

Budowa i ocena wykazu osobowo-adresowego wykorzystanego w NSP 2021

Dorota Szałtys
Grzegorz Gudaszewski
Departament Badań Demograficznych
Główny Urząd Statystyczny

MET2023
Sesja 6 Statystyka ludności

3 lipca 2023 r.

  Liczymy się
  **DLA POLSKI!**

Wykaz osobowo-adresowo-mieszkaniowy

1. Podstawa do zapewnienia kompletności badania oraz pokrycia populacji badanych jednostek.
2. Rola operacyjna - kluczowe narzędzie do zarządzania spisem, kontroli kompletności spisywanych osób, budynków i mieszkań, uwierzytelniania w samospisie internetowym, a także przepływu informacji o stopniu realizacji spisu pomiędzy kanałami gromadzenia danych.
3. Ważny element w procesie przetwarzania i opracowania wyników NSP 2021.

Źródła administracyjne wykorzystane w NSP 2021, w tym w budowie Wykazu

Dane administracyjne – wykorzystanych ok. 35 głównych rejestrów i systemów informacyjnych

w tym:

rejestr PESEL, systemy ZUS, KRUS, b. MRPiPS, NFZ, MF, b. MEN, b. MNiSW, UdSC, PFRON, PUP, EKSMOoN, EGiB, GUGiK

Inne – w tym:

Centralna Baza Danych Ksiąg Wieczystych, dane od operatorów publicznych sieci telekomunikacyjnych, dane od przedsiębiorców prowadzących sprzedaż i dystrybucję energii elektrycznej, dane od podmiotów prowadzących działalność w zakresie zbiorowego odprowadzania ścieków i zbiorowego zaopatrzenia w wodę, w zakresie dystrybucji lub obrotu paliwami gazowymi, w zakresie przesyłu i wytwarzania energii cieplnej, dla których odbiorcą końcowym dokonującym zakupu są gospodarstwa domowe.

Rejestry i bazy statystyczne – TERYT, BJS

Wykaz osobowo-adresowo-mieszkaniowy

Wykaz - to zbiory podstawowych danych dotyczących każdego rodzajów spisywanych jednostek. Tabele relacyjne: BUDYNKI, MIESZKANIA, OZZ, OSOBY oraz TELEFONY. Każda z tabel zawiera klucze łączenia z innymi tabelami wykazu.

Stanowi:

- operat do badania spisowego,
- źródło danych inicjalnych do zasilania systemu zarządzania spisem,
- źródło danych metodologiczno-organizacyjnych:
 - ✓ tzw. „flagi” – oznaczenia badanych jednostek, istotne z punktu widzenia zarządzania procesem zbierania danych od respondentów oraz ustalenia wstępnej wykazowej populacji docelowej.

Budowa – realizowana w trzech etapach:

- 1. Utworzenie najszerszego zbioru unikatowych adresów** (weryfikacja, aktualizacja i uzupełnienie przez gminy, aktualizacja w zakresie zmian w podziale terytorialnym kraju, nazewnictwie miejscowości oraz nazewnictwie ulic, jakie zostały wprowadzone wg stanu na dzień 1 stycznia 2021 r.).
- 2. Utworzenie operacyjnego wykazu osób** – według stanu na 31 grudnia 2020 (dostępność danych administracyjnych).
- 3. Integracja wykazu adresowo- mieszkaniowego z wykazem osób** -sprawdzenie zarówno jakości oraz poprawności ustalonych adresów, jak i adresów dołączonych do osób.

Unikatowy Numer Statystyczny (UNS) – klucz łączenia (zapewnienie unikatowości jednostek) → numer Pesel, inne klucze łączenia



(1) Scalenie wszystkich numerów UNS z wyselekcjonowanych zbiorów



(2) Weryfikacja szczególnych przypadków (oznaczenie flagą weryfikującą)



(3) Dołączenie imienia/imion i nazwiska



(4) Ustalenie i przypisanie adresu zamieszkania (referencyjność rejestru i aktualność adresu)

(1) Scalenie numerów UNS - wykorzystane źródła danych administracyjnych:

Zbiór	Gestor / źródło administracyjne	Zakres podmiotowy
PESEL	KPRM (b. Ministerstwo Cyfryzacji)/Powszechny Elektroniczny System Ewidencji Ludności	Osoby zameldowane na pobyt stały lub czasowy na terenie Polski, osoby, które wyemigrowały z Polski na pobyt stały
KEP	Ministerstwo Finansów/Centralny Rejestr Podmiotów – Krajowej Ewidencji Podatników	Osoby fizyczne (podatnicy) oraz osoby fizyczne prowadzące działalność gospodarczą (podatnicy)
ZUS	Zakład Ubezpieczeń Społecznych	Osoby fizyczne (Ubezpieczeni oraz płatnicy składek)
KRUS	Kasa Rolniczego Ubezpieczenia Społecznego/KRUSNAL	Rolnicy, ich małżonkowie i pracujący z nimi domownicy (ubezpieczeni) oraz osoby pobierające świadczenia rolnicze z KRUS
NFZ	Narodowy Fundusz Zdrowia/Centralny Wykaz Świadczeniobiorców	Osoby fizyczne (świadczeniobiorcy)
ARIMR	Agencja Restrukturyzacji i Modernizacji Rolnictwa/System informacyjny ARiMR	Osoby fizyczne - producenci rolni oraz posiadacze zwierząt gospodarskich: świnie, bydło, owce i kóz
PESEL_WYMELD	Powszechny Elektroniczny System Ewidencji Ludności	Osoby wymeldowane donikąd, które przed wymeldowaniem miały stałe miejsce pobytu w Polsce

