Estymacja wielkości populacji uchodźców wojennych z Ukrainy w Polsce w ujęciu lokalnym

Główny Urząd Statystyczny
Dorota Szałtys

Urząd Statystyczny w Poznaniu

Uniwersytet Ekonomiczny w Poznaniu
Maciej Beręsewicz

Urząd Statystyczny w Olsztynie
Wojciech Wasilewski

MET2023

Sesja 7

Statystka ludności

2023-07-03 Warszawa

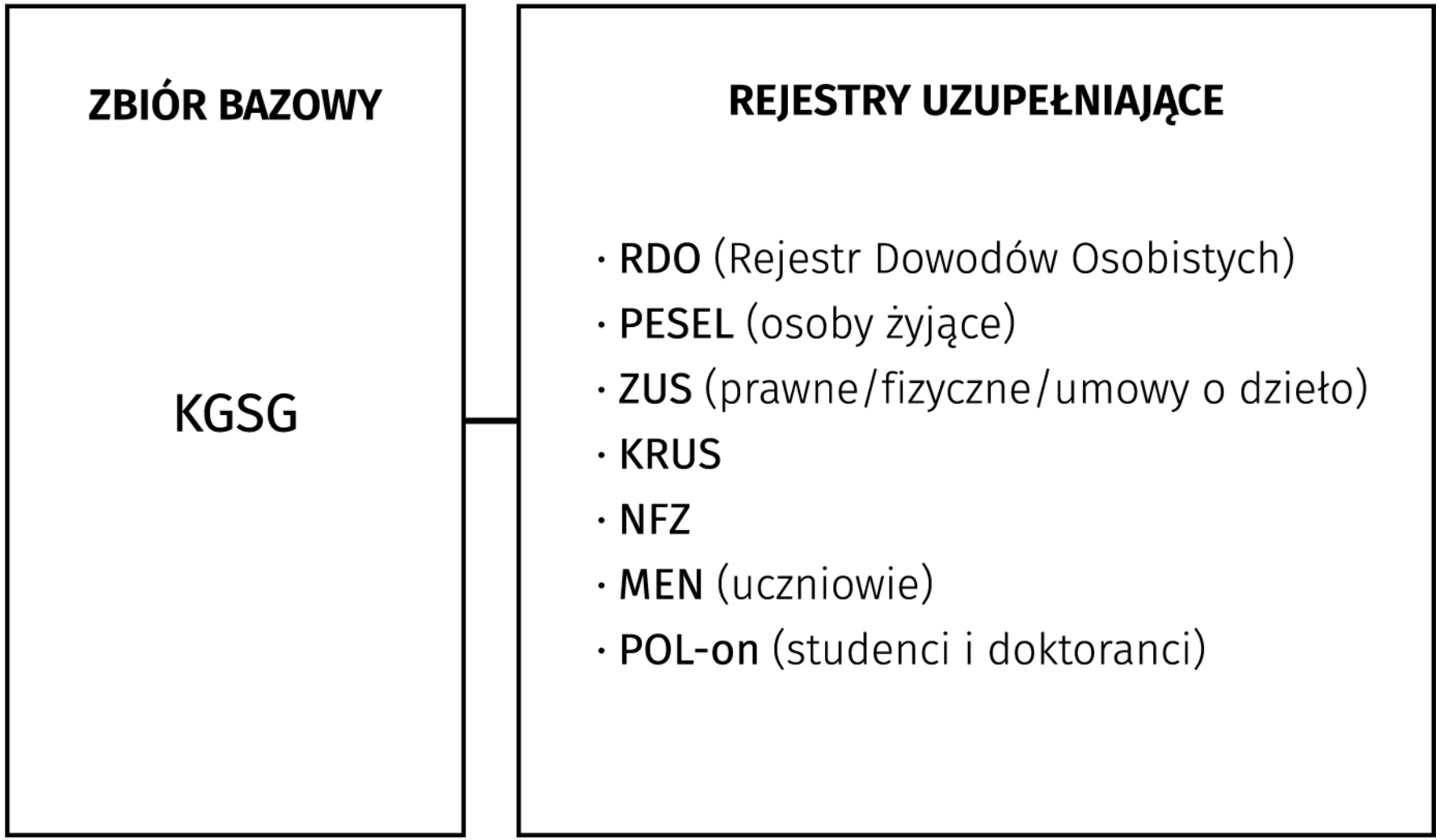
Plan prezentacji

- Wprowadzenie do problemu
- Integracja rejestrów administracyjnych
- Estymacja wielkości populacji na poziomie gmin
- Wyniki i podsumowanie

