

Ilu jest w Polsce cudzoziemców? Ujęcie statystyczne i ekonomiczne

Dr Maciej Beręsewicz, prof. UEP

1 Wprowadzenie

Jednym z kluczowych zagadnień w statystyce publicznej stanowiących istotne wyzwanie metodologiczne jest estymacja wielkości populacji. Informacja o liczbie osób mieszkających w danym kraju lub regionie, czy też stanowiących populację określoną w inny sposób aniżeli przetrzennie, ma bowiem istotny wpływ na wiele aspektów życia, takich jak podział funduszy europejskich, środków na ochronę zdrowia czy kalkulację PKB. Jednocześnie strategia i wizja rozwoju spisów powszechnych formułowana przez Eurostat kładzie główny nacisk na ideę odejścia od tradycyjnego spisu organizowanego w odstępach dziesięcioletnich na rzecz kroczącego spisu wirtualnego bazującego na źródłach administracyjnych uzupełnionych o badania reprezentacyjne. Powyższe działania są wynikiem funkcjonowania w statystyce publicznej nowego paradygmatu, w którym kładzie się nacisk na ponowne wykorzystanie wszystkich dostępnych źródeł danych zamiast tworzenia nowych.

Wykorzystanie „niestatystycznych” źródeł danych prowadzi jednak do powstania problemów natury metodologicznej. Jako przykład można podać zrealizowaną przez GUS pracę eksperymentalną „Populacja cudzoziemców w Polsce w czasie COVID-19” (Główny Urząd Statystyczny, 2020), w której w oparciu o dziewięć rejestrów administracyjnych oszacowano, że w dniu 31.12.2019 w Polsce przebywało 2,1 mln. obcokrajowców. Można zadać pytanie, jak bliska prawdy jest ta wartość? Jaki wpływ na ten szacunek mają błędy nielosowe i w jaki sposób możemy go ewentualnie skorygować? Ważna jest również kwestia precyzji uzyskanych szacunków i sposobu jej wyznaczania. Z ekonomicznego punktu widzenia kluczowa jest bowiem wiarygodność szacunków wskaźników, takich jak stopa bezrobocia czy PKB. Służy ona bowiem do oceny tego, w jakim stopniu poszerzają one wiedzę z zakresu sytuacji gospodarczej. Z tej perspektywy istotne jest zatem określenie, na ile informacje zawarte w rejestrach administracyjnych są wystarczające jako źródło wiedzy. Uzyskanie odpowiedzi na wyżej postawione pytania z jednej strony wypełni lukę metodologiczną, z drugiej zaś strony służyć będzie wypracowaniu metodyki estymacji charakterystyk populacji cudzoziemców w Polsce, która pozwoli poszerzyć wiedzę na temat tej populacji.

Kolejnym wyzwaniem jest określenie wielkości populacji uchodźców w przypadku konfliktów zbrojnych czy katastrof naturalnych. W wyniku napaści Rosji na Ukrainę zanotowano ponad 10 mln przekroczeń polskiej granicy ze strony Ukrainy (Komenda Główna Straży Granicznej, 2023), a w szczytowym momencie blisko 1,4 mln obywateli Ukrainy korzystało ze specjalnego statusu UKR (Kancelaria Prezesa Rady Ministrów, 2023), który m.in. umożliwia uzyskanie numeru PESEL, legalny pobyt, ułatwienia w kontaktach z urzędami oraz instytucjami państwowymi czy wykonywanie pracy. Status UKR zostaje odebrany w momencie przebywania poza Polską ponad miesiąc, a określenie faktycznej liczby uchodźców jest kluczowe do określenia m.in. populacji rezydującej, czyli przebywającej ponad 12 miesięcy na terenie Polski.

Celem niniejszego artykułu jest przedstawienie problemów związanych z określeniem liczby cudzoziemców w Polsce w świetle istniejących źródeł danych (statystycznych i niestatystycznych). Sekcja 2 zawiera podstawowe informacje o pomiarze wielkości populacji. Sekcja 3 omawia istniejące szacunki wielkości populacji cudzoziemców w Polsce w świetle dostępnych źródeł danych. Artykuł kończy się podsumowaniem oraz wybraną literaturą.

