

Tomasz KLIMANEK¹
Marcin SZYMKOWIAK^{2,3}
Tomasz JÓZEFOWSKI³

Badanie zjawiska niepełnosprawności w przekroju powiatów województwa wielkopolskiego z wykorzystaniem metod statystyki małych obszarów⁴

1. WPROWADZENIE

Zjawisko niepełnosprawności należy współcześnie do jednych z najpoważniejszych problemów społecznych. Liczba osób niepełnosprawnych wzrasta systematycznie, co jest w dużej mierze konsekwencją dostrzegalnego problemu związanego ze starzeniem się społeczeństw. Około 17% osób w Unii Europejskiej jest niepełnosprawnych, w stopniu od lekkiego do znacznego. Oznacza to, że około 80 mln Europejczyków często nie ma możliwości pełnego uczestniczenia w życiu społecznym i gospodarczym z powodu barier związanych ze środowiskiem i z postawami ich otoczenia. Również w Polsce niepełnosprawność stanowi poważny problem społeczny. Według danych Narodowego Spisu Powszechnego Ludności i Mieszkań (NSP 2011) w Polsce liczba osób niepełnosprawnych wynosiła około 4,7 mln. Stanowiło to 12% ludności kraju (GUS, 2012).

W związku z powyższym władze poszczególnych krajów, jak i władze szczebla lokalnego, będą zmuszone podejmować odpowiednie kroki celem niwelowania negatywnych skutków związanych z występowaniem zjawiska niepełnosprawności. Konieczność ta wynika również z zapisów Europejskiej Strategii w Sprawie Niepełnosprawności 2010–2020, zgodnie z którą na Unii Europejskiej i państwach członkowskich spoczywa odpowiedzialność za poprawę społecznej i ekonomicznej sytuacji osób niepełnosprawnych (Komisja Europejska, 2010).

¹ Urząd Statystyczny w Poznaniu, ul. Wojska Polskiego 27/29, 60–624 Poznań, Polska, autor prowadzący korespondencję – e-mail: t.klimanek@stat.gov.pl.

² Uniwersytet Ekonomiczny w Poznaniu, Wydział Informatyki i Gospodarki Elektronicznej, Katedra Statystyki, al. Niepodległości 10, 61–875 Poznań, Polska, Urząd Statystyczny w Poznaniu, ul. Wojska Polskiego 27/29, 60–624 Poznań.

³ Urząd Statystyczny w Poznaniu, ul. Wojska Polskiego 27/29, 60–624 Poznań, Polska.

⁴ Artykuł powstał w ramach grantu „Estymacja pośrednia w zakresie badania niepełnosprawności na podstawie NSP 2011”, który został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2013/11/B/HS4/01472.

Ogólnym celem tej strategii jest zwiększenie możliwości osób niepełnosprawnych, tak aby mogły one w pełni korzystać ze swoich praw i uczestniczyć w życiu społecznym oraz w europejskiej gospodarce, zwłaszcza dzięki jednolitemu rynkowi.

Podejmowanie właściwych decyzji oraz eliminowanie barier wymagać będzie jednak dokładnej, rzetelnej i szczegółowej informacji na temat niepełnosprawności. Również realizacja wspomnianej powyżej Europejskiej Strategii w Sprawie Niepełnosprawności 2010–2020 wymagać będzie dużego zaangażowania instytucji Unii Europejskiej i państw członkowskich, co wyrażać ma się nie tylko poprzez odpowiednie wsparcie finansowe i podnoszenie świadomości społecznej na temat niepełnosprawności, ale również poprzez gromadzenie odpowiednich statystyk i danych oraz ich monitorowanie. Infrastruktura statystyczna w zakresie niepełnosprawności stanowi zatem warunek *sine qua non* w procesie podejmowania odpowiednich decyzji związanych z tym zjawiskiem społecznym.

Obecnie dostępne w Polsce źródła danych na temat niepełnosprawności (badania reprezentacyjne czy spisy) pozwalają uzyskać podstawowe informacje na temat tego zjawiska jedynie na dość dużym stopniu agregacji przestrzennej – poziom kraju czy województwa. Pewne remedium stanowić może tutaj statystyka małych obszarów (SMO), dzięki której możliwa jest estymacja parametrów dla bardziej szczegółowych przekrojów terytorialnych czy odpowiednio zdefiniowanych domen (Rao, Molina, 2015).

Głównym celem artykułu jest estymacja odsetka liczby osób niepełnosprawnych prawnie i biologicznie w przekroju powiatów województwa wielkopolskiego i dla bardziej szczegółowych domen – z uwzględnieniem wykształcenia i z wykorzystaniem wybranych estymatorów jakie oferuje SMO. Umożliwi to przestrzenną ocenę zjawiska niepełnosprawności na niepublikowanym do tej pory poziomie szczegółowości. Realizacji głównego celu podporządkowany został odpowiedni układ prezentowanego artykułu.

W drugiej części artykułu omówiono kluczowe badania reprezentacyjne, w których poruszana jest problematyka niepełnosprawności. Mimo, że nie są one dedykowane ściśle temu zjawisku, a innym obszarom badawczym takim jak rynek pracy, ubóstwo czy budżety gospodarstw domowych, można w nich znaleźć informacje na temat osób niepełnosprawnych.

W trzeciej części artykułu opisana została niepełnosprawność w świetle danych zebranych w ramach NSP 2011. Przedstawiono najważniejsze definicje tego zjawiska, a także wskazano przekroje, dla których dostępne są informacje w NSP 2011 o osobach niepełnosprawnych. Dane pochodzące z NSP 2011 posłużyły autorom do oszacowania odsetka liczby osób niepełnosprawnych prawnie i biologicznie w przekroju powiatów województwa wielkopolskiego i dla określonych kategorii wykształcenia.

W kolejnym, czwartym rozdziale, przedstawiono dotychczasowe doświadczenia z wykorzystania SMO w obszarze niepełnosprawności. Następnie – w rozdziale piątym, opisano od strony teoretycznej najważniejsze estymatory klasy

SMO wykorzystane w artykule w procesie estymacji odsetka osób niepełnosprawnych w województwie wielkopolskim dla domeny zdefiniowanej jako przekrój powiatu i kategorii wykształcenia.

W ostatniej – szóstej części artykułu, opisano procedurę badawczą, przedstawiono wyniki estymacji z wykorzystaniem wybranych estymatorów klasy SMO, a także dokonano dyskusji związanej z jakością uzyskanych szacunków oraz z przestrzennym zróżnicowaniem badanego zjawiska. Artykuł kończy podsumowanie, w którym sformułowano najważniejsze wnioski oraz wskazano dalsze kierunki prac badawczych.

2. NIEPEŁNOSPRAWNOŚĆ W ŚWIETLE WYBRANYCH BADAŃ REPREZENTACYJNYCH

Prowadzenie badań ankietowych przez Główny Urząd Statystyczny w oparciu metodę reprezentacyjną jest nadal jednym z głównych sposobów pozyskiwania informacji statystycznej w zakresie kształtowania się w kraju procesów społecznych, demograficznych oraz gospodarczych. Dotyczy to również zagadnień związanych z niepełnosprawnością, które poddawane są cyklicznej analizie w ramach kilku badań reprezentacyjnych prowadzonych w resorcie statystyki.

Jednym z podstawowych źródeł statystycznych o osobach niepełnosprawnych na rynku pracy jest Badanie Aktywności Ekonomicznej Ludności (BAEL). Głównym celem badania jest pozyskanie informacji o osobach pracujących, bezrobotnych oraz biernych zawodowo. Natomiast dodatkowo, w ramach prowadzonego badania, została wyodrębniona zbiorowość osób niepełnosprawnych z ogółu ludności w wieku 15 lat i więcej, bazując na kryterium prawnym zdefiniowanym w ustawie z dnia 27 sierpnia 1997 r. o rehabilitacji zawodowej i społecznej oraz zatrudnianiu osób niepełnosprawnych (j. t. Dz. U. z 2011 r. Nr 127, poz. 721, ze zm.). W ramach kwartalnej publikacji Aktywność Ekonomiczna Ludności Polski jeden z działów tematycznych poświęcony jest analizie tej zbiorowości. Znajdują się tam podstawowe informacje na poziomie kraju o osobach niepełnosprawnych w podziale na pracujących, bezrobotnych i biernych zawodowo oraz w dodatkowych przekrojach. Tablice zawierają, między innymi, informacje o liczbie osób niepełnosprawnych według stopnia niepełnosprawności (znaczny, umiarkowany, lekki stopień niepełnosprawności lub równoważny), a także według typu gospodarstwa domowego czy też wybranych cech demograficznych (GUS, 2016a).

Kolejnym badaniem, w którym część pozyskiwanych informacji dotyczy również gospodarstw domowych z osobami niepełnosprawnymi wyłącznie prawnie, jest Badanie Budżetów Gospodarstw Domowych (BBGD). Na podstawie tego badania można uzyskać informacje o strukturze demograficzno-społecznej gospodarstw domowych, tj. o liczbie, wieku, płci, wykształceniu, niepełnosprawności, aktywności ekonomicznej osób wchodzących w skład badanego gospodar-

stwa domowego (GUS, 2011). W ramach tego badania można także uzyskać coroczną informację na poziomie kraju o przeciętnym miesięcznym przychodzie i rozchodzie na 1 osobę w gospodarstwach domowych z osobami niepełnosprawnymi. Podawane są również informacje o przeciętnym miesięcznym spożyciu niektórych artykułów żywnościowych oraz wyposażeniu w niektóre przedmioty trwałego użytkowania w gospodarstwach domowych z osobami niepełnosprawnymi (GUS, 2016b).

