

Spisy powszechne

Censuses



Narodowy Spis Powszechny Ludności i Mieszkań 2021
Ocena jakości danych

National Population and Housing Census 2021
Assessment of the data quality

Warszawa | Warsaw 2023

Spisy powszechne

Censuses



Narodowy Spis Powszechny Ludności i Mieszkań 2021

Ocena jakości danych

National Population and Housing Census 2021
Assessment of the data quality

Główny Urząd Statystyczny Statistics Poland
Warszawa Warsaw 2023



Narodowy Spis Powszechny Ludności i Mieszkań 2021

Ocena jakości danych

National Population and Housing Census 2021

Assessment of the data quality

Opracowanie merytoryczne

Content-related works

Główny Urząd Statystyczny,
Statistics Poland

pod kierunkiem
under the direction of

Dominika Rozkruta
Doroty Szaltys
Janusza Dygaszewicza

Zespół autorski

Author team

Agnieszka Matulska-Bachura, Maciej Beręsewicz, Piotr Filip, Emilia Gawińska-Druźba, Katarzyna Góral-Radziszewska, Grzegorz Gudaszewski, Ewa Kacperczyk, Paweł Murawski, Gabriela Nowakowska, Maciej Potyra, Dorota Szaltys, Marcin Szymkowiak, Kamil Waśkiewicz, Robert Wiczorkowski, Kamil Wilak, Krzysztof Woźnica, Anna Wysocka

Skład i opracowanie graficzne

Layout and graphic design

Łukasz Nowak, Katarzyna Stanecka

przy udziale

support

Ewy Morytz-Balskiej, Elżbiety Lisiak

ISBN 978-83-66466-95-1

Publikacja dostępna na stronie

Publication available at

stat.gov.pl

Przy publikowaniu danych GUS prosimy o podanie źródła.

When publishing Statistics Poland data – please indicate the source.

Przedmowa

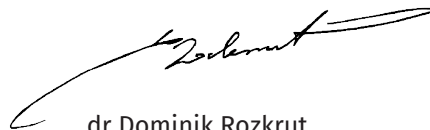
Ocena jakości spisu ludności i mieszkań jest nieodłącznym komponentem całego procesu realizacji badania uwzględniającym zarówno kwestie organizacyjne, jak i metodologiczne tego przedsięwzięcia. Dokonana ewaluacja powinna odpowiedzieć na pytanie, czy przyjęta strategia realizacji spisu ludności i mieszkań w kontekście poniesionych kosztów związanych z infrastrukturą, w tym informatyczną, zasobami ludzkimi, popularyzacją i promocją przyniosła oczekiwany rezultat. Dodatkowo powinna ona także być pomocna w zidentyfikowaniu słabych stron przedsięwzięcia i opracowaniu rekomendacji dla przyszłych spisów w celu poprawy ich efektywności i skuteczności. Ocena jakości wyników spisu powszechnego powinna natomiast dostarczyć potencjalnym użytkownikom tak krajowym, jak i zagranicznym odpowiednie miary jakości pozwalające na właściwą interpretację danych.

Niniejsza publikacja jest kontynuacją opracowania zatytułowanego „Narodowy Spis Powszechny Ludności i Mieszkań 2021. Metodologia i Organizacja badania”. W wydanym w styczniu 2022 r. raporcie przedstawiono kluczowe aspekty dotyczące realizacji spisu ludności i mieszkań w 2021 r. ze szczególnym uwzględnieniem przeprowadzonych konsultacji społecznych, metod realizacji spisu, wykorzystanych źródeł danych, systemów informatycznych oraz aplikacji wspierających realizację badania. W raporcie tym zaprezentowano również procesy przetwarzania danych, charakterystykę form i narzędzi udostępniania wyników, strategię promocji i popularyzacji spisu oraz ramowy harmonogram udostępniania wyników.

W obecnym opracowaniu skoncentrowano się na ocenie jakości wyników NSP 2021 m.in. z wykorzystaniem analizy demograficznej. Raport zawiera także wyniki badania kontrolnego, które dostarczyło informacji o błędach treści oraz poziomie zgodności dla analizowanych cech. Istotną częścią raportu jest także analiza i ocena Wykazu osobowo-adresowo-mieszaniowego, który stanowił podstawę do realizacji badania spisowego.

Przekazując do rąk Państwa publikację wyrażam nadzieję, że zaprezentowane w ramach niniejszego opracowania informacje dotyczące wybranych aspektów jakości wyników uzyskanych w ramach NSP 2021 będą pomocne w interpretacji kolejno udostępnianych zestawów danych wynikowych ze spisu, jak również w analizach porównawczych z wynikami opracowywanymi w ramach bieżących badań statystyki publicznej.

Prezes Głównego Urzędu Statystycznego
Generalny Komisarz Spisowy



dr Dominik Rozkrut

Warszawa, kwiecień 2023 r.

Preface

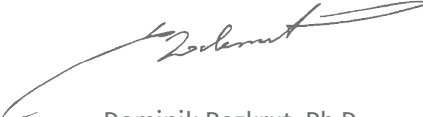
The quality assessment of the population and housing census is an inseparable component of the entire process of survey completion, taking into both organizational as well as methodological issues of this project. The conducted evaluation should answer the question whether the adopted strategy for the implementation of the population and housing census in the context of incurred costs related to infrastructure, including IT infrastructure, human resources, popularization and promotion, brought the expected result. In addition, it should also be helpful in identifying the weaknesses of project and developing recommendations for future censuses in order to improve their effectiveness and efficiency. The quality assessment of census results should, however, provide potential users, both domestic and foreign, with appropriate quality measures allowing for proper interpreting data.

This publication is the continuation of study entitled „National Population and Housing Census 2021. Research methodology and organization”. The report published in January 2022 covers the key aspects regarding the implementation of the population and housing census in 2021, with particular considering the public consultations, methods of census implementation, used data sources, IT systems and applications supporting the implementation of survey. The report also presents the processes of data processing, characteristics of the forms and tools used for disseminating the results, the strategy for promotion and popularization of census, as well as the framework schedule for release of result information.

The current study focuses on assessing the quality of 2021 Census results among others using demographic analysis. The report also contains the results of the control survey, which provided information on the content errors and the level of compliance for the analysed features. An important part of the paper is also the analysis and assessment of the List of persons, addresses and dwellings, which was the basis for the census study.

Handing over the publication to you, I hope that information presented in this study on selected aspects related to the quality of results obtained under the 2021 Census will be helpful in interpreting consecutively disseminated sets of census' results, as well as in comparative analysis with the results compiled as part of current surveys of official statistics.

President Statistics Poland
Census General Commissioner



Dominik Rozkrut, Ph.D.

Warsaw, April 2023

Spis treści

Contents

Przedmowa.....	5
Preface	6
Spis tablic.....	10
List of tables	10
Spis wykresów i schematów	13
List of charts and schemes	13
Spis map.....	15
List of maps	15
Ważniejsze skróty	16
Main abbreviations.....	16
Wprowadzenie.....	18
Introduction.....	21
I. Wykaz osobowo-adresowo-mieszkaniowy	24
I. List of persons, addresses and dwellings.....	24
1.1. Wykaz adresowo-mieszkaniowy	25
1.1. List of addresses and dwellings	25
1.2. Wykaz osób	32
1.2. List of persons.....	32
II. Ocena jakości wykazu mieszkań oraz osób.....	37
II. Assessing the quality of the housing and persons list.....	37
2.1. Mieszkania	37
2.1. Dwellings	37
2.1.1. Podsumowanie.....	41
2.1.2. Summary.....	42
2.2. Osoby	42
2.2. Persons.....	42
2.2.1. Podsumowanie i wnioski	48
2.2.1. Summary and conclusion	49
III. Dokładność wyników spisu ludności i mieszkań.....	51
III. Accuracy of the 2021 Census results	51
3.1. Klasyfikacje błędów	51
3.1. Classification Error	51
3.2. Wybrane metody oceny jakości spisów	51
3.2. Selected methods of assessing the quality of censuses.....	51

Spis treści (cd.)

Contents (cont.)

IV. Badanie kontrolne w spisie ludności i mieszkań 2021	54
IV. The control survey in the 2021 Census	54
4.1. Opis losowania próby do badania kontrolnego w NSP 2021	55
4.1. Description of drawing the sample for the control survey in the 2021 Census	55
4.2. Analiza wyników badania kontrolnego	55
4.2. Analysis of the results of the control survey	55
4.2.1. Przyjęte grupowania	57
4.2.1. Adopted groupings	57
4.2.2. Zmienne wybrane do analizy	59
4.2.2. Variables selected for analysis	59
4.2.3. Miary zgodności	62
4.2.3. Measures of consistency	62
4.2.4. Wnioski	77
4.2.5. Conclusions	78
V. Ocena kompletności wyników NSP 2021 z wykorzystaniem metod analizy demograficznej	80
V. Results completeness assessment of the 2021 Census using demographic analysis methods	80
5.1. Metodologia	81
5.1. Methodology	81
5.2. Wyniki	83
5.2. Results	83
5.2.1. Poziom krajowy	83
5.2.1. National level	83
5.2.1.1. Kompletność szacunków liczby niemowląt	88
5.2.1.1. Completeness estimation of the infants' number	88
5.2.1.2. Kompletność szacunków liczby dzieci w wieku 0–9 lat	90
5.2.1.2. Completeness estimation of the number of children aged 0–9 years	90
5.2.1.3. Kompletność szacunków liczby osób w wieku podwyższonej mobilności	91
5.2.1.3. Completeness estimation of the number of persons at the age with the higher mobility	91
5.2.1.4. Kompletność szacunków dla osób w wieku 90 lat i więcej	92
5.2.1.4. Completeness estimation for the persons aged 90 years and more	92
5.2.2. Poziom wojewódzki	93
5.2.2. Voivodship level	93
5.2.3. Podsumowanie	98
5.2.4. Summary	99

Spis treści (dok.) Contents (cont.)

VI. Ocena jakości wyników spisu w zakresie mieszkań	101
VI. Quality assessment of the Census 2021's results on housing.	101
6.1. Metodologia	101
6.1. Methodology	101
6.2. Wyniki	103
6.2. Results	103
6.3. Podsumowanie	107
6.3. Summary	108
Bibliografia	109
Bibliography	109

Spis tablic

List of tables

Tablica 1. Mieszkania wytypowane do Wykazu adresowo-mieszkaniowego na podstawie NOBC oraz źródeł administracyjnych (ZA) podlegające obowiązkowej weryfikacji.	27
Table 1. Dwellings selected for List of addresses and dwellings based on the NOBC as well as the administrative sources (ZA) subject to verification	
Tablica 2. Wyniki obowiązkowej weryfikacji mieszkań wytypowanych do Wykazu adresowo-mieszkaniowego na podstawie NOBC	29
Table 2. Results of mandatory verification of dwellings selected for List of addresses and dwellings based on the NOBC	
Tablica 3. Wyniki obowiązkowej weryfikacji mieszkań wytypowanych do Wykazu adresowo-mieszkaniowego na podstawie źródeł administracyjnych (ZA).....	30
Table 3. Results of mandatory verification of dwellings selected for List of addresses and dwellings based on the administrative sources (ZA)	
Tablica 4. Mieszkania dopisane do NOBC oraz usunięte z NOBC w ramach weryfikacji (obowiązkowa i dodatkowa łącznie).....	31
Table 4. Dwelling added to the NOBC and delated from the NOBC under the verification (mandatory and facultative combined)	
Tablica 5. Opis wykorzystanych źródeł danych administracyjnych	32
Table 5. Description of the administrative data sources used	
Tablica 6. Kolejność dołączania adresów	34
Table 6. Order of the adding addresses	
Tablica 7. Algorytmy modyfikacji danych podczas procesu parowania.....	36
Table 7. Data modification algorithms during the pair production process	
Tablica 8. Liczba mieszkań w tabeli MIESZKANIA Wykazu osobowo-adresowo-mieszkaniowego.....	37
Table 8. Number of dwellings in DWELLINGS table in List of persons, addresses and dwellings	
Tablica 9. Liczba mieszkań w populacji objętej spisem oraz docelowej populacji spisu.....	39
Table 9. Number of dwellings in census population as well as target population	
Tablica 10. Wykazowa populacja docelowa według kompletności adresu	43
Table 10. List's target population by completeness of address	
Tablica 11. Wykazowa populacja docelowa według pochodzenia adresu.....	44
Table 11. List's target population by origin of address	
Tablica 12. Oszacowana populacja docelowa według kategorii ludności oraz źródeł danych	45
Table 12. Estimated target population by population category and data source	
Tablica 13. Wykazowa populacja docelowa według pochodzenia adresu.....	47
Table 13. List's target population by origin of address	
Tablica 14. Oszacowana populacja docelowa według zgodności z wykazową populacją docelową.....	47
Table 14. Estimated target population by consistency with the list's target population	

Spis tablic (cd.)

List of tables (cont.)

Tablica 15. Kody kanałów realizacji spisu zasadniczego	57
Table 15. Modes used for collecting data in census	
Tablica 16. Słownik dla kodów województw	58
Table 16. Dictionary for voivodship codes	
Tablica 17. Słownik dla cechy PS_4	58
Table 17. Dictionary for the variable PS_4	
Tablica 18. Lista analizowanych zmiennych dotyczących osób	59
Table 18. List of analyzed variables concerning persons	
Tablica 19. Rozkład najliczniej reprezentowanych kategorii odpowiedzi na pytanie o narodowość w spisie zasadniczym (wśród respondentów badania kontrolnego)	60
Table 19. Distribution of the most frequently represented categories of answers to the question on nationality in the census (among respondents of the control survey)	
Tablica 20. Rozkład najliczniej reprezentowanych kategorii odpowiedzi na pytanie o narodowość w badaniu kontrolnym	60
Table 20. Distribution of the most frequently represented categories of answers to the question on the nationality in the control survey	
Tablica 21. Lista analizowanych zmiennych dotyczących mieszkań	60
Table 21. List of analyzed variables concerning dwellings	
Tablica 22. Miary zgodności spisu zasadniczego i badania kontrolnego na poziomie Polski ogółem dla cech osobowych	62
Table 22. Measures of consistency in the census and control survey at the country level for personal characteristics	
Tablica 23. Miary zgodności spisu zasadniczego i badania kontrolnego na poziomie Polski ogółem dla cech dotyczących mieszkań	62
Table 23. Measures of consistency in the census and control survey at the country level for dwelling characteristics	
Tablica 24. Wartości uogólnień na podstawie badania kontrolnego dla zmiennych dotyczących osób	63
Table 24. Estimates based on control survey for the variables concerning persons	
Tablica 25. Wartości uogólnień na podstawie badania kontrolnego dla zmiennych dotyczących mieszkań	64
Table 25. Estimates based on control survey for the variables concerning dwellings	
Tablica 26. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością obliczoną metodą szacunku wstecznego	84
Table 26. Comparison of the population balance (as of 31 December) with the population calculated using the retrospective estimation method	
Tablica 27. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością według płci obliczoną metodą szacunku wstecznego	86
Table 27. Comparison of the population balance (as of 31 December) with the population by sex calculated using the retrospective estimation method	

Spis tablic (dok.)

List of tables (cont.)

Tablica 28. Porównanie ludności bilansowej (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego z uwzględnieniem płci i 5-letnich grup wieku.	87
Table 28. Comparison of the population balance (as of 31 December 2020) with the population by sex and 5-years age groups calculated using the retrospective estimation method	
Tablica 29. Analiza kompletności liczby niemowląt dla NSP 2011 oraz NSP 2021.	89
Table 29. Completeness analysis of the infants' number in the 2011 Census and the 2021 Census	
Tablica 30. Analiza kompletności liczby dzieci w wieku 0–9 lat spisanych w NSP 2021.	90
Table 30. Completeness analysis of the number of children aged 0–9 enumerated in the 2021 Census	
Tablica 31. Porównanie ludności bilansowej w wieku 25–40 lat według płci (stan na 31 grudnia 2020 r.) – z ludnością obliczoną metodą szacunku wstecznego.	92
Table 31. Comparison of the population balance aged 25–40 by sex (as of 31 December 2020) with the population calculated using the retrospective estimation method	
Tablica 32. Porównanie ludności bilansowej (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego dla osób w wieku 90 lat i więcej według płci i wieku.	93
Table 32. Comparison of the population balance (as of 31 December 2020) with the population calculated using the retrospective estimation for the persons aged 90 years and more by sex and age	
Tablica 33. Porównanie ludności bilansowej według płci i województw (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego.	94
Table 33. Comparison of the population balance by sex and voivodships (as of 31 December 2020) with the population based on the retrospective estimation method	
Tablica 34. Różnica w tysiącach osób oraz różnica procentowa między ludnością bilansową a ludnością obliczoną metodą szacunku wstecznego z uwzględnieniem 15-letnich grup wieku oraz województw.	97
Table 34. The absolute (in thousands of persons) and the relative difference between the population balance and the population developed using the retrospective estimation method by 15-years age groups and voivodships	
Tablica 35. Oznaczenia.	101
Table 35. Indications	
Tablica 36. Liczba mieszkań (stan na 31 grudnia) opracowana na bazie wyników NSP 2011 oraz metodą szacunku wstecznego przy podstawie NSP 2021 dla Polski ogółem dla lat 2011–2020.	104
Table 36. Number of dwellings (as of 31 December) using the results of the 2011 Census and calculated using the retrospective estimation method on the basis of the 2021 Census in Poland for years 2011–2021	
Tablica 37. Liczba mieszkań (stan na 31 grudnia 2020 r.) opracowana na bazie wyników NSP 2011 oraz metodą szacunku wstecznego przy podstawie NSP 2021 według województw.	105
Table 37. Number of dwellings (as of 31 December 2020) obtained on the basis of the 2011 Census and calculated using the retrospective estimation method based on the 2021 Census at the voivodship level	

Spis wykresów i schematów

List of charts and schemes

Wykres 1. Odsetek mieszkań niedoszacowanych i przeszacowanych w stosunku do populacji mieszkań objętej spisem (w %)	40
Chart 1. Percentage of underestimated and overestimated dwellings in total census population (in %)	
Wykres 2. Struktura adresów mieszkań według kanałów zbierania danych (w %)	41
Chart 2. Structure of dwellings' addresses by channels of data collection (in %)	
Wykres 3. Wskaźniki zgodności BWZ dla wybranych zmiennych	65
Chart 3. Measures of consistency BWZ for selected variables	
Wykres 4. Wskaźniki zgodności BWZ dla wybranych zmiennych według kanałów realizacji spisu zasadniczego	66
Chart 4. Measures of consistency BWZ for selected variables according to modes of contact in census	
Wykres 5. Wskaźniki zgodności BWZ dla wybranych zmiennych według osoby odpowiadającej na pytania spisu zasadniczego	66
Chart 5. Measures of consistency BWZ for selected variables according to the person answering the census questions	
Wykres 6. Wskaźniki zgodności BWZ dla wybranych zmiennych według podziału na miasto/wieś	67
Chart 6. Measures of consistency BWZ for selected variables according to urban/rural division	
Wykres 7. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących osób według województw ..	67
Chart 7. Measures of consistency BWZ for selected personal variables by voivodships	
Wykres 8. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących mieszkań według województw	68
Chart 8. Measures of consistency BWZ for selected dwelling variables by voivodships	
Wykres 9. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących osób według grup wieku ...	68
Chart 9. Measures of consistency BWZ for selected personal variables by age groups	
Wykres 10. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących mieszkań według grup wieku	69
Chart 10. Measures of consistency BWZ for selected dwelling variables by age groups	
Wykres 11. Wskaźniki zgodności IZ dla wybranych zmiennych	70
Chart 11. Measures of consistency IZ for selected variables	
Wykres 12. Wskaźniki zgodności IZ dla wybranych zmiennych według kanałów realizacji spisu zasadniczego	71
Chart 12. Measures of consistency IZ for selected variables according to modes of contact in census	
Wykres 13. Wskaźniki zgodności IZ dla wybranych zmiennych według osoby odpowiadającej na pytania spisu zasadniczego	71
Chart 13. Measures of consistency IZ for selected variables according to the person answering the census questions	
Wykres 14. Wskaźniki zgodności IZ dla wybranych zmiennych według podziału na miasto/wieś	72
Chart 14. Measures of consistency IZ for selected variables according to urban/rural division	

Spis wykresów i schematów (dok.)

List of charts and schemes (cont.)

Wykres 15. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących osób według województw	72
Chart 15. Measures of consistency IZ for selected personal variables by voivodships	
Wykres 16. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących mieszkań według województw	73
Chart 16. Measures of consistency IZ for selected dwelling variables by voivodships	
Wykres 17. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących osób według grup wieku	73
Chart 17. Measures of consistency IZ for selected personal variables by age groups	
Wykres 18. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących mieszkań według grup wieku .	74
Chart 18. Measures of consistency IZ for selected dwelling variables by age groups	
Wykres 19. Wskaźniki zgodności CCC dla wybranych zmiennych.	75
Chart 19. Measures of consistency CCC for selected variables	
Wykres 20. Wskaźniki zgodności BWZ dla wybranych zmiennych według występowania braków dla cechy L05b (kod 1 na osi x odpowiada podzbiorowi danych z brakami dla cechy L05b, 0 – podzbiorowi bez braków), część 1.	76
Chart 20. Measures of consistency BWZ for selected variables according to the occurrence of missing data for variable L05b (code 1 on the x axis corresponds to the subset with missing data, 0 – the subset without missing data), part 1	
Wykres 21. Wskaźniki zgodności IZ dla wybranych zmiennych według występowania braków dla cechy L05b (kod 1 na osi x odpowiada podzbiorowi danych z brakami dla cechy L05b, 0 – podzbiorowi bez braków), część 2.	77
Chart 21. Measures of consistency BWZ for selected variables according to the occurrence of missing data for variable L05b (code 1 on the x axis corresponds to the subset with missing data, 0 – the subset without missing data), part 2	
Wykres 22. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością obliczoną metodą szacunku wstecznego.	84
Chart 22. Comparison of the population balance (as of 31 December) with the population calculated using the retrospective estimation method	
Wykres 23. Porównanie liczby mieszkań (stan na 31 grudnia) otrzymanej na bazie wyników NSP 2011 z liczbą mieszkań obliczoną metodą szacunku wstecznego przy podstawie NSP 2021.	105
Chart 23. Comparison of the number of dwellings (as of 31 December) using the results of the 2011 Census with the number of dwellings calculated using the retrospective estimation method on the basis of the 2021 Census	
Schemat 1. Mapowanie kluczy łączenia adresów z Wykazu osób z bazą NOBC	35
Scheme 1. Mapping of address linkage keys from the List of persons with the NOBC database	
Schemat 2. Schemat przeprowadzonego szacunku wstecznego	82
Scheme 2. Scheme of the retrospective estimation	

Spis map

List of maps

- Mapa 1. Różnica procentowa między ludnością bilansową (stan na 31 grudnia 2020 r.) a ludnością szacunkową (stan na 31 grudnia 2020 r.) według województw 96
- Map 1. Relative difference between the population balance (as of 31 December, 2020) and the estimated population (as of 31 December 2020) by voivodships
- Mapa 2. Różnica procentowa między liczbą mieszkań opracowaną na podstawie NSP 2011, a szacunkową liczbą mieszkań opracowaną przy podstawie NSP 2021 według województw według stanu na 31 grudnia 2020 r. 107
- Map 2. Relative difference between the number of dwellings based on the 2011 Census and the estimated number of dwellings based on the 2021 Census by voivodships as of 31 December 2020

Ważniejsze skróty

Main abbreviations

Skrót Abbreviation	Wyjaśnienie Explanation
ARIMR	Agencja Restrukturyzacji i Modernizacji Rolnictwa
BJS	Baza Jednostek Statystycznych.
CAPI, CAPI-PHONE, CAPI-PERSONAL	ang. Computer Assisted Personal Interview; wywiad bezpośredni prowadzony przez rachmistrza terenowego z użyciem interaktywnej aplikacji formularzowej na urządzeniu mobilnym. Wywiad mógł być prowadzony także jako wywiad telefoniczny.
CATI	ang. Computer Assisted Telephone Interview; wywiad telefoniczny prowadzony przez ankietera telefonicznego przy wykorzystaniu specjalistycznego oprogramowania komputerowego.
CATI2	ang. Computer Assisted Telephone Interview; wywiad telefoniczny prowadzony przez ankietera telefonicznego przy wykorzystaniu specjalistycznego oprogramowania komputerowego – na żądanie respondenta lub jako wsparcie CAPI.
CAWI	ang. Computer-Assisted Web Interview – wspomagany komputerowo wywiad przy pomocy strony WWW.
CAxI	łącznie: metody CAWI, CATI, CAPI i „Spis na żądanie”.
CRP-KEP	Centralny Rejestr Podmiotów – Krajowa Ewidencja Podatników
EMUiA	Ewidencja miejscowości, ulic i adresów
Eurostat	Europejski Urząd Statystyczny
GUS	Główny Urząd Statystyczny
UNS	Unikatowy numer statystyczny
KEP	Krajowa Ewidencja Podatników
KPRM	Kancelaria Prezesa Rady Ministrów
KRUS	Kasa Rolniczego Ubezpieczenia Społecznego
KRUS nSIU	nowy System Informatyczny Ubezpieczeń (nSIU) – Kasa Rolniczego Ubezpieczenia Społecznego
KSI CRU/ZUS	Kompleksowy System Informatyczny – Centralny Rejestr Ubezpieczonych Zakładu Ubezpieczeń Społecznych
NFZ	Narodowy Fundusz Zdrowia
CWU NFZ	Centralny Wykaz Ubezpieczonych – Narodowy Fundusz Zdrowia
NOBC	System identyfikacji adresowej ulic, nieruchomości, budynków i mieszkań
NSP 2021	Narodowy Spis Powszechny Ludności i Mieszkań 2021
OBM	Operacyjna Baza Mikrodanych
ONZ	Organizacja Narodów Zjednoczonych

Ważniejsze skróty (dok.)