2 Problematyka pomiaru wielkości populacji

2.1 Populacja trudna do zbadania

Beręsewicz, Gudaszewski i Szymkowiak (2019) w artykule poświęconym szacunkowi wielkości populacji cudzoziemców w Polsce na koniec 2015 i 2016r wymieniają następujące populacje trudne do zbadania:

- te których jednostki trudno wylosować do próby (ang. *hard-to-sample*) – w przypadku populacji trudnych do zbadania bardzo rzadko zdarza się, aby istniał właściwy operat losowania, z którego jednostki można byłoby wylosować do próby wykorzystując odpowiedni schemat jej pobierania. Przykładem tego typu populacji mogą być osoby bezdomne lub pracujący cudzoziemcy;
- te których jednostki trudno jest zidentyfikować (ang. *hard-to-identify*) – w niektórych przypadkach, szczególnie w odniesieniu do stygmatyzowanych grup społecznych, członkowie populacji mogą nie chcieć udostępnić swoich cech, co wiązać się może z lękiem przed ujawnieniem nielegalnego lub wstydliwego statusu społecznego. Przykład tego typu populacji stanowić mogą narkomani, alkoholicy czy różne mniejszości (na przykład osoby LGBT, wyznawcy określonych religii czy ideologii);
- te których jednostki trudno znaleźć i nawiązać z nimi kontakt (ang. *hard-to-find-and-contact*) – trudność w nawiązaniu kontaktu związana jest przede wszystkim z mobilnością tego typu populacji. Przykładem takiej populacji mogą być niezameldowani cudzoziemcy, członkowie koczowniczych kultur (Beduini z południowo-zachodniej Azji czy Tuaregowie z Afryki Północnej), mniejszości wędrownie (Romowie w Europie), osoby bezdomne;

- te której jednostki trudno namówić do wzięcia udziału w badaniu (ang. *hard-to-persuade*) – niechęć do wzięcia udziału w badaniu związana może być z drażliwością poruszanej tematyki, bądź z brakiem czasu. Przykładem tego typu populacji mogą być aktywni zawodowo, pracujący w szarej czy czarnej strefie, cudzoziemcy;
- te której jednostki można zachęcić do wzięcia udziału w badaniu, ale ciężko przeprowadzić wywiad (ang. *hard-to-interview*) – trudność w przeprowadzeniu wywiadu może wynikać z tego, że należy uzyskać na udział w badaniu danej jednostki zgody przełożonego, opiekuna prawnego czy rodzica. Przykład tego typu populacji stanowią więźniowie, osoby niepełnosprawne psychicznie, czy cudzoziemcy nie znający języka danego kraju.

Powyższe klasyfikacje odnoszą się do populacji cudzoziemców w Polsce, której jednostki trudno wylosować do próby (brak pełnego operatu losowania oraz kompleksowych źródeł danych statystycznych, z których można czerpać wiedzę na temat cudzoziemców), z którą trudno nawiązać kontakt (mobilność cudzoziemców na rynku pracy, brak stałego miejsca zamieszkania, czy też bariera językowa).

W Polsce, w przeciwieństwie do krajów anglosaskich (np. Wielkiej Brytanii czy USA), dysponujemy rejestrem populacji (PESEL), który z definicji powinien pokrywać całą populację cudzoziemców (por. Józefowski i Rynarzewska-Pietrzak, 2010; Gołata, 2018; Wallgren i Wallgren, 2014). Jednakże, nie wszyscy cudzoziemcy posiadają numer PESEL, a dane w rejestrze mogą być nieaktualne. Nawet gdybyśmy dysponowali wieloma rejestrami (np. ZUS, NFZ czy Urzędu ds. Cudzoziemców) to pewne subpopulacje cudzoziemców będą w nich nieobserwowane m.in. krótkookresowi migranci pracujący nielegalnie, czy migranci przebywający na terenie kraju bez ważnych dokumentów umożliwiających pobyt. W literaturze przedmiotu proponuje się jednak pewne rozwiązania, które stanowią swego rodzaju remedium na problemy związane z określeniem rzeczywistych rozmiarów populacji trudnych do zbadania. Należą one do grupy technik określanych wspólnym terminem *capture-recapture* (CR), które w głównej mierze opierają się na wykorzystaniu wielu zintegrowanych rejestrów do oszacowania całkowitej populacji (zarejestrowanej i niezarejestrowanej, por. Böhning, Bunge i Heijden (2018); Zhang i Chambers (2019)).

2.2 Estymacja wielkości populacji na podstawie źródeł administracyjnych

W literaturze wymienia się następujące założenia metod CR w zastosowaniach populacyjnych (Wolter, 1986):

- *populacja jest zamknięta* (brak zgonów/urodzeń) – rejestr wg stanu na określony dzień,
- *źródła muszą być niezależne* – uwzględnienie rejestrów policji, uwzględnienie co najmniej 3 źródeł i/lub estymatorów opracowanych dla źródeł zależnych,
- *homogeniczność przynależności do rejestrów* – wykorzystujemy zmienne społeczno-demograficzne i zakładamy homogeniczność w tych grupach,

- *brak błędów łączenia* – łączymy rejestry po kluczu, ewentualnie uwzględniamy błąd wynikający z probabilistycznego łączenia rekordów,
- *brak błędów pokrycia* – wykorzystujemy rejestry aktualizowane w ujęciu miesięcznym/kwartalnym.