Następnym badaniem reprezentacyjnym, w którym można znaleźć pewne informacje, dotyczące zbiorowości osób niepełnosprawnych biologicznie, jest Europejskie Badanie Warunków Życia Ludności (EU-SILC). W ramach wydawanej w corocznym cyklu publikacji pt. „Dochody i Warunki Życia Ludności” można znaleźć między innymi informacje o odsetku osób niepełnosprawnych według kryterium niepełnosprawności biologicznej na poziomie kraju, w podziale na płeć, grupy wieku oraz miejsce zamieszkania (miasto/wieś). Jest to jedyne badanie na podstawie którego w cyklu corocznym pozyskujemy informacje o osobach niepełnosprawnych biologicznie (GUS, 2017).

Kolejnym badaniem reprezentacyjnym przeprowadzonym metodą wywiadu bezpośredniego, które w swoim zakresie ujmuje zagadnienia z obszaru niepełnosprawności prawnej oraz biologicznej, jest Europejskie Ankietowe Badanie Zdrowia (EHIS). Badanie to umożliwia dostarczenie szerokiego spektrum informacji na poziomie kraju o zbiorowości osób niepełnosprawnych prawnie i biologicznie, w tym między innymi, według stopnia niepełnosprawności oraz poziomu wykształcenia i stanu cywilnego, a także statusu na rynku pracy (GUS, 2016c).

Jak wynika z dokonanego przeglądu, cyklicznie prowadzone badania reprezentacyjne dostarczają jedynie podstawowych informacji o osobach niepełnosprawnych na poziomie kraju. Najszerszy zakres danych o tym zjawisku można znaleźć we wspomnianym Europejskim Ankietowym Badaniu Zdrowia, które realizowane jest jednak zaledwie co pięć lat.

3. NIEPEŁNOSPRAWNOŚĆ W ŚWIETLE NSP 2011

Narodowy Spis Powszechny Ludności i Mieszkań w Polsce w 2011 roku dotyczył osób stale zamieszkałych (zameldowanych) na obszarze Polski bez względu na fakt, czy te osoby przebywały w kraju w czasie spisu czy też były za granicą oraz osób przebywających czasowo. Spis był przeprowadzony w budynkach, mieszkaniach, obiektach zbiorowego zakwaterowania i innych zamieszkałych pomieszczeniach niebędących mieszkaniami.

W odróżnieniu od poprzednich edycji spis ten został przeprowadzony metodą mieszaną, tzn. dane były pozyskiwane ze źródeł administracyjnych (rejestrów i systemów informacyjnych) oraz zbierane bezpośrednio od ludności w ramach badania reprezentacyjnego, a także tzw. badania pełnego. Ponadto przeprowadzone zostały dwa pełne badania obejmujące osoby przebywające

w obiektach zbiorowego zakwaterowania oraz osoby bezdomne. Zastosowane rozwiązania miały przede wszystkim na celu redukcję kosztów spisu oraz zmniejszenie obciążenia osób nim objętych, zachowując jednocześnie dobrą jakość jego wyników.

W badaniu pełnym realizowanym drogą internetową, a także wykorzystując informacje dostępne ze źródeł administracyjnych, pozyskano podstawowe dane demograficzno-społeczne i adresowe osób, które nie zostały objęte badaniem reprezentacyjnym lub spisem w obiektach zbiorowego zakwaterowania bądź też badaniem bezdomnych. Przeprowadzone w ramach NSP 2011 badanie reprezentacyjne dostarczyło danych, których nie można było pozyskać z rejestrów i systemów informacyjnych. Badanie to zostało przeprowadzone na próbie losowej około 20% mieszkań w skali kraju.

Zakres badania reprezentacyjnego w NSP 2011 uwzględniał sześć dużych obszarów tematycznych: ludność i jej charakterystyka demograficzno-społeczna, aktywność ekonomiczna, migracje wewnętrzne i zagraniczne ludności, narodowość i wyznanie, gospodarstwa domowe i rodziny oraz budynki i mieszkania. W ramach tych obszarów prowadzono również bardziej szczegółowe badania odpowiednio zdefiniowanych zbiorowości. Obszar ludność i jej charakterystyka demograficzno-społeczna obejmował również badanie zbiorowości osób niepełnosprawnych. Zgodnie z przyjętą w spisie definicją osoba niepełnosprawna – jest to taka osoba, która posiada odpowiednie orzeczenie wydane przez organ do tego uprawniony lub osoba, która takiego orzeczenia nie posiada, lecz odczuwa ograniczenie sprawności w wykonywaniu czynności podstawowych dla swojego wieku (zabawa, nauka, praca, samoobsługa).

Zgodnie z przyjętą definicją zbiorowość osób niepełnosprawnych została podzielona na dwie podstawowe grupy:

- osoby niepełnosprawne prawnie, tj. takie, które posiadały odpowiednie, aktualne orzeczenie wydane przez organ do tego uprawniony;
- osoby niepełnosprawne tylko biologicznie, tj. takie, które nie posiadały orzeczenia, ale miały (odczuwały) całkowicie lub poważnie ograniczoną zdolność do wykonywania czynności podstawowych stosownie do swojego wieku (GUS, 2012).

Zgromadzony zakres danych i przyjęta metodologia badania pozwala dostarczyć informacji o liczbie osób niepełnosprawnych prawnie i biologicznie w podziale na płeć, ekonomiczne grupy wieku, a także stopień niepełnosprawności do możliwie najniższego poziomu agregacji przestrzennej jakim był powiat. Jednak wprowadzenie dodatkowych przekrojów demograficzno-społecznych, w tym podziału tej zbiorowości ze względu na poziom wykształcenia, ze względu na zbyt małą liczebność próby obarczone jest zbyt dużym błędem szacunku. Dlatego też zasadnym wydaje się wprowadzenie narzędzi jakimi dysponuje statystyka małych obszarów w celu przeprowadzenia bardziej pogłębionych analiz tego zjawiska, czemu poświęcone jest niniejsze opracowanie.

4. DOŚWIADCZENIA Z WYKORZYSTANIA SMO W OBSZARZE NIEPEŁNOSPRAWNOŚCI

Potrzeby informacyjne odnoszące się do aktualnych i dokładnych danych w bardzo szczegółowych przekrojach oraz na niskim poziomie agregacji przestrzennej zgłaszane są przez różnego rodzaju instytucje, władze rządowe oraz lokalne. Dotyczy to w sposób szczególny tematyki niepełnosprawności. Kompleksowa informacja na temat tego bardzo ważnego zjawiska społecznego umożliwiłaby odpowiednim instytucjom i władzom poszczególnych szczebli podziału terytorialnego kraju podejmowanie właściwych decyzji związanych m.in. z alokacją środków na pomoc osobom niepełnosprawnym czy z odpowiednimi zmianami dotyczącymi rynku pracy wspierającymi zatrudnianie osób niepełnosprawnych. Jak to zostało wskazane w punkcie 2. i 3. niniejszego artykułu, badania prowadzone przez służby statystyki publicznej w Polsce dostarczają tego typu danych, jednak na dość wysokim poziomie agregacji przestrzennej bądź dla niezbyt szczegółowych domen. Podejmowanie właściwych decyzji wymaga jednak na ogół informacji szczegółowych na poziomie powiatu a nawet gminy i dla wąsko określonych domen (przykładowo w przekroju płci, grup wieku i wykształcenia).

Większość badań reprezentacyjnych, w których poruszana jest problematyka niepełnosprawności, projektowana jest tak, aby dostarczyć podstawowych szacunków na poziomie krajowym, ewentualnie regionalnym, z akceptowalną precyzją szacunku przy wykorzystaniu estymatora bezpośredniego. Jednocześnie obserwowana tendencja związana ze zmniejszaniem kosztów badań i obciążeń respondentów dodatkowo utrudnia prowadzenie dużych badań reprezentacyjnych zapewniających odpowiednią liczebność próby i wynikającą z niej akceptowalną precyzję szacunku dla bardziej szczegółowych przekrojów. Gdy liczebność próby jest zbyt mała w odpowiednich przekrojach dokonuje się szacunków jedynie na wyższych poziomach agregacji przestrzennej z wykorzystaniem estymatorów znanych z klasycznej metody reprezentacyjnej – przede wszystkim stosując tzw. estymator Horvitz-Thompsona, zwany również bezpośrednim. Niestety, w przypadku bardzo małej liczebności próby, estymatory bezpośrednie charakteryzują się niską precyzją szacunku (duża wariancja). Co więcej, w przypadku gdy jakiś przekrój nie jest reprezentowany w próbie, nie jest możliwe wykorzystanie tego estymatora.

Wychodząc naprzeciw potrzebom odbiorców informacji statystycznej dedykowanej niepełnosprawności można jednak dokonywać szacunków z akceptowalną precyzją w domenach, które nie są w ogóle reprezentowane w próbie (tzw. próba zerowa) bądź są reprezentowane jedynie przez niewielką liczbę jednostek. Remedium na niewielką liczebność próby w wyróżnionych domenach i niską precyzję szacunków estymatora bezpośredniego stanowi dynamicznie rozwijająca się gałąź statystyki tzw. statystyka małych obszarów. Metody jakie

oferuje statystyka małych obszarów są obecnie powszechnie wykorzystywane w wielu dziedzinach. Warto wspomnieć chociażby o zastosowaniu estymatorów klasy SMO w badaniu ubóstwa, w badaniach rolnych, z zakresu rynku pracy czy demografii przedsiębiorstw (Dehnel, 2010; Gołata, 2004; Marchetti i inni, 2018; Szymkowiak i inni, 2017).