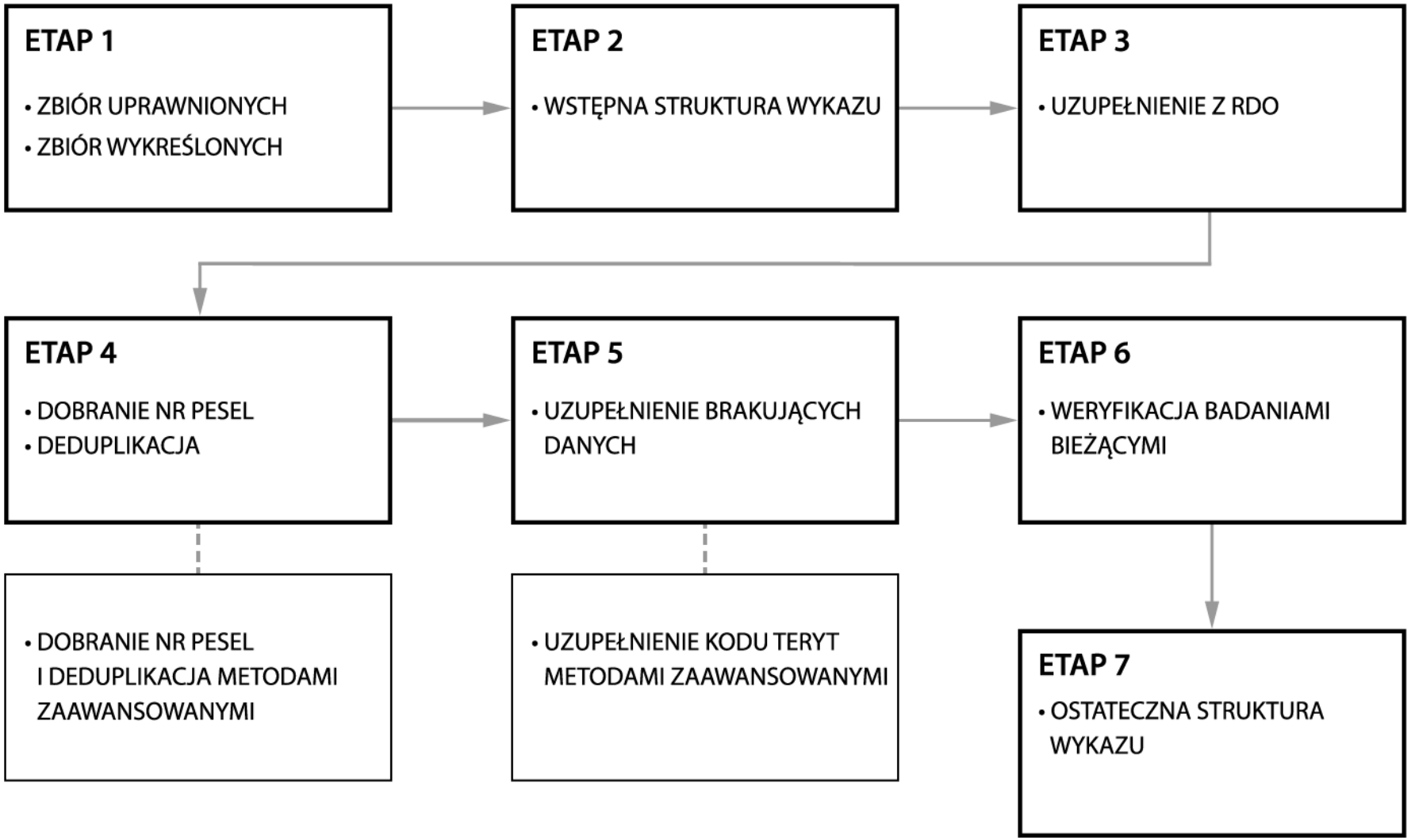
Wprowadzenie

- Potrzeba opracowywania i regularnego udostępniania przez statystykę publiczną danych statystycznych dotyczących mieszkańców Ukrainy przebywających na terytorium RP w związku z inwazją Federacji Rosyjskiej na Ukrainę (badania demograficzne, badania realizowane na próbach w gospodarstwach domowych, rachunki narodowe, badanie zdrowia).
- Uwzględnienie populacji w bilansie ludności rezydującej (w podziale na płeć, roczniki wieku i gminy imiennie) vs. sposób prezentacji danych
- Wyzwania metodologiczne – brak wiarygodnych i pełnych danych dotyczących miejsca zamieszkania.