Main abbreviations (cont.)

Skrót Abbreviation	Wyjaśnienie Explanation
OZZ	Obiekt zbiorowego zakwaterowania
PBA	Przestrzenna Baza Adresowa
PESEL	Powszechny Elektroniczny System Ewidencji Ludności
PESEL-WYMELD	Powszechny Elektroniczny System Ewidencji Ludności. Osoby wymeldowane donikąd, które przed wymeldowaniem miały stałe miejsce pobytu w Polsce
PPE	Punkt poboru energii zgodnie z ustawą z dnia 10 kwietnia 1997 r. Prawo energetyczne (z późniejszymi zmianami)
PRG	Państwowy rejestr granic i powierzchni jednostek podziałów terytorialnych kraju
SAS	Statistical Analysis System, pakiet oprogramowania statystycznego opracowany przez SAS Institute do zarządzania danymi, zaawansowanych analiz, analiz wielowymiarowych
TERYT	Krajowy rejestr urzędowy podziału terytorialnego
UG	Urząd gminy
US	Urząd statystyczny
WZDS	Wynikowy zbiór danych spisowych
ZA	Źródła administracyjne
Zbiór Energetyka, ZE	Systemy informacyjne przedsiębiorstw wykonujących działalność gospodarczą w zakresie sprzedaży energii elektrycznej
ZUS	Zakład Ubezpieczeń Spotecznych

Wprowadzenie

Spis ludności i mieszkań jest najlepszą okazją do zebrania pełnych informacji o liczbie i strukturze ludności, a także zasobach mieszkaniowo-budynkowych oraz tendencjach ich rozwoju. Dane stanowią podstawę do podejmowania niezliczonych kluczowych decyzji w zakresie programowania (planowania) społeczno-gospodarczego pomiędzy kolejnymi spisami zarówno na szczeblu krajowym, jak i międzynarodowym.

Każdy kraj podejmuje we własnym zakresie decyzję odnośnie do metody i wykorzystywanych źródeł dla realizacji badania. W Rozporządzeniu (WE) Parlamentu Europejskiego i Rady nr 763/2008 z dnia 9 lipca 2008 r. w sprawie spisów powszechnych ludności i mieszkań (Dz. Urz. UE L 218 z 13.8.2008)¹ określone zostały szczegółowo źródła, w oparciu o które państwa członkowskie mogą opracowywać dane spisowe. Artykuł 4 wskazuje, że są to:

- a) tradycyjne spisy powszechne,
 - b) spisy powszechne oparte na danych z rejestrów administracyjnych,
 - c) połączenie tradycyjnych spisów powszechnych z badaniami reprezentacyjnymi,
 - d) połączenie spisów powszechnych w oparciu o dane z rejestrów administracyjnych z badaniami reprezentacyjnymi,
 - e) połączenie spisów powszechnych w oparciu o dane z rejestrów administracyjnych z tradycyjnymi spisami powszechnymi,
 - f) połączenie spisów powszechnych w oparciu o dane z rejestrów administracyjnych z badaniami reprezentacyjnymi oraz tradycyjnymi spisami powszechnymi,
- oraz
- g) badania oparte na próbie rotacyjnej (tzw. „kroczące” spisy powszechne).

Ze względu na znaczenie spisu ludności i mieszkań stawiane są często pytania co do jakości otrzymanych wyników, od których w znacznym stopniu zależy ich przydatność (użyteczność) w różnych zastosowaniach. Dane są najbardziej użyteczne wówczas gdy mają wysoką jakość. Dlatego też ocena spisu ludności i mieszkań jest kluczowa przede wszystkim w celu zwiększenia zaufania użytkowników do udostępnianych wyników. Dodatkowo zarządzanie jakością w spisie powszechnym jest niezbędne nie tylko ze względu na potencjalne szerokie grono użytkowników danych, ale także dlatego, że wyniki opracowane na podstawie danych ze spisów ludności i mieszkań są wykorzystywane w krajowym i międzynarodowym systemie statystycznym. Co więcej, spis ludności jest często najlepszym, jeśli nie jedynym źródłem informacji na temat pewnych subpopulacji i małych obszarów.

Pomiar jakości w spisach ludności i mieszkań nie jest łatwym przedsięwzięciem, głównie dlatego, że obecnie nie ma standardowej, jednolitej metody oceny jakości, która miałaby zastosowanie do wszystkich metod spisu, jakie są wykorzystywane przez narodowe instytuty statystyczne.

Najpowszechniej stosowana definicja jakości koncentruje się na „użyteczności” informacji. Innymi słowy, dane same w sobie nie posiadają jakości, lecz jedynie potencjał, który służy do określonego celu (Bafour, Valente, 2012). Dlatego w kontekście spisu jakość należy rozpatrywać wielowymiarowo.

Europejski Urząd Statystyczny (Eurostat) przyjął definicję jakości według normy ISO 8402-1986. Jakość według niej jest zdefiniowana jako: „ogół cech i właściwości produktu lub usługi, który decyduje o zdolności zaspokojenia potrzeb zadeklarowanych lub domyślnych” (Eurostat, 2003). Określona się ją na podstawie 6 komponentów jakości, które obejmują: przydatność, dokładność, aktualność i terminowość, dostępność i przejrzystość, porównywalność oraz spójność.

Poniżej scharakteryzowano każdy z powyższych komponentów w kontekście spisu ludności i mieszkań.

1. **Przydatność informacji** odzwierciedla stopień, w jakim dane spisowe odpowiadają potrzebom użytkowników (m.in. osób prywatnych, instytucji rządowych i samorządowych, ośrodków naukowo-badawczych, insty-

¹ <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32008R0763&from=IT> (dostęp 15 marca 2023 r.)

tucji międzynarodowych). Najczęściej przydatność danych uzyskuje się dzięki przeprowadzonym szerokim konsultacjom zakresu informacyjnego badania, metodologii, projektu formularza, instrukcji itd. Z praktyki wynika jednak, że spełnienie wszystkich potrzeb użytkownika może być bardzo trudne, a czasem wręcz niemożliwe ze względu na ograniczone środki finansowe i organizacyjne (Kordos, 1987). Stąd wynika konieczność zachowania równowagi pomiędzy racjonalnym zaplanowaniem badania spisowego a życzeniami (często sprzecznymi) potencjalnych użytkowników danych spisowych.

2. **Dokładność wyników spisu** wyraża jej bliskość do wartości rzeczywistej (prawdziwej), którą otrzymano by gdyby dane zostały zebrane i opracowane bez żadnych błędów. W praktyce dokładny pomiar jakiegokolwiek zjawiska jest nieosiągalny ze względu na fakt, że żadne narzędzie służące do pomiaru nie jest idealne. Dokładność wyniku jest zatem mierzona za pomocą błędu pomiaru, który w uproszczeniu stanowi różnicę pomiędzy wartością uzyskaną z badania a wartością rzeczywistą, która podlega ocenie. W odniesieniu do spisu ludności i mieszkań, dokładność informacji zależy od zastosowanych metod identyfikacji i kontroli wszelkich błędów, które występują w całym procesie spisu.
3. **Komponent aktualności i terminowości** odnosi się do ram czasowych procesu realizacji spisu, jest to odstęp między datą odniesienia spisu a terminem, w którym wyniki zostają opracowane i upowszechnione. Zrozumiałym jest, że użyteczność wyników badania maleje wraz z upływem czasu od daty jego przeprowadzenia. Nie należy jednak zapominać, że realizacja tego komponentu jest zawsze bardzo czas- i pracochłonna, zwłaszcza gdy spis został zrealizowany łączonymi metodami i w oparciu o wiele źródeł, które wymagają odpowiedniego przetworzenia i integracji, przy jednoczesnym spełnieniu jak najwyższych standardów jakości opracowanych wyników. Skrócenie czasu potrzebnego na odpowiednie przetworzenie, walidację i wdrożenie niezbędnych korekt czy zastosowanie imputacji danych, wreszcie analizę wyników w celu opublikowania wyników w czasie, który dla użytkowników byłby najbardziej akceptowalny (tj. najszybszy), może wpłynąć na jakość udostępnianych wyników. Należy zatem znaleźć kompromis pomiędzy wymiarem dokładności a aktualności i terminowości danych.
4. **Dostępność danych spisowych** odnosi się do dostępności i przejrzystości z jaką dane są udostępniane. Komponent ten obejmuje m.in. formy i narzędzia udostępniania danych spisowych, procedury składania zamówień, czas potrzebny na realizację zamówienia, politykę cenową, dostępność mikro- i makrodanych, otoczenie informacyjne danych, czyli m.in. łączenie ich z odpowiednimi metadanymi (informacje tekstowe, wyjaśnienia, dokumentacja itp.) i formami graficznymi (wykresy, diagramy i mapy), prawa autorskie oraz zakres pomocy dodatkowej oferowanej przez instytucję statystyczną. Dodatkowo w tym wymiarze uwzględnia się strategię kontroli ujawniania danych (zachowanie tajemnicy statystycznej), bez pogarszania jakości wyników.
5. **Komponent porównywalności** ma na celu pomiar wpływu różnic w stosowanych pojęciach i definicjach spisowych przy porównywaniu danych z różnych obszarów geograficznych, dziedzin (tematów spisowych) lub okresów bazowych (data referencyjna spisu).
6. **Spójność danych spisowych** odzwierciedla stopień, w jakim wyniki mogą być łączone z innymi istniejącymi danymi statystycznymi. Zastosowanie standardowych pojęć, definicji i klasyfikacji, w tym przyjętych na szczeblu międzynarodowym, przyczynia się do zwiększonej spójności. Poziomą jakość w wymiarze spójności można ocenić za pomocą programu certyfikacji i walidacji wyników spisowych, dokonując porównania z odpowiednimi danymi pochodzącymi z historycznych (wcześniejszych) spisów ludności i mieszkań, badań statystycznych (reprezentacyjnych) lub źródeł administracyjnych.

Ocena spisu nie musi obejmować wszystkich wymienionych sześciu komponentów by spełnić warunek jakości. Należy jednak mieć świadomość, że działania dążące do poprawy jednego wymiaru mogą skutkować negatywnym wpływem na inny komponent (Baffour, Valente, 2012).

Niniejsze opracowanie stanowi komplementarną część wydanej w styczniu 2022 r. publikacji: „Narodowy Spis powszechny Ludności i Mieszkań 2021. Metodologia i Organizacja badania”, w której przedstawiono kluczowe aspekty dotyczące realizacji spisu ludności i mieszkań w 2021 r. ze szczególnym uwzględnieniem przeprowadzonych konsultacji społecznych, metod realizacji spisu, wykorzystanych źródeł danych, systemów informatycznych oraz aplikacji wspierających realizację badania. W raporcie tym zaprezentowano również procesy przetwarzania

danych, charakterystykę form i narzędzi udostępniania wyników, strategię promocji i popularyzacji spisu oraz ramowy harmonogram udostępniania wyników.

W obecnym opracowaniu skoncentrowano się na dokonaniu oceny jakości wyników spisu ludności i mieszkań 2021, w tym z wykorzystaniem analizy demograficznej.

Publikacja składa się z 6 rozdziałów.

Pierwszy rozdział dotyczy zagadnień związanych z budową wykazu osobowo-adresowo-mieszkaniowego oraz obiektów zbiorowego zakwaterowania.

W drugim rozdziale skoncentrowano się na ocenie jakości wykazu osobowo-adresowo-mieszkaniowego, zarówno w odniesieniu do populacji objętej spisem, jak i mieszkań i budynków.

Trzeci rozdział – o naturze teoretycznej – prezentuje źródła błędów oraz opisuje wybrane metody oceny jakości spisów powszechnych.

W rozdziale czwartym zaprezentowano wyniki badania kontrolnego, w szczególności ocenę błędów zawartości i poziom zgodności dla badanych cech. Do oceny wykorzystano dane zebrane na podstawie specjalnie zaprojektowanego badania reprezentacyjnego. W wyniku analizy opracowano szereg wskaźników statystycznych oraz sformułowano wnioski.

Rozdział piąty poświęcony został ocenie kompletności wyników NSP 2021 z wykorzystaniem metod analizy demograficznej. Badaniem kompletności objęto całą populację z uwzględnieniem noworodków i dzieci, osób najbardziej mobilnych oraz osób najstarszych. Przy wykorzystaniu metod analizy demograficznej wskazano różnice między ludnością otrzymaną na podstawie NSP 2021, a ludnością opracowaną na podstawie poprzedniego spisu (NSP 2011) oraz badań bieżących z zakresu urodzeń, zgonów, migracji i bilansów opracowywanych w okresach międzyspisowych.

Publikację zamyka rozdział szósty, w którym zaprezentowano ocenę jakości wyników spisu uzyskanych w zakresie mieszkań w aspekcie ich spójności z informacjami dostępnymi w statystyce publicznej. W tym celu porównano dane na temat liczby mieszkań pozostających w zasobach mieszkaniowych opracowane na podstawie NSP 2011 i NSP 2021.

Wszystkie tablice, mapy, schematy i wykresy zaprezentowane w publikacji są opracowaniem własnym.

Introduction

The Population and Housing Census is the best opportunity to collect full information on the number and structure of population as well as on housing stock and tendencies of its development. The data are fundamental for making numerous key decisions in the field of socio-economic programming (planning) between subsequent censuses at both national and international levels.

Each country makes a decision on its own regarding the method and the sources used for the implementation of the survey. The Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (Official Journal of the European Union, L 218, 13.8.2008)² specifies in detail the sources, based on which Member States can compile the census data. Article 4 states that these are:

- a) traditional censuses,
 - b) censuses based on data from administrative registers,
 - c) combinations of traditional censuses with sample surveys,
 - d) combinations of censuses based on data from administrative registers with sample surveys,
 - e) combinations of censuses based on data from administrative registers with traditional censuses,
 - f) combinations of censuses based on data from administrative registers with sample surveys and traditional censuses,
- and
- g) surveys based on rotating samples ('rolling' censuses).

Due to the importance of the population and housing census, questions about the quality of the obtained results are often asked, which largely determine their usefulness (relevance) in various applications. Data are mostly useful when they are of high quality. Therefore, the assessment of the population and housing census is essential, mostly in order to increase users' confidence in the results that are made available. Moreover, quality management in the census is necessary not only because of the potential wide group of data users, but also because the results developed on the basis of the population and housing census are used in the national and international statistical system. Furthermore, the census is often the best, if not the only, source of information on some sub-populations and small areas.

Quality measurement in population and housing censuses is a difficult task, mostly because there is currently no standard, uniform quality assessment method, which applies to all census methods used by national statistical institutes.

The most commonly used definition of quality focuses on "usefulness" of information. In other words, data itself have no quality, but only the potential which is used for a specific purpose (Bafour, Valente, 2012). Therefore, in the context of the census, quality should be considered in a multidimensional manner.

The European Statistical Office (Eurostat) adopted definition of quality according to the ISO 8402-1986 standard. Due to it quality is defined as: the totality of characteristics and properties of a product or service, which determines the ability to meet declared or implied needs (Eurostat, 2003). It is determined on the basis of 6 quality components, which include: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence.

Each of the above components is characterized below in the context of the population and housing census.

1. **The relevance of information** reflects the extent to which census data meet the users' needs (including private persons, government and local government institutions, scientific and research centers, international institutions). Most often, the relevance of the data is obtained thanks to extensive consultations on the information scope of the survey, methodology, questionnaire design, instruction etc. However, as practice

² <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:218:0014:0020:en:PDF> (available: March 15, 2023)

shows, meeting all the user's needs can be very difficult, and sometimes even impossible due to limited financial and organizational resources (Kordos, 1987). Hence, it is necessary to maintain a balance between the rational planning of a census survey and the wishes (often contradictory) of potential users of census data.

2. **The accuracy of the census results** expresses its closeness to the actual (true) value that would be obtained if the data were collected and processed without any errors. In practice, accurate measurement of any phenomenon is unachievable due to the fact that no measurement tool is perfect. The accuracy of the result is therefore measured by the measurement error, which, in simple terms, is the difference between the value obtained from the survey and the actual value that is being assessed. With regard to the population and housing census, the accuracy of the information depends on the methods used to identify and control any errors that occur throughout the census process.
3. **The timeliness and punctuality component** refers to the time frames of the census implementation process, it is the interval between the reference date of the census and the date when the results are compiled and disseminated. It is understandable that the usefulness of the results of the survey decreases over time from the date of its implementation. However, one should not forget that the implementation of this component is always very time-consuming and laborious, especially when the census was conducted using combined methods and based on many sources which require appropriate processing and integration, while meeting the highest quality standards of the developed results. Reducing the time needed for appropriate processing, validation and implementation of the necessary corrections or the use of data imputation, and finally the analysis of the results in order to publish the results in a time that would be most acceptable for users (i.e. the fastest), may affect the quality of the shared results. Therefore, a compromise should be found between the dimension of accuracy and timeliness and punctuality of data.
4. **The census data accessibility** refers to the accessibility and clarity with which data are made available. This component includes, among others, forms and tools of sharing census data, ordering procedures, time needed to process the order, pricing policy, micro- and macro-data availability, information environment of the data, i.e. combining them with appropriate metadata (text information, explanations, documentation, etc.) and graphical forms (charts, diagrams and maps), copyright and the scope of additional assistance offered by the statistical institution. In addition, this dimension takes into account the strategy of data disclosure control (statistical confidentiality), without compromising the quality of the results.
5. **The comparability component** aims at measuring the impact of differences in applied census terms and definitions when comparing data from different geographical areas, domains (census topics) or base periods (census reference date).
6. **The coherence of census data** reflects the extent to which the results can be combined with other existing statistical data. The use of standard terms, definitions and classifications, including internationally accepted ones, contributes to increased coherence. The level of quality in terms of cohesion can be evaluated using a program of certification and validation of census results by comparison with relevant data from historical (previous) population and housing censuses, statistical surveys (representative) or administrative sources.

The evaluation of the survey does not have to include all six components to meet the quality requirement. However, one should be aware that actions aimed at improving one dimension may result in a negative impact on another component (Baffour, Valente, 2012).

This study is a complementary part of the publication entitled „National Population and Housing Census 2021. Research methodology and organization” published in January 2022, which presents the key aspects regarding the implementation of the 2021 Census, with particular considering the public consultations, methods of census implementation, used data sources, IT systems and applications supporting the implementation of survey. The report also presents the data processing, the characteristics of the forms and tools used for disseminating the results, the strategy for promotion and popularization of census, as well as the framework schedule for release of result information.

The current study focuses on assessing the quality of the 2021 Census results i.a using demographic analysis.

The publication consists of six chapters.

The first chapter deals with issues related to the construction of the list of persons, addresses, dwellings and collective living quarters.

The second chapter focuses on assessing the quality of the list of persons, addresses and dwellings, both in relation to the population covered by the census, as well as dwellings and buildings.

The third chapter – of a theoretical nature – presents sources of errors and describes selected methods for assessing the quality of censuses.

The fourth chapter presents the results of the control survey, in particular in regards to the assessment of content errors and the level of compliance for the examined features. Data collected on the basis of a specially designed representative survey were used for this assessment. As a result of the analysis, a number of statistical indicators were developed and conclusions were formulated.

The fifth chapter is devoted to the assessment of the completeness of the Census 2021 data using the methods of demographic analysis. The completeness study covered the entire population, including newborns and children, the most mobile persons and the oldest persons. Using the methods of demographic analysis, the differences between the population obtained on the basis of the 2021 Census and the population compiled on the basis of the previous census (the 2011 Census) and current surveys in the field of births, deaths, migrations and balances developed in the periods between the both censuses were indicated.

The publication ends with the sixth chapter, which presents the assessment of the quality of the census results obtained in the field of housing in terms of their consistency with the information available in official statistics. For this purpose, data on the number of dwellings remaining in the housing stocks, prepared on the basis of the 2011 Census and the 2021 Census, were compared.

All tables, maps, schemes and charts presented in the publication are of own compilation.

I. Wykaz osobowo-adresowo-mieszkaniowy

I. List of persons, addresses and dwellings

Wykaz jednostek statystycznych, które powinny zostać objęte spisem powszechnym, jest jednym z podstawowych warunków umożliwiających prawidłowe przeprowadzenie badania spisowego i uzyskanie jak najlepszych jakościowo informacji wynikowych, w tym w szczególności pod względem zapewnienia kompletności badania oraz pokrycia populacji badanych jednostek.

Zakres podmiotowy badania spisowego wskazany został w ustawie o narodowym spisie powszechnym ludności i mieszkań w 2021 r. (NSP 2021) i obejmował on:

- a) osoby fizyczne stale zamieszkane i czasowo przebywające w mieszkaniach, budynkach i innych zamieszkałych pomieszczeniach niebędących mieszkaniami,
- b) osoby fizyczne niemające miejsca zamieszkania,
- c) mieszkania, budynki, obiekty zbiorowego zakwaterowania oraz zamieszkane pomieszczenia niebędące mieszkaniami.

Prace związane z utworzeniem wykazu jednostek objętych obserwacją skoncentrowały na przygotowaniu Wykazu osobowo-adresowo-mieszkaniowego oraz Wykazu obiektów zbiorowego zakwaterowania.

Wykaz osobowo-adresowo-mieszkaniowy przygotowany został jako zestaw relacyjnych tablic, z których każda dotyczyła innego rodzaju jednostki podlegającej obserwacji (tj. osób, mieszkań i budynków). Podstawą do utworzenia Wykazu były dane adresowe mieszkań i budynków pochodzące z systemu NOBC oraz dane dotyczące cech identyfikacyjnych oraz miejsca zamieszkania osób, ustalone w oparciu o informacje zgromadzone w funkcjonujących rejestrach administracyjnych i statystycznych.

Przy budowie Wykazu osobowo-adresowo-mieszkaniowego wykorzystano wiele heterogenicznych rejestrów, systemów informacyjnych administracji publicznej, a także zbiorów statystycznych oraz systemów informatycznych podmiotów prywatnych. Ze względu na przyjętą metodę realizacji spisu powszechnego, tj. z uwagi na pandemię COVID-19 głównie poprzez samospis internetowy i wywiad telefoniczny, założono, że Wykaz powinien w swoim zakresie podmiotowym uwzględnić jak najbardziej pełną populację osób (obywateli polskich oraz cudzoziemców) oraz mieszkań wraz z budynkami, w których były one zlokalizowane.

Wykaz był zatem głównym źródłem informacji o jednostkach podlegających spisaniu, ale ze względu na cel badania również i kluczowym narzędziem do zarządzania spisem, kontroli kompletności spisywanych osób, budynków i mieszkań, uwiarytelniania w samospisie internetowym, a także przepływu informacji o stopniu realizacji spisu pomiędzy kanałami gromadzenia danych. Stanowił także bardzo ważny element w procesie przetwarzania i opracowania wyników NSP 2021.

Mając na względzie powyższe funkcjonalności Wykazu można stwierdzić, że spełniał on w spisie przede wszystkim rolę operacyjną. Niemniej jednak, zarówno w części budynkowo-mieszkaniowej, jak i osobowej Wykazu wyznaczono wstępne populacje osób i mieszkań, które powinny być objęte spisem. W istocie rzeczy stanowiły one próbę przybliżenia docelowych populacji¹ jednostek statystycznych ustalonych na podstawie danych rejestrowych dostępnych na etapie przygotowywania Wykazu.

Zaprezentowane w niniejszym rozdziale analizy będą zatem dotyczyć wstępnej wykazowej populacji docelowej w odniesieniu do budynków i mieszkań oraz osób.

¹ Zgodnie z Rozporządzeniem (EU) Nr 1151/2010z dnia 8 grudnia 2010 r. w sprawie wykonania rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 763/2008 w sprawie spisów powszechnych ludności i mieszkań w odniesieniu do ustaleń dotyczących raportów jakości i ich struktury oraz formatu technicznego przekazywania danych „Populacja docelowa” oznacza zbiór wszystkich jednostek statystycznych na określonym obszarze geograficznym w dniu referencyjnym, które kwalifikują się do badania dotyczącego jednego lub większej liczby określonych tematów. Populacja docelowa obejmuje każdą ważną jednostkę statystyczną dokładnie raz. Jest to koncepcja teoretyczna, która w zasadzie nigdy nie osiąga doskonałego pokrycia w praktyce, a w toku badań dąży się jedynie do możliwie najlepszego jej oszacowania.