(2) Weryfikacja rekordów (oznaczenie specjalnymi znacznikami/flagami) dotyczyła następujących przypadków:

- stałych mieszkańców innych krajów, zameldowanych na pobyt czasowy, którzy zgłosili wyjazd z Polski lub którym wygaś deklарowany okres pobytu w Polsce,
- osób mieszkających i pobierających świadczenia poza Polską – dotyczyło to głównie sytuacji, gdzie świadczenia pobierane są w Polsce, ale miejsce zamieszkania świadczeniobiorcy jest poza Polską oraz sytuacji, w których w adresie zamieszkania osoby w rejestrach pojawia się „adres zagraniczny” świadczący, iż dana osoba nie mieszka aktualnie w Polsce,
- osób, które wymeldowały się z pobytu stałego w związku z wyjazdem za granicę na pobyt stały,
- osób występujących tylko w jednym rejestrze,
- osób mających wyłącznie adres zamieszkania poza granicami Polski.

Przy tworzeniu wykazu osób brano pod uwagę **tzw. ślady życia** (ang. sign of administrative life) czyli potwierdzenie obecności osoby poprzez stwierdzenie występowania faktu aktualizacji informacji o tej osobie w co najmniej dwóch rejestrach.

(3) Dołączanie imienia/imion i nazwiska → personalizacja danych, uwierzytelnianie



Deduplikacja przypadków różnych imion i nazwisk dla tego samego nr Pesel



Sprawdzenie powtórzonych wartości w poszczególnych rejestrach administracyjnych i zdublowanych rekordów pomiędzy rejestrami

(4) Kolejność dołączania adresów

Źródło	Rodzaj adresu
KEP	adres zamieszkania (adres zameldowania)
ZUS	adres zamieszkania (adres zameldowania, korespondencji)
PESEL	adres zameldowania na pobyt czasowy
PESEL	adres zameldowania na pobyt stały
KRUS	adres zamieszkania (adres korespondencji)
ARiMR	adres zamieszkania (adres korespondencji)
NFZ	adres zamieszkania

3. Integracja wykazu osób z danymi adresowymi zaktualizowanej bazy NOBC - sprawdzenie zarówno jakości oraz poprawności ustalonych adresów, jak i adresów dołączonych do osób.

Aktualizacja wykazu osobowo-adresowo-mieszkańcowego na dzień krytyczny spisu

Ze względu na opóźnienie stanów wykazu osobowo-adresowo-mieszkańcowego w stosunku do momentu krytycznego badania, dane z wybranych źródeł administracyjnych, pozyskanych wg stanu na 31 marca 2021, posłużyły do zaktualizowania wykazu budynków, mieszkań i osób w zakresie faktów, jakie miały miejsce w okresie 01.01.2021 – 31.03.2021 rok dla zbioru:

- osób zmarłych,
- osób urodzonych,
- imigrantów zameldowanych w Polsce na pobyt stały,
- emigrantów wymeldowanych z pobytu stałego w Polsce w związku z wyjazdem za granicę na pobyt stały,
- budynków/mieszkań oddanych do użytku,
- ubytków budynków/mieszkań (m.in. rozbiórka, przeznaczone do wyburzenia).

Ocena wykazu osób w kontekście wyników spisu ludności

Podstawowa idea: ocena wykazu nie tylko jako operatu spisu, ale przede wszystkim – jako wstępnego wyniku w zakresie zasobów ludnościowych („wynik przed-spisowy”).

Cel: sprawdzenie możliwości oraz trafności wyznaczenia populacji ludnościowych jedynie na podstawie zastanych zasobów danych (zbiorów rejestrów) – tj. bez dedykowanego badania terenowego.

Sposób: poprzez porównanie wykazu osób z oszacowaniami NSP 2021 (spisowym zbiorem wyników).

Wyznaczenie ram porównania wykazu osób ze spisowym zbiorem wynikowym

- Operacyjny i techniczny sposób wykorzystania wykazu osób w ankietowym badaniu spisowym implikował brak ograniczeń w doborze jednostek (wszystkie unikatowe jednostki zidentyfikowane w aktualnych zbiorach rejestrowych);
tym samym – techniczny wykaz osób z założenia mógł cechować się nadmiernym pokryciem w stosunku do badanych populacji.
- Dlatego jednocześnie – podjęto próbę zastosowania w obrębie wykazu kryteriów wyznaczania w nim jakiejś zbiorowości odniesienia – poprzez utworzenie wektora cech-oznaczników (tzw. „flag”) zaliczająco-wykluczających.
- Na podstawie kombinacji flag wyznaczono z ogółu jednostek wykazu osób zbiorowość referencyjną, tzn. taką, która według najlepszej wiedzy przed spisem, miała znaleźć się również w zbiorze jednostek podlegających badaniu i zbadanych w NSP.
- Ramy porównań wykazu osób z wynikami spisu określa właśnie ta zbiorowość referencyjna – tzw. **„Populacja docelowa”**.