Również w zakresie niepełnosprawności można znaleźć wiele praktycznych zastosowań statystyki małych obszarów. Jak podaje Żądło (2015), jednym z pierwszych zastosowań statystyki małych obszarów w statystyce publicznej było badanie niepełnosprawności w ramach tzw. Health Interview Survey, które przeprowadzone zostało w 1968 roku przez National Center for Health Statistics w USA. Także Australijski Urząd Statystyczny (ang. *Australian Bureau of Statistics*) wykorzystuje wybrane techniki jakie oferuje SMO w kontekście badań nad niepełnosprawnością (Elazar, Conn, 2004). Na potrzeby estymacji liczby osób niepełnosprawnych wykorzystywane są w Australii proste estymatory syntetyczne oraz bardziej złożone uogólnione modele liniowe bazujące na rozkładzie Poissona czy Bernoulliego.

Bardzo ciekawą aplikację przygotowaną w programie R z wykorzystaniem pakietu Shiny dedykowanemu tworzeniu interaktywnych stron internetowych z wizualizacją wyników poświęconych zjawisku niepełnosprawności stworzył Urząd Statystyczny Nowej Zelandii⁵. Korzystając z metod jakie oferuje statystyka małych obszarów i łącząc dane pochodzące z dwóch badań: Badania Niepełnosprawności w Gospodarstwach Domowych 2013 (ang. *Household Disability Survey*) oraz Spisu Powszechnego 2013 dokonał on szacunku odsetka i liczby osób niepełnosprawnych w bardzo szczegółowych przekrojach, dla których tradycyjne estymatory charakteryzowały się zbyt niską precyzją szacunków. Oszacowania, które powstały z wykorzystaniem estymatorów klasy SMO, prezentowane są w przekroju pięcioletnich grup wieku, regionu oraz rodzaju niepełnosprawności w zakresie słuchu, wzroku, umysłu czy ciała.

Ciekawe zastosowanie hierarchicznych modeli klas ukrytych dla małych obszarów w kontekście szacowania liczby osób niepełnosprawnych przedstawili Fabrizi i inni (2015). Autorzy rozważają problem estymacji liczby osób niepełnosprawnych we Włoszech z wykorzystaniem danych pochodzących z badania reprezentacyjnego Health Conditions and Appeal to Medicare oraz odpowiednich danych z rejestrów administracyjnych (w charakterze zmiennych pomocniczych). W badaniu tym zjawisko niepełnosprawności jest mierzone pośrednio poprzez wykorzystanie zmiennych, które opisują zdolność osoby do wykonywania codziennych zadań i czynności. Co więcej, badanie to jest w taki sposób zaprojektowane, aby dostarczać wiarygodnych szacunków na wysokim poziomie agregacji przestrzennej tj. regionów. Oszacowanie liczby osób niepełnosprawnych, jak wskazują autorzy, dla bardziej szczegółowych domen z wykorzysta-

⁵ Strona dostępna pod linkiem <https://statisticsnz.shinyapps.io/Disability/>.

niem klasycznych metod estymacji obarczone jest zbyt dużym błędem. Ze względu na małe liczebności próby w bardziej szczegółowo zdefiniowanych domenach⁶ a także na fakt, że zmienna celu (niepełnosprawność) nie jest bezpośrednio mierzona w badaniu, proponują wykorzystanie bayesowskich hierarchicznych modeli klas ukrytych dla małych obszarów celem estymacji liczby osób niepełnosprawnych.

W polskiej literaturze z zakresu statystyki małych obszarów kwestia niepełnosprawności nie była do tej pory szeroko analizowana (Dehnel, Klimanek, 2016, czy Klimanek, 2017). W wielu pracach poświęconych zastosowaniom estymacji pośredniej uwaga skupiona była raczej na takich obszarach badawczych jak ubóstwo (Urząd Statystyczny w Poznaniu 2013; Marchetti i inni, 2018; Szymkowiak i inni, 2017), rynek pracy (Gołata, 2004; Józefowski, Szymkowiak, 2013) czy statystyka przedsiębiorstw (Dehnel, 2010). Impuls do podjęcia prac badawczych związanych z zastosowaniem metod statystyki małych obszarów w niepełnosprawności stanowił zatem brak kompleksowych opracowań w tym zakresie. Pierwsze wyniki dyskutowane były na międzynarodowych konferencjach naukowych (Józefowski i inni 2016; Szymkowiak i inni, 2016). Podjęcie prac poświęconych kwestii niepełnosprawności, z wykorzystaniem metod jakie oferuje statystyka małych obszarów, przez polskie środowisko naukowe było zatem konsekwencją dużego popytu na informacje odnośnie tego zjawiska społecznego na niskich poziomach agregacji przestrzennej. Dotychczasowa infrastruktura statystyczna w tym zakresie była bowiem stosunkowo skromna.

W artykule autorzy skupili swoją uwagę na estymacji odsetka osób niepełnosprawnych w przekroju powiatów województwa wielkopolskiego i dla bardziej szczegółowych domen – z uwzględnieniem wykształcenia tj. na niepublikowanym do tej pory poziomie szczegółowości. Wykorzystali w tym celu dane pochodzące z NSP 2011 i wybrane estymatory klasy SMO. W dalszej części artykułu opisane zostały od strony teoretycznej rozważane estymatory. W części empirycznej przedstawione zostały z kolei najważniejsze wyniki z przeprowadzonych prac badawczych.

5. WYBRANE METODY ESTYMACJI POŚREDNIEJ

Na potrzeby estymacji odsetka osób niepełnosprawnych w przekroju powiatów województwa wielkopolskiego i odpowiednich kategorii wykształcenia wybranych zostało sześć różnych estymatorów klasy SMO. Skorzystano przy tym z estymatorów, które były przedmiotem rozważań w ramach projektu Eurarea (ONS, 2004): estymator bezpośredni, GREG, estymator syntetyczny typu A, estymator syntetyczny typu B, EBLUP typu A oraz EBLUP typu B. W projekcie

⁶ Oszacowania liczby osób niepełnosprawnych autorzy dokonują w przekroju obszarów określanych mianem „health districts” w liczbie 12 oraz 3 grup wieku: 50–64, 65–74 i 75+.

tym przedstawiono konstrukcję estymatorów różnego typu dla parametru będącego średnią w domenie⁷. Kompleksowy opis wszystkich estymatorów statystyki małych obszarów, również z uwzględnieniem sześciu wziętych do dalszej analizy, można znaleźć w monografii Rao, Molina (2015).

Estymator bezpośredni – Horvitz-Thompsona

Załóżmy, że s oznacza próbę wylosowaną z populacji U , a s_d część próby, która należy do domeny d . Niech n_d oznacza liczebność próby w domenie d a N_d liczebność populacji w domenie d . Estymator Horvitz-Thompsona wartości średniej stanowi jeden z podstawowych estymatorów wykorzystywanych przez krajowe urzędy statystyczne w procesie estymacji i wyraża się wzorem:

$$\hat{Y}^{(1)} = \hat{Y}_d^{HT} = \frac{1}{\hat{N}_d} \sum_{i \in s_d} w_{id} y_{id}, \quad (1)$$

gdzie \hat{Y}_d^{HT} oznacza oszacowanie średniej wartości zmiennej y w domenie d , y_{id} to wartość zmiennej y w domenie d dla i -tej jednostki badania, w_{id} to waga wynikająca z planu losowania próby dla i -tej jednostki w domenie d a $\hat{N}_d = \sum_{i \in s_d} w_{id}$. Estymator ten jest nieobciążony oraz bardzo prosty w zastosowaniu. Nie jest bowiem w jego przypadku wykorzystywana żadna informacja spoza próby w charakterze zmiennych pomocniczych. Jego wadą jest natomiast to, że w przypadku domen, które są reprezentowane w próbie przez niewielką liczbę jednostek charakteryzuje się małą efektywnością. W przypadku gdy dana domena nie jest w ogóle reprezentowana przez żadną jednostkę (próba zerowa) nie jest możliwe oszacowanie wartości średniej zmiennej y w tej domenie.

Uogólniony estymator regresyjny – GREG

Kolejnym rozpatrywanym estymatorem jest tzw. uogólniony estymator regresyjny GREG. Jest on dosyć często wykorzystywany w badaniach próbkowych przez krajowe urzędy statystyczne. Stanowi on szczególny przypadek tzw. estymatora kalibracyjnego. Wyraża się on następującym wzorem:

$$\hat{Y}^{(2)} = \hat{Y}_d^{GREG} = \frac{1}{\hat{N}_d} \sum_{i \in s_d} w_{id} y_{id} + \left(\bar{\mathbf{x}}_d - \frac{1}{\hat{N}_d} \sum_{i \in s_d} w_{id} \mathbf{x}_{id} \right)^T \hat{\boldsymbol{\beta}}, \quad (2)$$

gdzie $\bar{\mathbf{x}}_d = (\bar{x}_{d,1}, \dots, \bar{x}_{d,p})^T$ jest wektorem złożonym z p średnich wartości zmiennych pomocniczych, $\mathbf{x}_{id} = (x_{id,1}, \dots, x_{id,p})^T$ jest wektorem złożonym z jednostko-

⁷ W artykule opisany jest problem estymacji odsetka osób niepełnosprawnych w odpowiednio zdefiniowanej domenie. Dla zmiennej binarnej przyjmującej dwa warianty: 1 – osoba niepełnosprawna, 0 – w przeciwnym wypadku, estymacja średniej jest równoznaczna z estymacją odsetka jednostek ze względu na wyróżnioną cechę.

wych obserwacji p zmiennych pomocniczych, a $\hat{\beta} = (\sum_{i \in s_d} w_{id} \mathbf{x}_{id} \mathbf{x}_{id}^T)^{-1} \sum_{i \in s_d} w_{id} \mathbf{x}_{id} y_{id}$ jest wektorem oszacowań współczynników regresji liniowej otrzymanych ważoną metodą najmniejszych kwadratów. GREG nie jest estymatorem nieobciążonym, ale jego obciążenie maleje wraz ze wzrostem liczebności próby w szacowanym przekroju.