Budowa Wykazu była realizowana w kilku głównych etapach, tj. przygotowanie wykazu adresowo-mieszkaniowego ze współrzędnymi x, y, aktualizacja gminna adresów, przygotowanie wykazu osób w oparciu o dane administracyjne, integracja przygotowanych wykazów jednostek oraz dołączanie numerów telefonów do osób. Poniżej opisane zostały szczegółowo etapy tworzenia wykazu oraz działania aktualizujące, wykonane dodatkowo w trakcie trwania spisu zasadniczego.

Należy podkreślić, że efektywność wykazu osobowo-adresowo-mieszkaniowego była testowana w dwóch spisach próbnych, a udoskonalone rozwiązania zostały wdrożone w spisie zasadniczym.

1.1. Wykaz adresowo-mieszkaniowy

1.1. List of addresses and dwellings

Etap pierwszy obejmował przygotowanie Wykazu adresowo-mieszkaniowego. Dla utworzenia najszerszego zbioru unikatowych adresów oraz poprawy ich jakości (poprzez uspoźnienie, uzupełnienie, aktualizację) wykorzystano:

- 1) krajowy rejestr urzędowy podziału terytorialnego TERYT – system identyfikacji adresowej ulic, nieruchomości, budynków i mieszkań (NOBC)²,
- 2) systemy informacyjne administracji publicznej i rejestry urzędowe:
 - a) Powszechny Elektroniczny System Ewidencji Ludności (PESEL),
 - b) Centralny Rejestr Podmiotów – Krajowa Ewidencja Podatników (CRP-KEP),
 - c) Centralny Wykaz Ubezpieczonych – Narodowy Fundusz Zdrowia (NFZ),
 - d) Kompleksowy System Informatyczny – Centralny Rejestr Ubezpieczonych (KSI CRU)/ZUS,
 - e) nowy System Informatyczny Ubezpieczeń (nSIU) – Kasa Rolniczego Ubezpieczenia Społecznego (KRUS),
 - f) państwowy rejestr granic i powierzchni jednostek podziałów terytorialnych kraju (PRG), uzupełniająco ewidencja miejscowości, ulic i adresów (EMUiA),
- 3) niepubliczne systemy informacyjne – systemy informacyjne przedsiębiorstw wykonujących działalność gospodarczą w zakresie sprzedaży energii elektrycznej (zbiór Energetyka ZE),
- 4) operat statystyczny dla jednostek prowadzących działalność gospodarczą, tj. Bazę Jednostek Statystycznych (BJS) – wykorzystany do przygotowania wykazu obiektów zbiorowego zakwaterowania.

Ze zbiorów wymienionych w punktach 2–4 wykorzystane zostały następujące rodzaje adresów:

- a) adres zameldowania na pobyt stały (PESEL),
- b) adres zameldowania na pobyt czasowy (PESEL),
- c) adres zameldowania/zamieszkania (ZUS),
- d) adres zamieszkania (KEP, NFZ, ZUS, KRUS),
- e) adres budynku o statusie „budynek istniejący” (EMUiA),
- f) adresy lokali, w których zlokalizowany jest licznik energii elektrycznej (ZE),
- g) adres siedziby jednostki prawnej lub lokalnej lub lokalnej umownej (BJS) dla potrzeb wykazu obiektów zbiorowego zakwaterowania.

Do wykazu zostały dołączone dane o lokalizacji przestrzennej pochodzące z Przestrzennych Baz Adresowych (PBA) – obiekty ze zbioru NOBC charakteryzujące się pełnym pokryciem współrzędnych. Natomiast w przypadku adresów pochodzących z pozostałych zbiorów, uzupełnienie współrzędnych zostało dokonane w oparciu o dane państwowego rejestru granic i powierzchni jednostek podziałów terytorialnych kraju (PRG) i uzupełniająco w oparciu o dane ewidencji miejscowości, ulic i adresów (EMUiA).

² W ramach systemu identyfikacji adresowej ulic, nieruchomości, budynków i mieszkań NOBC prowadzony jest Centralny Katalog Ulic obejmujący nazwy ulic zgodne z brzmieniem uchwał rad gmin o ich nadaniu, identyfikatory nazw oraz przynależność ulic do województwa, powiatu, gminy i miejscowości.

W drugim etapie prac weryfikacja Wykazu obejmowała obiekty bezpośrednio związane z faktem przebywania/zamieszkiwania w nich ludności. Weryfikacja prowadzona była przez pracowników urzędów gmin (UG), a wnoszone przez nich zmiany podlegały następnie analizie i akceptacji przez pracowników urzędów statystycznych (US). W przypadku wątpliwości, finalny kształt rekordu danych wynikał z wzajemnych uzgodnień pomiędzy US a UG. W proces weryfikacji Wykazu adresowo-mieszkaniowego zaangażowani byli również pracownicy Głównego Urzędu Statystycznego i Centrum Informatyki Statystycznej, którzy stanowili wsparcie merytoryczne i techniczne w pracach weryfikacyjnych.

Celem prowadzonej weryfikacji było zapewnienie kompletności budynków, mieszkań, obiektów zbiorowego zakwaterowania i innych obiektów zamieszkiwanych przez ludność (w szczególności oddanych do użytku w ostatnim czasie) oraz ustalenie poprawności ich adresów.

Przedmiotem weryfikacji, aktualizacji i uzupełnienia Wykazu adresowo-mieszkaniowego były dane dotyczące budynków, mieszkań i obiektów zbiorowego zakwaterowania, tj. następujące informacje:

- a) w zakresie budynków: nazwa miejscowości, nazwa ulicy, numer porządkowy albo numery porządkowe budynku, dodatkowa numeracja budynków położonych na tej samej nieruchomości i oznaczonych tym samym numerem porządkowym, rodzaj budynku (mieszkalny, niemieszkalny, obiekt zbiorowego zakwaterowania), opis budynku niemieszkalnego (bliższe określenie budynku, np. szkoła, biuro, magazyn), opis adresu, położenie budynku w państwowym systemie odniesień przestrzennych,
- b) w zakresie mieszkań: typ (mieszkanie, część mieszkania wchodzącego w skład mieszkaniowego zasobu gminy zajmowaną na podstawie umowy najmu, mieszkanie położone w obiekcie zamkniętym, zamieszkałe pomieszczenie niebędące mieszkaniem), numer mieszkania, przyczyna niezamieszkania mieszkania wchodzącego w skład mieszkaniowego zasobu gminy,
- c) w zakresie obiektów zbiorowego zakwaterowania: nazwa obiektu, liczba budynków zajmowanych przez obiekt.

Proces weryfikacji Wykazu adresowo-mieszkaniowego wspomagany był przygotowaną do tego celu aplikacją obejmującą:

- a) część opisową uwzględniającą osobno adresy obiektów z rejestru TERYT oraz adresy pochodzące ze źródeł administracyjnych i poza administracyjnych (odrębnie dla mieszkań i obiektów zbiorowego zakwaterowania),
- b) mapę cyfrową zawierającą następujące warstwy: podział administracyjny (granice województw, powiatów, gmin), podział statystyczny (granice i numery rejonów statystycznych i obwodów spisowych), podział ewidencyjny, lokalizację i nazwy miejscowości, sieć ulic i dróg, punkty adresowe. Podkładem dla ww. warstw były: ortofotomapa i ogólnie dostępne dane topograficzne.

Aplikacja umożliwiała następujące operacje: modyfikację parametrów rekordu (w tym zmianę położenia na mapie), potwierdzenie poprawnego rekordu, dodawanie nowych pozycji, oznaczanie pozycji do usunięcia oraz łączenie adresów ze zbiorów administracyjnych z odpowiadającymi im mieszkaniami z rejestru TERYT. Istotnym elementem działań poprzedzających sam etap weryfikacji były szkolenia realizowane dwuetapowo jako szkolenia dla: pracowników urzędów statystycznych prowadzone przez pracowników Głównego Urzędu Statystycznego i Centrum Informatyki Statystycznej oraz pracowników urzędów gmin prowadzone przez pracowników urzędów statystycznych.

Weryfikacja Wykazu adresowo-mieszkaniowego przez urzędy gmin została przeprowadzona w terminie od 2 listopada do 31 grudnia 2020 r. Prace urzędów statystycznych polegające na kontroli wprowadzanych przez urzędy gmin danych zakończyły się 8 stycznia 2021 r. W celu uniknięcia spiętrzenia prac w ostatnim czasie przewidzianego okresu wprowadzone zostało zróżnicowanie w terminach prowadzenia weryfikacji, które było uzależnione od wielkości jednostki i liczebności danych podlegających weryfikacji. Na przyporządkowanie gminy do podgrupy wpływ miała liczba mieszkań wg stanu po 1 kwartale 2020 r. z systemu NOBC. Szacowanie wielkości odbywało się oddzielnie w obrębie każdego województwa. Pomimo, iż zastosowany podział miał charakter rekomendacji, a urzędy statystyczne mogły przyjąć inne podejście, został on wykorzystany w 15 spośród 16 województw.

Postęp prac weryfikacyjnych podlegał bieżącemu monitorowaniu z wykorzystaniem raportów i meldunków zdawanych przez poszczególne urzędy statystyczne oraz z wykorzystaniem narzędzia „targetowania”, tzn. przekazywania do US informacji o liczbie rekordów, która w ujęciu dziennym powinna zostać zweryfikowana, aby finalnie weryfikacja mogła zostać zakończona w ustalonym terminie. Dzięki narzędziu „targetowania” postęp prac weryfikacyjnych monitorowany był w ujęciu sprawności dziennej i sprawności w podziale na województwa.

Zgodnie z przyjętą metodologią weryfikacja Wykazu adresowo-mieszkaniowego dotyczyła tylko części adresów w nim występujących – obowiązkowej weryfikacji podlegały tylko te pozycje z systemu NOBC oraz ze zbiorów administracyjnych, które w wyniku prowadzonego parowania tych zbiorów oraz zbiorów uzupełniających (Państwowego Rejestru Granic, zbiorów energetycznych) budziły wątpliwości co do ich poprawności. Weryfikacja mogła być również prowadzona wobec danych spoza grupy obowiązkowej, jeśli dana jednostka uznała to za zasadne. Wyniki weryfikacji posłużyły do aktualizacji bazy NOBC, z której został wygenerowany zbiór adresów wykorzystany następnie do połączenia z wykazem osób.

Biorąc pod uwagę udział procentowy gmin (miejskich, wiejskich i miejsko-wiejskich) według liczebności rekordów podlegających obowiązkowej weryfikacji, największy odsetek stanowiły gminy, dla których wielkość ta mieściła się w przedziale 201–500 rekordów mieszkaniowych (67,2%). Gminy, w których weryfikacji podlegało nie więcej niż 1000 rekordów, stanowiły w sumie niemal 90% wszystkich gmin w Polsce. Największa liczebność danych podlegających obowiązkowej weryfikacji przypadła na duże ośrodki miejskie i jednostki z nimi sąsiadujące.

Na 15485301 rekordów występujących w Wykazie (łącznie danych z systemu NOBC rejestru TERYT – 15068185 rekordów i zbiorów administracyjnych – 417126 rekordów) obowiązkowej weryfikacji podlegało 1522850 mieszkań. W populacji tej uwzględniono 1105724 adresy mieszkań z NOBC (7,3%) oraz wszystkie adresy pochodzące ze źródeł administracyjnych, tj. 417126 rekordy. Były to mieszkania wyselekcjonowane według opracowanych warunków uwzględniających występowanie ich adresów w rejestrach administracyjnych, zbiorach energetycznych oraz rejestrach zawierających dane przestrzenne. Znalazły się wśród nich również mieszkania niezamieszkałe. Najwięcej mieszkań podlegających obowiązkowej weryfikacji wystąpiło w województwie pomorskim (9,5%), a najmniej w województwie opolskim (4,4%).

Tablica 1. Mieszkania wytypowane do Wykazu adresowo-mieszkaniowego na podstawie NOBC oraz źródeł administracyjnych (ZA) podlegające obowiązkowej weryfikacji

Table 1. Dwellings selected for List of addresses and dwellings based on the NOBC as well as the administrative sources (ZA) subject to verification

Województwo Voivodship		NOBC		Źródła administracyjne Administrative sources
		liczba mieszkań number of dwellings	udział w ogólnej liczbie mieszkań w NOBC share in total number of dwellings in NOBC	liczba mieszkań number of dwellings
02	dolnośląskie	76466	6,2%	50476
04	kujawsko-pomorskie	55231	7,1%	31768
06	lubelskie	53722	6,7%	19418
08	lubuskie	21551	5,6%	13931
10	łódzkie	88861	8,5%	42888
12	małopolskie	116658	9,2%	32639
14	mazowieckie	180139	7,5%	63684

Tablica 1. Mieszkania wytypowane do Wykazu adresowo-mieszkaniowego na podstawie NOBC oraz źródeł administracyjnych (ZA) podlegające obowiązkowej weryfikacji (dok.)**Table 1. Dwellings selected for List of addresses and dwellings based on the NOBC as well as the administrative sources (ZA) subject to verification (cont.)**

Województwo Voivodship		NOBC		Źródła administracyjne Administrative sources
		liczba mieszkań number of dwellings	udział w ogólnej liczbie mieszkań w NOBC share in total number of dwellings in NOBC	liczba mieszkań number of dwellings
16	opolskie	15802	4,4%	12572
18	podkarpackie	39888	5,8%	18631
20	podlaskie	25534	5,5%	9350
22	pomorskie	86972	9,5%	49173
24	śląskie	140127	7,8%	33122
26	świętokrzyskie	42797	9,3%	1821
28	warmińsko-mazurskie	34072	6,4%	4409
30	wielkopolskie	91145	7,1%	28858
32	zachodniopomorskie	36759	5,4%	4386
00	POLSKA	1105724	7,3%	417126

W trakcie tego etapu zweryfikowanych zostało 1504297 (98,8%) pozycji. Wśród 18553 mieszkań będących różnicą powyższych wartości 25% stanowiły mieszkania, dla których weryfikacja nie została przeprowadzona przez gminę, zaś pozostałe 75% to mieszkania, które zostały zweryfikowane, ale nie uzyskały akceptacji danego urzędu statystycznego. Z puli mieszkań z systemu NOBC podlegających obowiązkowej weryfikacji potwierdzonych zostało 80,2% rekordów – najwięcej w województwie małopolskim (87,4%), najmniej w województwie kujawsko-pomorskim (69,3%). Pozostałe mieszkania (19,8%) podlegające obowiązkowej weryfikacji zostały zweryfikowane przez gminy negatywnie, tj. oznaczone do usunięcia jako nieistniejące.

Tablica 2. Wyniki obowiązkowej weryfikacji mieszkań wytypowanych do Wykazu adresowo-mieszkańowego na podstawie NOBC**Table 2. Results of mandatory verification of dwellings selected for List of addresses and dwellings based on the NOBC**

Województwo Voivodship		Udział mieszkań Share of dwellings	
		potwierdzonych approved	oznaczonych do usunięcia marked for deletion
		w ogólnej liczbie mieszkań wytypowanych do Wykazu na podstawie NOBC podlegających weryfikacji in total number of dwellings selected for List based on the NOBC subject to verification	
02	dolnośląskie	78,0%	22,0%
04	kujawsko-pomorskie	69,3%	30,7%
06	lubelskie	79,2%	20,8%
08	lubuskie	85,1%	14,9%
10	łódzkie	82,2%	17,8%
12	małopolskie	87,4%	12,6%
14	mazowieckie	74,6%	25,4%
16	opolskie	75,3%	24,7%
18	podkarpackie	82,2%	17,8%
20	podlaskie	76,7%	23,3%
22	pomorskie	83,5%	16,5%
24	śląskie	84,9%	15,1%
26	świętokrzyskie	81,7%	18,3%
28	warmińsko-mazurskie	79,0%	21,0%
30	wielkopolskie	78,4%	21,6%
32	zachodniopomorskie	83,6%	16,4%
00	POLSKA	80,2%	19,8%

W przypadku mieszkań wytypowanych do Wykazu adresowo-mieszkańowego na podstawie źródeł administracyjnych 31,8% adresów odnalazło się w NOBC, a 32,6% adresów mieszkań zostało zweryfikowanych pozytywnie i dodano je do NOBC. Jednocześnie 35,2% adresów z tej populacji zostało oznaczonych do usunięcia. Najwięcej mieszkań w stosunku do liczby mieszkań zidentyfikowanych na podstawie źródeł administracyjnych dodano do NOBC w województwach: wielkopolskim – 46,5% oraz kujawsko-pomorskim – 41,8%. Jednocześnie największy udział mieszkań oznaczonych do usunięcia odnotowano w województwach: lubelskim – 54,6% oraz opolskim – 51,2%.

Tablica 3. Wyniki obowiązkowej weryfikacji mieszkań wytypowanych do Wykazu adresowo-mieszkaniowego na podstawie źródeł administracyjnych (ZA)**Table 3. Results of mandatory verification of dwellings selected for List of addresses and dwellings based on the administrative sources (ZA)**

Województwo Voivodship		Udział mieszkań Share of dwellings		
		połączonych z NOBC linked with NOBC	dopisanych do NOBC added to NOBC	oznaczonych do usunięcia marked for deletion
		w ogólnej liczbie mieszkań wytypowanych do Wykazu na podstawie ZA podlegających weryfikacji in total number of dwellings selected for List based on the ZA subject to verification		
02	dolnośląskie	24,0	39,0	37,0
04	kujawsko-pomorskie	23,1	41,8	35,1
06	lubelskie	20,2	24,5	54,6
08	lubuskie	30,2	28,1	41,8
10	łódzkie	38,5	25,6	35,9
12	małopolskie	46,1	34,4	19,4
14	mazowieckie	31,6	36,9	31,4
16	opolskie	17,1	31,7	51,2
18	podkarpackie	34,2	32,3	33,5
20	podlaskie	30,1	25,1	44,8
22	pomorskie	38,3	27,1	34,6
24	śląskie	39,2	22,5	38,3
26	świętokrzyskie	47,8	25,8	26,4
28	warmińsko-mazurskie	33,6	23,2	43,2
30	wielkopolskie	20,3	46,5	28,8
32	zachodniopomorskie	47,3	15,0	37,8
00	POLSKA	31,8	32,6	35,2

W wyniku przeprowadzonej weryfikacji i to zarówno tzw. obowiązkowej, jak i dodatkowej, liczba mieszkań dopisanych do systemu NOBC stanowiła 1,4% ogólnej liczby mieszkań w NOBC. Grupa ta obejmuje zarówno adresy mieszkań występujące w rejestrach administracyjnych, które podlegały obowiązkowej weryfikacji, jak i dodatkowe mieszkania dopisane przez gminy i niewykazane we wstępnym wykazie adresowo-mieszkaniowym (z systemu NOBC oraz ze zbiorów administracyjnych). Najwięcej procentowo mieszkań dopisano w województwach: kujawsko-pomorskim (2,3%) oraz dolnośląskim i pomorskim (po 2,1%), najmniej natomiast w województwach: warmińsko-mazurskim (0,6%) i śląskim (0,7%).

Tablica 4. Mieszkania dopisane do NOBC oraz usunięte z NOBC w ramach weryfikacji (obowiązkowa i dodatkowa łącznie)**Table 4. Dwelling added to the NOBC and delated from the NOBC under the verification (mandatory and facultative combined)**

Województwo Voivodship		Udział mieszkań Share of dwellings	
		dopisanych do NOBC added to NOBC	oznaczonych do usunięcia z NOBC marked for deletion from NOBC
		w ogólnej liczbie mieszkań wytypowanych do Wykazu in total number of dwellings selected for List	
02	dolnośląskie	2,1%	1,6%
04	kujawsko-pomorskie	2,3%	2,7%
06	lubelskie	1,0%	1,7%
08	lubuskie	1,4%	1,1%
10	łódzkie	1,5%	2,1%
12	małopolskie	1,4%	1,2%
14	mazowieckie	1,6%	2,2%
16	opolskie	1,4%	1,6%
18	podkarpackie	1,3%	1,3%
20	podlaskie	1,0%	1,7%
22	pomorskie	2,1%	1,9%
24	śląskie	0,7%	1,3%
26	świętokrzyskie	1,0%	2,0%
28	warmińsko-mazurskie	0,6%	1,6%
30	wielkopolskie	1,7%	1,8%
32	zachodniopomorskie	0,9%	1,1%
00	POLSKA	1,4%	1,7%

Jednocześnie spośród wszystkich mieszkań pochodzących z systemu NOBC i ujętych w wykazie adresowo-mieszkaniowym do usunięcia oznaczonych zostało 1,7% mieszkań. Procentowo najwięcej ubytków odnotowano w województwie kujawsko-pomorskim (2,7%), najmniej zaś w województwie lubuskim i w województwie zachodniopomorskim (po 1,1%). W wyniku weryfikacji do ostatecznego Wykazu adresowo-mieszkaniowego, który został wykorzystany w dalszych pracach spisowych, dopisano 78576 nowych mieszkań, a usunięto 248718 pozycji. Ostateczna liczba rekordów w Wykazie adresowo-mieszkaniowym wyniosła 15070130. Wynik ten stanowi rezultat zarówno prac weryfikacyjnych przeprowadzanych w gminach, jak i kwartalnej aktualizacji systemu NOBC.

Ważnym aspektem prac związanych z budową Wykazu było utworzenie Wykazu Obiektów Zbiorowego Zakwaterowania (OZZ).

Przygotowanie tego wykazu miało na celu ustalenie populacji obiektów zbiorowego zakwaterowania o ściśle określonej typologii i zakresie działalności. Wykaz ten był tworzony wieloetapowo w celu osiągnięcia możliwie najpełniejszego zakresu podmiotowego obiektów zbiorowego zakwaterowania. Wykaz obiektów zbiorowego zakwaterowania został utworzony na podstawie danych pochodzących z:

- systemu identyfikacji adresowej ulic, nieruchomości, budynków i mieszkań (NOBC), wchodzącego w skład krajowego rejestru urzędowego podziału terytorialnego kraju TERYT,
- rozproszonych licznych rejestrów i wykazów OZZ publikowanych na stronach internetowych (m.in. rejestrów i wykazów prowadzonych przez urzędy wojewódzkie),
- wykazów OZZ pochodzących z zasobów statystyki publicznej, w tym m.in. z Bazy Jednostek Statystycznych (BJS).

Proces przygotowania Wykazu OZZ obejmował analizę źródeł danych i wybór obiektów zbiorowego zakwaterowania, które finalnie znalazły się w Wykazie OZZ, kontrolę i standaryzację adresów OZZ. Ponadto, analogicznie jak dla budynków i mieszkań, przeprowadzono parowanie adresów OZZ z adresami lokali, do których dostarczana jest energia elektryczna (PPE³) od przedsiębiorstw prowadzących działalność gospodarczą w zakresie sprzedaży lub dystrybucji energii elektrycznej oraz z danymi z PRG (uzupełniająco z EMUiA). Dodatkowo do przypisania współrzędnych x, y posłużyły przestrzenne bazy adresowe (PBA) i PRG (Państwowy Rejestr Granic)/ EMUiA dla obiektów uwzględnionych w Wykazie OZZ. W efekcie parowania Wykazu OZZ i tabeli OZZ z bazy NOBC ze źródłami administracyjnymi tj. PRG, EMUiA, PBA oraz z PPE, otrzymano wyjściową wersję OZZ. W dalszej kolejności niesparowane z bazą NOBC adresy obiektów podlegały weryfikacji gminnej, w tym samym okresie co weryfikacja Wykazu adresowo-mieszkaniowego. Dane adresowe ujęte w Wykazie adresowo-mieszkaniowym, dotyczące obiektów zbiorowego zakwaterowania, zweryfikowane w urzędach gmin zostały poddane analizie w urzędach statystycznych, w zakresie zmian wnoszonych przez UG. Każda ze zmian podlegała akceptacji urzędu statystycznego, a w przypadku wątpliwości, finalny kształt rekordu danych podlegał wzajemnym uzgodnieniom pomiędzy urzędem statystycznym a urzędem gminy.

1.2. Wykaz osób

1.2. List of persons

Kolejnym etapem w procesie tworzenia wykazu spisowego było przygotowanie operacyjnego wykazu osób. Do budowy wykazu osób wybrane zostały źródła, które ze względu na zakres podmiotowy i przedmiotowy oraz zweryfikowaną jakość były optymalne do realizacji założonego celu.

Tablica 5. Opis wykorzystanych źródeł danych administracyjnych

Table 5. Description of the administrative data sources used

Lp. No.	Zbiór Data set	Gestor/źródło administracyjne Owner/administrative source	Zakres podmiotowy Scope of entities
1.	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
2.	KEP	Ministerstwo Finansów/Centralny Rejestr Podmiotów – Krajowej Ewidencji Podatników	Osoby fizyczne (podatnicy) oraz osoby fizyczne prowadzące działalność gospodarczą (podatnicy)

³ Punkt poboru energii – punkt pomiarowy w instalacji lub sieci, dla którego dokonuje się rozliczeń oraz dla którego może nastąpić zmiana sprzedawcy zgodnie z ustawą z dnia 10 kwietnia 1997 r. Prawo energetyczne.

