych integrujących PESEL z innymi rejestrami administracyjnymi, zgodnie z dobrymi praktykami statystyki europejskiej.

Słowa kluczowe (PL)

operat PESEL; błąd pokrycia operatu; niedostępność adresowa; ważenie poststratyfikacyjne; jakość statystyki publicznej

6th Polish Statistics Congress – Abstract (English version)

Thematic block: Statistical surveys – methodology and applications

Title (EN)

Erosion of the PESEL Sampling Frame: Structural Non-randomness of Inaccessibility and Its Consequences for Inference about the Polish Population, 2010–2025

Author(s)

Dominik Batorski, Uniwersytet Warszawski *(autor korespondencyjny)*

Zbigniew Marczewski, CBOS

OBJECTIVE

The PESEL register provides the backbone of probability-based social research in Poland, including key projects of official statistics as well as academic and international surveys. The aim of this analysis is to empirically assess two layers of degradation of its address component – missing addresses in the register and outdated addresses among formally complete records – and to answer the question of the extent to which the resulting inaccessibility of the adult population is structurally non-random and whether it can be effectively corrected through standard post-stratification weighting.

METHODS

The analysis draws on data from twelve monthly omnibus surveys conducted by CBOS each year between 2010 and 2025. Samples are drawn through a probability scheme directly from the PESEL register and cover the adult population of Poland. The total gross sample size over the analysed period amounts to approximately 540,000 sampled individuals (around 36,000 per year on average). Results are generalised to the population through post-stratification weighting by age, sex, locality size, and region, based on Statistics Poland's current population estimates and the 2021 Polish Census.

Inaccessibility was operationalised in two layers: (1) missing address in the list released for sampling – treated as a frame coverage error; (2) address-based inaccessibility confirmed during fieldwork – defined as cases in which the interviewer established that the sampled person did not reside at the PESEL address, and the invitation could not be effectively delivered even through alternative channels (CAWI, CATI, PAPI). This category is distinct from temporary noncontact and captures cases of structural unreachability. The structure of inaccessibility was analysed by age, sex, region, and locality size; its scale was compared against the resident population figures published by Statistics Poland. Confidence intervals for proportions were estimated, and the effectiveness of standard calibration in removing bias from estimates was examined.

RESULTS

Total inaccessibility of the adult population through the PESEL frame currently stands at approximately 20%, comprising about 6% of records with no address and approximately 14–15% of records with an address that does not lead to effective contact even in mixed-mode designs. In earlier years, before CAWI and CATI channels were introduced into mixed-mode schemes, the proportion of fully unreachable persons was higher – mixed-mode mitigates the problem only partially, not fully. Inaccessibility is strongly non-random: it more often affects younger persons, residents of large urban centres, and regions of high mobility. Standard post-stratification weighting only partially reduces estimation bias.

CONCLUSIONS

Inaccessibility within the PESEL frame is no longer an operational issue of individual research projects – it is a systemic problem of official statistics and the social sciences in Poland. Mixed-mode designs reduce its impact but do not eliminate the hard core of structurally unreachable persons. Standard calibration is reaching its limits, since the inaccessible differ from the accessible also within weighting strata. The necessary direction of development is the integration of PESEL with other administrative registers, in line with established practices of European official statistics.

Keywords (EN)

PESEL register; sampling frame coverage error; address-based inaccessibility; post-stratification weighting; official statistics quality