Estymatory syntetyczne

W procesie estymacji wykorzystano również dwa estymatory syntetyczne typu A i B w zależności od rodzaju modelu, który wykorzystany został w ich konstrukcji.

Estymator syntetyczny typu A wyraża się następującym wzorem:

$$\hat{Y}^{(3)} = \hat{Y}_d^{SyntA} = \bar{\mathbf{X}}_d^T \hat{\beta}^{ind} \quad (3)$$

przy założeniu liniowego dwupoziomowego modelu dla danych indywidualnych postaci $y_{id} = \mathbf{x}_{id}^T \beta + u_d + e_{id}$, gdzie $u_d \sim iidN(0, \sigma_u^2)$, $e_{id} \sim iidN(0, \sigma_e^2)$, przy czym σ_u^2 oraz σ_e^2 oznaczają odpowiednio wariancję składnika resztowego u_d na poziomie obszaru i wariancję składnika resztowego e_{id} na poziomie jednostki obserwacji. Parametry te są zazwyczaj nieznane i należy je oszacować metodą największej wiarygodności z ograniczeniami (ang. *Restricted Maximum Likelihood Method – REML*). Wektor parametrów β w powyższym modelu jest szacowany ważoną metodą najmniejszych kwadratów, a jego estymator wyraża się wzorem:

$$\hat{\beta}^{ind} = \left(\sum_d \mathbf{x}_d^T \hat{\mathbf{V}}_d^{-1} \mathbf{x}_d \right)^{-1} \left(\sum_d \mathbf{x}_d^T \hat{\mathbf{V}}_d^{-1} \mathbf{y}_d \right), \quad (4)$$

przy czym \mathbf{x}_d jest $(n_d \times p)$ – wymiarową macierzą złożoną z p wartości zmiennych pomocniczych w domenie d , \mathbf{y}_d jest $(n_d \times 1)$ – wymiarowym wektorem złożonym z wartości zmiennej objaśnianej y w domenie d . $\hat{\mathbf{V}}_d$ jest natomiast oszacowaną macierzą wariancji-kowariancji, którą można wyrazić następującym wzorem:

$$\hat{\mathbf{V}}_d = \hat{\sigma}_e^2 \mathbf{I}_{n_d} + \hat{\sigma}_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}^T, \quad (5)$$

gdzie \mathbf{I}_{n_d} oznacza macierz jednostkową o wymiarach $(n_d \times n_d)$ a $\mathbf{1}_{n_d}$ jest $(n_d \times 1)$ – wymiarowym wektorem złożonym z samych 1, $\hat{\sigma}_e^2$ oraz $\hat{\sigma}_u^2$ są z kolei oszacowaniami nieznanymi wariancji σ_e^2 i σ_u^2 odpowiednio.

Estymator syntetyczny typu B wyraża się z kolei następującym wzorem:

$$\hat{Y}^{(4)} = \hat{Y}_d^{SyntB} = \bar{\mathbf{X}}_d^T \hat{\beta}^{obszar}, \quad (6)$$

przy założeniu liniowego modelu dla zmiennych pomocniczych na poziomie obszaru postaci $y_d = \bar{\mathbf{X}}_d^T \beta + u_d + \bar{e}_d$, gdzie $u_d + \bar{e}_d \sim iidN\left(0, \sigma_u^2 + \frac{\sigma_e^2}{n_d}\right)$. Wektor parametrów β w powyższym modelu jest szacowany ważoną metodą najmniejszych kwadratów i jego estymator wyraża się wzorem:

$$\hat{\beta}^{obszar} = \left(\sum_d \bar{\mathbf{X}}_d^T \hat{\mathbf{V}}_d^{-1} \bar{\mathbf{X}}_d \right)^{-1} \left(\sum_d \bar{\mathbf{X}}_d^T \hat{\mathbf{V}}_d^{-1} \bar{\mathbf{y}}_d \right). \quad (7)$$

Empiryczne najlepsze liniowe nieobciążone predyktory – EBLUP

W artykule na potrzeby estymacji odsetka osób niepełnosprawnych wzięto również pod uwagę dwa empiryczne najlepsze liniowe nieobciążone predyktory – EBLUP (ang. *Empirical Best Linear Unbiased Predictor*) postaci:

$$\hat{Y}^{(5)} = \hat{Y}_d^{EBLUPA} = w_d^{EBLUPA} \hat{Y}_d^{Greg} + (1 - w_d^{EBLUPA}) \hat{Y}_d^{SyntA}, \quad (8)$$

przy czym:

$$w_d^{EBLUPA} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \frac{\hat{\sigma}_e^2}{n_d}} \quad (9)$$

oraz

$$\hat{Y}^{(6)} = \hat{Y}_d^{EBLUPB} = w_d^{EBLUPB} \hat{Y}_d^{HT} + (1 - w_d^{EBLUPB}) \hat{Y}_d^{SyntB}, \quad (10)$$

gdzie:

$$w_d^{EBLUPB} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \frac{\hat{\sigma}_e^2}{n_d}}. \quad (11)$$

Estymatory typu EBLUP są zatem kombinacjami liniowymi odpowiednich estymatorów syntetycznych oraz bezpośrednich.

Jako miarę oceny jakości precyzji wszystkich sześciu wziętych pod uwagę estymatorów przyjęto średni względny błąd szacunku CV . Jest to względna miara jakości szacunku wykorzystująca w swej konstrukcji błąd średniokwadratowy – MSE (ang. *Mean Squared Error*). Wyraża się ona następującym wzorem:

$$CV_d = \frac{\sqrt{MSE_d}}{\hat{Y}_d}. \quad (12)$$

Miara ta opisuje jaki jest udział błędu estymacji w wartości szacowanej zmiennej na poziomie domeny d . Zgodnie z wytycznymi, jakie przyjmuje Główny Urząd Statystyczny podczas publikowania wyników z badania reprezentacyjnego, oszacowania, dla których $CV < 10\%$ można uznać za wiarygodne. W odniesieniu do oszacowań, dla których CV przyjmuje wartości z przedziału 10–20%,

należy interpretować je ze szczególną ostrożnością. Z kolei do oszacowań, dla których $CV > 20\%$, należy podchodzić sceptycznie i powinny być one tylko publikowane w postaci zagregowanej (GUS, 2013).

Kluczową sprawą w wyznaczeniu średniego względnego błędu szacunku CV odgrywa sposób oszacowania MSE. Szczegółowe rozważania na temat błędów średniokwadratowych estymatorów rozpatrywanych w artykule na potrzeby estymacji odsetka osób niepełnosprawnych można znaleźć w dokumentacji projektu Eurarea (ONS, 2004).

6. ESTYMACJA POŚREDNIA LICZBY OSÓB NIEPEŁNOSPRAWNYCH W WOJEWÓDZTWIE WIELKOPOLSKIM W PRZEKROJU WYBRANYCH DOMEN

Jak to zostało wspomniane we wprowadzeniu do niniejszego artykułu, celem przeprowadzonego postępowania badawczego była estymacja odsetka osób niepełnosprawnych według poziomu wykształcenia w przekroju powiatów województwa wielkopolskiego z wykorzystaniem wybranych estymatorów jakie oferuje SMO. Założono ponadto, że uzyskane oszacowania odsetka osób niepełnosprawnych według poziomu wykształcenia umożliwią zweryfikowanie sformułowanych dwóch hipotez badawczych. Pierwsza z nich głosi, że estymacja pośrednia dla zdefiniowanych w niniejszym artykule domen charakteryzuje się lepszą precyzją szacunku niż estymacja oparta na podejściu Horvitz-Thompsona. Druga hipoteza ma charakter poznawczy i stanowi, że przestrzenny rozkład odsetka osób niepełnosprawnych prawnie i biologicznie w domenach zdefiniowanych jako przekrój powiatu i poziomu wykształcenia na obszarze województwa wielkopolskiego nie jest równomierny.

W celu weryfikacji postawionych hipotez i osiągnięcia założonego celu badawczego wykorzystano dane z NSP 2011, przy czym na etapie ich przygotowania ograniczono populację do osób w wieku 15 lat i więcej oraz tych, którzy w spisie 2011 zostali spisani w ramach badania reprezentacyjnego, tzn. mieli przypisaną tzw. wagę z losowania⁸. Wyodrębniono także subpopulację osób niepełnosprawnych jednocześnie biologicznie i prawnie, która była przedmiotem zainteresowania w procedurze badawczej. Ponadto przekształcono wyjściową kategoryzację poziomów wykształcenia na nową, składającą się z czterech grup: osób z wykształceniem wyższym, osób z wykształceniem policealnym, średnim zawodowym i średnim ogólnokształcącym, osób z wykształceniem zasadniczym zawodowym

⁸ Narodowy Spis Powszechny Ludności i Mieszkań – NSP 2011, jak to zostało wcześniej zasygnalizowane, to badanie, w którym po raz pierwszy w historii polskiej statystyki publicznej została zastosowana tzw. metoda mieszana, tzn. wykorzystano dane pochodzące ze źródeł administracyjnych, a także dane zbierane od ludności w ramach przeprowadzonego na szeroką skalę badania reprezentacyjnego (np. dotyczące stanu cywilnego faktycznego), w którym liczebność próby była określona na poziomie 20% mieszkań dla Polski.

oraz osób z wykształceniem podstawowym ukończonym, podstawowym nieukończonym, a także bez wykształcenia. Ostatecznie zbiór dla województwa wielkopolskiego liczył 655 381 osób i oprócz zmiennej binarnej, która była głównym przedmiotem zainteresowań badawczych, określonej jako BIOLIPRAW⁹, zawierał następujące charakterystyki osób:

- domenę – zmienną utworzoną jako kombinacja kodu powiatu i kodu wykształcenia,
- fakt posiadania orzeczenia o niepełnosprawności wydanego przez organ do tego uprawniony, czyli zmienną binarną określającą, tzw. niepełnosprawność prawną,
- płeć (mężczyzna, kobieta),
- miejscowość faktycznego zamieszkania (w podziale na miasto i wieś),
- wykształcenie,
- identyfikatory pięcioletnich grup wieku (jako zestaw zmiennych binarnych),
- wagę z losowania.