Tablica 5. Opis wykorzystanych źródeł danych administracyjnych (dok.)
Table 5. Description of the administrative data sources used (cont.)

Lp. No.	Zbiór Data set	Gestor/źródło administracyjne Owner/administrative source	Zakres podmiotowy Scope of entities
3.	ZUS	Zakład Ubezpieczeń Społecznych	Osoby fizyczne (ubezpieczeni oraz płatnicy składek)
4.	KRUS	Kasa Rolniczego Ubezpieczenia Społecznego/ KRUSNAL	Rolnicy, ich małżonkowie i pracujący z nimi domownicy (ubezpieczeni) oraz osoby pobierające świadczenia rolnicze z KRUS
5.	NFZ	Narodowy Fundusz Zdrowia/Centralny Wykaz Świadczeniobiorców	Osoby fizyczne (świadczeniobiorcy)
6.	ARIMR	Agencja Restrukturyzacji i Modernizacji Rolnictwa/System informacyjny ARiMR	Osoby fizyczne – producenci rolni oraz posiadacze zwierząt gospodarskich: świń, bydła, owiec i kóz
7.	PESEL_WYMELD	Powszechny Elektroniczny System Ewidencji Ludności	Osoby wymeldowane donikąd, które przed wymeldowaniem miały stałe miejsce pobytu w Polsce

Wykaz osób został przygotowany według stanu na dzień 31 grudnia 2020 r. z uwagi na dostępność danych administracyjnych. Do zdefiniowania populacji posłużył identyfikator UNS (unikatowy numer statystyczny). Numer UNS nadawany był automatycznie każdej osobie posiadającej numer PESEL, a także – za pomocą specjalnych, złożonych algorytmów – przy wykorzystaniu wielu źródeł danych, osobom, które tego numeru nie posiadały. UNS jest stałym symbolem numerycznym, który identyfikuje jednoznacznie osobę fizyczną i poprzez swoją unikatowość pozwala na odróżnienie wielu osób noszących to samo imię i nazwisko. Założenie związane z nadawaniem UNS powstało z powodu konieczności uwzględnienia podczas NSP 2021 również cudzoziemców przebywających na terytorium Polski, część z nich nie posiadała numeru PESEL.

Pierwszym etapem budowy wykazu osób było zgromadzenie i scalenie wszystkich numerów UNS z wyselekcjonowanych rejestrów administracyjnych. Ze względu na to, że zakres podmiotowy w rejestrach może częściowo pokrywać się – tj. kilka rejestrów może zawierać informacje o tych samych jednostkach – przyjęto metodę inkrementalnego zasilenia operatu numerami UNS, wybierając identyfikator jednokrotnie według określonej algorytmem hierarchii zbiorów (PESEL, KEP, ZUS, KRUS, NFZ, ARiMR, PESEL_WYMELD).

Drugim etapem prac była weryfikacja szczególnych przypadków, które zostały oznaczone w operacyjnym wykazie osób specjalnymi znacznikami (flagami). To działanie miało na celu umożliwienie identyfikacji rekordów i usprawnienie procesu ich dokładnej weryfikacji w spisie.

Weryfikacja rekordów dotyczyła następujących przypadków:

- stałych mieszkańców innych krajów, zameldowanych na pobyt czasowy, którzy zgłosili wyjazd z Polski lub którym wygaśnię deklarowany 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.

Istotnym elementem tworzenia całego wykazu na potrzeby NSP 2021 w kontekście personalizacji danych oraz realizacji spisu zasadniczego poprzez aplikację formularzową (uwierzytelnianie) było dołączanie nazwiska i imion do osób oraz ustalenie ich adresu zamieszkania. Proces przypisania nazwiska i imienia/imion poprzedziło wyszukanie przypadków różnych nazwisk/imion dla tego samego numeru PESEL, w tym sprawdzenie powtórzonych wartości w poszczególnych rejestrach administracyjnych i zdublowanych rekordów pomiędzy rejestrami. W ramach poszczególnych rejestrów mogą bowiem wystąpić pojedyncze przypadki takich błędów, głównie literowych i łatwo weryfikowalnych poprzez odniesienie do innego rejestru.

Bardziej skomplikowanym procesem było ustalenie i przypisanie osobie adresu zamieszkania. Należy podkreślić, że nie istnieje w Polsce żaden przepis prawny, który reguluje numerowanie kondygnacji i lokali, co oznacza dowolność w ustalaniu numeracji lokali w budynku, jak i w sposobie zapisu numeru lokalu. Ponadto gestorzy rejestrów zasadniczo nie stosują standardów klasyfikacyjnych, numeru identyfikacyjnego i oznaczeń kodowych przyjętych w rejestrze terytorialnym TERYT. W efekcie, zapisy miejscowości, ulic i zapisów numerów budynków i lokali w rejestrach są traktowane dowolnie. Przyjęte w procesie przetwarzania zbiorów administracyjnych reguły standaryzujące i walidujące – ze względu na stosowany automatyczny proces – nie zawsze są w stanie uwzględnić pojedynczych przypadków lub specyficznych zapisów.

Przypisanie adresu zamieszkania osobie polegało w pierwszej kolejności na wyborze najlepszego źródła danych adresowych spośród wytypowanych rejestrów administracyjnych. Przy konstruowaniu algorytmu wyboru adresu kierowano się referencyjnością danego rejestru i aktualnością adresu, ale także kompletnością adresu i zgodnością jego poszczególnych składowych z jednostkami rejestru TERYT. Przyjęta kolejność wyboru źródła adresu (według jego rodzaju) przedstawiona została poniżej.

Tablica 6. Kolejność dołączania adresów
Table 6. Order of the adding addresses

Lp. No.	Źródło Source	Rodzaj adresu Type of the address
1.	KEP	adres zamieszkania (adres zameldowania)
2.	ZUS	adres zamieszkania (adres zameldowania, korespondencji)
3.	PESEL	adres zameldowania na pobyt czasowy
4.	PESEL	adres zameldowania na pobyt stały
5.	KRUS	adres zamieszkania (adres korespondencji)
6.	ARiMR	adres zamieszkania (adres korespondencji)
7.	NFZ	adres zamieszkania

Kolejna faza prac związanych z adresami zamieszkania polegała na zintegrowaniu wykazu osób z przypisanymi adresami zamieszkania pochodzącymi z rejestrów z danymi adresowymi zaktualizowanej bazy NOBC. Przeprowadzenie tego etapu było istotne ze względu na fakt sprawdzenia zarówno jakości oraz poprawności ustalonych adresów, jak i adresów dołączonych do osób.

W ramach integracji przeprowadzono następujące działania:

- 1) Przygotowanie danych z NOBC – w celu połączenia wykazu osób z bazą NOBC zdefiniowano tzw. pełny klucz adresowy, w skład którego weszły:
 - nazwa gminy,
 - nazwa miejscowości,

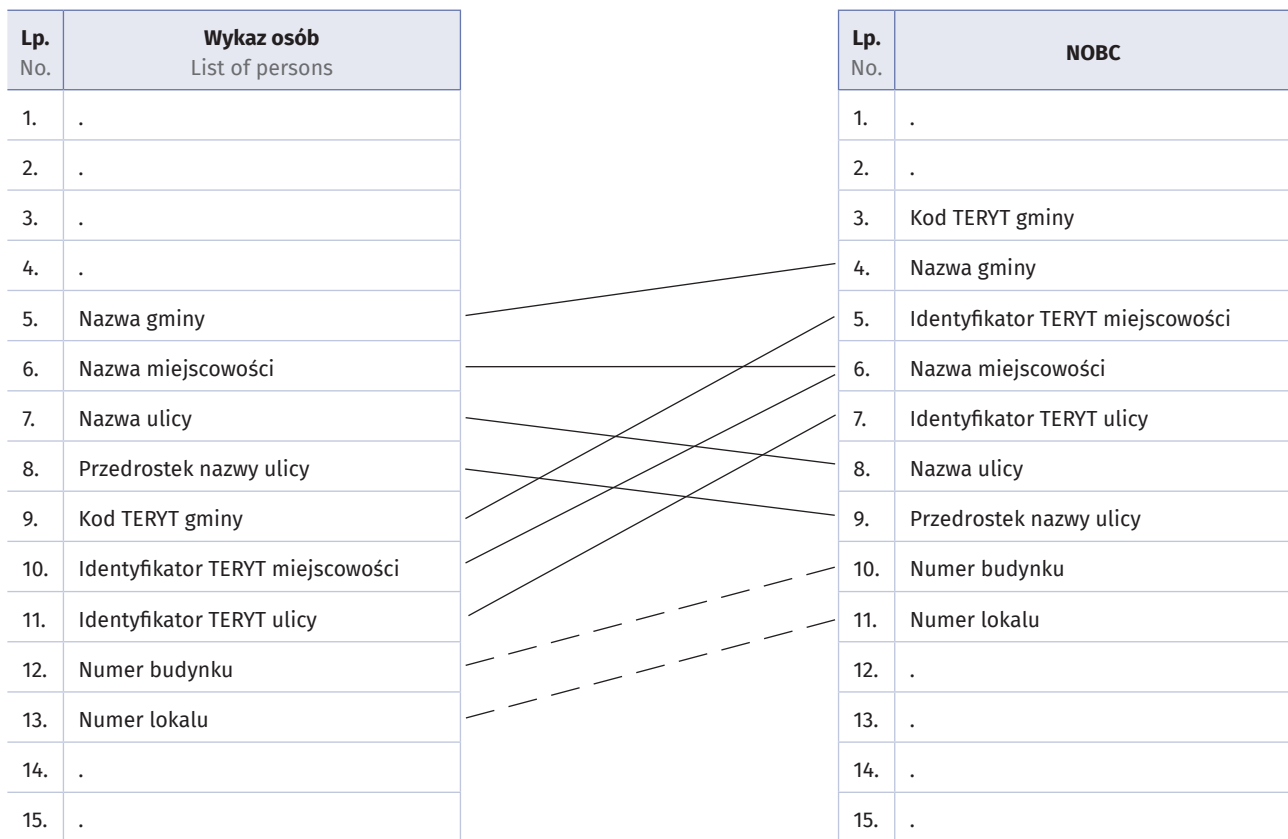
- nazwa ulicy,
- kod TERYT gminy,
- kod TERYT miejscowości,
- kod TERYT ulicy,
- numer budynku,
- numer lokalu.

W systemie NOBC sprawdzana była unikatowość klucza poprzez zastosowanie metody deduplikacji.

- 2) Klucze łączenia – różne sposoby zapisu numeru budynku i lokalu wymusiły etapowe łączenie wykazu osób z NOBC w oparciu o modyfikowane wersje tych danych, celem działania była próba uzyskania spójnych zapisów. Na poniższym schemacie mapowanie wymienionych zmiennych zaznaczono linią przerywaną.

Schemat 1. Mapowanie kluczy łączenia adresów z Wykazu osób z bazą NOBC

Scheme 1. Mapping of address linkage keys from the List of persons with the NOBC database



———— Dane niepodlegające modyfikacji w trakcie łączenia - - - - Dane podlegające modyfikacji w trakcie łączenia

- 3) Algorytmy łączenia – przyjęto siedem algorytmów łączenia. W każdy algorytmie wystąpiły zapisy danych oryginalnych, a nie zmodyfikowanych przez poprzedni algorytm. W celu identyfikacji algorytmu sparowania danych utworzono dodatkowe kolumny. W zbiorze wyników zachowano oryginalne wpisy numeru budynku i lokalu, jak i te, po których nastąpiło połączenie.

Tablica 7. Algorytmy modyfikacji danych podczas procesu parowania
Table 7. Data modification algorithms during the pair production process

Algorytm Algorithm	Opis modyfikacji danych z Wykazu osób Description of data modification from the List of persons	Opis modyfikacji danych z NOBC Description of data modification from the NOBC	Przykłady modyfikacji danych: wykaz osób/NOBC Examples of data modification: list of persons/NOBC	
1.	brak (oryginalna postać danych)	brak (oryginalna postać danych)		
2.	podniesienie do standardu wielkich/małych liter (numer budynku, numer lokalu)	brak (oryginalna postać danych)	28a → 28A 3b → 3B	28A → 28a 3B → 3b
3.	usunięcie spacji pomiędzy liczbą, a literą (numer budynku, numer lokalu)	brak (oryginalna postać danych)	17 a → 17a 7 C → 7C 124 Ab → 124Ab	
4.	<ul style="list-style-type: none"> usunięcie spacji pomiędzy liczbą, a literą (numer budynku, numer lokalu) podniesienie do standardu wielkich/małych liter (numer budynku, numer lokalu) 	brak (oryginalna postać danych)	19 a → 19A 54 Ab → 54AB 101 c → 101C	19 A → 19a 54 Ab → 54ab 101 C → 101c
5.	zamiana znaku "-" na "/" i odwrotnie (numer budynku)	brak (oryginalna postać danych)	54-Ab → 54/Ab 10/Cd → 10-Cd	
6.	<ul style="list-style-type: none"> zamiana znaku "-" na "/" i odwrotnie (numer budynku) podniesienie do standardu wielkich/małych liter (numer budynku, numer lokalu) 	brak (oryginalna postać danych)	15-17a → 15/17A 74/80 → 74-80 27-31VI → 27/31VI	15-17A → 15/17a 74/80 → 74-80 27-31VI → 27/31VI
7.	<ul style="list-style-type: none"> pominięcie niedozwolonych znaków (numer budynku, numer lokalu) podniesienie do standardu wielkich liter 	<ul style="list-style-type: none"> pominięcie niedozwolonych znaków (numer budynku, numer lokalu) podniesienie do standardu wielkich liter 	21. → 21 1BL.3 → 1BL3 113* → 113	

4) Etapy łączenia:

- z wykazu osób wybrano osoby posiadające adres (nie ograniczono selekcji spójnością adresu oraz wypełnieniem numeru budynku czy lokalu). Dla tych obserwacji zdefiniowano trzy główne etapy łączenia:
 - oryginalne dane z wykazu osób,
 - próba aktualizacji numeru budynku i lokalu w wykazu osób rejestrem PESEL,
 - łączenie po kodzie miejscowości podstawowej w NOBC.
- każdy etap przechodził przez ten sam zestaw algorytmów transformacji danych.

W przypadku etapu I pozostawiono dane wejściowe w niezmienionej postaci. Dla dużej liczby obserwacji w wykazie osób występujących w zbiorze PESEL, ale z przypisanym tym samym adresem z innego zbioru stwierdzono lepszą jakość wpisów w polach numer budynku i lokalu. Na potrzeby procesu łączenia w etapie II zdecydowano się na zmapowanie tych zmiennych z rejestru PESEL. Następnie w etapie III zmieniono symbol miejscowości na kod miejscowości podstawowej NOBC jako jeden z kluczy łączenia.

II. Ocena jakości wykazu mieszkań oraz osób

II. Assessing the quality of the housing and persons list

2.1. Mieszkania

2.1. Dwellings

Mieszkanie stanowi jedną z podstawowych jednostek statystycznych, która była poddana obserwacji w ramach NSP2021. Zgodnie z Rozporządzeniem wykonawczym Komisji (UE) 2017/881 z dnia 23 maja 2017 r. w sprawie wykonania rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 763/2008 w sprawie spisów powszechnych ludności i mieszkań w odniesieniu do ustaleń dotyczących raportów jakości i ich struktury oraz formatu technicznego przekazywania danych⁴, zmieniającym rozporządzenie (UE) nr 1151/2010 kategoria mieszkania wykorzystywana na potrzeby sporządzenia krajowego raportu jakości NSP2021 obejmuje zarówno mieszkania konwencjonalne, jak również zamieszkałe pomieszczenia mieszkalne niebędące mieszkaniami.

Z jednej strony mieszkanie stanowi odrębną jednostkę obserwacji, a informacje zebrane w ramach spisu na temat wielkości mieszkań, okresu wybudowania budynku, w którym zlokalizowane jest mieszkanie, wyposażenia mieszkań w instalacje i urządzenia techniczno-sanitarne umożliwiają ocenę standardu zasobów mieszkaniowych. Z drugiej strony, dopiero łączna analiza informacji charakteryzujących mieszkania, w tym w szczególności mieszkania konwencjonalne, z informacjami dotyczącymi osób, które je zamieszkują, pozwala na określenie warunków mieszkaniowych ludności pozostającej na terenie kraju 31 marca 2021 r.

W tabeli MIESZKANIA Wykazu osobowo-adresowo-mieszkaniowego przygotowanego według stanu na 31 grudnia 2020 r. znajdowało się 15073130 rekordów, z których każdy odpowiadał jednemu mieszkaniu. Mając na celu zapewnienie równoczesności źródeł danych zasilających badanie spisowe, tj. pozyskanie informacji według stanu na 31 marca 2021 r., dokonano aktualizacji adresów mieszkań danymi pochodzącymi z NOBC.

Po aktualizacji Wykazu osobowo-adresowo-mieszkaniowego stanem NOBC na 31 marca 2021 r. w tabeli MIESZKANIA dopisano 101234 rekordów, w wyniku czego w Wykazie według stanu rejestrowego znalazło się 15174364 mieszkań. Jednocześnie ustalono, że 3094 mieszkania nie zostały odnalezione w NOBC, ale ich adresy zostały wskazane w rejestrach jako adresy zamieszkania osób. Dlatego też, rekordy dla tych mieszkań nie zostały usunięte z Wykazu.

Tablica 8. Liczba mieszkań w tabeli MIESZKANIA Wykazu osobowo-adresowo-mieszkaniowego
Table 8. Number of dwellings in DWELLINGS table in List of persons, addresses and dwellings

Województwo Voivodship	Ogółem według stanu na 31 grudnia 2020 Grand total as of 31 December 2020	Ogółem według stanu na 31 marca 2021 ^a Grand total as of 31 March 2021 ^a	Z liczby ogółem – mieszkania: Of total number – dwellings:		
			zweryfikowane NOBC verified with NOBC		niezweryfikowane NOBC razem non-verified with NOBC of total
			razem total	w tym dopisane na podstawie NOBC of which added based on NOBC	
POLSKA OGÓŁEM POLAND OF TOTAL	15073130	15174364	15171270	101234	3094
Dolnośląskie	1241702	1251941	1251679	10239	262

a Wykaz osobowo-adresowo-mieszkaniowy zaktualizowany NOBC według stanu na 31 marca 2021 r.

a List of persons, addresses and dwellings updated with the NOBC as of 31 March 2021.

⁴ <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A32017R0881> (dostęp: 15 marca 2023 r.)

Tablica 8. Liczba mieszkań w tabeli MIESZKANIA Wykazu osobowo-adresowo-mieszkaniowego (dok.)
Table 8. Number of dwellings in DWELLINGS table in List of persons, addresses and dwellings (cont.)

Województwo Voivodship	Ogółem według stanu na 31 grudnia 2020 Grand total as of 31 December 2020	Ogółem według stanu na 31 marca 2021 ^a Grand total as of 31 March 2021 ^a	Z liczby ogółem – mieszkania: Of total number – dwellings:		
			zweryfikowane NOBC verified with NOBC		niezweryfikowane NOBC razem non-verified with NOBC of total
			razem total	w tym dopisane na podstawie NOBC of which added based on NOBC	
Kujawsko-pomorskie	771289	775447	775325	4158	122
Lubelskie	793535	796428	796293	2893	135
Lubuskie	387874	390647	390584	2773	63
Łódzkie	1038434	1043862	1043460	5428	402
Małopolskie	1267299	1276144	1275997	8845	147
Mazowieckie	2389009	2409927	2409412	20918	515
Opolskie	357796	359186	359107	1390	79
Podkarpackie	693458	698847	698721	5389	126
Podlaskie	462945	466348	466289	3403	59
Pomorskie	921640	930984	930837	9344	147
Śląskie	1798575	1805813	1805397	7238	416
Świętokrzyskie	455726	457368	457285	1642	83
Warmińsko-mazurskie	526428	529295	529237	2867	58
Wielkopolskie	1286375	1297006	1296637	10631	369
Zachodniopomorskie	681045	685121	685010	4076	111

a Wykaz osobowo-adresowo-mieszkaniowy zaktualizowany NOBC według stanu na 31 marca 2021 r.

a List of persons, addresses and dwellings updated with the NOBC as of 31 March 2021.

Populacja mieszkań licząca 15174364 jednostki stanowiła populację objętą spisem zgodnie z definicją zawartą w Rozporządzeniu wykonawczym Komisji (UE) 2017/881 z dnia 23 maja 2017 r. w sprawie wykonania rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 763/2008 w sprawie spisów powszechnych ludności i mieszkań w odniesieniu do ustaleń dotyczących raportów jakości i ich struktury oraz formatu technicznego przekazywania danych, zmieniające rozporządzenie (UE) nr 1151/2010.

Metoda realizacji spisu zakładała wykorzystanie zarówno danych zebranych w ramach wywiadów spisowych, jak i ze źródeł pozastatystycznych. W przypadku zasobów mieszkaniowych rozwiązaniem to zapewniło pokrycie informacji w zakresie co najmniej jednego szczegółowego tematu dla każdego mieszkania uwzględnionego w Wykazie. W związku z tym szacunkowa docelowa populacja mieszkań została utożsamiona z populacją docelową i wyniosła 15228932 mieszkań, w tym 15227927 stanowiły mieszkania konwencjonalne.

Tablica 9. Liczba mieszkań w populacji objętej spisem oraz docelowej populacji spisu**Table 9. Number of dwellings in census population as well as target population**

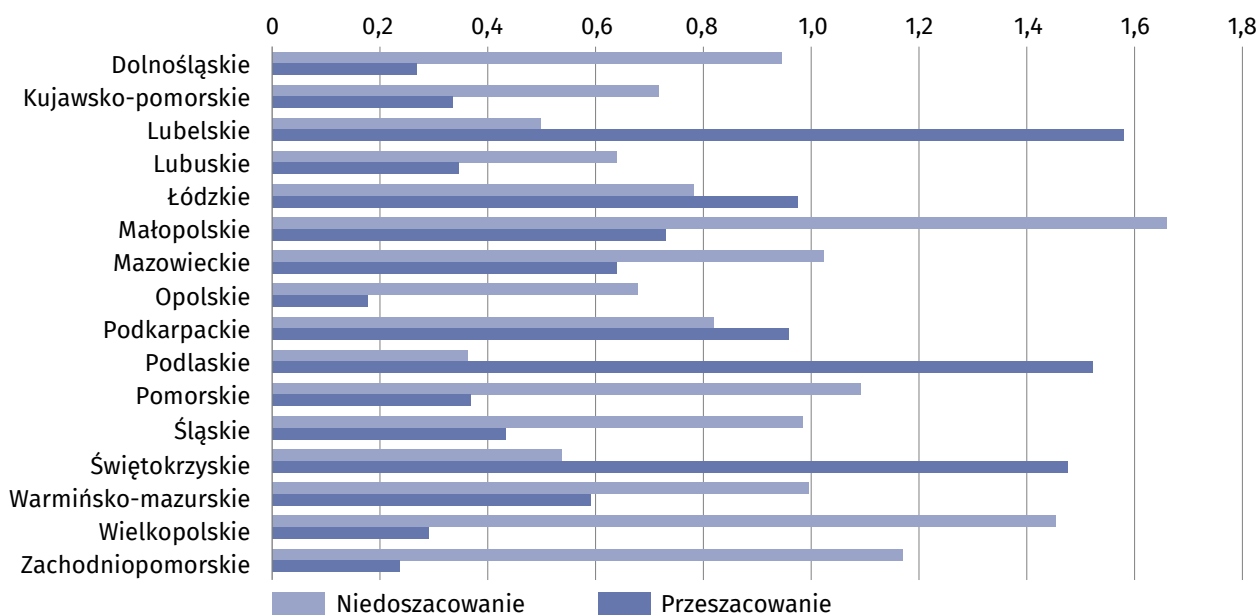
Województwo Voivodship	Populacja objęta spisem (A) Census population (A)	Mieszkania w ramach badania spisowego Dwellings under census		Populacja docelowa (A+B-C) ^a Target population (A+B-C) ^a	
		dopisane (B) added (B)	usunięte (C) deleted (C)	ogółem total	w tym mieszkania konwencjonalne of which conventio- nal dwellings
POLSKA OGÓŁEM	15174364	150204	95636	15228932	15227927
Dolnośląskie	1251941	11849	3342	1260448	1260423
Kujawsko-pomorskie	775447	5553	2597	778403	778325
Lubelskie	796428	3962	12575	787815	787813
Lubuskie	390647	2493	1351	391789	391756
Łódzkie	1043862	8165	10170	1041857	1041795
Małopolskie	1276144	21186	9313	1288017	1288017
Mazowieckie	2409927	24636	15373	2419190	2419081
Opolskie	359186	2433	632	360987	360972
Podkarpackie	698847	5718	6694	697871	697871
Podlaskie	466348	1689	7101	460936	460930
Pomorskie	930984	10166	3421	937729	937670
Śląskie	1805813	17763	7829	1815747	1815709
Świętokrzyskie	457368	2453	6753	453068	453065
Warmińsko-mazurskie	529295	5269	3120	531444	531444
Wielkopolskie	1297006	18857	3744	1312119	1311580
Zachodniopomorskie	685121	8012	1621	691512	691476

a Populację mieszkań docelową stanowią mieszkania ujęte w populacji objętej spisem (A) powiększoną o mieszkania dopisane w trakcie spisu (B) i pomniejszoną o mieszkania wykluczone z zasobów mieszkaniowych (C).