Populacja docelowa w spisie – w odniesieniu do ludności – definicje

W wymiarze teoretycznym:

Populacja docelowa – w najprostszym ujęciu – to zbiorowość osób, które z uwagi na przyjęty zakres przedmiotowy i podmiotowy badania spisowego podlegają spisowi (powinny zostać spisane).

W wymiarze praktycznym NSP 2021:

Zakres podmiotowy docelowej populacji osób wyznaczają trzy rodzaje niezależnie zdefiniowanych (sub)populacji, które częściowo są ze sobą wspólne, a częściowo rozłączne (suma logiczna), tzn.:

- 1) ludność według definicji krajowej,
- 2) ludność rezydująca,
- 3) imigranci.

Zatem populacja docelowa to łączny zbiór obejmujący wszystkie rodzaje badanych zbiorowości (kategorii) osób i dlatego **NIE należy jej utożsamiać z ludnością kraju** jako taką – rozumianą jako populacja zasadnicza.

Populacja docelowa w spisie – w odniesieniu do ludności – definicje (cd)

W odniesieniu do wykazu:

Wykazowa (wstępna) populacja osób - stanowi przybliżenie rzeczywistej (teoretycznej) populacji docelowej.

W wykazie do spisu (w obrębie wykazowej populacji docelowej) - nie wyodrębniono ww. szczegółowych populacji

Źródła danych dla wykazowej populacji docelowej to:

- wyłącznie na podstawie danych administracyjnych.

W odniesieniu do wyników spisu:

Oszacowana populacja osób - stanowi przybliżenie rzeczywistej (teoretycznej) populacji docelowej, dokonane jako ostatecznie opracowany wynik spisu (w zakresie stanów ludności), reprezentowany fizycznie w postaci rekordów danych tabeli osób, utworzonej w ramach Wynikowego Zbioru Danych Spisowych (WZDS).

Źródła danych dla oszacowanej populacji docelowej to:

- badanie ankietowe (kanały CAxI),
- dane z rejestrów administracyjnych.

Podstawowe pytania problemowe

1. W jakim stopniu wykazowa (wstępna) populacja docelowa pokrywa się z oszacowaną (wynikową) populacją docelową?
2. W jakim stopniu dane adresowe wykazowej populacji docelowej są zgodne z tymi jakie ustalono w obrębie oszacowanej populacji docelowej?

Zgodność pokrycia populacji docelowych - wykazowej i oszacowanej

Struktura oraz źródła danych o populacjach

Wykazowa populacja docelowa według zgodności z oszacowaną populacją docelową

Wyszczególnienie	Ogółem	W %
Wykazowa populacja docelowa – ogółem	39566532	100,00
Jednostki zgodne z oszacowaną populacją docelową	39014242	98,60
Jednostki niezgodne z oszacowaną populacją docelową	552290	1,40

Oszacowana populacja docelowa według zgodności z wykazową populacją docelową

Wyszczególnienie	Ogółem	W %
Oszacowana populacja docelowa – ogółem	39534391	100,00
Jednostki zgodne z wykazową populacją docelową	39014242	98,68
Jednostki niezgodne z wykazową populacją docelową	520149	1,32

Oszacowana populacja docelowa według kategorii ludności oraz źródeł danych

Rodzaj źródła danych (sposób spisania)	Ogółem	w tym: kategorie ludności			
		ludność wg definicji krajowej	ludność rezydująca	imigranci	
Ogółem	39534391	38036118	37019327	1433779	
		w %			
		100,00	100,00	100,00	100,00
Badanie ankietowe - razem	35134955	88,87	91,25	91,57	25,40
kanały CAxI - razem	35000342	88,53	90,91	91,23	25,00
CAWI	20386182	51,57	52,99	53,19	13,35
CAPI	9791267	24,77	25,63	25,58	2,41
CATI	4822893	12,20	12,28	12,46	9,24
badanie OZZ	134613	0,34	0,34	0,34	0,40
Dane z rejestrów	4399436	11,13	8,75	8,43	74,60

Wykazowa populacja docelowa według kategorii ludności oraz źródeł danych

Rodzaj źródła danych (sposób spisania)	Ogółem	w tym: kategorie ludności			
		ludność wg definicji krajowej	ludność rezydująca	imigranci	
Ogółem	39566532	?			
		w %			
	100,00	100,00	100,00	100,00	
Dane z rejestrów	39566532	100,00	100,00	100,00	100,00

Analiza danych adresowych

Wykazowa populacja docelowa według pochodzenia adresu

Źródło pochodzenia adresu	Ogółem	W %
Polska ogółem	39566532	100,00
KEP (Krajowa Ewidencja Podatkowa)	29696417	75,05
PESEL	8669874	21,91
ZUS	406159	1,03
NFZ	326486	0,83
KRUS	61955	0,16
ARiMR	731	0,00
Brak danych adresowych	404910	1,02