W następnej kolejności korzystając z sekwencyjnej metody doboru zmiennych objaśniających do modelu – regresji krokowej (procedura PROC REG w programie SAS z wykorzystaniem opcji *selection=stepwise*), dokonano redukcji wyjściowego zbioru zmiennych objaśniających. Należy przy tym dodać, że procedura ta została przeprowadzona na danych zagregowanych do poziomu domen, dla których wyznaczono średnie dla zdychotomizowanych wariantów wymienionych powyżej zmiennych. Ostatecznie w estymacji pośredniej opartej na modelach (jednostkowych i obszarowych) znalazły się następujące zmienne pomocnicze:

- płeć (jako zmienna zerojedynkowa: 1 – mężczyzna, 0 – kobieta),
- miejscowość faktycznego zamieszkania (jako zmienna zerojedynkowa: 1 – miasto, 0 – wieś),
- fakt bycia osobą niepełnosprawną prawnie (jako zmienna zerojedynkowa: 1 – osoba niepełnosprawna prawnie, 0 – osoba niebędąca niepełnosprawna prawnie),
- fakt bycia osobą przynależącą do odpowiedniej grupy wiekowej (jako zmienna zerojedynkowa: 1 – osoba w wieku 25–34, 0 – osoba w innym wieku),
- fakt bycia osobą przynależącą do odpowiedniej grupy wiekowej (jako zmienna zerojedynkowa: 1 – osoba w wieku 70 lat i więcej, 0 – osoba w innym wieku).

Tak jak wspomniano, w celu wyznaczenia ocen estymatorów oraz ich precyzji wykorzystano oprogramowanie projektu EURAREA¹⁰. Zamieszczone poniżej kartogramy (rysunek 1 – rysunek 4) prezentują kształtowanie się oszacowań odsetka osób niepełnosprawnych prawnie i biologicznie w układzie powiatów

⁹ Zmienna BIOLIPRAW przyjmowała wartość 1, kiedy osoba należała do populacji osób niepełnosprawnych jednocześnie biologicznie i prawnie oraz wartość 0 w przeciwnym wypadku.

¹⁰ Kody języka 4GL programu SAS wykorzystane w artykule są do pobrania na stronie internetowej pod linkiem <http://webarchive.nationalarchives.gov.uk/20160128190222/http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-and-modelling/eurarea/index.html>.

województwa wielkopolskiego dla poszczególnych kategorii wykształcenia uzyskanych z wykorzystaniem omówionych powyżej estymatorów. Celem pełniejszego przedstawienia wyników zamieszczono również w formie tabelarycznej najważniejsze statystyki opisowe uzyskanych oszacowań (por. tabela 1).

Zastosowane w kartogramach i w tabelach skróty nazw estymatorów są następujące:

- HT – estymator bezpośredni – Horvitz Thompsona, opisany wzorem (1),
- GREG – uogólniony estymator regresyjny, opisany wzorem (2),
- SYNTA – estymator syntetyczny typu A, opisany wzorem (3),
- SYNTB – estymator syntetyczny typu B, opisany wzorem (6),
- EBLUPA – najlepszy liniowy nieobciążony predyktor typu A, opisany wzorem (8),
- EBLUPB – najlepszy liniowy nieobciążony predyktor typu B, opisany wzorem (10).

Odsetek osób niepełnosprawnych prawnie i biologicznie z wykształceniem wyższym w ogólnej liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej waha się od 1,4% do 6,7% w zależności od zastosowanego estymatora (por. rysunek 1 i tabela 1). Podkreślić natomiast należy, że obraz tego zjawiska, uzyskany za pomocą różnych podejść, jest bardzo podobny. Współczynnik korelacji między ocenami estymatora bezpośredniego i tych wspomaganých albo opartych na modelu waha się od 0,946 w przypadku estymatora SYNTHA do wartości bliskiej 1 w przypadku estymatora GREG.

TABELA 1. WYBRANE STATYSTYKI OPISOWE OSZACOWAŃ ODSETKA OSÓB NIEPEŁNOSPRAWNYCH PRAWNIE I BIOLOGICZNIE W LICZBIE LUDNOŚCI POWIATÓW WOJ. WIELKOPOLSKIEGO W WIEKU 15 LAT I WIĘCEJ W ZALEŻNOŚCI OD WYKSZTAŁCENIA I ESTYMATORA

Estymator	Wykształcenie	Min	Max	\bar{x}	S_x	Q ₁	Q ₂	Q ₃
HT	Wyższe	1,52	5,87	3,42	1,06	2,71	3,52	4,03
	Policealne i średnie	3,83	11,88	7,35	1,87	5,86	7,16	8,55
	Zasadnicze zawodowe	6,40	16,56	10,21	2,30	8,65	10,16	11,30
	Podstawowe i brak	6,64	16,37	11,03	2,51	8,89	10,56	12,94
GREG	Wyższe	1,52	5,87	3,42	1,06	2,71	3,52	4,03
	Policealne i średnie	3,83	11,88	7,35	1,87	5,86	7,16	8,55
	Zasadnicze zawodowe	6,40	16,56	10,21	2,30	8,65	10,16	11,30
	Podstawowe i brak	6,64	16,37	11,03	2,51	8,89	10,56	12,94
SYNTA	Wyższe	1,72	6,68	3,71	1,31	2,64	3,68	4,41
	Policealne i średnie	3,88	12,70	7,54	2,18	5,76	7,23	9,23
	Zasadnicze zawodowe	6,36	16,24	10,17	2,45	8,38	9,57	12,22
	Podstawowe i brak	6,29	16,72	10,64	2,61	8,76	9,93	12,10
SYNTB	Wyższe	1,38	5,99	3,34	1,28	2,30	3,25	4,02
	Policealne i średnie	3,84	12,35	7,33	2,06	5,54	6,98	8,86
	Zasadnicze zawodowe	6,58	16,02	10,24	2,30	8,65	9,67	12,16
	Podstawowe i brak	6,73	16,52	11,00	2,38	9,29	10,24	12,34

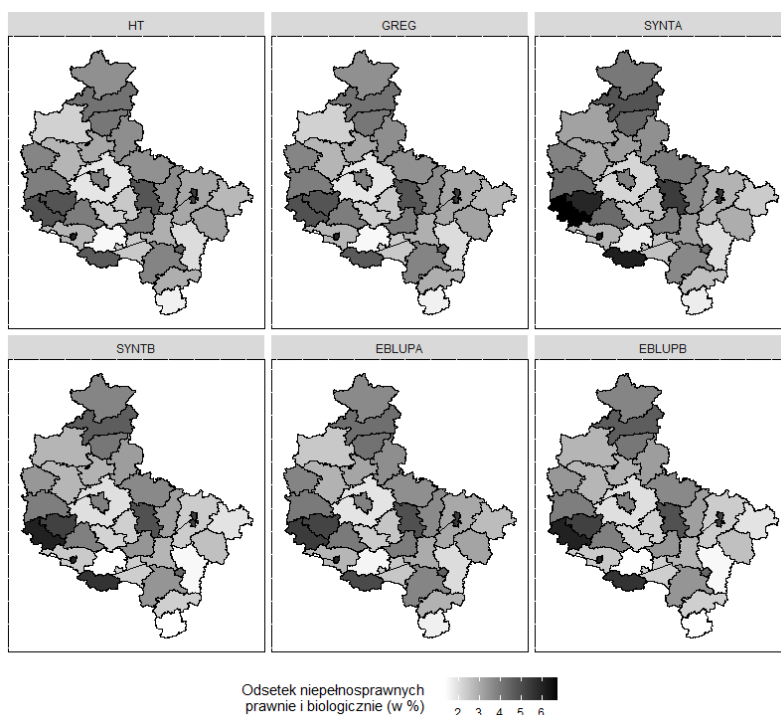
TABELA 1. WYBRANE STATYSTYKI OPISOWE OSZACOWAŃ ODSETKA OSÓB NIEPEŁNOSPRAWNYCH PRAWNIE I BIOLOGICZNIE W LICZBIE LUDNOŚCI POWIATÓW WOJ. WIELKOPOLSKIEGO W WIEKU 15 LAT I WIĘCEJ W ZALEŻNOŚCI OD WYKSZTAŁCENIA I ESTYMATORA (dok.)

Estymator	Wykształcenie	Min	Max	\bar{x}	S_x	Q ₁	Q ₂	Q ₃
EBLUPA	Wyższe	1,57	5,93	3,49	1,11	2,64	3,54	4,14
	Policealne i średnie	3,85	11,95	7,37	1,90	5,84	7,14	8,62
	Zasadnicze zawodowe	6,39	16,51	10,21	2,31	8,66	10,11	11,30
	Podstawowe i brak	6,63	16,42	10,99	2,52	8,89	10,48	12,86
EBLUPB	Wyższe	1,38	5,99	3,34	1,28	2,30	3,25	4,02
	Policealne i średnie	3,84	12,35	7,33	2,06	5,54	6,98	8,86
	Zasadnicze zawodowe	6,58	16,02	10,24	2,30	8,65	9,67	12,16
	Podstawowe i brak	6,73	16,52	11,00	2,38	9,29	10,24	12,34

Źródło: opracowanie własne.