Należy jednak zaznaczyć, że istnieje wiele metod, które uwzględniają złamanie tych założeń. W szczególności należy tutaj wymienić: Zhang (2019), który rozważa przypadek rejestru administracyjnego oraz pospisowego badania kontrolnego (ang. *post-enumeration survey*) ograniczając wymienione założenia, jak również Nour (1982), Wolter (1986), Chatterjee i Bhuyan (2017), Chatterjee i Mukherjee (2021), Chatterjee i Bhuyan (2019), którzy rozważają przypadek dwóch lub więcej skorelowanych zmiennych czy też Ding i Fienberg (1994), Di Consiglio i Tuoto (2015), Di Cecco i in. (2018), Di Consiglio, Tuoto i Zhang (2019), którzy uwzględniają błędy łączenia między rejestrami w estymacji wielkości populacji.

W najprostszym przypadku, gdy dysponujemy dwoma źródłami danych i wszystkie powyższe założenia są spełnione, do oszacowania wielkości populacji możemy zastosować estymator Lincolna-Petersena dany wzorem

$$\hat{N} = \frac{n_1 n_2}{m},$$

gdzie n_1 będzie wielkością pierwszego źródła, n_2 będzie wielkością drugiego źródła, a m będzie częścią wspólną obydwu źródeł (por. Brittain i Böhning (2009)).

Przykładowo, Gerritse, van der Heijden i Bakker (2015), Gerritse (2016) zastosowali metodę CR do oszacowania liczby obywateli Polski przebywających w Holandii w 2009 roku. W tym celu wykorzystali dwa źródła danych: Rejestr Populacji (*Gemeentelijke BasisAdministratie*; GBA) oraz rejestr Policji (*HerKenningsdienst Systeem*; HKS) uzyskując następującą poniższą tablicę

Tablica 1: Liczba obywateli Polski w dwóch rejestrach w Holandii w 2009r.

		HKS	
		Tak	Nie
GBA	TAK	374	39 488
	NIE	1 445	?

Źródło: Opracowanie własne na podstawie Gerritse (2016).

Na podstawie tablicy 1 Gerritse (2016) oszacowała wielkość populacji obywateli Polski Holandii na blisko 193 tys. Dla porównania, według Narodowego Spisu Powszechnego Ludności i Mieszkań 2011 (NSP 2021) Polskich emigrantów przebywających czasowo w Holandii było 106 tys.

Niestety, aktualnie w Polsce niemożliwe jest dokonanie podobnego szacunku na podstawie danych pochodzących z Policji (Krajowy System Informacji Policji; KSIP) ponieważ regulacje prawne uniemożliwiają przekazanie identyfikowalnych danych z KSIP Głównemu Urzędowi Statystycznemu. Jedyną możliwością zastosowania metod CR jest wykorzystanie i zintegrowanie wszelkich dostępnych źródeł administracyjnych, a następnie zastosowania zaproponowanych w literaturze rozwiązań z zakresu CR. Pewne sugestie, jak można tego dokonać przedstawili Zhang i Dunne (2017), którzy dokonali szacunku wielkości populacji w Irlandii nie posiadając odpowiednika rejestru PESEL. Wykorzystali w tym celu *Person Activity Register* (PAR), który składa się z 10 zintegrowanych rejestrów do określenia aktywności ludności (tzw. śladów życia; ang. *signs-of-life*) oraz rejestr praw jazdy (RPJ). Jako drugie źródło w metodzie CR wykorzystali fakt odnawiania praw jazdy lub ubieganie się o nowe w danym roku na podstawie RPJ.

Zastosowanie tego podejścia w praktyce jest ściśle związane z jakością danych administracyjnych, a błędy pokrycia czy też łączenia będą przekładały się na obciążenie szacunków. Kluczowe jest również porównanie uzyskanych szacunków z wszelkimi dostępnymi źródłami danych zarówno tymi ze statystyki publicznej (jak spisy), jak i spoza (np. big data).

3 Populacja cudzoziemców w Polsce w świetle dostępnych danych

3.1 Rejestry administracyjne

Zgodnie z wiedzą autora, jedynym opracowaniem zawierającym ocenę polskich rejestrów administracyjnych w odniesieniu do populacji cudzoziemców jest wspomniana praca eksperymentalna Główny Urząd Statystyczny (2020). Wcześniejsze szacunki wykorzystujące metodę CR w Polsce zawarte były w pracy metodologicznej pt. „Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym” (Główny Urząd Statystyczny, 2019) oraz publikacji Beręsewicz, Gudaszewski i Szymkowiak (2019).

W tej pracy Główny Urząd Statystyczny (2020) zespół GUS podjął problem wyznaczenia populacji cudzoziemców na podstawie zintegrowanych rejestrów administracyjnych oszacowując, że na koniec grudnia 2019 roku w Polsce przebywało 2,1 mln cudzoziemców, z czego obywatele Ukrainy, Białorusi oraz Rosji stanowili odpowiednio: 1 351 tys., 105 tys. oraz 37 tys.

Integracji danych dokonano z wykorzystaniem identyfikatorów tj. numerów PESEL, a tablica 2 przedstawia informacje o liczbie cudzoziemców w poszczególnych rejestrach omawianych w tym projekcie.