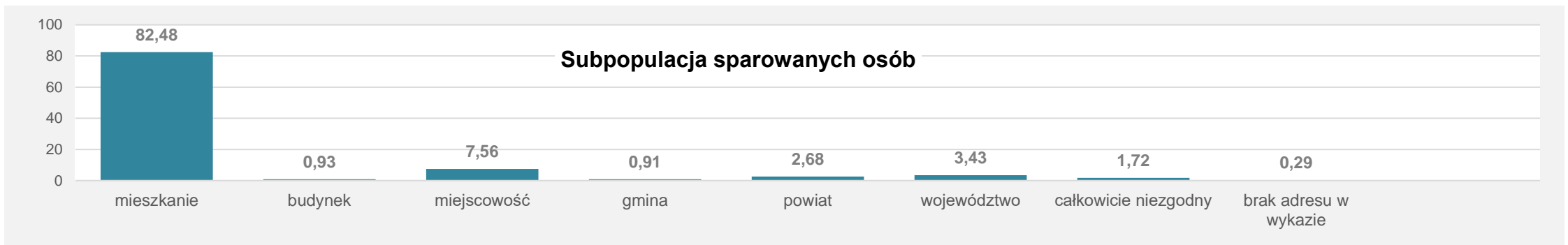
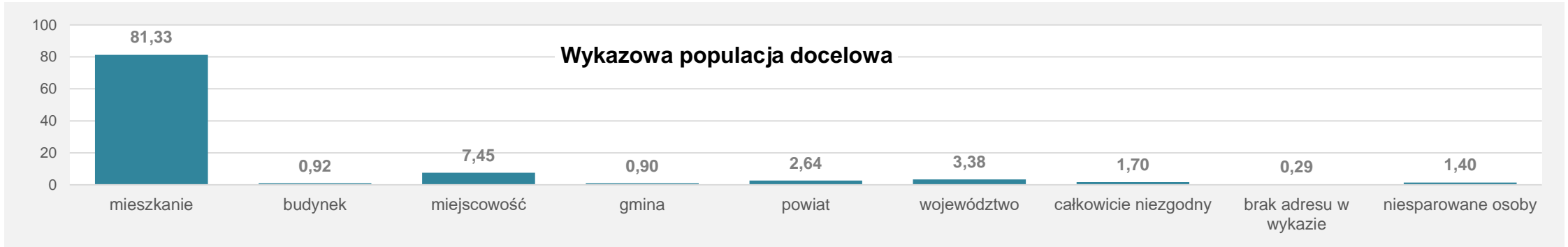
Wykazowa populacja docelowa według kompletności adresu

Kompletność adresu do poziomu:	Ogółem	W %	% skumulowany
Polska ogółem	39566532	100,00	x
Mieszkania	37957943	95,93	95,93
Budynku	188557	0,48	96,41
Ulicy	979470	2,48	98,89
Miejscowości	33264	0,08	98,97
Gminy	2388	0,01	98,98
Brak danych adresowych	404910	1,02	100,00

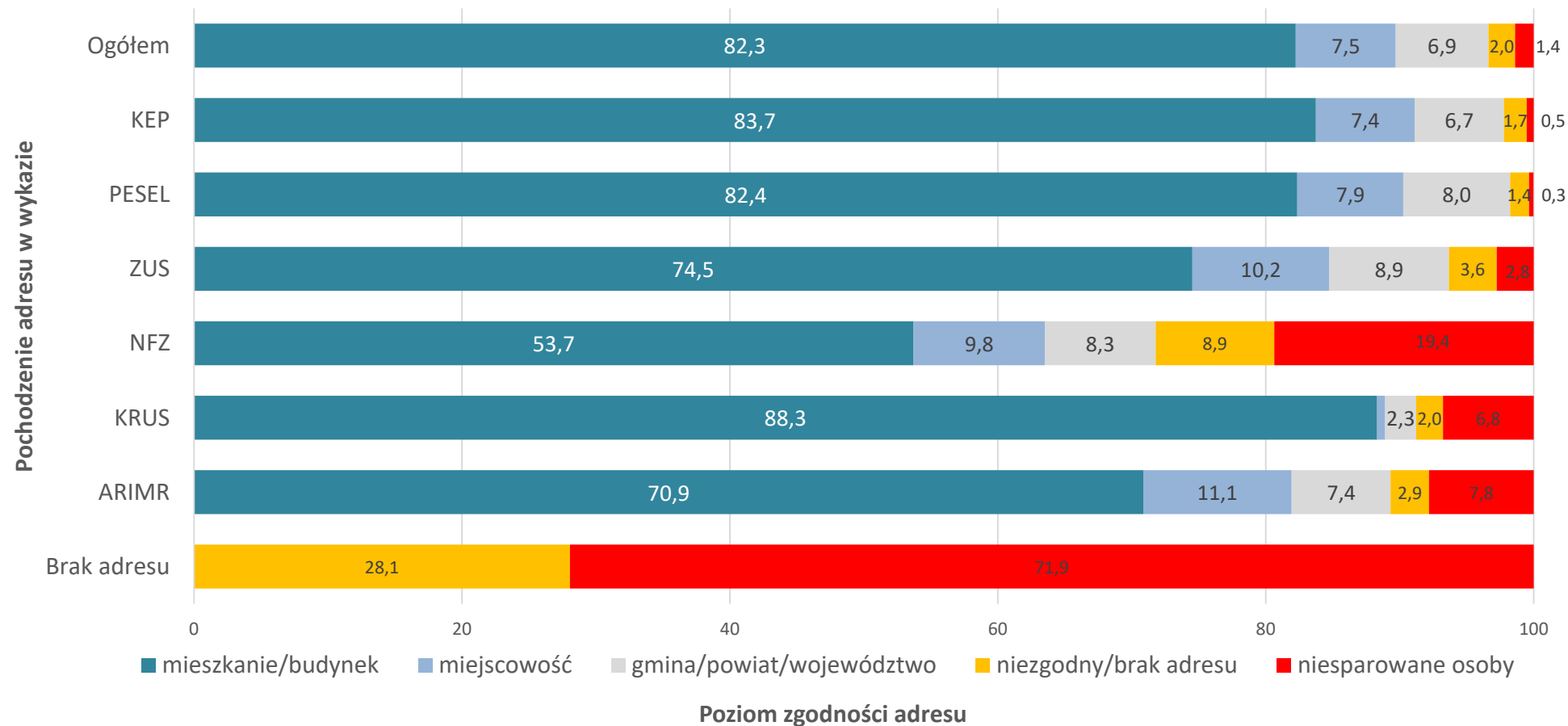
Wykazowa populacja docelowa według zgodności adresu z oszacowaną populacją docelową