Największy udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem wyższym występuje w powiatach grodzkich: Leszno i Konin oraz w powiecie wolsztyńskim, natomiast najmniejszym udziałem takich osób charakteryzują się powiaty kępiński, kaliski oraz poznański.

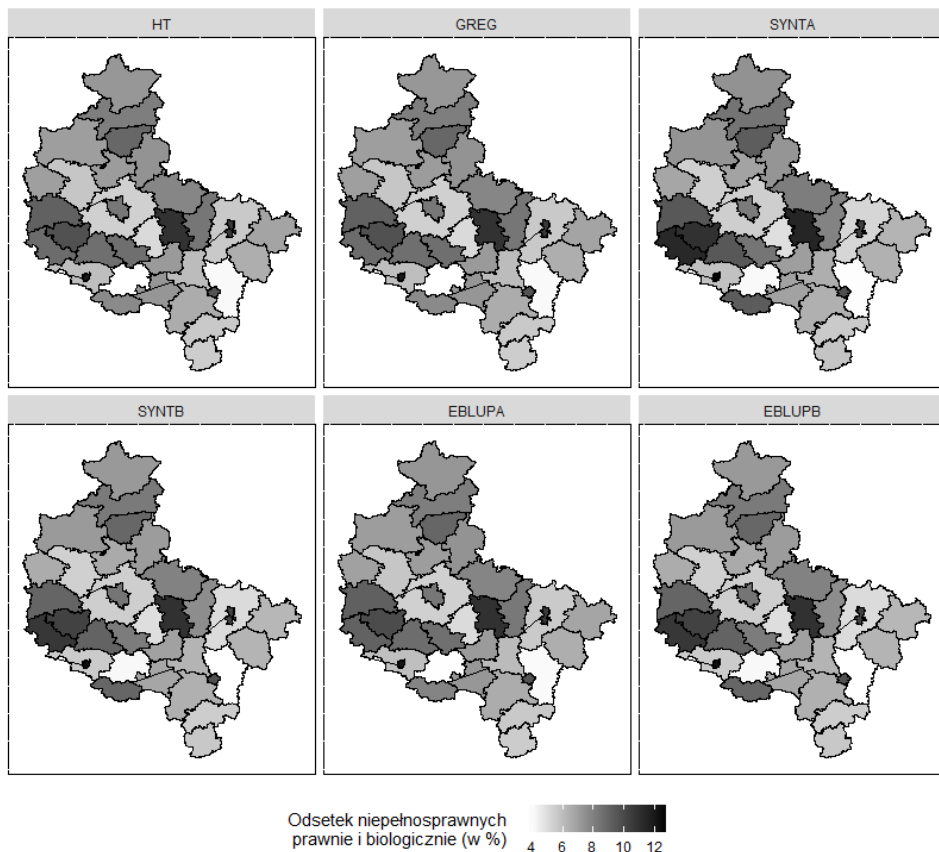
Rysunek 1. Udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem wyższym w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej



Źródło: opracowanie własne.

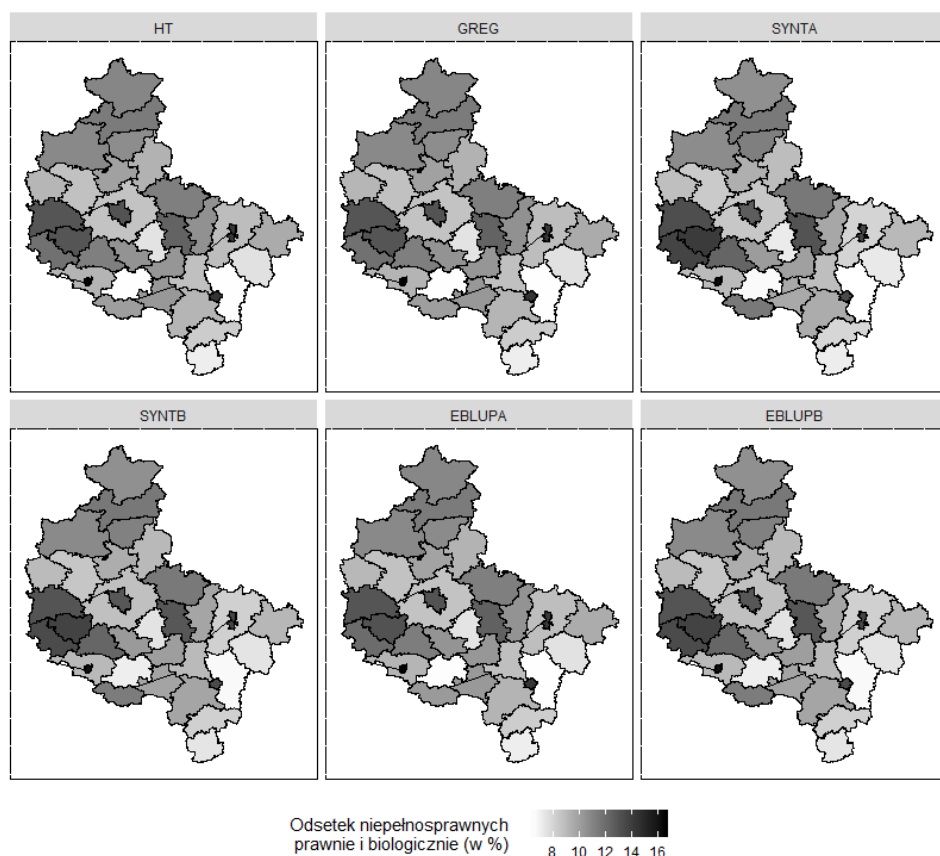
Stopień zgodności między oszacowaniami odsetka osób niepełnosprawnych prawnie i biologicznie z wykształceniem policealnym i średnim w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat uzyskanymi za pomocą estymacji typu Horvitz-Thompsona, a pozostałymi rozważanymi estymatorami, mierzony współczynnikiem korelacji liniowej, jest na podobnym poziomie jak w przypadku wykształcenia wyższego i waha się od 0,963 do wartości bardzo bliskich 1. Należy przy tym zauważyć, że zarówno poziom badanego zjawiska, w przypadku tej kategorii wykształcenia, jak i jego zróżnicowanie, są znacznie wyższe (por. rysunek 2 i tabela 1). Największy udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem policealnym i średnim występuje w powiatach grodzkich: miasto Konin, miasto Leszno oraz w powiecie wrzesińskim – od 10,6% do niecałych 12%, natomiast najmniejszym udziałem takich osób charakteryzują się powiaty gostyński i kaliski – około 4%.

Rysunek 2. Udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem policealnym i średnim w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej

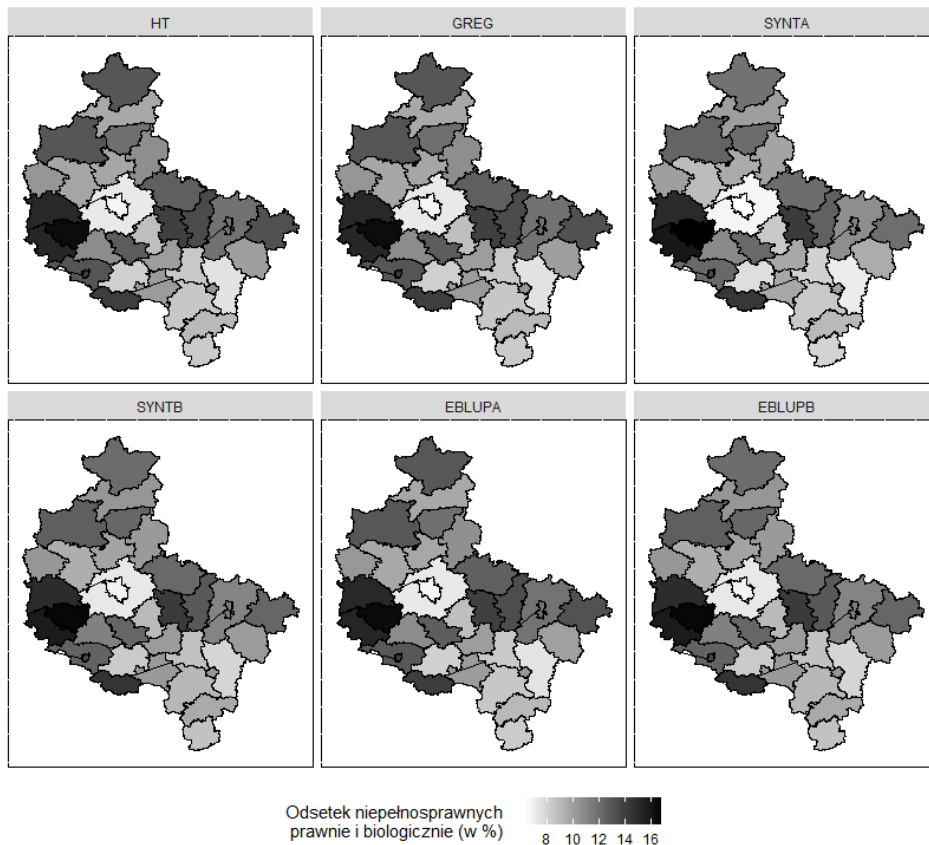


Zdecydowanie największy udział osób niepełnosprawnych prawnie i biologicznie w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej dotyczy pozostałych dwóch kategorii wykształcenia, tzn. zasadniczego zawodowego (por. rysunek 3 i tabela 1) oraz podstawowego, nieukończonego podstawowego i osób bez wykształcenia (por. rysunek 4 i tabela 1). W przypadku wykształcenia zasadniczego zawodowego udział ten waha się od nieco ponad 6% w powiatach gostyńskim i kaliskim, do ponad 16% w Lesznie. W przypadku ostatniej kategorii wykształcenia udział ten kształtuje się podobnie od nieco ponad 6% w Poznaniu (w powiecie grodzkim) do ponad 16% w powiecie grodziskim. Podobnie jak w przypadku wcześniej omawianych kategorii wykształcenia występuje duża zgodność uzyskiwanych ocen estymatora bezpośredniego oraz estymatorów wspomaganym modelem i opartych na modelu. Odpowiednie współczynniki korelacji liniowej uzyskują wartości przekraczające 0,96.