Na poziomie Polski ogółem populacja mieszkań objęta spisem została przeszacowana o 95636 jednostek (tj. o 0,63%), które zostały wykluczone z zasobów mieszkaniowych oraz niedoszacowana o 150204 jednostek (tj. o 0,99%), które zostały dopisane do populacji mieszkań w ramach realizacji wywiadów spisowych.

Wykres 1. Odsetek mieszkań niedoszacowanych i przeszacowanych w stosunku do populacji mieszkań objętej spisem (w %)

Chart 1. Percentage of underestimated and overestimated dwellings in total census population (in %)



W ujęciu terytorialnym najwyższy odsetek niedoszacowania mieszkań odnotowano w przypadku województw: małopolskiego (1,66%) oraz wielkopolskiego (1,45%), podczas gdy najmniejsze wartości tego wskaźnika wystąpiły w województwach: podlaskim (0,36%), lubelskim (0,50%) i świętokrzyskim (0,54%).

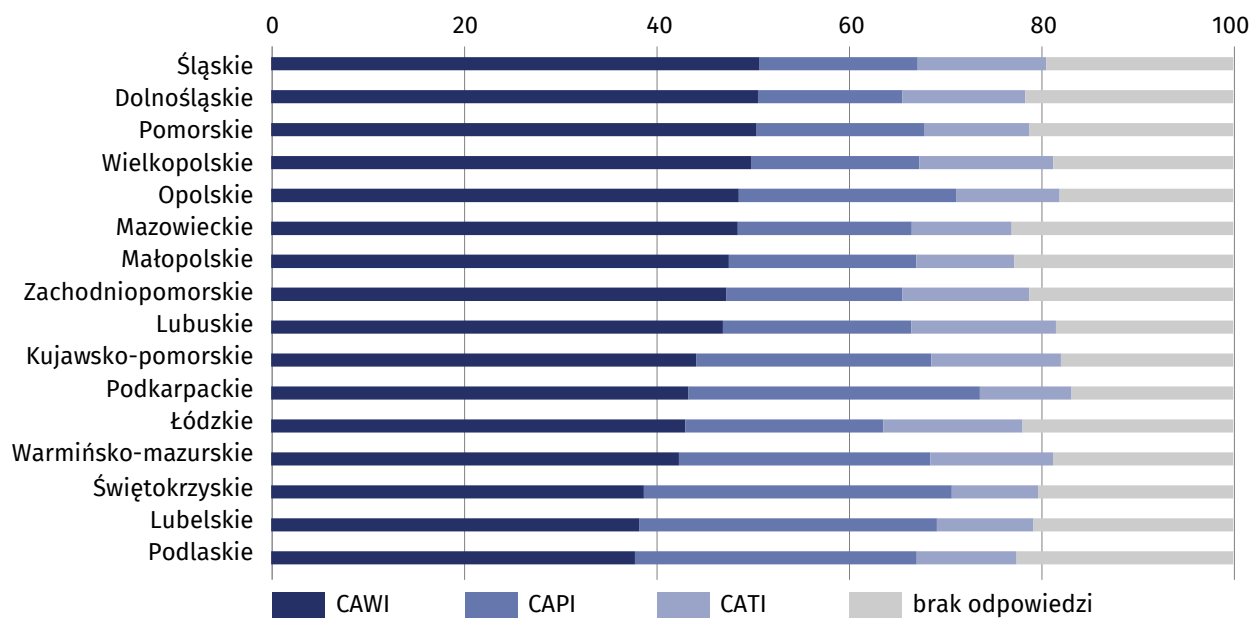
Wyniki spisu wskazują jednocześnie, że najwyższy odsetek mieszkań, które zostały wykluczone z zasobów mieszkaniowych objętych spisem, zaobserwowano w województwach: lubelskim (1,58%), podlaskim (1,52%) i świętokrzyskim (1,48%). Najmniejszy udział takich mieszkań odnotowano natomiast w województwach: opolskim (0,18%) oraz zachodniopomorskim (0,24%).

Istotnym elementem weryfikacji wykazu adresów mieszkań, które znalazły się w populacji docelowej, było potwierdzenie ich istnienia w ramach wywiadów spisowych. Na 15228932 adresów mieszkań (tj. zarówno mieszkań konwencjonalnych, jak i pomieszczeń mieszkalnych niebędących mieszkaniem), które znalazły się w docelowej populacji mieszkań, 12080515 adresów (79,3%) zostało wskazanych w ramach wywiadów spisowych jako miejsce stałego bądź czasowego zamieszkania, lub też ostatniego miejsca zamieszkania przed wyjazdem za granicę

W ramach wywiadów spisowych przeprowadzonych przy wykorzystaniu różnych kanałów zbierania danych najwyższy odsetek adresów mieszkań zostało wskazanych w województwach: podkarpackim – 83,2%, kujawsko-pomorskim – 82,1%, czy też opolskim – 81,9%, najmniej natomiast w województwach: mazowieckim – 72,0%, małopolskim – 77,2% oraz podlaskim – 77,5%.

Najczęściej wybieranym kanałem przekazywania danych był samospis internetowy (CAWI). Przy wykorzystaniu aplikacji internetowej zamieszczonej na stronie GUS wskazanych zostało 46,8% adresów mieszkań zlokalizowanych na terenie kraju 31 marca 2021 r. Wywiad bezpośredni (CAPI) znalazł zastosowanie w przypadku 20,6% adresów, a wywiad telefoniczny – 11,9%. Istnienie pozostałych 20,7% adresów mieszkań, które nie zostały wskazane w ramach wywiadów spisowych jako miejsce stałego lub czasowego zamieszkania, czy też adres ostatniego miejsca przebywania przed wyjazdem za granicę, zostało potwierdzonych w oparciu o źródła administracyjne.

Wykres 2. Struktura adresów mieszkań według kanałów zbierania danych (w %)
Chart 2. Structure of dwellings' addresses by channels of data collection (in %)



Samospis internetowy (CAWI) odegrał dominującą rolę w województwach: śląskim, dolnośląskim oraz pomorskim, w których przy wykorzystaniu tego kanału wskazanych zostało ponad połowa adresów mieszkań. Najrzadziej ten sposób wskazania miejsca zamieszkania wystąpił natomiast w województwach: podlaskim – 37,8% , lubelskim – 38,3% oraz świętokrzyskim – 38,7%.

Wywiady przeprowadzane bezpośrednio przez rachmistrzów spisowych (CAPI) najczęstsze zastosowanie znalazły w przypadku województw: świętokrzyskiego – 32,0% potwierdzonych adresów mieszkań, lubelskiego – 30,9% oraz podkarpackiego – 30,3%, a wywiad telefoniczny w województwach: lubuskim – 15,0% oraz łódzkim – 14,5%.

Samospis internetowy był też głównym źródłem informacji o adresach nowo dopisanych mieszkań, które nie były wcześniej zidentyfikowane i ujęte w Wykazie osobowo-adresowo-mieszkaniowym. Na 150204 adresów takich mieszkań poprzez internetową aplikację spisową dopisanych zostało 105989 mieszkań (tj. 70,6% dopisanych adresów), a w ramach wywiadu telefonicznego – 35379 mieszkań (23,6%).

2.1.1. Podsumowanie

1. W tabeli MIESZKANIA Wykazu osobowo-adresowo-mieszkaniowego według stanu NOBC na 31 marca 2021 r. znalazły się 15174364 mieszkania, które stanowiły populację mieszkań objętych spisem.
2. Wyniki uzyskane w ramach przeprowadzonego spisu wskazują, że na poziomie Polski ogółem populacja mieszkań wyznaczona w Wykazie została przeszacowana o 0,63% oraz niedoszacowana o 0,99%.
3. Spośród 15228932 mieszkań znajdujących się w populacji docelowej 79,3% stanowiły mieszkania, których adresy zostały wskazane w ramach wywiadów spisowych jako miejsce zamieszkania osób (stałego, czasowego lub też ostatniego miejsca zamieszkania przed wyjazdem za granicę).
4. Najwyższy odsetek adresów mieszkań z populacji docelowej (46,8%) został potwierdzony poprzez samospis internetowy (CAWI). 20,6% adresów mieszkań zostało wskazanych w ramach bezpośrednich wywiadów przeprowadzonych przez rachmistrzów spisowych (CAPI), a 11,9% poprzez wywiad telefoniczny (CATI). W przypad-

ku pozostałych mieszkań ich adresy zostały potwierdzone na podstawie informacji pochodzących ze źródeł administracyjnych.

5. Samospis internetowy był również głównym źródłem informacji o adresach nowo dopisanych mieszkań. W ten sposób identyfikowano 70,6% mieszkań spośród 150 204, które nie były wcześniej ujęte w Wykazie osobowo-adresowo-mieszaniowym.

2.1.2. Summary

1. The DWELLING table of the List of persons, addresses and dwellings according to the state of NOBC as at 31 March 2021 includes 15,174,364 dwellings, which constituted the population of dwellings covered by the census.
2. The results obtained as part of the conducted census indicate that at the level of Poland the population of dwellings designated in the List was overestimated by 0.63% and underestimated by 0.99%.
3. Of the 15,228,932 dwellings in the target population, 79.3% were dwellings whose addresses were indicated during census interviews as the place of residence of persons (permanent, temporary or last place of residence before going abroad).
4. The highest percentage of dwellings' addresses from the target population (46.8%) was confirmed by the on-line self-enumeration (CAWI). 20.6% of housing addresses were indicated as part of direct interviews conducted by census enumerators (CAPI), and 11.9% through telephone interviews (CATI). In the case of the remaining dwellings their addresses were confirmed on the basis of information from administrative sources.
5. The on-line self-enumeration was also the main source of information about the addresses of newly added dwellings. In this way, 70.6% of the 150,204 dwellings that were not previously included in the List of persons, addresses and dwellings were identified.

2.2. Osoby

2.2. Persons

Jak już zaznaczono wyżej (rozdział 1.2) wykaz osób pełnił przede wszystkim funkcję operacyjną w badaniu spisowym, tzn. że został wykorzystany w fazie kwestionariuszowego badania spisowego do celów techniczno-organizacyjnych, m.in. jako podstawa (narzędzie) do logowania i uwierzytelniania się respondentów oraz do obserwacji przebiegu i stopnia realizacji spisu. Z perspektywy takiej roli wykazu osób celowym wydawało się wykorzystanie całej tabeli wykazu osób, czyli obejmującej rekordy danych dotyczących wszystkich zidentyfikowanych w rejestrach osób, w tym również takich, które według wstępnej analizy weryfikacyjnej przyznano niskie prawdopodobieństwo przynależenia do populacji podlegającej spisowi⁵.

Niezależnie od powyższego w obrębie tabeli wykazu osób podjęto próbę wyznaczenia i oznakowania rekordów danych dla osób, które z wysokim prawdopodobieństwem odpowiadają populacji osób podlegających spisowi – czyli określono tzw. „wykazową populację docelową”. Wyznaczenie populacji docelowej w obrębie tabeli wykazu osób dokonano na podstawie kombinacji opisanych uprzednio przygotowanych flag weryfikacyjnych (rozdział 1.2).

Wykazowa populacja docelowa stanowi przybliżenie rzeczywistej (teoretycznej) populacji docelowej, którą z kolei najprościej można określić jako zbiorowość osób, które z uwagi na przyjęty zakres przedmiotowy i podmiotowy badania spisowego podlegają spisowi, a co za tym idzie – powinny zostać spisane.

⁵ Innymi słowy – użycie w systemie obsługi i kontroli przebiegu spisu szerokiego wykazu osób (obejmującego m.in. rekordy danych osób nienależących już do żadnej badanej zbiorowości) nie przysparzało żadnych negatywnych skutków dla przebiegu spisu, a jednocześnie dawało możliwość zainicjowania udziału w samospisie (autentykacji i zalogowania się) osobom niestety niezaliczonych do populacji docelowej.

W konkretnym wymiarze praktyczno-merytorycznym ostatniego spisu 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 krajowej⁶,
2. ludność rezydująca⁷ oraz
3. imigranci⁸.

Z merytorycznego punktu widzenia można wyróżnić jeszcze co najmniej kilka innych różnych populacji osób, ale każda z nich zawiera się już w którejś z tych trzech głównych wyżej wymienionych (np. populacja tzw. „emigrantów”, czyli mieszkańców Polski przebywających czasowo za granicą, zawiera się w ludności według definicji krajowej). W tym miejscu należy podkreślić, że na etapie przygotowań do spisu zdołano określić wykazową populację docelową jako całość (jeden podzbiór tabeli wykazu osób) – czyli bez wyróżniania w niej każdej z tych trzech wyżej wymienionych elementarnych populacji (kategorii ludności).

Należy zauważyć, że wyznaczenie populacji docelowej w tabeli wykazu osób miało również pewne organizacyjne zastosowanie w fazie badania ankietowego – jako dodatkowe kryterium obserwacji przebiegu spisu, ale przede wszystkim miało istotny walor w szerszej perspektywie metodologii demograficznej – jako sposobność do sprawdzenia możliwości skutecznego (trafnego) wyznaczenia badanych populacji ludnościowych wyłącznie na podstawie zastanych danych (zbiorów rejestrowych).

Wydaje się bowiem, że zdefiniowanie wykazowej populacji docelowej i jej weryfikacja poprzez badanie spisowe – pomimo braku podziału na kategorie ludności – stanowi ważny krok ku wypracowaniu alternatywnego źródła danych o ludności w okresach międzyspisowych.

Przedmiotem opisów i analiz w kolejnych akapitach tego podrozdziału jest zatem tak zdefiniowana (jeszcze przed badaniem spisowym) w obrębie danych tabeli wykazu osób wykazowa populacja docelowa⁹. W szczególności analizy te dotyczą jakości cech adresowych wykazowej populacji docelowej oraz jej relacji do ustalonej jako ostateczny wynik NSP 2021 populacji spisanych osób, czyli tzw. „oszacowanej populacji docelowej”.

Tablica 10. Wykazowa populacja docelowa według kompletności adresu

Table 10. List's target population by completeness of address

Kompletność adresu do poziomu: Completeness of address up to the level:	Ogółem Total	W % In %	% skumulowany % cumulated
POLSKA OGÓŁEM	39566532	100,00	x
Mieszkania Dwelling	37957943	95,93	95,93

6 Zgodnie z definicją krajową do ludności danej gminy zalicza się osoby tam zamieszkujące przez okres co najmniej 3 miesiące, zatem do ludności danej gminy zaliczani są stali mieszkańcy (osoby zameldowane) z wyjątkiem tych mieszkańców, którzy wyjechali na ponad 3 miesiące do innej gminy w kraju, a także osoby przybyłe z innego miejsca w kraju na okres ponad 3 miesiące. Natomiast do ludności gminy nie są zaliczani imigranci przebywający w Polsce czasowo, z kolei stali mieszkańcy Polski przebywający czasowo za granicą (bez względu na okres ich nieobecności) są uwzględniani w stanie ludności danej gminy

7 Według definicji międzynarodowej, do rezydentów danej gminy zalicza się wszystkie osoby mieszkające lub zamierzające mieszkać w tej jednostce przez co najmniej 1 rok. Na rezydentów składają się zatem: stali mieszkańcy (osoby zameldowane lub stale mieszkające bez zameldowania), z wyjątkiem tych mieszkańców, którzy wyjechali na okres co najmniej 12 miesięcy do innego miejsca w kraju lub za granicę; osoby przybyłe z innego miejsca w kraju lub z zagranicy (imigranci bez karty pobytu) na okres co najmniej 12 miesięcy.

8 Osoby przybyłe z zagranicy do kraju w celu osiedlenia się (zamieszkania na stałe) lub na pobyt czasowy.

9 Wykazowa populacja docelowa jest tu analizowana jako rezultat zastosowania alternatywnej wobec badania spisowego metody wyznaczania populacji docelowej, ale jednocześnie jako (potencjalny) operat osób. Jakkolwiek bowiem sposób wykorzystania wykazu osób w badaniu spisowym – jak wykazano wyżej – nie wymuszały restrykcyjnego określania operatu, np. poprzez usunięcie rekordów danych osób, które z dużym prawdopodobieństwem nie podlegają spisowi, to gdyby zaistniała taka potrzeba (użycia ścisłego operatu), to właśnie wyznaczona w wykazie populacja docelowa stanowiłaby ten operat.

Tablica 10. Wykazowa populacja docelowa według kompletności adresu (dok.)**Table 10. List's target population by completeness of address (cont.)**

Kompletność adresu do poziomu: Completeness of address up to the level:	Ogółem Total	W % In %	% skumulowany % cumulated
Budynku Building	188557	0,48	96,41
Ulicy Street	979470	2,48	98,89
Miejscowości Locality	33264	0,08	98,97
Gminy Gmina	2388	0,01	98,98
Brak danych adresowych No address data available	404910	1,02	100,00

Na podstawie Tablicy 10, która przedstawia stan kompletności adresów przyporządkowanych jednostkom z wykazowej populacji docelowej można stwierdzić, że populacja ta obejmowała 39566532 osoby, a dla 95,9% przypadków z tej liczby zdołano ustalić pełen adres, co – z technicznego punktu widzenia – wyrażało się przyporządkowaniem identyfikatora mieszkania z tabeli wykazu mieszkań. Natomiast w kontekście konieczności przekazywania liczby ludności na poziomie siatki kilometrowej¹⁰ istotna wydaje się konstatacja, że w sumie w odniesieniu do 96,4% z opisywanej populacji zdołano ustalić położenie z dokładnością do co najmniej budynku – co daje możliwość bezpośredniego nadania osobie koordynatów przestrzennych – a wśród pozostałej części populacji, cechującej się mniej precyzyjnymi danymi adresowymi (ale z możliwością pośredniego – szacunkowego przyporządkowania do kwadratów siatki) wyraźną większość (2,48%) stanowiły jednostki opatrzone danymi z dokładnością do ulicy. Pewnym wyzwaniem metodologicznym pozostają natomiast przypadki bez żadnych danych adresowych, które jednak można uznać za relatywnie niewielką część populacji, bo stanowiące w niej niemal dokładnie 1%.

Tablica 11. Wykazowa populacja docelowa według pochodzenia adresu**Table 11. List's target population by origin of address**

Źródło pochodzenia adresu Address source	Ogółem Total	W % In %
POLSKA OGÓŁEM	39566532	100,00
KEP	29696417	75,05
PESEL	8669874	21,91
ZUS	406159	1,03
NFZ	326486	0,83
KRUS	61955	0,16

10 Rozporządzenie Wykonawcze Komisji (UE) 2018/1799 z dnia 21 listopada 2018 r. w sprawie ustanowienia tymczasowego bezpośredniego działania w dziedzinie statystyki dotyczącego rozpowszechniania wybranych tematów spisu ludności i mieszkań z 2021 r. geokodowanych w siatce kilometrowej.

Tablica 11. Wykazowa populacja docelowa według pochodzenia adresu (dok.)**Table 11. List's target population by origin of address (cont.)**

Źródło pochodzenia adresu Address source	Ogółem Total	W % In %
ARIMR	731	0,00
Brak danych adresowych No address data available	404910	1,02

Na podstawie Tablicy 11 opisującej źródła pochodzenia informacji adresowych wykazowej populacji docelowej należy przede wszystkim podkreślić, że głównym, tym samym najważniejszym – a według hierarchii referencyjności opisanej w rozdziale 1.2 – także najbardziej pożądanym źródłem danych o adresie jest przede wszystkim rejestr KEP, z którego otrzymało adresy ¾ populacji, a następnie rejestr PESEL, z którego pozyskano dane adresowe dla kolejnych prawie 22% populacji. Poza tym można w uproszczeniu zauważyć, że ilościowy porządek źródeł adresów w znacznym stopniu oddaje opisaną w rozdziale 1.2 hierarchię referencyjności rejestrów pod względem danych adresowych, wszakże z dwoma widocznymi wyjątkami: wyższą pod względem ilościowym pozycją (miejsce 2) adresów pochodzących z rejestru PESEL od bardziej poświadczonych z rejestru ZUS (3) oraz ilościowo wyższą (o dwa miejsca) od oczekiwanej pozycją adresów z NFZ, sytuowanego pod kątem referencyjności adresów na ostatnim miejscu wśród źródeł.

W kontekście analizy relacji wykazowej populacji docelowej do oszacowanej populacji docelowej warto na wstępie wskazać jak należy rozumieć tę drugą z wymienionych populacji, wyjaśniając, że oszacowana populacja docelowa to również przybliżenie rzeczywistej (teoretycznej) populacji docelowej (osób podlegających spisowi), dokonane jako ostatecznie opracowany wynik spisu (w zakresie stanów ludności), reprezentowany fizycznie w postaci rekordów danych tabeli osób, utworzonej w ramach WZDS (Wynikowy Zbiór Danych Spisowych).

Tablica 12. Oszacowana populacja docelowa według kategorii ludności oraz źródeł danych**Table 12. Estimated target population by population category and data source**

Rodzaj źródła danych (sposób spisania) Type of data source (enumeration method)	Ogółem Grand total	W tym: kategorie ludności In which: population categories		
		ludność według definicji krajowej population by national definition	ludność rezydująca residents	imiigranci immigrants
Ogółem Grand total	39534391	38036118	37019327	1433779
		w % in %		
		100,00	100,00	100,00
Badanie ankietowe – razem Field survey – total	35134955	88,87	91,25	25,40
kanaly CAxI – razem CAxI channels – total	35000342	88,53	90,91	25,00
CAWI	20386182	51,57	52,99	13,35
CAPI	9791267	24,77	25,63	2,41

Tablica 12. Oszacowana populacja docelowa według kategorii ludności oraz źródeł danych (dok.)**Table 12. Estimated target population by population category and data source (cont.)**

Rodzaj źródła danych (sposób spisania) Type of data source (enumeration method)	Ogółem Grand total		W tym: kategorie ludności In which: population categories		
			ludność według definicji krajowej population by national definition	ludność rezydująca residents	imigranci immigrants
CATI	4822893	12,20	12,28	12,46	9,24
badanie OZZ CLQs survey	134613	0,34	0,34	0,34	0,40
Dane z rejestrów Data from registers	4399436	11,13	8,75	8,43	74,60

Oszacowana populacja docelowa różni się od wykazowej – powstałej wyłącznie na podstawie danych z rejestrów – przede wszystkim złożonością typów źródeł pochodzenia tworzących (określających) ją danych. Na podstawie Tablicy 12, która prezentuje oszacowaną populację docelową oraz wchodzące w jej skład, a jednocześnie konstytuujące ją, rodzaje populacji podstawowych (kategorie ludności) według źródeł danych (sposobu spisania osób), można odnotować, że oszacowana jako wynik spisu populacja docelowa liczy 39534391 osób, a tym samym jest bardzo zbliżona – pod względem liczności – do opisywanej wyżej wykazowej populacji docelowej (różnica wynosi nieco ponad 32 tys.).

Ponadto dane Tablicy 12 wskazują, że w zdecydowanej większości, bo w prawie 88,9% przypadków, rekordy w wynikowym zbiorze danych o osobach, wyznaczające oszacowaną populację docelową, pochodzą z zainicjowanych w ramach spisu różnych metod badania ankietowego. Przy tym warto podkreślić, że w głównej mierze są to dane z tzw. kanałów CAxI, czyli technik badania bazujących na pełnowymiarowym indywidualnym kwestionariuszu elektronicznym, dedykowanym głównie osobom w mieszkaniach – 88,53%, a tylko w niewielkiej części informacje uzyskane za pomocą metod przeznaczonych dla OZZ – 0,34%. Z kolei pod względem rodzaju konkretnej techniki spisywania należy zauważyć, że najbardziej powszechnym źródłem danych o osobach okazał się samospis internetowy (CAWI), zasilający 51,6% rekordów danych wynikowych. Stosunkowo niewielka część opisujących populację docelową osób, pochodzi wyłącznie z rejestrów.

Porównując w tym kontekście poszczególne kategorie ludności można stwierdzić, że pod względem udziału danych z badania ankietowego minimalnie lepiej prezentuje się ludność rezydująca (91,57%) od ludności według definicji krajowej (91,25%). Wyraźnie natomiast pod tym względem odstaje relatywnie niewielka (1433,8 tys.) zbiorowość imigrantów, wśród której udział uczestniczących w spisowym badaniu kwestionariuszowym stanowił zaledwie 25,4%, co wynika z faktu, że w jej skład wchodzi w głównej mierze cudzoziemcy, którzy – jak się zdaje – są na ogół mniej skłonni do wzięcia udziału w badaniu spisowym w kraju czasowego pobytu.