Integracja danych



Integracja danych



Integracja – deduplikacja

Tablica 1 – Dane umowne będące wynikiem integracji danych

PESEL	Nazwisko	Imię	Data urodzenia	Nr dokumentu	PESEL opiekun	Nr dokumentu
22222222222	ZELENSKY	JULIA	20200901		3333333333	
	ZELENSKY	UJLIA	20200901			AA123456
22222222221	ZELENSKI	JULIA	20200501	AA123456		
	ZELENSKII	JULIA	20200901	AA123456 MAMA	3333333333	
	ZIELENSKY	JULIA	20200910	123456		

Procedura:

1. Blokowanie rekordów po współwystępowaniu numerów PESEL, nr dokumentów (nr dowodu, nr paszportu etc.)
2. Imputacja dedukcyjna oraz probabilistyczna nr PESEL
3. Deduplikacja rekordów

Integracja - deduplikacja

- Zbiór wejściowy – 1 859 509:
  - z PESEL-ami – 934 655 (50,26%)
  - bez PESEL-i – 924 854 (49,73%)
- Po probabilistycznej deduplikacji:
  - z PESEL-ami – 934 408 (51,36%)
  - bez PESEL-i – 885 057 (48,64%; ok. 40 tys.)
- Po probabilistycznej integracji ze zbiorem UKR:
  - z PESEL-ami – 952 103 (52,33%)
  - bez PESEL-i – 867 362 (47,67%)

Miejsce zamieszkania

Tablica 2 – Podstawowe informacje o danych adresowych

Posiada PESEL	Posiada adres rejestracji	Posiada adres zamieszkania	Liczba osób	Zgodność adresów
TAK	TAK	TAK	902 412	62,8%
		NIE	16 403	--
	NIE	TAK	18 885	--
		NIE	144 003	--
NIE	TAK	TAK	104	80,8%
		NIE	--	--
	NIE	TAK	13 011	--
		NIE	854 247	--

Źródło: opracowanie własne na podstawie zintegrowanych danych

Ustalenie miejsca zamieszkania

Tablica 3 – Źródło pochodzenia informacji o miejscu zamieszkania

Rejestr	Liczba	%
RDO	365 242	39,1
ZUS	269 594	31,7
MEN	143 873	15,4
NFZ	121 346	13,0
KRUS	4 379	0,5
PESEL	2 727	0,3
MRiPS	216	<0,1
POL-on	36	<0,1

Imputacja miejsca zamieszkania

- W przypadku danych z rejestru dowodów osobistych (RDO) informacja o miejscu zamieszkania jest taka sama jak miejsca rejestracji.
- W przypadku pozostałych rejestrów adresy w większości przypadków (51-78%) jest inny od adresu rejestracji.
- W związku z tym podjęto decyzję o imputacji miejsca pobytu w przypadku RDO stosując:
  - Metodę ***fractional counting*** (sumujemy *P(adres|.)*);
  - Imputujemy wektor prawdopodobieństw P(adres|.) na podstawie **wielomianowej regresji LASSO** (*glmnet*).

Miejsce zamieszkania (top 10)

Tablica 4 – Estymacja wielkości populacji UKR na poziomie gmin

Gmina	Oszacowanie punktowe	Odsetek z populacji
Warszawa (1465011)	89 150	9,54
Wrocław (0264011)	41 612	4,45
Kraków (1261011)	33 930	3,63
Poznań (3064011)	26 907	2,88
Łódź (1061011)	26 521	2,84
Gdańsk (2261011)	19 032	2,04
Szczecin (3262011)	17 939	1,92
Bydgoszcz (0461011)	11 885	1,27
Katowice (2469011)	10 401	1,11
Lublin (0663011)	9 004	0,96
Pozostałe	648 027	69,35

Źródło: opracowanie własne na podstawie zintegrowanych danych

Podsumowanie

- Okresy referencyjne ograniczają możliwości bieżącej statystyki dot. uchodźców.
- Jakość rejestrów administracyjnych znacząco wydłuża prace (błędy, duplikaty).
- Probabilistyczne łączenie rekordów umożliwiło deduplikację i integrację rejestrów.
- Przedstawione szacunki są obarczone błędami pokrycia, połączenia oraz imputacji przez co rekomendowane jest raportowanie przedziałów ufności dla liczby.

Główny Urząd Statystyczny
Dorota Szałtys

Urząd Statystyczny w Poznaniu

Uniwersytet Ekonomiczny w Poznaniu
Maciej Beręsewicz

Urząd Statystyczny w Olsztynie
Wojciech Wasilewski

stat.gov.pl

2023-07-03 Warszawa