Tablica 2: Liczba cudzoziemców według występowania w źródłowych rejestrach administracyjnych według stanu na 31.12.2019 r. Skróty odnoszą się do ministerstw będących gestorami rejestrów administracyjnych

Rejestr źródłowy	Liczba cudzoziemców
PESEL	263 976
UdSC	426 266
ZUS	696 434
MF	919 491
MRPiPS	767 033
MNiSW	66 824
MEN	18 032
KRUS	60 198
ZUS	1 230 717

Źródło: Opracowanie własne na podstawie Główny Urząd Statystyczny (2020).

Liczba cudzoziemców zidentyfikowana między rejestrami znacząco się różni. Liczba cudzoziemców wykazana w rejestrze PESEL jest bardzo niska co wynika to z faktu uwzględnienia subpopulacji cudzoziemców zameldowanych na pobyt czasowy lub stały.

Występowanie cudzoziemców w rozważanych rejestrach przedstawiono w tablicy 3.

Tablica 3: Liczba cudzoziemców według liczby rejestrów

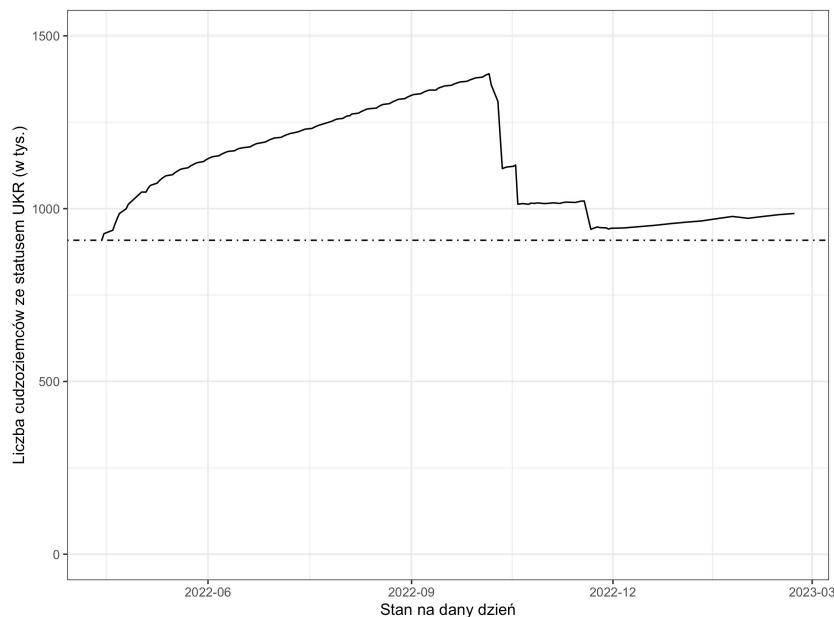
Współwystępowanie	Liczba cudzoziemców
W 1 zbiorze	988 235
W 2 zbiorach	461 233
W 3 zbiorach	303 767
W 4 zbiorach	222 911
W 5 zbiorach	111 896
W 6 zbiorach	17 921
W 7 zbiorach	138
W 8 zbiorach	0
W 9 zbiorach	0

Źródło: Opracowanie własne na podstawie Główny Urząd Statystyczny (2020).

Od 2019 roku sytuacja w Polsce uległa zmianie. Napaść Rosji na Ukrainę spowodowała znaczący wzrost liczby obywateli Ukrainy (głównie kobiet i dzieci) w Polsce.

Rysunek 1 przedstawia liczbę cudzoziemców posiadających status UKR (w tys.) w okresie od 14 kwietnia 2022 do 21 lutego 2023. Początkowa liczba wynosiła ponad 908,5 tys. a końcowa

986 tys. W szczytowym momencie (6 października 2020) liczba osób ze statusem UKR wynosiła blisko 1,4 mln. Gwałtowny spadek związany jest m.in. z rozpoczęciem roku szkolnego oraz innymi przyczynami, a opóźnienie w raportowaniu wskazuje na znaczną zmianę w październiku oraz listopadzie. Widać także lekki wzrost liczby osób ze statusem UKR w 2023r., ale dostępne dane nie pozwalają określić, czy są to osoby nowe czy uzyskujące status UKR po raz kolejny.



Rysunek 1: Liczba obywateli Ukrainy ze statusem UKR (w tys.) w okresie od 14 kwietnia 2022 do 21 lutego 2023

Źródło: Opracowanie własne na podstawie Kancelaria Prezesa Rady Ministrów (2023).

Jeżeli rejestry administracyjne mają zastąpić spisy powszechne to ocena ich jakości oraz integracja nie tylko z wykorzystaniem numeru PESEL jest kluczowa. Należy jednak pamiętać, że uzyskane wyniki mogą być porównywalne w czasie, dlatego należy również odnieść się do wyników najnowszego spisu z marca 2021 roku.