Poziom zgodności adresu z oszacowaną populacją docelową	Ogółem	W %	% skumulowany
Ogółem	39566532	100,00	x
Mieszkanie	32179026	81,33	81,33
Budynek	362586	0,92	82,25
Miejscowość	2948110	7,45	89,70
Gmina	355923	0,90	90,60
Powiat	1046006	2,64	93,24
Województwo	1338262	3,38	96,62
Całkowicie niezgodny	670669	1,70	98,32
Brak adresu w wykazie	113660	0,29	98,60
Niesparowane osoby	552290	1,40	100,00

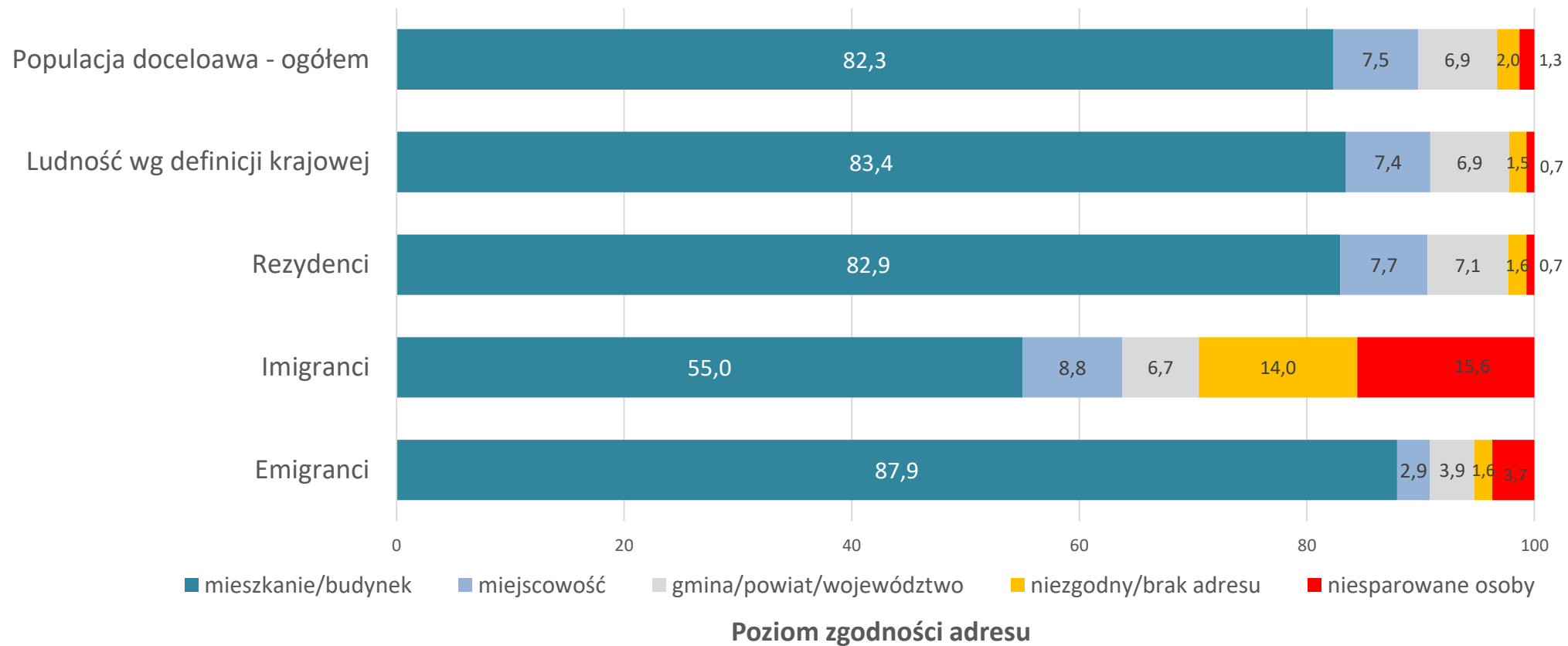
Poziom zgodności adresów wykazowych z wynikowymi z perspektywy różnych populacji