Z punktu widzenia realizowanej tu krytycznej analizy wykazowej populacji docelowej kluczowe wydaje się określenie jej zgodności z oszacowaną jako wynik spisu populacją docelową, którą przyjmuje się tu jako kryterium poprawności – jako najlepsze odwzorowanie rzeczywistej populacji docelowej.

Stwierdzone wyżej podobieństwo liczebnościowe obu populacji nie stanowi jeszcze o ich podobieństwie czy zbieżności, a tym samym o trafności wyznaczenia populacji docelowej na podstawie danych rejestrowych, czy – innymi słowy – jej jakości. W tym kontekście istotne jest przede wszystkim sprawdzenie zgodności obu omawianych populacji na poziomie wchodzących w ich składy poszczególnych jednostek, którego wyniki prezentują Tablice 13 i 14.

Tablica 13. Wykazowa populacja docelowa według pochodzenia adresu**Table 13. List's target population by origin of address**

Wyszczególnienie Specification	Ogółem Total	W % In %
Wykazowa populacja docelowa – ogółem List's target population	39566532	100,00
Jednostki zgodne z oszacowaną populacją docelową (wg NSP 2021) Units consistent with estimated target population (according to the 2021 Census)	39014242	98,60
Jednostki niezgodne z oszacowaną populacją docelową Units not consistent with estimated target population (according to 2021 Census).	552290	1,40

Tablica 14. Oszacowana populacja docelowa według zgodności z wykazową populacją docelową**Table 14. Estimated target population by consistency with the list's target population**

Wyszczególnienie Specification	Ogółem Total	W % In %
Oszacowana w NSP 2021 populacja docelowa – ogółem The 2021 Census estimated target population – total	39534391	100,00
Jednostki zgodne z wykazową populacją docelową Units consistent with list's target population	39014242	98,68
Jednostki niezgodne z wykazową populacją docelową Units not consistent with list's target population	520149	1,32

Na ich podstawie można stwierdzić, że zbiór wspólnych (sparowanych) jednostek w obu populacjach obejmuje 39014242 osoby, co stanowi 98,60% wykazowej populacji docelowej i 98,68% oszacowanej populacji docelowej. W konsekwencji stwierdzonej zgodności jednostek w obu populacjach, 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 (ang. over-coverage). 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 (ang. under-coverage).

Wydają się zatem, że przywołane wyżej wielkości upoważniają do stwierdzenia, iż rozbieżności w pokryciu obu populacji są ogólnie biorąc niezbyt duże, a co za tym idzie – że obrany kierunek postępowania w zakresie określania populacji docelowej (jako całości) na podstawie danych rejestrowych można uznać za trafny.

Warto w tym miejscu jeszcze dodać, że z ogólnej liczby 520,1 tys. osób należących do oszacowanej populacji docelowej a niezaliczonych uprzednio do wykazowej populacji docelowej większość, bo 424,1 tys., występowała

w szerokim (operacyjnym) wykazie osób (osoby te znaleziono w rejestrach, ale błędnie nie zakwalifikowano do populacji docelowej). Natomiast zupełnie poza przygotowanym do spisu wykazem było w sumie 96,1 tys. spośród zaliczonych ostatecznie do oszacowanej populacji docelowej osób, z czego większą zbiorowość, obejmującą 56 tys. osób, stanowili odnalezieni w ramach wyliczania wyników spisu w rejestrach imigranci, w przeważającej mierze cudzoziemcy, którzy najprawdopodobniej do momentu spisu nie zaistnieli w rejestrach, bądź też pominęły ich procedury aktualizujące wykaz z 31 grudnia 2020 r. do stanu na 31 marca 2021 r., zaś pozostałą część, liczącą 40,1 tys. osób, zidentyfikowano wyłącznie w kanałach badania spisowego CAxI.

Uprawnione wydaje się przekonanie, że mając do dyspozycji oszacowaną na podstawie spisu populację docelową będzie można dość łatwo doprecyzować (uszczelnić) jeszcze algorytmy wyznaczania w przyszłości wykazowej populacji docelowej (jako całości), zwiększając trafność jej oszacowania, np. poprzez doprecyzowanie weryfikacyjnych oznaczników, znuansowanie niektórych flag wykluczających lub dodanie nowych.

Więszym wyzwaniem natomiast pozostaje ciągle jeszcze wyznaczenie konkretnych kategorii ludności w obrębie wykazowej populacji docelowej, które jednak, przy odpowiedniej analizie dostępnych w wykazie (rejestrach) cech w relacji z danymi spisowymi, jawi się jako zdanie jak najbardziej realne.

2.2.1. Podsumowanie i wnioski

1. W celu dokonania oceny jakości wykazu w odniesieniu do zbioru osób wyznaczono w obrębie Wykazu osób tzw. wykazową populację docelową. Populacja ta stanowi przybliżenie rzeczywistej (teoretycznej) populacji docelowej, będącej – z uwagi na przyjęty zakres przedmiotowy i podmiotowy badania spisowego – zbiorowością osób podlegających spisowi (tj. która powinna zostać spisana).
2. Przeprowadzona analiza dotyczy jakości cech adresowych wykazowej populacji docelowej oraz jej relacji do ustalonej jako ostateczny wynik NSP 2021 populacji spisanych osób, czyli tzw. „oszacowanej populacji docelowej”.
3. Oszacowana populacja docelowa to również przybliżenie rzeczywistej (teoretycznej) populacji docelowej (osób podlegających spisowi), 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).
4. 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 zdołano ustalić położenie z dokładnością do co najmniej budynku.
5. 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.
6. Oszacowana jako wynik spisu populacja docelowa liczy 39534391 osób, a tym samym jest bardzo zbliżona – pod względem liczności – do wykazowej populacji docelowej (różnica wynosi nieco ponad 32 tys.).
7. W zdecydowanej większości, bo w prawie 88,9% przypadków, rekordy w wynikowym zbiorze danych o osobach, wyznaczające oszacowaną populację docelową, pochodzą z zainicjowanych w ramach spisu różnych metod badania ankietowego (kanały CAxI). Najbardziej powszechnym źródłem danych o osobach okazał się samospis internetowy (CAWI), zasilający 51,6% rekordów danych wynikowych.
8. Analiza pokazuje, że zbiór wspólnych (sparowanych) jednostek w obu populacjach obejmuje 39014242 osoby, co stanowi 98,60% wykazowej populacji docelowej i 98,68% oszacowanej populacji docelowej. W konsekwencji stwierdzonej zgodności jednostek w obu populacjach, 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. Rozbieżności w pokryciu obu populacji są ogólnie biorąc niezbyt duże, a co za tym idzie można uznać, że obrany kierunek postę-

powania w zakresie określania populacji docelowej (jako całości) na podstawie danych rejestrowych można uznać za trafny.

9. Uprawnione wydaje się przekonanie, że mając do dyspozycji oszacowaną na podstawie spisu populację docelową będzie można dość łatwo doprecyzować (uszczelnić) jeszcze algorytmy wyznaczania w przyszłości wykazowej populacji docelowej (jako całości), zwiększając trafność jej oszacowania, np. poprzez doprecyzowanie weryfikacyjnych oznaczników, zniuansowanie niektórych flag wykluczających lub dodanie nowych.
10. Większym wyzwaniem natomiast, biorąc pod uwagę nową perspektywę realizacji spisów ludności i mieszkań oraz projektowane przez Komisję Europejską coroczne przekazywanie przez kraje członkowskie UE liczby i struktury ludności z uwzględnieniem siatki kilometrowej oraz szerokiego zestawu charakterystyk demograficzno-społecznych i migracyjnych, pozostaje ciągle jeszcze wyznaczenie kategorii ludności rezydującej, w obrębie wykazowej populacji docelowej. Jednakże przy odpowiedniej analizie dostępnych w wykazie (rejestrach) cech w relacji z danymi spisowymi jest to zadaniem, które ma szansę być zrealizowane z sukcesem.

2.2.1. Summary and conclusion

1. In order to assess the quality of the list in relation to the set of persons, the so-called list's target population has been designated. This population is an approximation of the actual (theoretical) target population, which is – due to the adopted scope and subject matter of the census survey – the set of persons to be enumerated.
2. The analysis carried out concerns the quality of the address characteristics of the list's target population and its relationship to the census enumeration population established as the final result of the 2021 Census, the so-called „estimated target population”.
3. The estimated target population is also an approximation of the actual (theoretical) target population (people, who should be enumerated), made as the final compiled census result (in terms of population states), physically represented in the form of data records of the table of persons, created within the Resulting Census Data Set.
4. The list's target population comprised 39566532 persons, a full address was established for 95.9%. For 96.4% of this population, a location was established with accuracy to at least a building.
5. The source of address data was primarily the Tax register (KEP) from which addresses were obtained for ¾ of the population, followed by the PESEL register, from which address data was obtained for a further almost 22% of the population.
6. The target population estimated as a result of the 2021 Census is 39534391 persons and is thus very close – in terms of numbers – to the list's target population (the difference is just over 32 thousand).
7. The vast majority, almost 88.9%, of the records in the Resulting Census Data Set, delineating the estimated target population, came from the various survey methods (CAXI channels) initiated as part of the census. The most common source of data on individuals was the online self-enumerated channel (CAWI), feeding 51.6% of the resulting data records.
8. The analysis shows that the set of common (paired) units in the two populations comprises 39014242 individuals, representing 98.60% of the list's target population and 98.68% of the estimated target population. The scale of discrepancy considered from the perspective of the list's target population is 552,290 persons included in it, i.e. 1.40% of its number, who were not ultimately included in the estimated target population – which is a measure of the over-coverage error (overestimation) of the list's target population. In turn, 520,149 people, i.e. 1.32% of those belonging to the estimated target population, were not included in the list's target population, which is expressed by the size of the error of insufficient coverage (underestimation) by the list's target population. The discrepancies in the coverage of both populations are generally not very large, and thus it can be considered that the chosen course of action in determining the target population (as a whole) on the basis of registry data can be considered accurate.
9. It seems justified to believe that having the target population estimated on the basis of the census, it will be quite easy to specify (tighten) the algorithms for determining the target population (as a whole) in the

II. ASSESSING THE QUALITY OF THE HOUSING AND PERSONS LIST

future, increasing the accuracy of its estimation, e.g. by specifying the verification markers (flags), nuanced some exclusion flags or added new ones.

10. A greater challenge, however (taking into account the new perspective of the implementation of population and housing censuses and the planned annual submission by the EU Member States of the number and structure of population by the kilometer grid and a wide set of demographic, social and migration characteristics) is still to be determined resident population within the list's target population. However, with proper analysis of variables available in the list (registers) in relation to census data, it is a task that has a chance to be successfully completed.

III. Dokładność wyników spisu ludności i mieszkań

III. Accuracy of the 2021 Census results

3.1. Klasyfikacje błędów

3.1. Classification Error

Klasyfikacja błędów, które występują w różnego rodzaju badaniach statystycznych obejmuje dwa zasadnicze ich rodzaje: błędy losowe i nielosowe. Pierwszy rodzaj błędów występuje zazwyczaj w badaniach reprezentacyjnych bazujących na odpowiednio dobranej próbie losowej. Wielkość tego błędów jest uzależniona od wielu czynników, do których można zaliczyć liczebność próby, zróżnicowanie badanej cechy, czy przyjęty na potrzeby badania schemat losowania próby. Ten rodzaj błędów na ogół nie występuje w spisach powszechnych, które są badaniami pełnymi. Nie można ich jednak pomijać, zwłaszcza w nowoczesnych spisach wykorzystujących wiele źródeł danych (spis oparty na metodach mieszanych) czy w spisach kroczących¹¹. Ten rodzaj błędów jest dobrze rozpoznany i opisywany w literaturze z zakresu metody reprezentacyjnej i badań próbkowych.

Znacznie poważniejsze konsekwencje w kontekście jakości oszacowań, w węższym ujęciu, i jakości całego badania w szerszym kontekście, mają błędy nielosowe, tj. których źródłem nie jest mechanizm odpowiedzialny za losowy dobór jednostek do próby (Szreder, 2010). Występują one zarówno w badaniach reprezentacyjnych, jak i w badaniach pełnych takich jak spisy, a także w źródłach administracyjnych. Wśród błędów nielosowych wyróżnia się ich dwie zasadnicze kategorie: błędy pokrycia (zwane również błędami kompletności) oraz błędy treści.

Błędy pokrycia dotyczą sytuacji, w której pominięto pewne jednostki w procesie spisywania, co prowadzi zazwyczaj do niedoszacowania. Mogą one również odnosić się do powtórzeń tj. sytuacji, w której ta sama jednostka jest spisywana więcej niż raz. Zazwyczaj prowadzi to do zjawiska przeszacowania w kontekście spisywanej populacji. Błędy pokrycia odnosić się mogą wreszcie do błędnego zakwalifikowania pewnych jednostek tj. spisania jednostek, które nie powinny zostać ujęte w badaniu spisowym. W odniesieniu do badań spisowych, bez względu na metodę realizacji spisu, kluczową kategorię błędów wśród błędów pokrycia odgrywają również błędy wynikające z nieuzyskania odpowiedzi (na przykład odmowa udziału w badaniu czy nieobecność podczas badania) oraz błędy w wykazie jednostek badania związane z ich aktualnością czy kompletnością.

Drugą wspomnianą kategorię błędów nielosowych są błędy treści, które mogą być konsekwencją niewłaściwych odpowiedzi, opracowania, analizy i prezentacji czy niepoprawnego zakwalifikowania wariantów rozpatrywanych cech (Gołata, 2018).

3.2. Wybrane metody oceny jakości spisów

3.2. Selected methods of assessing the quality of censuses

W literaturze przedmiotu wyróżnia się szereg metod oceny jakości spisów powszechnych w zależności od stopnia ich złożoności, rodzaju wykorzystywanych mierników czy źródeł danych wziętych pod uwagę w procesie ewaluacji. Jedną z najczęściej wykorzystywanych klasyfikacji odnosi się do źródeł informacji uwzględnionych na etapie oceny jakości spisu (Kordos, 2007; Gołata, 2018). Zgodnie z nią wyróżnia się trzy typy metod oceny jakości danych spisowych:

- oparte wyłącznie na danych z analizowanego spisu (jedno źródło),
- oparte na porównaniu wyników spisowych z wynikami z innych badań statystycznych (wiele źródeł danych),
- oparte na wynikach z dodatkowo przeprowadzonych badań.

¹¹ Łączenie źródeł, w oparciu o które kraje UE opracowują dane spisowe, określa artykuł 4 Rozporządzenia (WE) Parlamentu Europejskiego i Rady nr 763/2008 z dnia 9 lipca 2008 r. w sprawie spisów powszechnych ludności i mieszkań (Dz. Urz. UE L 218 z 13.8.2008).

Każda z tych metod może być wykorzystana na potrzeby oceny skali wspomnianych błędów pokrycia i treści. W pierwszej i drugiej grupie metod na szczególną uwagę zasługuje analiza demograficzna danych spisowych, która jest jedną z najczęściej wykorzystywanych technik w ocenie jakości spisu. Jest ona obok spisu kontrolnego metodą rekomendowaną przez Dywizję Statystyczną Organizacji Narodów Zjednoczonych (United Nations, 2017).

Analiza demograficzna jest ważnym narzędziem oceny danych spisowych, szczególnie w krajach, w których brakuje niezależnych źródeł danych, takich jak rejestracja ludności i badania reprezentacyjne, lub w których nie prowadzi się spisu kontrolnego, będącego repliką spisu zasadniczego.

W obszarze analizy demograficznej można wyróżnić szereg podejść bazujących na różnych rozwiązaniach metodologicznych różniących się wymaganiami dotyczącymi danych, jakością wyników czy zaawansowaniem aparatu matematycznego wymaganego do jego zastosowania. Może ona przykładowo odnosić się do prostych rozwiązań związanych z porównaniem struktur spisowych według płci czy wieku. Można również rozpatrywać w analizie demograficznej uwzględnienie różnych współczynników demograficznych umożliwiających szacowanie liczby ludności czy skomplikowanych modeli matematycznych, które mogą być źródłem informacji o zgodności i poprawności wyników spisowych. Metody analizy demograficznej wymagają na ogół wiarygodnych danych na temat urodzeń, zgonów czy migracji, a ich pewną słabością jest to, że generalnie nie dostarczają wystarczających informacji, pozwalających na oddzielenie błędów pokrycia od błędów treści. Do ogólnej oceny jakości spisu ludności, metody stosowane do oceny błędu pokrycia wykorzystują dane dotyczące grup wieku i płci lub kohort według wieku. Piramida wieku i płci jest tutaj standardową metodą, podobnie jak sumaryczne wskaźniki, takie jak indeks Whipple'a i Indeks Myersa. Przeprowadza się również analizę ludności ustabilizowanej, o ile spełnione są pewne założenia, takie jak stały współczynnik płodności, umieralności oraz brak migracji. W obszarze analizy demograficznej wyniki spisu mogą być porównywane z danymi z innych źródeł informacji, takich jak rejestracja urodzeń i zgonów oraz saldo migracji. Z kolei podejście kohortowe w analizie demograficznej wykorzystujące dane z dwóch kolejnych spisów powszechnych, a także wskaźniki przeżycia, współczynniki dzietności zależne od wieku i szacunkowe poziomy migracji międzynarodowej między spisami powszechnymi umożliwia porównania między wynikami obydwu rozważanych spisów. Populacja wyliczona w pierwszym spisie jest prognozowana do daty referencyjnej drugiego spisu na podstawie szacunkowych poziomów i rozkładów wieku płodności, umieralności i migracji, a „przewidywana” populacja jest porównywana z populacją wyliczoną w drugim spisie.

Generalnie w obszarze analizy demograficznej wykorzystywanej w ocenie jakości spisu wyróżnia się cztery zasadnicze grupy metod (United Nations, 2017; Gołata, 2018):

1. polegające na międzyspisowych szacunkach stanu i struktury populacji na podstawie bieżącej rejestracji ruchu naturalnego i wędrownego ludności (urodzenia, zgonu i migracje),
2. polegające na wykorzystaniu projekcji demograficznej bazującej na strukturze ludności z poprzedniego spisu z uwzględnieniem płodności, umieralności oraz migracji na podstawie różnych źródeł oraz porównanie wyników szacunków z wynikami z poprzedniego spisu,
3. polegające na porównaniu wyników analizy kohortowej na podstawie danych poprzedniego spisu z wynikami spisu podlegającego ocenie,
4. polegających na szacunku korygujących współczynników pokrycia z wykorzystaniem odpowiednio dobranych modeli regresji, które mają zapewnić spójność struktur ludności w dwóch kolejnych spisach.

Drugą z metod ewaluacji jakości spisów powszechnych szeroko rekomendowaną przez ONZ jest spis kontrolny. Wpisuje się on w grupę metod oceny jakości spisów tj. opartych na wynikach z dodatkowo przeprowadzonych badań. Jest to specjalne badanie reprezentacyjne przeprowadzane w sposób niezależny po zakończeniu właściwego spisu i będące jego repliką. Głównym celem badania kontrolnego jest:

- ocena stopnia pokrycia podczas właściwego spisu, tj. oszacowanie błędu pokrycia na poziomie całego kraju, jak i w różnych przekrojach terytorialnych i domenach,
- ocena wpływu braków w zakresie pokrycia, jeśli występują, na użyteczność danych spisowych,
- uzyskanie informacji do projektowania przyszłych spisów powszechnych i badań reprezentacyjnych,
- zbadanie cech osób, które mogły zostać pominięte podczas spisu.

Dobrze zaplanowane badanie kontrolne powinno umożliwić porównanie precyzyjnych oszacowań z wynikami uzyskanymi w spisie w odniesieniu do kluczowych cech i w konsekwencji otrzymanie informacji o kompletności, tj. o skali błędów pokrycia. Badanie to powinno również dostarczyć najważniejszych informacji o błędach treści w odniesieniu do wybranych charakterystyk.

Ze względu na fakt, że spis kontrolny powinien być repliką spisu ludności i mieszkań przy jego organizacji i realizacji powinno przestrzegać się kilku istotnych założeń:

- czas między spisem powszechnym a spisem kontrolnym powinien być jak najkrótszy, aby uniknąć błędów pamięci respondentów oraz zmian zachodzących w analizowanej populacji (urodzenia, zgonu i migracje),
- spis kontrolny powinien być niezależny względem spisu ludności i mieszkań (ankieterzy nie powinni mieć informacji o udzielonych przez respondentów odpowiedziach w spisie zasadniczym spis kontrolny powinien być zrealizowany przez odrębną, niezależną organizację),
- aby zachować niezależność spisu kontrolnego, proces gromadzenia danych i operacje przetwarzania muszą być całkowicie oddzielone od gromadzenia i przetwarzania danych z zasadniczego spisu.

IV. Badanie kontrolne w spisie ludności i mieszkań 2021

IV. The control survey in the 2021 Census

Badanie kontrolne w spisie ludności i mieszkań 2021 składało się z dwóch faz.

Pierwsza faza działań kontrolujących wykonywana była podczas przebiegu spisu zasadniczego i polegała na rutynowej kontroli poprawności pracy wybranej losowo lub celowo 2% grupy rachmistrzów, innej w każdym miesiącu. Ze względu na realizację spisu metodami teleinformatycznymi, bieżąca kontrola tej fazy odbywała się w zakresie danych wprowadzanych do formularzy spisowych przez respondentów i rachmistrzów spisowych w spisie zasadniczym, w oparciu o założenia zaimplementowane do systemu, który nie pozwalał na kontynuację spisu w przypadku wykrycia błędu podczas wypełniania ankiety.

Zasadniczym celem drugiej fazy badania kontrolnego była ocena jakości spisu ludności i mieszkań, w szczególności ocena błędów zawartości dla badanych cech. Do oceny wykorzystano dane zebrane na podstawie specjalnie zaprojektowanego badania reprezentacyjnego. W wyniku analizy opracowano szereg wskaźników statystycznych oraz sformułowano wnioski.

Ze względu na istniejące ograniczenia wynikające z COVID-19, przy planowaniu badania kontrolnego podjęto szereg decyzji, które w istotny sposób ograniczyły badaną populację mieszkań, do około 64% mieszkań zamieszkałych (brano pod uwagę głównie dostępność numeru telefonu; szczegółowe założenia podano w dalszej części). Z tego powodu wnioski uzyskane na podstawie badania kontrolnego mogły dotyczyć tylko tej ograniczonej populacji. Rozszerzenie tych wniosków na całą populację będącą przedmiotem spisu zasadniczego należy traktować z dużą ostrożnością, w szczególności nie ma uzasadnienia metodologicznego do wyznaczenia dokładnych ocen liczbowych błędów pokrycia (czyli m.in. określenia oszacowania liczby osób, które nie zostały ujęte w spisie zasadniczym). Jak wynika z literatury przedmiotu oraz doświadczeń historycznych z wielu krajów przeprowadzających spisy kontrolne, aby w pełni ocenić błędy pokrycia i zawartości należy przeprowadzać dodatkowy niezależny spis kontrolny (najczęściej z wykorzystaniem specjalnie losowanych prób), który musi spełniać szereg wymagań metodologicznych (kluczowym założeniem teoretycznym jest niezależność działań dotyczących prac w ramach spisu zasadniczego i kontrolnego), co podkreślono w poprzednim rozdziale.

W Polsce nie ma tradycji w stosowaniu takiego pełnego podejścia. Głównym czynnikiem ograniczającym w polskich warunkach są nadmierne koszty oraz ograniczone zasoby personelu realizującego prace w ramach systemu statystyki publicznej, a także panujące warunki sanitarne związane z COVID-19. W związku z tym opisane poniżej podejście do kontroli jakości w ramach NSP 2021 jest wynikiem koniecznego z praktycznego punktu widzenia kompromisu. Zastosowane uproszczone podejście (stosowane również w poprzednich spisach) umożliwia ocenę jakości spisu jedynie pod kątem błędów zawartości (treści). Problematyka badań kontrolnych w poprzednich spisach ludności w Polsce została omówiona m.in. w artykule Kordosa (2007) oraz w opracowaniu Gołaty (2018).

W zaproponowanym rozwiązaniu skorzystano z metodyki badań w zakresie błędów pomiarów i wykorzystania określonej podpróby celem ich oceny. W literaturze to podejście można znaleźć pod pojęciem dwufazowego badania z wykorzystaniem próby walidacyjnej (ang. two-phase validation sampling; Shepherd i in. 2022). Metoda ta stosowana jest w praktyce oceny jakości rejestrów administracyjnych (np. rejestry administracyjne służby zdrowia, rejestry bezrobotnych, por. Schenkel i Zhang 2022).