3.2 Narodowy Spis Powszechny Ludności i Mieszkań 2021

Tablica 4 zawiera wyniki Narodowego Spisu Powszechnego Ludności i Mieszkań 2021 (NSP 2021) w zakresie liczby cudzoziemców według obywatelstwa oraz długości pobytu. Kolumna *stali* odnosi się do tzw. ludności wg definicji krajowej¹, natomiast kolumna *stali i czasowi* odnosi się do ludności bez względu na okres przebywania. Należy zaznaczyć, że dane zawarte

¹Ludność według definicji krajowej – stali mieszkańcy Polski - w tym osoby, które przebywają czasowo za granicą (bez względu na okres przebywania), ale zachowały stałe zameldowanie w Polsce. Do ludności nie są natomiast zaliczani imigranci przebywający w Polsce czasowo.

w tablicy 4 są oznaczone przez GUS jako wstępne i mogące ulec zmianie. Zgodnie z tymi szacunkami liczba cudzoziemców w Polsce wynosiła blisko 1,5 mln, z czego większość stanowili obywatele krajów Europy (bez Unii Europejskiej). Niestety, w momencie przygotowywania artykułu GUS nie udostępnił szczegółowych informacji o obywatelstwie. Jednak można domniemywać, że większość z grupy 1,1 mln cudzoziemców stanowią obywatele i obywatelki Ukrainy. We wszystkich przypadkach wymienionych w tablicy 4 relacja liczby cudzoziemców przebywających czasowo do stałych mieszkańców była wyższa niż 5, a w przypadku *pozostałych krajów Europy* wynosiła ponad 12.

Tablica 4: Cudzoziemcy wg grup krajów obywatelstwa oraz długości pobytu według NSP 2021 (w tys., stan na 31.03.2021)

obywatelstwo	stali	stali i czasowi	relacja
ogółem	111,8	1 459,2	12,1
Pozostałe kraje Europy	80,7	1 132,2	13,0
Azja	10,6	145,6	12,7
Kraje Unii Europejskiej	15,5	140,2	8,0
Afryka	2,8	18,5	5,6
Ameryka Północna i Środkowa	1,5	14,7	8,8
Ameryka Południowa	0,6	5,7	8,5
Australia i Oceania	0,1	1,2	11,0
Bezpaństwowcy	0,1	1,0	9,0
Obywatelstwo obce - kraj nieustalony	0,0	0,0	0,0

Źródło: Opracowanie własne na podstawie danych NSP 2021 (dane pobrane zgodnie ze stanem na 23.02.2023).

Dane z NSP 2021 mogą nam ukazać pewne zależności, które występowały w marcu 2021, ale po rozpoczęciu agresji Rosji na Ukrainę 24 lutego 2022 i fali migracji uchodźców, liczby te zdecydowanie różnią się od bieżącej sytuacji.

3.3 Źródła big data

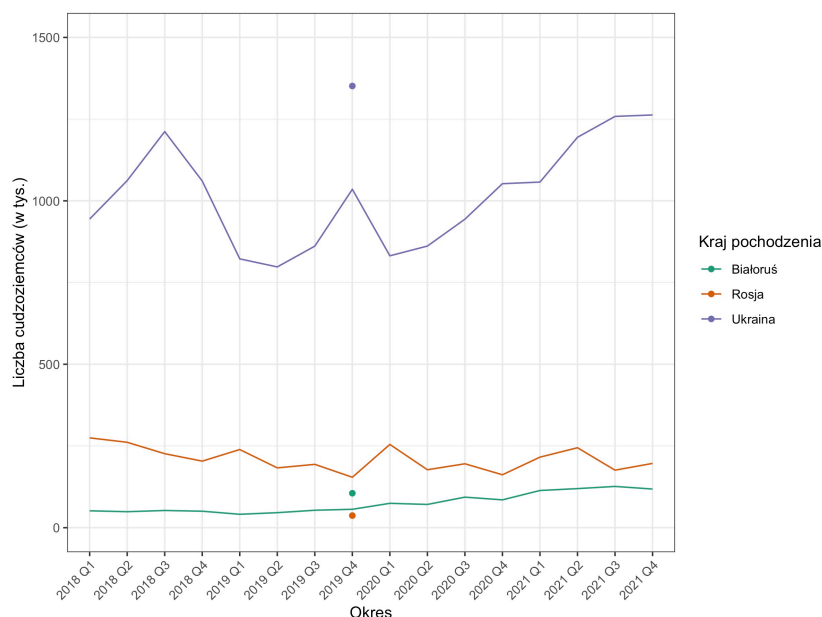
Statystyka publiczna coraz częściej sięga po alternatywne, do rejestrów administracyjnych, źródła danych, które można zdefiniować jako big data (Daas i in., 2015; Tam i Clarke, 2015; Beręsewicz i Szymkowiak, 2015). Dotyczy to głównie wykorzystania danych pochodzących z różnych systemów informacyjnych sieci telekomunikacyjnych, nadzoru ruchu drogowego i wodnego, czy aktywności ludności i podmiotów gospodarczych w sieci.