Rysunek 3. Udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem zasadniczym zawodowym w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej



Rysunek 4. Udział osób niepełnosprawnych prawnie i biologicznie z wykształceniem podstawowym, nieukończonym podstawowym i bez wykształcenia w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej



Źródło: opracowanie własne.

Jako miarę oceny jakości precyzji wszystkich sześciu wziętych pod uwagę estymatorów przyjęto średni względny błąd szacunku CV , opisany wzorem (12). Oceny precyzji estymatorów (CV) wykorzystanych do oszacowania udziału osób niepełnosprawnych prawnie i biologicznie w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej dla czterech kategorii wykształcenia przedstawiono w tabeli 2 oraz w formie wykresów pudełkowych (por. rysunek 5). Analiza kształtowania się tego miernika jakości precyzji dla wszystkich estymatorów i czterech kategorii wykształcenia wskazuje, że najlepszą jakością pod tym względem charakteryzują się estymatory SYNTB i EBLUPB. Dobrymi własnościami charakteryzują się także estymatory typu GREG i bezpośredni.

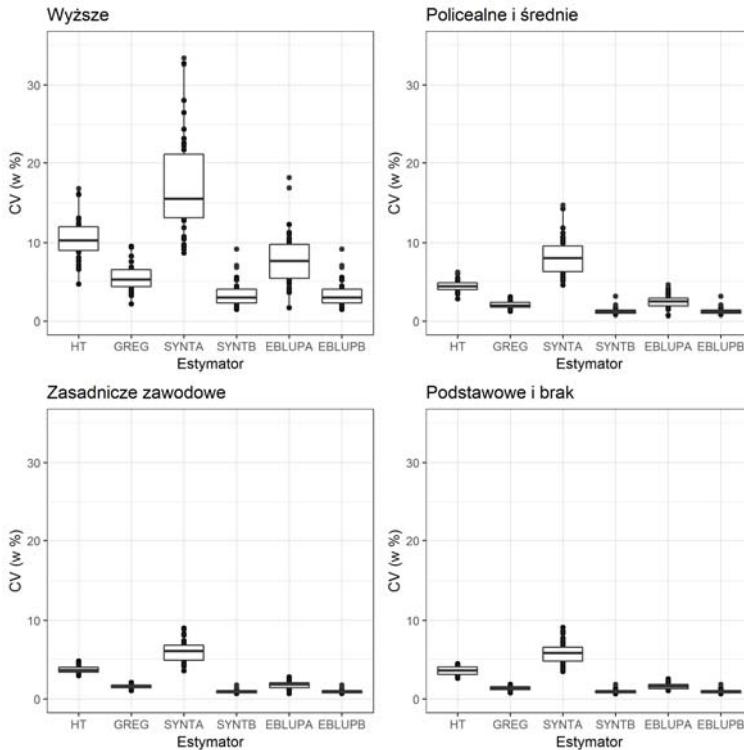
TABELA 2. WYBRANE STATYSTYKI OPISOWE CV ODSETKA OSÓB NIEPEŁNOSPRAWNYCH PRAWNIE I BIOLOGICZNIE W LICZBIE LUDNOŚCI POWIATÓW WOJ. WIELKOPOLSKIEGO W WIEKU 15 LAT I WIĘCEJ W ZALEŻNOŚCI OD WYKSZTAŁCENIA I ESTYMATORA

Estymator	Wykształcenie	Min	Max	\bar{x}	S_x	Q ₁	Q ₂	Q ₃
HT	Wyższe	4,63	16,77	10,56	2,71	8,87	10,28	12,13
	Policealne i średnie	2,74	6,11	4,44	0,76	3,86	4,34	4,80
	Zasadnicze zawodowe	2,91	4,75	3,72	0,47	3,36	3,62	3,98
	Podstawowe i brak	2,59	4,40	3,54	0,48	3,06	3,55	4,02
GREG	Wyższe	2,11	9,55	5,35	1,71	4,27	5,17	6,45
	Policealne i średnie	1,17	3,00	2,01	0,43	1,66	1,88	2,30
	Zasadnicze zawodowe	1,04	2,05	1,56	0,23	1,36	1,54	1,72
	Podstawowe i brak	0,74	1,80	1,33	0,25	1,16	1,36	1,54
SYNTA	Wyższe	8,57	33,33	17,43	6,34	12,98	15,53	21,67
	Policealne i średnie	4,51	14,75	8,25	2,47	6,20	7,91	9,93
	Zasadnicze zawodowe	3,53	8,99	5,95	1,42	4,68	5,98	6,83
	Podstawowe i brak	3,42	9,10	5,70	1,41	4,73	5,76	6,53
SYNTB	Wyższe	1,42	9,15	3,34	1,73	2,21	2,90	3,99
	Policealne i średnie	0,76	3,06	1,23	0,45	0,92	1,13	1,37
	Zasadnicze zawodowe	0,66	1,72	0,95	0,26	0,75	0,88	1,07
	Podstawowe i brak	0,64	1,82	0,94	0,24	0,74	0,93	1,06
EBLUPA	Wyższe	1,61	18,13	7,92	3,53	5,15	7,54	10,08
	Policealne i średnie	0,68	4,55	2,49	0,90	1,76	2,43	2,82
	Zasadnicze zawodowe	0,68	2,73	1,77	0,51	1,38	1,80	2,11
	Podstawowe i brak	1,05	2,50	1,62	0,40	1,30	1,58	1,83
EBLUPB	Wyższe	1,42	9,15	3,34	1,73	2,21	2,90	3,99
	Policealne i średnie	0,76	3,06	1,23	0,45	0,92	1,13	1,37
	Zasadnicze zawodowe	0,66	1,72	0,95	0,26	0,75	0,88	1,07
	Podstawowe i brak	0,64	1,82	0,94	0,24	0,74	0,93	1,06

Źródło: opracowanie własne.

W przypadku estymatora bezpośredniego wyjątkiem jest kategoria wykształcenia wyższego, bowiem widać wyraźnie, że występują powiaty, dla których wnioskowanie wymaga pewnej ostrożności, z tego względu, że w ponad połowie powiatów średni względny błąd szacunku estymatora bezpośredniego dla oszacowania udziału osób niepełnosprawnych prawnie i biologicznie z wykształceniem wyższym w liczbie ludności w wieku 15 lat i więcej przekracza postulowane 10%. Estymatorem charakteryzującym się zdecydowanie najgorszymi charakterystykami jakości precyzji był SYNTA (dotyczy to w zasadzie kategorii wykształcenia wyższego).

Rysunek 5. Ocena precyzji estymatorów (CV) zastosowanych do oszacowania udziału osób niepełnosprawnych prawnie i biologicznie w liczbie ludności powiatów woj. wielkopolskiego w wieku 15 lat i więcej dla czterech kategorii wykształcenia



Źródło: opracowanie własne.

7. WNIOSKI

W artykule przedstawiono propozycję aplikacji statystyki małych obszarów, w szacowaniu niepełnosprawności. Jest to ten obszar badawczy, w którym wykorzystanie estymacji pośredniej w polskiej literaturze jest wciąż niewystarczające. Wyniki przeprowadzonego badania wskazują, że zastosowanie estymatorów klasy SMO umożliwia oszacowanie odsetka osób niepełnosprawnych w przekroju powiatów województwa wielkopolskiego i dla określonych kategorii wykształcenia z wyższą precyzją. Na uwagę zasługuje fakt, że jest to poziom agregacji, który nie był do tej pory przedmiotem zainteresowania statystyki publicznej ze względu na nieakceptowalne błędy szacunków estymatora bezpośredniego dla tak szczegółowo zdefiniowanej domeny. Wyniki badania empirycznego wskazały, że poziom tego zjawiska mierzony odsetkiem osób niepełnosprawnych biologicznie i prawnie w grupie wieku 15+ w przekroju powiatów województwa wielkopolskiego i dla odpowiednich kategorii wykształcenia jest zbliżony dla estymatora bezpośredniego i rozpatrywanych estymatorów pośrednich. Zauważone różnice dotyczą jednak średnich względnych błędów szacunku (CV) estymatorów uwzględnionych w badaniu. Z wyjątkiem estymatora syntetycznego typu A

wszystkie wzięte pod uwagę estymatory klasy SMO odznaczały się lepszą precyzją (mniejsze *CV*) w porównaniu z estymatorem bezpośrednim stosowanym na ogół w tego typu badaniach prowadzonych przez krajowe urzędy statystyczne. Warto również zwrócić uwagę, że średnia ocena rozpatrywanego w artykule parametru jakości estymatorów pośrednich (z wyjątkiem estymatora syntetycznego typu A) nie przekracza progu 10%, a jest to zwyczajowo przyjmowana granica wiarygodnych szacunków. Jak wykazało badanie empiryczne w odniesieniu do estymatora bezpośredniego, nie dla wszystkich domen (przekrój powiatu i kategorii wykształcenia wyższego), postulat ten był spełniony. Statystyka małych obszarów może zatem stanowić swego rodzaju remedium na niską precyzję szacunków charakterystyczną dla estymatora bezpośredniego dla szczegółowo zdefiniowanych domen. Dzięki jej zastosowaniu możliwe jest ponadto zapewnienie pokrycia informacyjnego w obszarze niepełnosprawności, który stanowi kluczowy element prowadzenia odpowiedniej polityki społecznej, zwłaszcza w kontekście popytu zgłaszanego przez władze szczebla lokalnego.