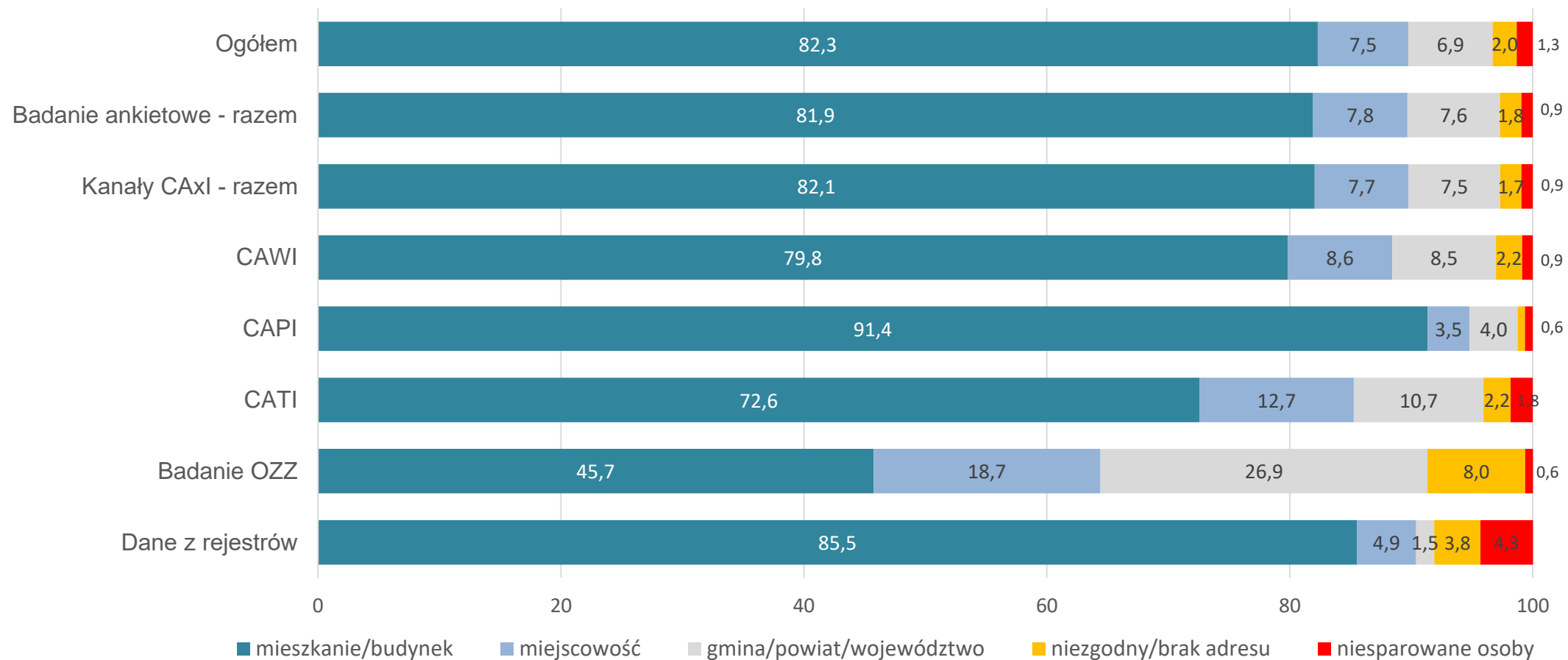
Wykazowa populacja docelowa według pochodzenia adresu oraz jego zgodności z adresem w oszacowanej populacji docelowej (w %)



Oszacowana populacja docelowa według kategorii ludności oraz zgodności adresu z wykazową populacją docelową (w %)