Pomimo istotnej przewagi źródeł big data nad rejestrami administracyjnymi w kontekście aktualności i automatycznego sposobu zbierania informacji, główne problemy z punktu widzenia

estymacji pozostają te same: błędy pokrycia oraz pomiaru. Aby móc wykorzystać te źródła do celów statystycznych należy je skrupulatnie ocenić.

W tej sekcji przedstawione zostaną dane pochodzące od firmy Selectivv², która działa na polskim i międzynarodowym rynku mobilnej reklamy internetowej tzw. reklamy programatycznej (ang. *programmatic advertisement*). W skrócie polega ona na dobraniu odpowiedniej informacji dla danego użytkownika smartfona na podstawie zebranych o nim informacji takich, jak zainstalowane aplikacje, system operacyjny, marka smartfona, geolokalizacja czy, o ile takie informacje są dostępne, cech społeczno-demograficznych. Firma Selectivv zbiera informacje o ponad 30 mln użytkowników smartfonów w Polsce, z czego w swojej bazie miała w latach 2018-2021 ponad 1,2 mln „obywateli Ukrainy”, 50 tys. „obywateli Białorusi” oraz blisko 200 tys. „obywateli Rosji”. Użycie cudzysłowu służy podkreśleniu faktu, że informacje dotyczące obywatelstwa są ustalane na podstawie określonych reguł decyzyjnych oraz algorytmów, które biorą pod uwagę język systemu telefonu oraz dane geolokalizacyjne. Dane Selectivv były wykorzystane m.in. w raporcie Unii Metropolii Polskich (2022) do określenia liczby Ukraińców w marcu i maju 2022, która wynosiła odpowiednio 3,85 i 3,37 mln osób, a w samych 12 największych polskich miastach 1,43 mln.

Rysunek 2 przedstawia liczbę cudzoziemców określoną na podstawie danych ze smartfonów między 2018 i 2021 rokiem, a kropka oznacza liczbę cudzoziemców określoną na podstawie pracy metodologicznej GUS (Główny Urząd Statystyczny, 2020).

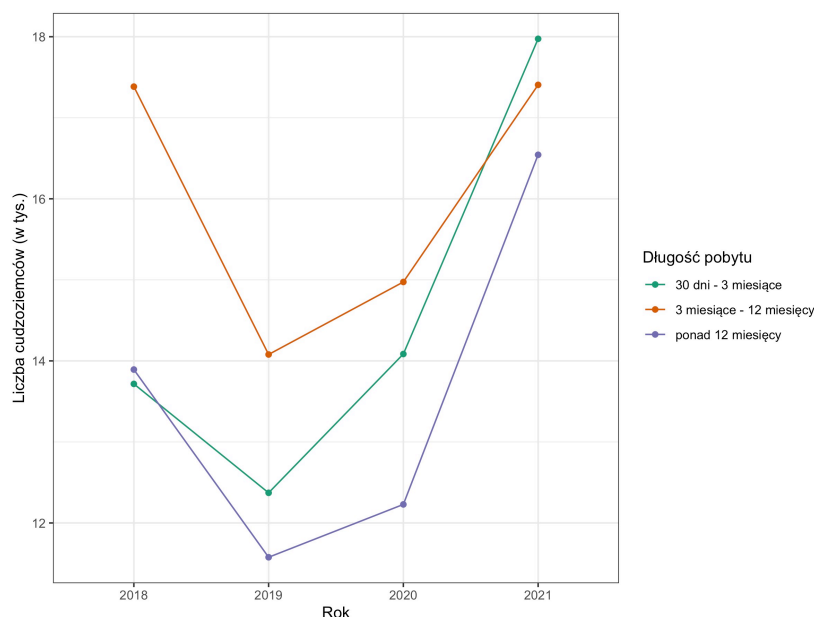


Rysunek 2: Liczba cudzoziemców określona na podstawie danych ze smartfonów między 2018, a 2021 rokiem

²<https://selectivv.com/>

Źródło: Opracowanie własne na podstawie danych Selectivv. Punkty na wykresie przedstawiają szacunek omówiony w sekcji 3.1.

Dane pochodzące ze smartfonów umożliwiają m.in. określenie długości pobytu w danym miejscu na podstawie aktywności oraz geolokalizacji. Rysunek 3 przedstawia liczbę „obywateli Ukrainy” na podstawie danych ze smartfonów według długości pobytu w Poznaniu w latach 2018-2021. Według danych Selectivv w 2018 roku liczba „obywateli Ukrainy” wynosiła blisko 45 tys. z czego 13 tys. przebywało od 30 do 3 miesięcy, 17 tys. od 3 miesięcy do 12, a blisko 14 tys. ponad 12 miesięcy. Liczba ta spadła w 2019 roku do 38 tys., a w kolejnych wzrosła do blisko 51 tys. Zgodnie z tymi danymi ze smartfonów struktura pobytu „obywateli Ukrainy” była stała i wynosiła około 30% dla pobytu ponad 12 miesięcy i 60% dla pobytów krótszych.