W dalszych pracach badawczych należy rozważyć problem, który nie był przedmiotem niniejszego opracowania, ale w kontekście badań nad niepełnosprawnością jest szczególnie istotny (dotyczył zwłaszcza danych z NSP 2011). Chodzi tutaj o wysoki udział braków danych oraz tych sytuacji, w których respondent odmawia odpowiedzi na pytania kwestionariusza spisowego. Z pewnością w kontekście uogólniania wyników NSP 2011 należy podjąć prace badawcze, które uwzględnią problem estymacji poziomu niepełnosprawności w sytuacji istnienia różnych mechanizmów generowania braków danych (ang. *nonresponse*). Kolejnym problemem badawczym, który należałoby podjąć w dalszych rozważaniach jest estymacja innych aniżeli frakcja parametrów populacji generalnych. Przykładowo, szczególnie interesująca byłaby informacja o liczbie osób niepełnosprawnych (wartość globalna) w przekroju powiatów z uwzględnieniem kategorii wykształcenia. W tym celu warto rozważyć zastosowanie estymatorów typu SPREE (ang. *Structure Preserving Estimation*), które nie tylko umożliwiają uzyskanie szacunków dla niższych poziomów przestrzennych bądź subpopulacji z akceptowalną precyzją, ale również zapewniają spełnienie często wymaganego w statystyce małych obszarów postulatu tzw. benchmarkingu. Mówi on o tym, że uzyskane oszacowania dla niższego poziomu sumują się do szacunków uzyskanych z wykorzystaniem estymatora bezpośredniego dla wyższego poziomu. Wyniki estymacji typu SPREE są już publikowane przez niektóre krajowe urzędy statystyczne w ramach tzw. statystyki oficjalnej. W przypadku estymatorów klasy SMO, również rozważanych w tym artykule, postulat ten nie zawsze jest spełniony. W kontekście przyszłych prac warto również rozważyć inne domeny (na przykład przekrój powiatu i aktywności zawodowej osób niepełnosprawnych). Informacja taka byłaby szczególnie pożądana przez instytucje samorządowe zajmujące się lokalnymi rynkami pracy. Kluczowym zagadnieniem jest również ocena samego obciążenia rozpatrywanych estymatorów klasy SMO, która nie była przedmiotem rozważań w niniejszym artykule.

LITERATURA

- Dehnel G., (2010), *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*, Uniwersytet Ekonomiczny w Poznaniu, Poznań.
- Dehnel G., Klimanek T., (2016), Disability in the National Censuses of 2002 and 2011 – a Comparison of Information Scope, *Acta Universitatis Lodzianis, Folia Oeconomica*, 5 (325), 127–141.
- Elazar D., Conn L., (2004), Small Area Estimation of Disability in Australia, *Research Paper*, Australian Bureau of Statistics.
- Fabrizi E., Montanari G. E., Ranalli G. M., (2015), A Hierarchical Latent Class Model for Predicting Disability Small Area Counts from Survey Data, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 179 (1), 1–29.
- Gołata E., (2004), *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, Wydawnictwo Akademii Ekonomicznej w Poznaniu.
- GUS (2011), *Metodologia badania Budżetów Gospodarstw Domowych*, Warszawa.
- GUS (2012), *Raport z wyników. Narodowy Spis Powszechny Ludności i Mieszkań 2011*, Warszawa.
- GUS (2013), *Ludność. Stan i struktura demograficzno-społeczna. Narodowy Spis Powszechny Ludności i Mieszkań 2011*, Zakład Wydawnictw Statystycznych, Warszawa.
- GUS (2016a), *Aktywność ekonomiczna ludności Polski III kwartał 2016*, Główny Urząd Statystyczny, Warszawa.
- GUS (2016b), *Budżety Gospodarstw Domowych w 2015 roku*, Główny Urząd Statystyczny, Warszawa.
- GUS (2016c), *Stan zdrowia ludności Polski w 2014 r.*, Główny Urząd Statystyczny, Warszawa.
- GUS (2017), *Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2015)*, Główny Urząd Statystyczny, Warszawa.
- Józefowski T., Grygiel G., Klimanek T., (2016), Indirect Estimation of Disabled Persons – a Simulation Study, Small Area Estimation Conference (SAE 2016), Maastricht, Holandia.
- Józefowski T., Szymkowiak M. (2013), Zastosowanie estymatora typu SPREE w szacowaniu liczby osób bezrobotnych w przekroju podregionów, *Studia Oeconomica Posnaniensia*, 1, 10 (259), 120–135.
- Klimanek T., (2017), Estymacja pośrednia niepełnosprawności – badanie symulacyjne, *Taksonomia 29 Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 469, 91–99.
- Komisja Europejska (2010), *Europejska strategia w sprawie niepełnosprawności 2010–2020: Odnowione zobowiązanie do budowania Europy bez barier*, Bruksela.
- Marchetti S., Beręsewicz M., Salvati N., Szymkowiak M., Wawrowski Ł., (2018), The Use of a Three Level M-quantile Model to Map Poverty at LAU 1 in Poland, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 181 (4), 1077–1104.
- ONS (2004), *Eurarea Project – Enhancing Small Area Estimation Techniques to Meet European Needs*, Project documentation.
- Rao J. N. K., Molina I., (2015), *Small Area Estimation*, 2nd Edition, Wiley Series in Survey Methodology.
- Szymkowiak M., Józefowski T., Klimanek T., (2016), An Application of Log-linear Models to Disability in Poland, ICEA – FAA 2016, Bukareszt, Rumunia.
- Szymkowiak M., Młodak A., Wawrowski Ł., (2017), Mapping Poverty at the Level of Subregions in Poland Using Indirect Estimation, *Statistics in Transition - new series*, 18 (4), 609–635.
- Urząd Statystyczny w Poznaniu (2013), *Mapy ubóstwa na poziomie podregionów w Polsce z wykorzystaniem estymacji pośredniej*, raport badawczy.
- Żądło T., (2015), *Statystyka małych obszarów w badaniach ekonomicznych, Podejście modelowe i mieszane*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

BADANIE ZJAWISKA NIEPEŁNOSPRAWNOŚCI W PRZEKROJU POWIATÓW WOJEWÓDZTWA WIELKOPOLSKIEGO Z WYKORZYSTANIEM METOD STATYSTYKI MAŁYCH OBSZARÓW

Streszczenie

Głównym źródłem informacji w statystyce publicznej na temat niepełnosprawności są badania reprezentacyjne oraz spisy powszechne realizowane przez Główny Urząd Statystyczny. Ze względu na fakt, że liczebności próby w odpowiednich domenach, w odniesieniu do których wyrażane są potrzeby informacyjne dotyczące zjawiska niepełnosprawności przez różnych odbiorców, są zbyt małe aby skorzystać z klasycznych technik estymacji, wyniki publikowane są zazwyczaj na dość wysokim poziomie agregacji (kraj czy województwo) bądź dla dość ogólnie zdefiniowanych przekrojów.

W artykule autorzy wychodząc naprzeciw rosnącym oczekiwaniom na szczegółową informację na temat niepełnosprawności podejmują próbę zastosowania metod jakie oferuje statystyka małych obszarów na potrzeby estymacji odsetka osób niepełnosprawnych biologicznie i prawnie w powiatach województwa wielkopolskiego z uwzględnieniem wykształcenia. Na potrzeby realizacji tak postawionego celu wykorzystano dane z Narodowego Spisu Powszechnego Ludności i Mieszkań z 2011 roku oraz wybrane techniki estymacji pośredniej.

Wyniki przeprowadzonych analiz pozwoliły ocenić przestrzenne zróżnicowanie zjawiska niepełnosprawności dla tak zdefiniowanej domeny z większą precyzją. Na uwagę zasługuje fakt, że jest to poziom agregacji, który nie był do tej pory przedmiotem zainteresowania statystyki publicznej ze względu na nieakceptowalne błędy szacunków estymatora bezpośredniego.

Słowa kluczowe: niepełnosprawność, statystyka małych obszarów, EURAREA, błąd średniokwadratowy

A STUDY OF DISABILITY ACROSS DISTRICTS OF THE PROVINCE OF WIELKOPOLSKA USING SMALL AREA ESTIMATION METHODS

Abstract

Surveys and censuses conducted by the Central Statistical Office in Poland are the main sources of information about disability for official statistics. Because sample sizes for relevant cross-classification domains are too small to employ classical estimation methods, results are usually published at a relatively high level of aggregation (at country or province level) or for very broadly defined domains.

To meet the growing demand for detailed information about disability, the authors present an attempt of applying the methodology of small area estimation to estimate the percentage of disabled people, in the legal and biological sense, across districts (NUTS 4/LAU 1 units) of the province of Wielkopolska cross-classified by the level of education. This methodological exercise is based on data from the 2011 census and employs selected techniques of indirect estimation.

Estimates obtained in the study provide an indication of the spatial variation of disability in the target domains with greater precision. It is worth noting that this level of aggregation has not been considered for purposes of official statistical outputs because of unacceptably high estimation errors of the direct estimator.

Keywords: disability, small area estimation, EURAREA, mean squared error