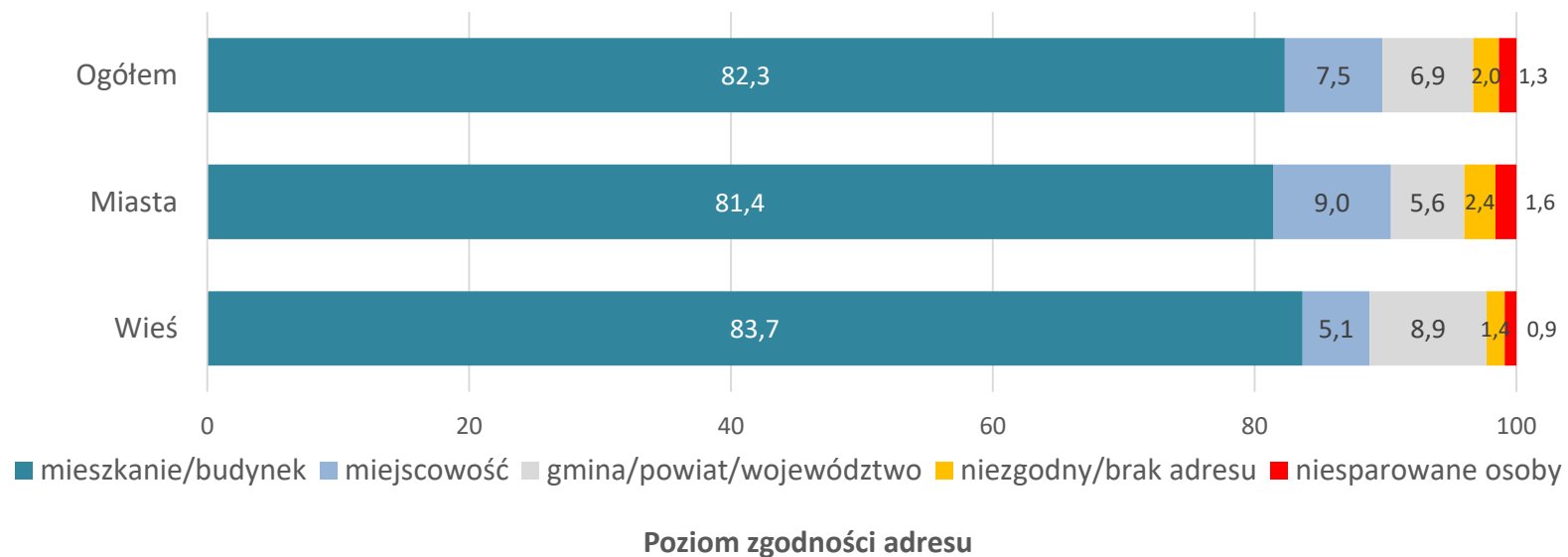


Oszacowana populacja docelowa według źródeł danych oraz zgodności adresu z wykazową populacją docelową (w %)