Rysunek 3: Liczba „obywateli Ukrainy” na podstawie danych ze smartfonów według długości pobytu w Poznaniu w latach 2018-2021

Źródło: Opracowanie własne na podstawie danych Selectivv.

Należy jednak zaznaczyć, że długość pobytu określana jest dla danego użytkownika smartfona na podstawie unikalnych numerów reklamowych. Oznacza to, że zmiana smartfona powoduje brak ciągłości w danych i osoba, która przebywała ponad 12 miesięcy po zmianie telefonu może zostać zaklasyfikowana jako migrant krótkookresowy (do 3 miesięcy).

4 Podsumowanie

Celem niniejszego artykułu było przedstawienie problemów związanych z określeniem liczby cudzoziemców w Polsce w świetle istniejących źródeł danych (statystycznych i niestatystycznych). Skupiono się na problemie populacji cudzoziemców, która jest *populacją trudną do zbadania*, niewystarczająco pokrytą przez istniejące, rozproszone źródła danych w Polsce a przez to niemożliwe jest aktualnie precyzyjne określenie jej wielkości.

Remedium na omawiane problemy stanowią metody *capture-recapture*, które wykorzystując wiele zintegrowanych źródeł danych, przy odpowiednich założeniach, mogą dostarczyć szacunków zarówno części zarejestrowanej jak i poza rejestrami. Należy jednak podkreślić, że jakość szacunków zależy w głównej mierze od jakości danych oraz ich łączenia. Jest to o tyle ważne, że nie wszyscy cudzoziemcy w Polsce posiadają numer PESEL i łączenie wyłącznie po tym identyfikatorze może zawyżać szacunki wielkości populacji. Prace z wykorzystaniem probabilistycznego łączenia rekordów umożliwią połączenie większej liczby cudzoziemców między rejestrami co przyczyni się do poprawy szacunków wielkości populacji cudzoziemców w Polsce.

5 Podziękowania

Autor dziękuje prof. dr hab. Markowi Świtońskiemu za zaproszenie do wygłoszenia referatu pt. „Ilu jest w Polsce cudzoziemców? Ujęcie statystyczne i ekonomiczne” w ramach Dwugłosu Nauki „Migracje – historia i teraźniejszość” 21.11.2022.

Badania autora są realizowane dzięki wsparciu Narodowego Centrum Nauki (grant NCN OPUS 20 nr 2020/39/B/HS4/00931 pt. „Statystyka cudzoziemców bez spisu powszechnego – jakość, integracja danych i estymacja”).

Literatura

- Beręsewicz, Maciej, Grzegorz Gudaszewski i Marcin Szymkowiak (28 paź. 2019). “Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture-recapture”. W: *Wiomości Statystyczne. The Polish Statistician* 64.10, s. 7–35. DOI: [10.5604/01.3001.0013.7589](https://doi.org/10.5604/01.3001.0013.7589). URL: <http://dx.doi.org/10.5604/01.3001.0013.7589>.
- Beręsewicz, Maciej i Marcin Szymkowiak (2015). “Big data w statystyce publicznej–nadzieje, osiągnięcia, wyzwania i zagrożenia”. W: *Econometrics. Ekonometria. Advances in Applied Data Analytics* 2, s. 9–22.
- Böhning, Dankmar, John Bunge i Peter GM Heijden (2018). *Capture-recapture methods for the social and medical sciences*. CRC Press Boca Raton.
- Brittain, Sarah i Dankmar Böhning (2009). “Estimators in capture–recapture studies with two sources”. W: *AStA Advances in Statistical Analysis* 93.1, s. 23–47.