4.1. Opis losowania próby do badania kontrolnego w NSP 2021

4.1. Description of drawing the sample for the control survey in the 2021 Census

W ramach realizacji badania kontrolnego przyjęto następujące założenia:

1. Badanie kontrolne przeprowadzono w terminie od 12 do 24 listopada 2021 r.
2. Badanie kontrolne realizowano jako badanie reprezentacyjne.
3. Operat do badania kontrolnego zdefiniowano jako zbiór zbadanych w spisie zasadniczym mieszkań z dostępnym numerem telefonu, dla których zrealizowano spis (metodami CAWI, CATI i CAPI), przy czym brana była pod uwagę dostępność numeru telefonu dla którejkolwiek z osób spośród zbadanych pod danym adresem, natomiast jednostką badania była osoba, która w spisie zasadniczym bezpośrednio udzielała wywiadu (tzw. pierwszy respondent). Tym samym liczba mieszkań w operacie badania kontrolnego była tożsama z liczbą reprezentantów.
4. Badaniu kontrolnemu nie podlegały:
 - a) mieszkania zamieszkane wyłącznie przez cudzoziemców,
 - b) osoby małoletnie,
 - c) osoby bezdomne,
 - d) obiekty zbiorowego zakwaterowania (OZZ),
 - e) pomieszczenia niebędące mieszkaniami,
 - f) osoby spisane poza granicami Polski,
 - g) mieszkania w budynkach bez ustalonego numeru mieszkania.
5. Wielkość próby do badania kontrolnego ustalono na poziomie ok. 1% liczby adresów w operacie; ze względów organizacyjnych ograniczono liczebność próby do 100 tysięcy adresów (w rezultacie uzyskano frakcję próby na poziomie 1,2%).
6. Jako główne kryterium warstwowania przyjęto kanał pozyskania danych w spisie zasadniczym.
7. Alokacji zaplanowanej liczby jednostek dokonano wstępnie pomiędzy kanałami, przy zastosowaniu reguły alokacji proporcjonalnej do pierwiastka kwadratowego z liczby mieszkań. W ramach ustalonych kanałów zdefiniowano podział na głębsze warstwy w każdym powiecie, z uwzględnieniem części miejskich i wiejskich, podziału dużych miast na dzielnice oraz z wykorzystaniem dostępnej w operacie informacji o liczbie osób w mieszkaniu. Ostateczny rozdział próby pomiędzy głębsze poziomy warstw w danym kanale wykonano za pomocą metody proporcjonalnej.
8. Na podstawie ustalonej alokacji pomiędzy warstwy, wykonano losowanie, przy zastosowaniu algorytmu losowania prostego bez zwracania, niezależnie w każdej z warstw. Obliczenia zrealizowano w systemie SAS w środowisku OBM.

4.2. Analiza wyników badania kontrolnego

4.2. Analysis of the results of the control survey

Spis kontrolny został zrealizowany dla 76073 mieszkań, co stanowiło ok. 76% wylosowanej próby mieszkań do tego badania. Głównymi przyczynami nie zrealizowania badania były braki kontaktu oraz odmowy. W celu poprawnego uogólniania danych zastosowano korektę wag z losowania, oddzielnie w grupach zdefiniowanych jako przekroje powiatów i kanałów kontaktu, uwzględniając podział na części miejskie i wiejskie (czyli użyto grupowań wykorzystanych przy warstwowaniu próby). Skorygowane wagi były używane w dalszych obliczeniach.

Do opracowania wyników badania kontrolnego wykorzystano klasyczną teorię metody reprezentacyjnej, traktując wylosowaną próbę jako przykład tzw. badania powtarzalnego, gdzie badane są te same cechy w dwóch okresach.

W przypadku badania kontrolnego po NSP 2021 za pierwszy okres przyjęto moment spisu zasadniczego, a za drugi – moment przeprowadzenia badania kontrolnego. Przyjęto, że wielkość próby reprezentacyjnej jest taka sama w obu okresach. Jak pokazuje teoria (patrz np. Bracha (1996)) badanie w dwóch okresach tej samej próby pozwala w sposób optymalny (z najmniejszym błędem losowym) oceniać różnice średnich dla obu okresów – a to jest jeden z głównych elementów określających jakość danych zebranych w spisie (tzn. brak systematycznego obciążenia wyników). Zastosowanie przy losowaniu próby schematu losowania warstwowego może dodatkowo zwiększyć precyzję estymacji, czyli z większym prawdopodobieństwem wykryć potencjalne błędy pomiarów badanych cech. Ponieważ z różnych powodów badanie reprezentacyjne nie zawsze udaje się zrealizować dla wszystkich wylosowanych jednostek (np. z powodu odmów), konieczna jest odpowiednia korekta wag z losowania, co teoretycznie wprowadza do estymacji opartej na badaniu reprezentacyjnym dodatkowy błąd nielosowy (który trudno ocenić bez specjalnych dodatkowych i kosztownych badań). Dlatego w dalszym wnioskowaniu założono, że tego typu błąd można pominąć, co ma swoje uzasadnienie w tym, że poziom braków odpowiedzi był stosunkowo niski (na tle innych typowych badań reprezentacyjnych realizowanych przez GUS) oraz rozsądne było przyjęcie założenia o „losowym” charakterze braków – w przyjętych do korekty wag z losowania grupach warstw.

Efektom realizacji badania kontrolnego była ocena jakości danych przy wykorzystaniu następujących miar:

1. Indeks zgodności (**IZ**) uwzględniający łącznie korelację wyników obu badań oraz wariancje wyników każdego z badań;

Miara ta obliczana jest według wzoru:

$$IZ = 1 - (S^2(X - Y)) / (S^2(X) + S^2(Y)),$$

gdzie X oznacza wartości danej cechy z badania zasadniczego, Y – wartości cechy z badania kontrolnego, a $S^2(X)$ – wariancję z próby dla danej cechy X . Wariancja dla n -elementowego ciągu wartości x_1, x_2, \dots, x_n (czyli zebranych w badaniu pomiarów dla cechy X) wyznaczana jest według wzoru: $S^2(x) = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i (x_i - \bar{x})^2$, gdzie $\bar{x} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i$. Użyte we wzorach wielkości oznaczają wagi do uogólnień wyznaczone zgodnie z opisaną wcześniej metodą, jako odpowiednio skorygowane wagi z losowania.

2. Błąd względny (**BWZ**) średnich z pomiarów każdej z cech;

Miara zdefiniowana jako:

$$BWZ = 100\% * (\bar{X} - \bar{Y}) / \bar{X},$$

gdzie \bar{X} oznacza wartość średnią z próby dla danej cechy X z badania zasadniczego, \bar{Y} – wartość średnią z próby dla odpowiedniej cechy Y z badania kontrolnego; $\bar{X} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i$, $\bar{Y} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i y_i$.

3. Współczynnik korelacji zgodności Lina (**CCC**, w terminologii angielskiej tzw. *concordance correlation coefficient*);

Jest to miara zgodności, będąca modyfikacją współczynnika korelacji liniowej Pearsona, która może być wykorzystana do oceny stopnia zgodności dla zmiennych ciągłych z badania zasadniczego i badania kontrolnego. Współczynnik **CCC** definiowany jest za pomocą wzoru:

$$CCC = \frac{2Cov(X, Y)}{S^2(X) + S^2(Y) + (\bar{X} - \bar{Y})^2},$$

gdzie $Cov(X, Y)$ oznacza kowariancję z próby dla cech X i Y obliczaną według wzoru:

$$Cov(X, Y) = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i (x_i - \bar{x})(y_i - \bar{y}), \text{ gdzie } \bar{x} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i, \bar{y} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i y_i.$$

Indeks zgodności (**IZ**) jest uogólnieniem znanego powszechnie w statystyce współczynnika korelacji, ale mierzy nie tylko zależność dwóch zmiennych, ale wychwytuje dodatkowo różnice w odchyleniach standardowych charakteryzujących zmienność poszczególnych rozkładów jednowymiarowych. Wskaźnik **IZ** nie mierzy natomiast poziomu przesunięcia, czyli obciążenia wyników, stąd jako uzupełnienie analizy oblicza się proste statystyki w postaci błędów względnych (**BWZ**); takie oceny razem z wykresami rozkładów różnic pozwalają na wykrycie ewentualnych systematycznych obciążeń.

Współczynnik **CCC**, podobnie jak miara **IZ**, jest uogólnieniem klasycznego współczynnika korelacji liniowej, przy czym w odróżnieniu od **IZ** uwzględnia dodatkowy czynnik związany z możliwym obciążeniem wynikającym z różnicy wartości przeciętnych analizowanych cech. Obie rozważane miary zachowują podstawową własność klasycznego współczynnika korelacji liniowej tzn. ich wartości mogą należeć do przedziału od do. Współczynnik **CCC** można również wykorzystać w przypadku zmiennych kategoryalnych mierzonych na skali porządkowej (w szczególności dla cech binarnych), gdyż w takich przypadkach miara ta pokrywa się z popularnym ważonym współczynnikiem **kappa Cohena**. Współczynnik **kappa** oraz ważony współczynnik **kappa** są często wykorzystywane do pomiaru stopnia zgodności dwóch pomiarów w przypadku zmiennych dyskretnych.

Szczegółowe informacje na temat miary **CCC** (jak również miary **kappa Cohena**) można znaleźć w artykule (Lin i inni, Survey Methodology, 2012).

Proponowane miary jakości: **IZ** oraz **BWZ** były wcześniej zastosowane do analizy spisu kontrolnego po Spisie Rolnym 1996, po Mikrospisie 1995, po NSP 2002 i PSR 2002 oraz po NSP 2011 i PSR 2020. Aspekty teoretyczne użytego indeksu zgodności opisano w artykule: Szablowski P.J., Wesołowski J., Wieczorkowski R., Indeks zgodności jako miara jakości danych (na podstawie wyników spisu kontrolnego do Mikrospisu 1995), Wiadomości Statystyczne, nr 4, 1996, s. 43–49.

4.2.1. Przyjęte grupowania

4.2.1. Adopted groupings

Indeksy zgodności wyznaczano na poziomie kraju (oznaczenie grupowania jako **PL**) oraz w podziale na: rodzaj kanału przeprowadzenia spisu zasadniczego w wylosowanym mieszkaniu (cecha oznaczana jako **kanal**), podział na miasto/wieś (**mw**), województwa (**woj**), sposób udzielania odpowiedzi na pytania spisu zasadniczego (cecha **PS_4** według zestawu tzw. pytań sondujących) oraz według grup wieku (dla pierwszego respondenta).

Uzyskane wyniki wskazują na występowanie zróżnicowania wartości indeksów zgodności ze względu na niektóre zastosowane grupowania, co zostanie dokładniej omówione w dalszej części podrozdziału.

Definicja kanałów przeprowadzenia spisu zasadniczego wynika z kodowania przyjętego na etapie losowania próby. Poniżej podano rozkład 4 wariantów kanałów na podstawie informacji z operatu losowania (gdzie wykorzystano cechę *spis_realizacja*). Jak wynika z tablicy częstości, rodzaj kanału był związany z metodą realizacji ankiety tzn. spis przez internet (metoda CAWI), przez telefon (CATI) lub tradycyjnie (CAPI).

Tablica 15. Kody kanałów realizacji spisu zasadniczego

Table 15. Modes used for collecting data in census

Kod kanału w badaniu kontrolnym Modes in control survey	Kod kanału w spisie zasadniczym Modes in census	Liczba przypadków Number of cases	Frakcja (w %) Fraction (in %)
1	CAWI	4846642	56,21
2	CAPI-PHONE	1847681	21,43
3	CAPI-PERSONAL	660731	7,66
4	CATI	35314	0,41
4	CATI2	1232334	14,29

W przypadku podziału na części miejską i wiejską stosowano umowne kodowanie tzn. wartość 1 dla miast oraz wartość 2 dla wsi.

W ramach grupowania według województw stosowano następujące oznaczenia:

Tablica 16. Słownik dla kodów województw

Table 16. Dictionary for voivodship codes

Kod województwa Voivodship code	Opis Description
02	dolnośląskie
04	kujawsko-pomorskie
06	lubelskie
08	lubuskie
10	łódzkie
12	małopolskie
14	mazowieckie
16	opolskie
18	podkarpackie
20	podlaskie
22	pomorskie
24	śląskie
26	świętokrzyskie
28	warmińsko-mazurskie
30	wielkopolskie
32	zachodniopomorskie

Słownik dla grupowania według cechy **PS_4** był następujący:

Tablica 17. Słownik dla cechy PS_4

Table 17. Dictionary for the variable PS-4

PS_4	Opis Description
1	Odpowiedzi na pytania spisowe udzielane osobiście
2	Odpowiedzi przy współudziale innych osób
3	Odpowiedzi udzielał ktoś inny

Informację o wieku dla tzw. pierwszego respondenta, dostępną w danych dla spisu zasadniczego, wykorzystano do konstrukcji dodatkowego grupowania (oznaczanego symbolem **GW**) tzn. 5 grup wieku:

1. poniżej 35 lat
2. od 35 do 44 lat
3. od 45 do 54 lat
4. od 55 do 64 lat
5. wiek 65 i więcej lat.

4.2.2. Zmienne wybrane do analizy

4.2.2. Variables selected for analysis

W ramach badania kontrolnego zadawano pytania dotyczące mieszkań oraz osób. Każde z tych pytań występowało również w spisie zasadniczym, co dało możliwość porównywania dwóch wartości dla każdej z badanych cech. Na podstawie ankiety badania kontrolnego wybrano następujący zestaw cech dotyczących ludności, dokonując jednocześnie odpowiedniego przekodowania wartości kodów liczbowych.

Tablica 18. Lista analizowanych zmiennych dotyczących osób
Table 18. List of analyzed variables concerning persons

Zmienna Variable	Opis Description	Uwagi Remarks
L05	Czy w dniu 31 marca 2021 r. o godz. 24:00 pod tym adresem mieszkały także inne osoby?	Zmienna o wartości 1 (odpowiedź TAK) lub 0 (odpowiedź NIE)
L05b	Proszę podać ile osób mieszkało pod tym adresem w dniu 31 marca 2021 r. o godz. 24:00, nie licząc Pana/Pani	Wartość 0 lub liczba całkowita większa od 0
L08	Czy adres, który Pan/Pani wskazał(a) jako adres zamieszkania w dniu 31 marca 2021 r. był Pana(i) stałym czy czasowym miejscem zamieszkania?	Zmienna o wartości 1 (odpowiedź TAK tzn. adres stały) lub 0 (odpowiedź NIE)
L14	Jaki jest Pana/Pani najwyższy poziom wykształcenia?	Wartości od 1 do 11, kodujące poziom wykształcenia
L15_001	Jaka jest Pana/Pani narodowość?	Zmienna o wartości 1 w przypadku zaznaczenia narodowości polskiej, 0 – w przeciwnym przypadku
L15_236	Jaka jest Pana/Pani narodowość?	Zmienna o wartości 1 w przypadku zaznaczenia narodowości śląskiej, 0 – w przeciwnym przypadku
L15_037	Jaka jest Pana/Pani narodowość?	Zmienna o wartości 1 w przypadku zaznaczenia narodowości białoruskiej, 0 – w przeciwnym przypadku
L15_257	Jaka jest Pana/Pani narodowość?	Zmienna o wartości 1 w przypadku zaznaczenia narodowości ukraińskiej, 0 – w przeciwnym przypadku
L16	Czy odczuwa Pan/Pani przynależność także do innego narodu lub wspólnoty etnicznej?	Zmienna o wartości 1 w przypadku odpowiedzi NIE, 0 – w przeciwnym przypadku

Zmienne: **L15_001**, **L15_236**, **L15_037** oraz **L15_257** zostały wyznaczone na podstawie zmiennej **L15**, która zawierała kody zaznaczonych przez respondentów narodowości. Do dalszych analiz wybrano warianty o największej liczbie wystąpień w spisie zasadniczym oraz badaniu kontrolnym, co wynika z zamieszczonych poniżej zestawień:

Tablica 19. Rozkład najliczniej reprezentowanych kategorii odpowiedzi na pytanie o narodowość w spisie zasadniczym (wśród respondentów badania kontrolnego)

Table 19. Distribution of the most frequently represented categories of answers to the question on nationality in the census (among respondents of the control survey)

Narodowość Nationality	Liczebność Number of cases	Frakcja (w %) Frequency (in %)	Frakcja skumulowana (w %) Cumulated frequency (in %)
Polska	75198	98,85	98,85
Śląska	432	0,57	99,42
Białoruska	98	0,13	99,55
Ukraińska	80	0,11	99,65

Tablica 20. Rozkład najliczniej reprezentowanych kategorii odpowiedzi na pytanie o narodowość w badaniu kontrolnym

Table 20. Distribution of the most frequently represented categories of answers to the question on the nationality in the control survey

Narodowość Nationality	Liczebność Number of cases	Frakcja (w %) Frequency (in %)	Frakcja skumulowana (w %) Cumulated frequency (in %)
Polska	75692	99,50	99,50
Śląska	69	0,09	99,59
Białoruska	65	0,09	99,68
Ukraińska	57	0,07	99,75

Tablica 21. Lista analizowanych zmiennych dotyczących mieszkań

Table 21. List of analyzed variables concerning dwellings

Zmienna Variable	Opis Description	Uwagi Remarks
M02_1	Wskazanie własności mieszkania – osoby/ osób fizycznych (dotyczy także współwłasności np. małżeńskiej)	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_2	Wskazanie własności mieszkania – spółdzielni mieszkaniowej (mieszkanie własnościowe)	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_3	Wskazanie własności mieszkania – spółdzielni mieszkaniowej (mieszkanie lokatorskie)	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku

Tablica 21. Lista analizowanych zmiennych dotyczących mieszkań (dok.)**Table 21. List of analyzed variables concerning dwellings (cont.)**

Zmienna Variable	Opis Description	Uwagi Remarks
M02_4	Wskazanie własności mieszkania – gminy	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_5	Wskazanie własności mieszkania – Skarbu Państwa	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_6	Wskazanie własności mieszkania – zakładu pracy	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_7	Wskazanie własności mieszkania – towarzystwa budownictwa społecznego	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M02_8	Wskazanie własności mieszkania – inne	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M03a	Mieszkanie zamieszkuje(ą) – przynajmniej jedna osoba będąca właścicielem całego lub części mieszkania	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M03b	Mieszkanie zamieszkuje(ą) – przynajmniej jedna osoba mająca spółdzielcze prawo (własnościowe lub lokatorskie) do mieszkania	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M03c	Mieszkanie zamieszkuje(ą) – osoba(y) z tytułu najmu całego mieszkania	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M0d	Mieszkanie zamieszkuje(ą) – przynajmniej jedna osoba z tytułu podnajmu całego lub części mieszkania (dotyczy także najmu części mieszkania)	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku
M0e	Mieszkanie zamieszkuje(ą) – osoba(y) na podstawie innego tytułu	Zmienna o wartości 1 w przypadku wybrania tej kategorii, 0 – w przeciwnym przypadku

Zmienne mieszkaniowe: od **M02_1** do **M02_8** zostały wyznaczone na podstawie cechy **M02**, która występowała w formularzach spisu oraz badania kontrolnego i zawierała kody odpowiednich 8 kategorii.

4.2.3. Miary zgodności

4.2.3. Measures of consistency

Przykładowe wartości wybranych miar zgodności obliczone dla całego kraju zostały zaprezentowane poniżej.

Tablica 22. Miary zgodności spisu zasadniczego i badania kontrolnego na poziomie Polski ogółem dla cech osobowych

Table 22. Measures of consistency in the census and control survey at the country level for personal characteristics

Zmienna Variable	IZ	CCC	BWZ
L05	0,877	0,877	0,1
L05b	0,878	0,887	-10,1
L08	0,457	0,456	-0,5
L14	0,965	0,965	-0,6
L15_001	0,243	0,242	-0,8
L15_236	0,182	0,181	85,6
L15_037	0,210	0,210	30,2
L15_257	0,583	0,583	30,9
L16	0,486	0,486	-0,5

Tablica 23. Miary zgodności spisu zasadniczego i badania kontrolnego na poziomie Polski ogółem dla cech dotyczących mieszkań

Table 23. Measures of consistency in the census and control survey at the country level for dwelling characteristics

Zmienna Variable	IZ	CCC	BWZ
M02_1	0,513	0,512	1,9
M02_2	0,407	0,406	-11,5
M02_3	0,417	0,417	-6,3
M02_4	0,744	0,744	-6,0
M02_5	0,299	0,299	-0,5
M02_6	0,491	0,491	-31,5
M02_7	0,716	0,716	20,7
M02_8	0,130	0,130	28,8
M03a	0,575	0,575	2,8

Tablica 23. Miary zgodności spisu zasadniczego i badania kontrolnego na poziomie Polski ogółem dla cech dotyczących mieszkań (dok.)**Table 23. Measures of consistency in the census and control survey at the country level for dwelling characteristics (cont.)**

Zmienna Variable	IZ	CCC	BWZ
M03b	0,419	0,419	-11,2
M03c	0,720	0,720	0,3
M03d	0,187	0,187	-12,3
M03e	0,283	0,283	-23,5

W celu bardziej szczegółowych analiz (z uwzględnieniem dodatkowych grupowań) zdecydowano o ograniczeniu zestawu zmiennych do 10 tzn. do listy cech: **L05, L05b, L08, L14, L15_001, L16, M02_1, M02_2, M03a, M03b**. Taki wybór wynika z tego, że pozostałe zmienne dotyczą bardzo małych części populacji, co pokazują dwie poniższe tablice, w których zamieszczono wyniki estymacji średnich na podstawie danych z realizacji badania kontrolnego (po połączeniu z danymi ze spisu zasadniczego). Oceny frakcji elementów z cechą wyróżnioną (czyli średnich dla zmiennych o wartościach 0 lub 1; a taki charakter ma większość analizowanych cech) w przypadku, gdy frakcje są na poziomie poniżej kilku procent, obciążone są bardzo dużymi względnymi błędami losowymi, co utrudnia wnioskowanie o występowaniu istotnych statystycznie różnic.

W przypadku dwóch zmiennych: **L05b** i **L14**, które w obliczeniach były traktowane jako cechy ciągłe, należy zwrócić uwagę, że podane w poniższej tabeli wartości liczbowe są wartościami średnimi po umownym przemnożeniu przez 100%; zatem na przykład dla cechy **L05b** w spisie zasadniczym i badaniu kontrolnym odpowiednio średnie liczby osób w mieszkaniu wynosiły 1,540 i 1,695.

Tablica 24. Wartości uogólnień na podstawie badania kontrolnego dla zmiennych dotyczących osób**Table 24. Estimates based on control survey for the variables concerning persons**

Zmienna Variable	Estymator średniej (w %) – spis zasadniczy Estimate of mean (in %) – census	Estymator średniej (w %) – badanie kontrolne Estimate of mean (in %) – control survey
L05	74,27	74,19
L05b	153,97	169,51
L08	96,17	96,69
L14	558,10	561,27
L15_001	97,52	98,01
L15_236	98,68	99,48
L15_037	0,69	0,10
L15_257	0,12	0,08
L16	0,12	0,08

Tablica 25. Wartości uogólnień na podstawie badania kontrolnego dla zmiennych dotyczących mieszkań
Table 25. Estimates based on control survey for the variables concerning dwellings

Zmienna Variable	Estymator średniej (w %) – spis zasadniczy Estimate of mean (in %) – census	Estymator średniej (w %) – badanie kontrolne Estimate of mean (in %) – control survey
M02_1	76,8	75,3
M02_2	14,8	16,5
M02_3	1,8	1,9
M02_4	3,5	3,8
M02_5	0,2	0,2
M02_6	0,3	0,3
M02_7	1,0	0,8
M02_8	1,6	1,1
M03a	69,3	67,4
M03b	15,2	16,9
M03c	9,4	9,4
M03d	1,4	1,5
M03e	6,5	8,1

Otrzymane wyniki świadczą ogólnie o dobrej zgodności obu przeprowadzonych badań dla większości analizowanych cech. Jednak dla kilku cech indeks zgodności jest zbyt niski, dotyczy to w szczególności zmiennych dotyczących narodowości. Wyjaśnić to można specyfiką konstrukcji miary zgodności oraz charakterystyką analizowanych cech, które mają w większości wartość 1 jako wariant odpowiedzi. Niektóre przypadki można również próbować wytłumaczyć złożonością definicji rozważanych cech, co mogło spowodować trudności w interpretacji pytań przez ankietowane osoby.

Należy podkreślić, że dla użytych wskaźników zgodności nie ma jednoznacznych teoretycznych zasad określających w jakich zakresach liczbowych wartości tych wskaźników są „poprawne”. Podobnie jak w przypadku klasycznego współczynnika korelacji liniowej istnieje pewna dowolność w określaniu progów od których uznajemy, że uzyskaliśmy większy lub mniejszy stopień „zgodności”; w literaturze przedmiotu spotyka się jedynie sugestie ekspertów w danej dziedzinie, które wynikają z doświadczeń analizy wielu przykładowych zbiorów danych. W celu ogólnej jakościowej oceny można przyjąć np. zasadę traktowania sytuacji, gdy uzyskano miary powyżej poziomu 0,8 jako sygnału bardzo dobrej zgodności wyników, natomiast przypadki z wartościami poniżej 0,2 można uznać za wskazanie problemów ze zgodnością.