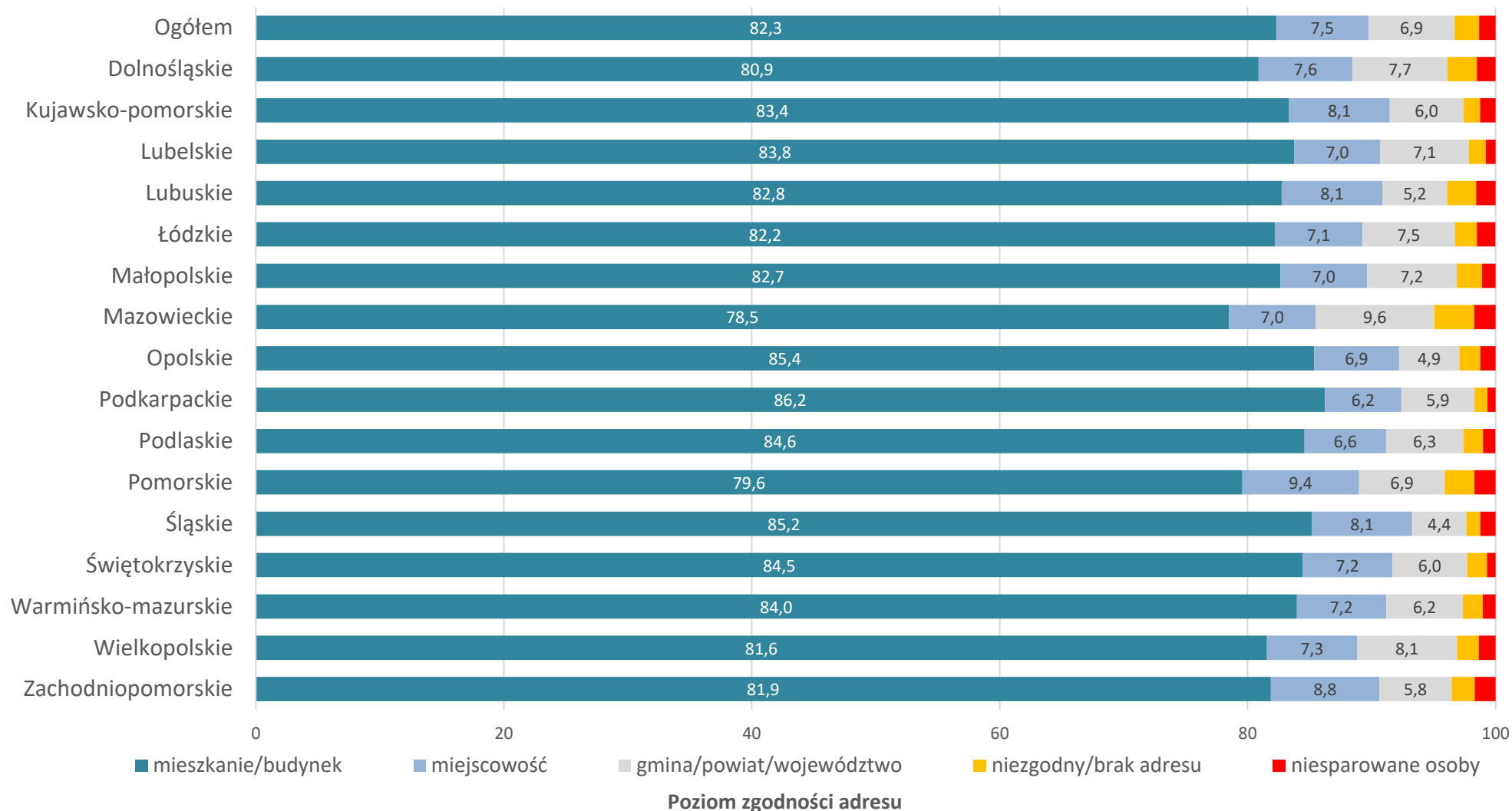


Poziom zgodności adresu

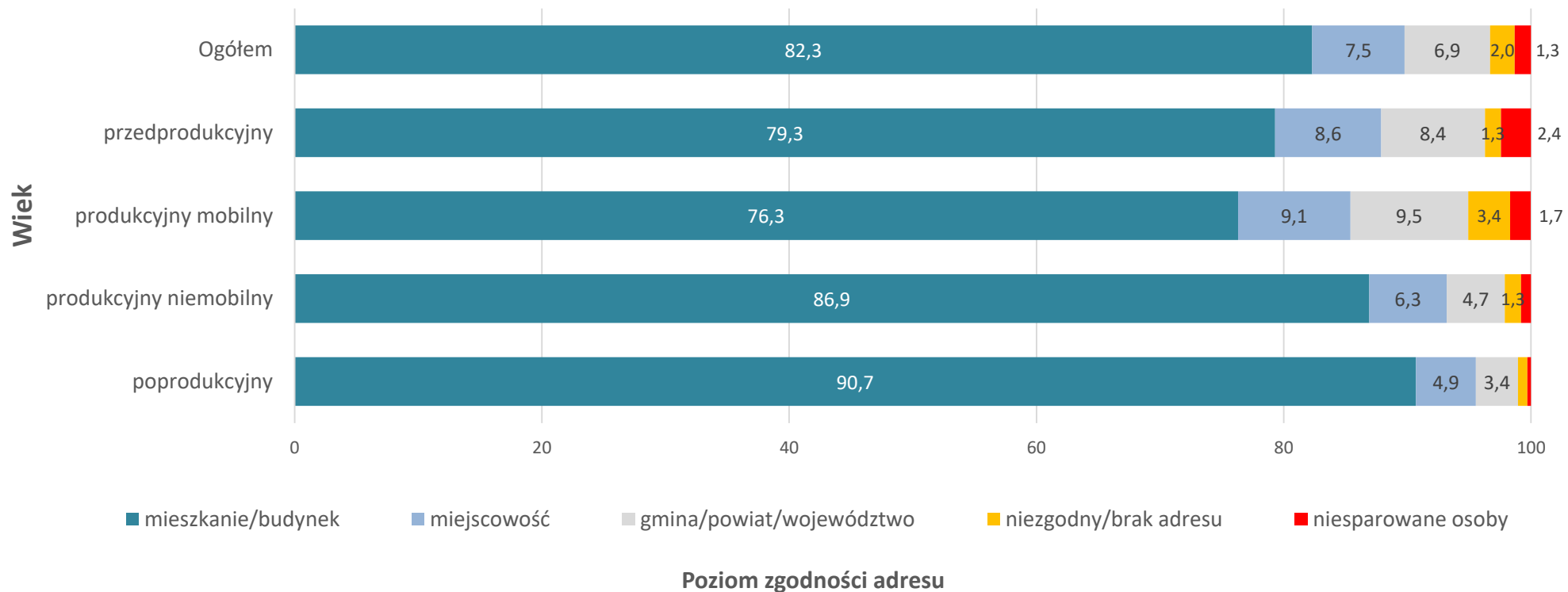
Oszacowana populacja docelowa według miejsca zamieszkania oraz zgodności adresu z wykazową populacją docelową w miastach i na wsi (w %)



Oszacowana populacja docelowa według województw oraz zgodności adresu z wykazową populacją docelową (w %)



Oszacowana populacja docelowa według ekonomicznych grup wieku oraz zgodności adresu z wykazową populacją docelową w miastach i na wsi (w %)



Podsumowanie

- Wykazowa populacja docelowa obejmowała 39566532 osoby, dla 95,9% przypadków z tej liczby ustalony został pełen adres. W odniesieniu do 96,4% z opisywanej populacji udało się ustalić położenie z dokładnością do co najmniej budynku.
- Źródłem danych o adresie był przede wszystkim rejestr KEP, z którego otrzymało adresy $\frac{3}{4}$ populacji, a następnie rejestr PESEL, z którego pozyskano dane adresowe dla kolejnych prawie 22% populacji.
- Oszacowana jako wynik spisu populacja docelowa liczy 39534391 osób (różnica w odniesieniu do wykazowej populacji docelowej wynosi nieco ponad 32 tys.).
- Zbiór wspólnych (sparowanych) jednostek w obu populacjach obejmuje 39014242 osoby, co stanowi 98,60% wykazowej populacji docelowej i 98,68% oszacowanej populacji docelowej.
- Skala rozbieżności rozpatrywana z perspektywy wykazowej populacji docelowej to 552290 zaliczonych do niej osób, czyli 1,40% jej stanu, które nie znalazły się ostatecznie w oszacowanej populacji docelowej – co stanowi miarę błędu nadmiernego pokrycia (przeszacowania) wykazowej populacji docelowej.
- Z kolei 520149 osób, czyli 1,32% spośród należących do oszacowanej populacji docelowej nie zaliczono do wykazowej populacji docelowej, co wyraża wielkość błędu niedostatecznego pokrycia (niedoszacowania) przez wykazową populację docelową.
- Osiągnięto wysoki poziom zgodności adresów wykazowych (poziom mieszkania): wykazowa populacja docelowa – 81,33%, oszacowana populacja docelowa – 81,4%, subpopulacja sparowanych osób – 82,48%.

Dziękujemy za uwagę

d.szaltys@stat.gov.pl
g.gudaszewski@stat.gov.pl

 Liczymy się
DLA POLSKI!

Więcej na spis.gov.pl



Narodowy Spis Powszechny
Ludności i Mieszkań 2021