- Chatterjee, Kiranmoy i Prajamitra Bhuyan (2017). “On the estimation of population size from a post-stratified two-sample capture–recapture data under dependence”. W: *Journal of Statistical Computation and Simulation* 90, s. 819–838.
- (2019). “On the estimation of population size from a dependent triple-record system”. W: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182.4, s. 1487–1501.
- Chatterjee, Kiranmoy i Diganta Mukherjee (2 wrz. 2021). “On the estimation of population size under dependent dual-record system: an adjusted profile-likelihood approach”. W: *Journal of Statistical Computation and Simulation* 91.13, s. 2740–2763. DOI: [10.1080/00949655.2021.1908284](https://doi.org/10.1080/00949655.2021.1908284). URL: <https://www.tandfonline.com/doi/full/10.1080/00949655.2021.1908284>.
- Daas, Piet JH i in. (2015). “Big data as a source for official statistics”. W: *Journal of Official Statistics* 31.2, s. 249–262.
- Di Cecco, Davide i in. (2018). “Population Size Estimation Using Multiple Incomplete Lists with Overcoverage”. W: *Journal of Official Statistics* 34.2, s. 557–572.
- Di Consiglio, Loredana i Tiziana Tuoto (2015). “Coverage evaluation on probabilistically linked data”. W: *Journal of Official Statistics* 31.3, 415–429.
- Di Consiglio, Loredana, Tiziana Tuoto i Li-Chun Zhang (2019). “Capture-recapture methods in the presence of linkage errors”. W: *Analysis of Integrated Data*. Chapman i Hall/CRC, s. 39–71.
- Ding, Y i SE Fienberg (1994). “Dual system estimation of Census undercount in the presence of matching error”. W: *Survey Methodology* 20.2, 149–158.
- Gerritse, Susanna C, Peter GM van der Heijden i Bart FM Bakker (2015). “Sensitivity of population size estimation for violating parametric assumptions in log-linear models”. W: *Journal of official statistics* 31.3, s. 357–379.
- Gerritse, Susanna Charlotte (2016). “An application of population size estimation to official statistics: Sensitivity of model assumptions and the effect of implied coverage”. Prac. dokt. Utrecht University.
- Główny Urząd Statystyczny (2019). *Cudzoziemcy na krajowym rynku pracy w ujęciu regionalnym*. URL: <https://stat.gov.pl/statystyki-eksperymentalne/kapital-ludzki/cudzoziemcy-na-krajowym-ryнку-pracy-w-ujęciu-regionalnym,6,1.html>.
- (2020). *Populacja cudzoziemców w Polsce w czasie COVID-19*. URL: <https://stat.gov.pl/statystyki-eksperymentalne/kapital-ludzki/populacja-cudzoziemcow-w-polsce-w-czasie-covid-19,12,1.html>.
- Gołata, Elżbieta (2018). *Koniec ery tradycyjnych spisów ludności*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Józefowski, Tomasz i Beata Rynarzewska-Pietrzak (2010). “Ocena możliwości wykorzystania rejestru PESEL w spisie ludności”. W: *Zeszyty Naukowe/Uniwersytet Ekonomiczny w Poznaniu* 149, s. 67–84.
- Kancelaria Prezesa Rady Ministrów (2023). *Dane statystyczne dot. wniosków o nadanie statusu cudzoziemca UKR w związku z konfliktem na Ukrainie. Szczegółowe statystyki dot. osób zarejestrowanych w rejestrze obywateli Ukrainy i członków ich rodzin, którym nadano status cudzoziemca na podstawie specustawy. Stan na 20.02.2023 r.* URL: <https://dane.gov.pl/pl/dataset/2715,zarejestrowane-wnioski-o-nadanie-statusu-ukr>.

- Komenda Główna Straży Granicznej (2023). *Dane statystyczne dotyczące sytuacji na granicy z Ukrainą. Baza ruchu granicznego (osoby, cudzoziemcy, środki transportu, wszystkie odcinki zewn. granicy) - dane za okres: styczeń - grudzień 2022 r. / dane z ZSE6 / Straż Graniczna.* URL: <https://dane.gov.pl/pl/dataset/2705,dane-statystyczne-dotyczace-sytuacji-na-granicy-z-Ukraina>.
- Nour, El-Sayed (1982). "On the Estimation of the Total Number of Vital Events with Data from Dual Collection Systems". W: *Journal of the Royal Statistical Society: Series A (General)* 145.1, s. 106–116. DOI: <https://doi.org/10.2307/2981424>.
- Tam, Siu-Ming i Frederic Clarke (2015). "Big data, official statistics and some initiatives by the Australian Bureau of Statistics". W: *International Statistical Review* 83.3, s. 436–448.
- Unia Metropolii Polskich (2022). *3,4 mln Ukraińców w Polsce. Nowy raport Unii Metropolii Polskich.* URL: <https://metropolie.pl/artukul/34-mln-ukraincow-w-polsce-nowy-raport-unii-metropolii-polskich-1>.
- Wallgren, Britt i Anders Wallgren (2014). *Register-based Statistics: statistical methods for administrative data.* John Wiley & Sons.
- Wolter, Kirk M (1986). "Some coverage error models for census data". W: *Journal of the American Statistical Association* 81.394, s. 337–346.
- Zhang, Li-chun (2019). "A Note on Dual System Population Size Estimator". W: *Journal of Official Statistics* 35.1, s. 279–283.
- Zhang, Li-Chun i Raymond L. Chambers, red. (18 kw. 2019). *Analysis of Integrated Data.* Chapman i Hall/CRC. DOI: [10.1201/9781315120416](https://doi.org/10.1201/9781315120416). URL: <http://dx.doi.org/10.1201/9781315120416>.
- Zhang, Li-Chun i John Dunne (2017). "Trimmed dual system estimation". W: *Capture-recapture methods for the social and medical sciences*, s. 237–257.