Dla cech, dla których wskaźnik **IZ** przyjął wartości bliskie jedności, zaś wskaźnik **BWZ** pomiędzy -5% do 5%, sytuacja jest bardzo dobra tzn. wyniki spisu zasadniczego są najbardziej wiarygodne. Natomiast nie wydają się być wiarygodne te wyniki spisu zasadniczego, dla których zaobserwowano wartości współczynnika **IZ** znacznie różniące się od 1 lub współczynnika **BWZ** większe co do wartości bezwzględnej niż 5% – świadczy to o stosunkowo dużych błędach systematycznych bądź zależnej, nieobiektywnej pracy ankietatorów (bądź nieprawdziwych odpowiedziach respondentów na dane pytanie).

W interpretacji uzyskanych miar do wyników badania kontrolnego należy dodatkowo uwzględniać specyfikę badanych cech i np. porównywać uzyskane wyniki z dostępnymi wynikami z analogicznych badań realizowanych w latach poprzednich oraz analizować różnice wartości miar w ramach różnych grupowań.

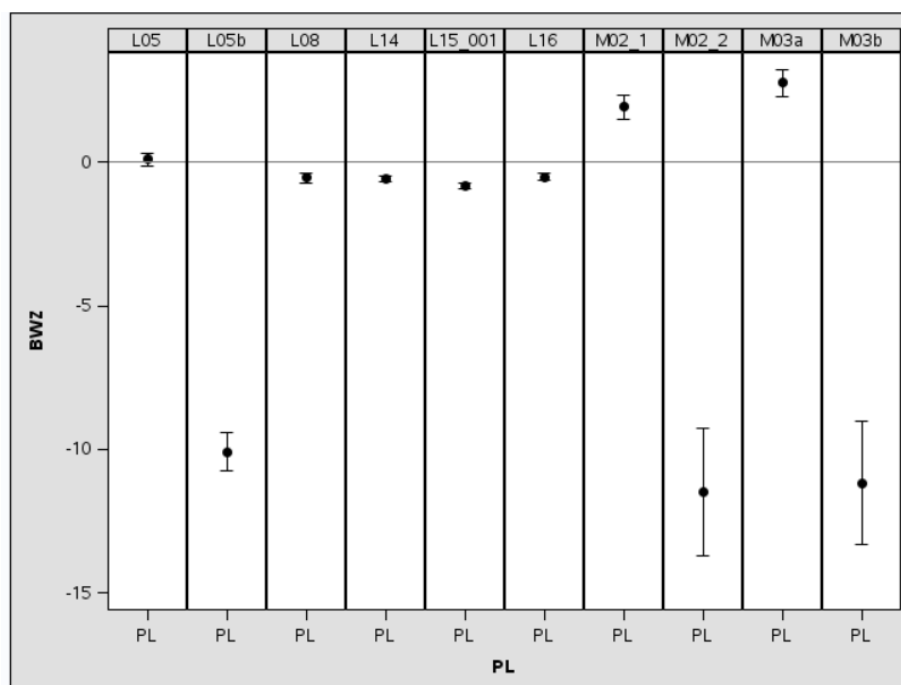
Dokładniejsza analiza wskaźnika zgodności wymagała dodatkowego oszacowania błędów losowych. W tym celu zastosowano wariant metody bootstrap opisany w pracy: Chipperfield J., Preston J. 2007. Rozważany w artykule algorytm można wykorzystać do efektywnej implementacji obliczeń w przypadku jednostopniowych schematów losowania, w przypadku małych frakcji losowania algorytm redukuje się do znanej metody powtarzalnych próbek zrównoważonych. Zamieszczone poniżej wykresy prezentują przedziały ufności (na poziomie 95%) wyznaczone na podstawie oszacowanych wskaźników oraz ich precyzji, dla całego kraju oraz dla ustalonych grupowań.

Prezentacja wyników estymacji punktowej razem z przedziałami ufności daje możliwość czytelnej oceny wartości wskaźników z uwzględnieniem ich błędów losowych oraz wnioskowania o istotnych statystycznie różnicach. Jeżeli na przykład przedział dla wskaźnika **BWZ** nie obejmuje wartości zerowej, to możemy wnioskować o występowaniu istotnego statystycznie przesunięcia dla średnich cech badanych w dwóch okresach. Ocena czy takie przesunięcie dla danej cechy ma znaczący wpływ na jakość uzyskanych ze spisu uogólnień powinna uwzględniać też skalę przesunięcia (wykresy pokazują przypadki, gdzie poziom różnicy jest minimalny, rzędu kilku procent).

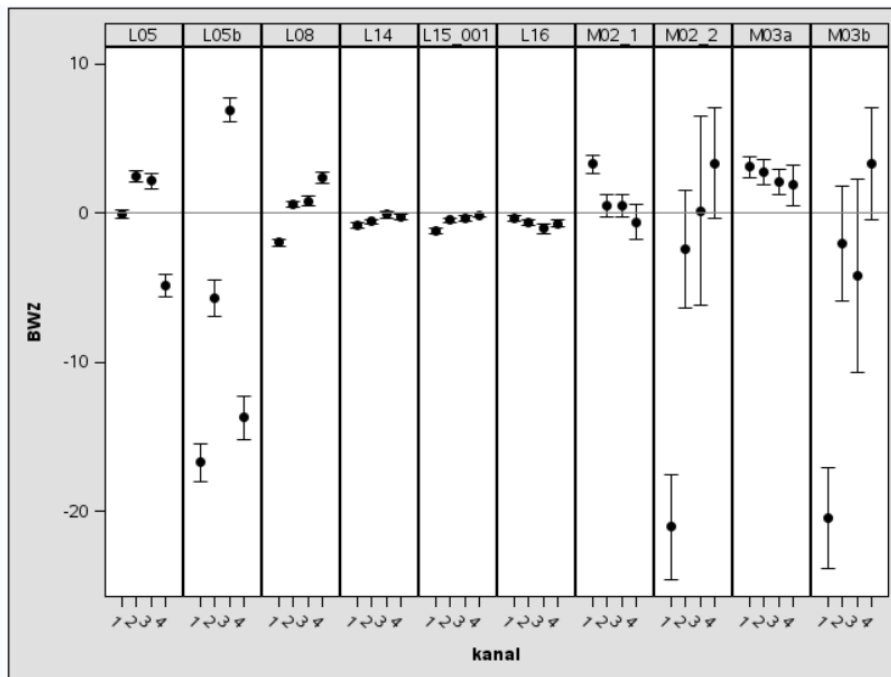
Wskaźniki BWZ

Wykres 3 Wskaźniki zgodności BWZ dla wybranych zmiennych

Chart 3. Measures of consistency BWZ for selected variables

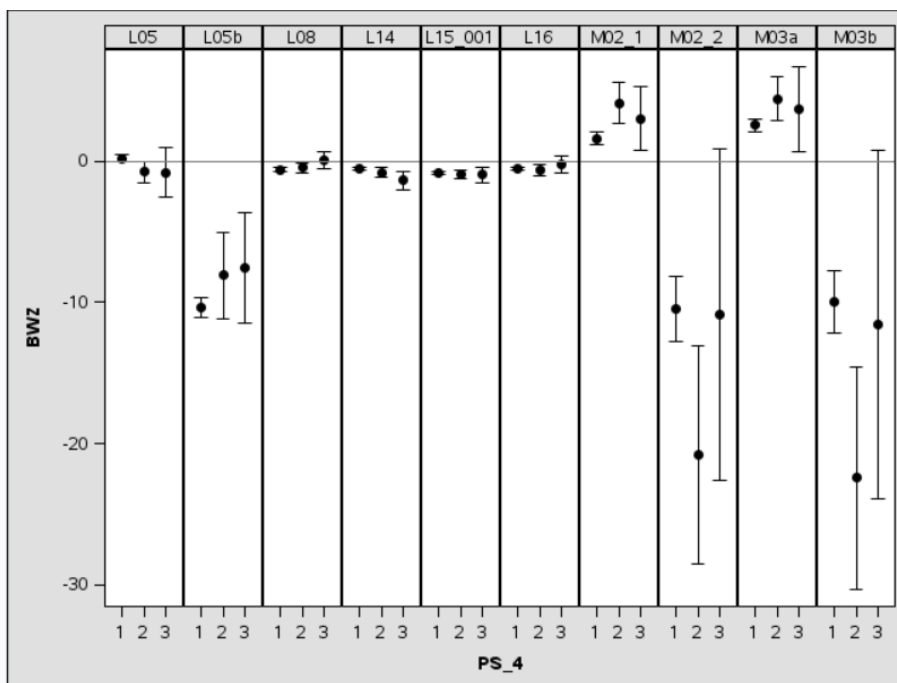


Wykres 4. Wskaźniki zgodności BWZ dla wybranych zmiennych według kanałów realizacji spisu zasadniczego
Chart 4. Measures of consistency BWZ for selected variables according to modes of contact in census

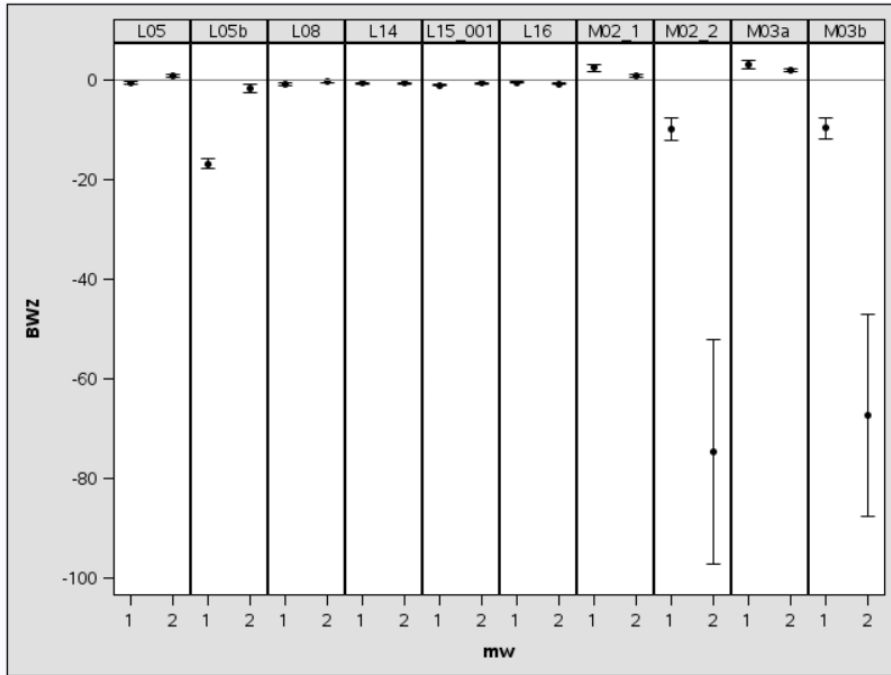


Wykres 5. Wskaźniki zgodności BWZ dla wybranych zmiennych według osoby odpowiadającej na pytania spisu zasadniczego

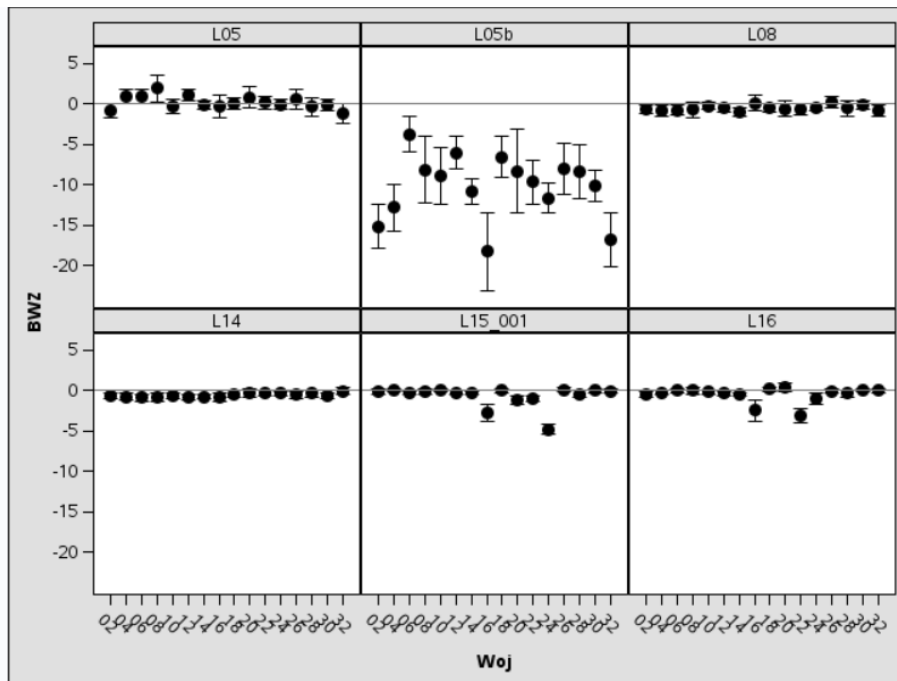
Chart 5. Measures of consistency BWZ for selected variables according to the person answering the census questions



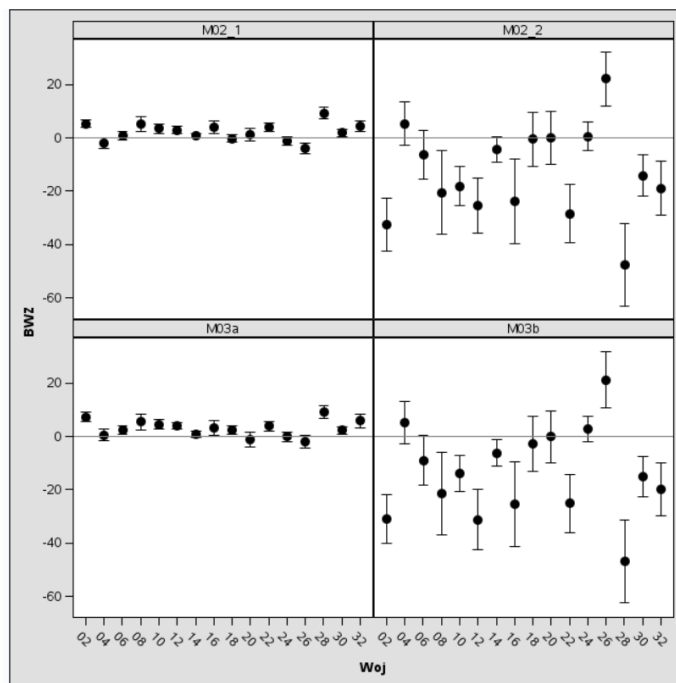
Wykres 6. Wskaźniki zgodności BWZ dla wybranych zmiennych według podziału na miasto/wieś
Chart 6. Measures of consistency BWZ for selected variables according to urban/rural division



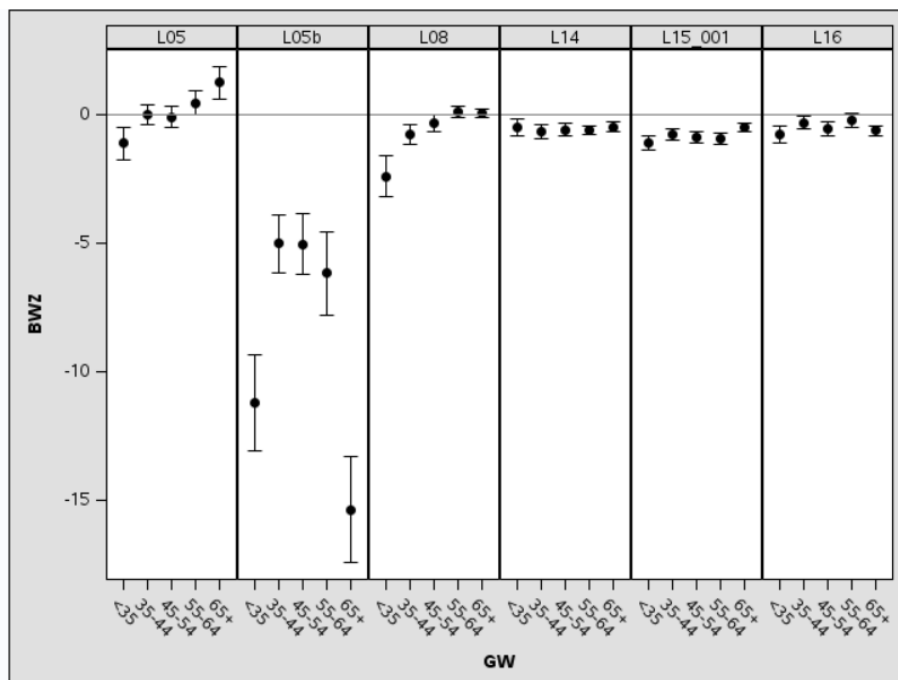
Wykres 7. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących osób według województw
Chart 7. Measures of consistency BWZ for selected personal variables by voivodships



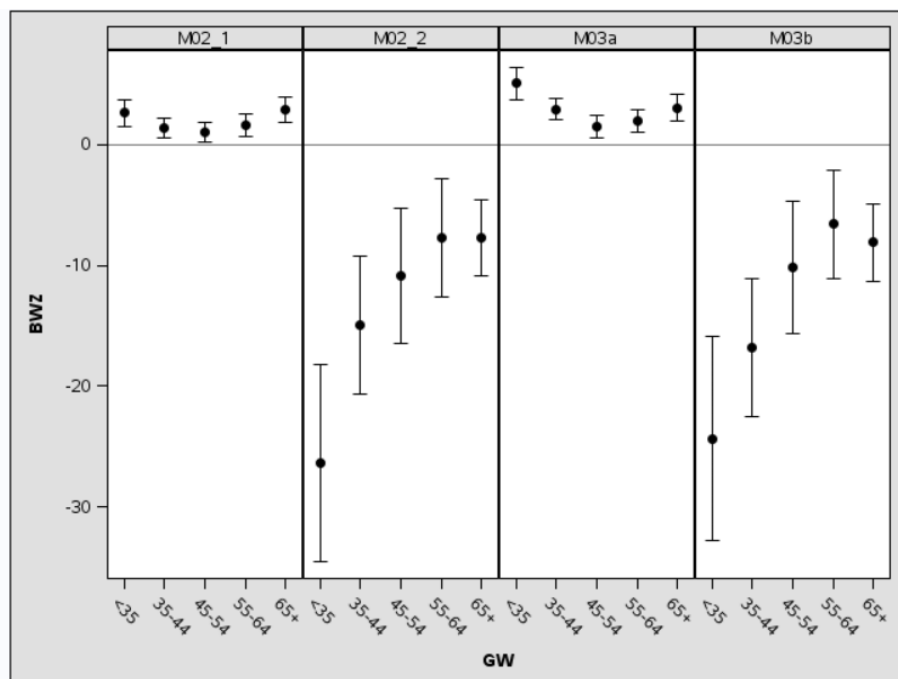
Wykres 8. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących mieszkań według województw
Chart 8. Measures of consistency BWZ for selected dwelling variables by voivodships



Wykres 9. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących osób według grup wieku
Chart 9. Measures of consistency BWZ for selected personal variables by age groups



Wykres 10. Wskaźniki zgodności BWZ dla wybranych zmiennych dotyczących mieszkań według grup wieku
Chart 10. Measures of consistency BWZ for selected dwelling variables by age groups



Zamieszczone wykresy ze wskaźnikami **BWZ** pozwalają na graficzną analizę i wnioskowanie o statystycznie istotnych przesunięciach wyników pomiędzy spisem zasadniczym i badaniem kontrolnym. Analiza błędów względnych (**BWZ**) wykazała pewne istotne różnice systematyczne dla niektórych badanych cech.

Zgodnie z klasycznym podejściem metody reprezentacyjnej, przy postulowanej zgodności wyników przedziały ufności powinny obejmować wartości zerowe (tzn. poprawna statystyczna interpretacja przedziału ufności mówi, że przy hipotetycznym wielokrotnym przeprowadzeniu wszystkich etapów opisanego badania reprezentacyjnego średnio w 95 na 100 przypadków zero będzie należeć do kolejno wyznaczanych przedziałów). Zatem w sytuacji, gdy w naszym (jednorazowym) badaniu uzyskujemy dla danego wskaźnika przedział nie zawierający zera, to jest to dla nas „sygnał” występowania istotnego statystycznie przesunięcia wyników.

Analizując kolejne wykresy można stwierdzić, że:

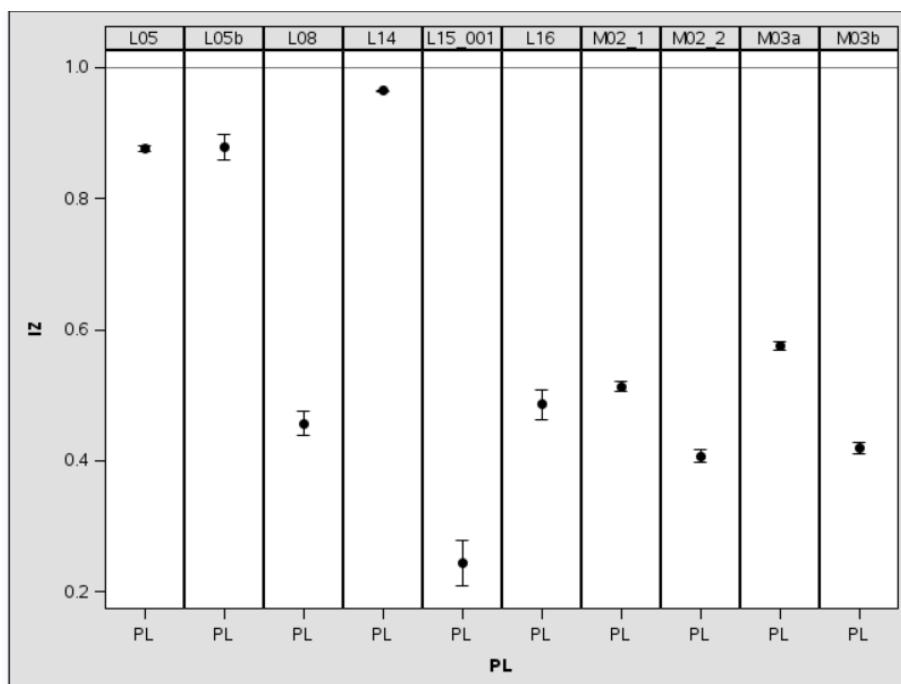
1. dla większości badanych cech nie można wnioskować o braku statystycznie istotnej różnicy, chociaż tylko dla trzech zmiennych (**L05b**, **M02_2**, **M03b**) miara **BWZ** na poziomie kraju jest poniżej -10% (przy czym znak wskazuje w tych przypadkach na zniżenie wartości uzyskanych po stronie spisu zasadniczego),
2. w zakresie badanych cech osobowych problematyczna wydaje się jedynie ocena dla zmiennej **L05b** (co można tłumaczyć występowaniem dla tej zmiennej bardzo dużego odsetka braków w zbiorze z danymi ze spisu zasadniczego),
3. wykresy z grupowaniem według kanałów kontaktu pokazują, że problem z dużym zniżeniem wyników w spisie zasadniczym dla zmiennych **M02_2** i **M03b** jest wyraźnie widoczny dla kanału realizacji przez samospis internetowy; podobny problem dla tych dwóch zmiennych mieszkaniowych występuje w przypadku, gdy odpowiedzi były udzielane przy współudziale innych osób,
4. grupowanie z podziałem na miasto/wieś wykazuje również, że w badaniu kontrolnym dla dwóch cech mieszkaniowych (**M02_2** oraz **M03b**) widoczne jest wyraźne zawyżenie uogólnień w badaniu kontrolnym (w stosunku do spisu zasadniczego),

- miary **BWZ** z podziałem według województw dla cech dotyczących osób dla większości zmiennych (poza problematyczną **L05b**) pokazują poziomy błędów skupione blisko wartości zerowej; natomiast dla cech dotyczących mieszkań ponownie potwierdza się problem z cechami **M02_2** oraz **M03b**. W szczególności bardzo duże ujemne wartości (rzędu -40%) wystąpiły dla województwa warmińsko-mazurskiego, natomiast dosyć duże wartości dodatnie (rzędu 20% dotyczyły województwa świętokrzyskiego),
- analiza wskaźnika w ramach grup wieku tzw. pierwszego respondenta (czyli osoby udzielającej odpowiedzi w spisie zasadniczym) pokazała dla większości rozważanych zmiennych brak istotnego statystycznie różnicowania; jednak dla zmiennej **L05** widoczne jest pewne minimalne przesunięcie pomiędzy najniższą i najwyższą grupą wieku. Natomiast dla zmiennej **L05b** w skrajnych grupach wieku ujawnia się wyraźne zwiększenie średnich wartości w stosunku do spisu zasadniczego (oceny **BWZ** niższe niż -10%), w przypadku „problematycznych” zmiennych mieszkaniowych **M02_2** i **M03b** obserwuje się szczególnie niskie ujemne wartości dla najmłodszej grupy wieku.

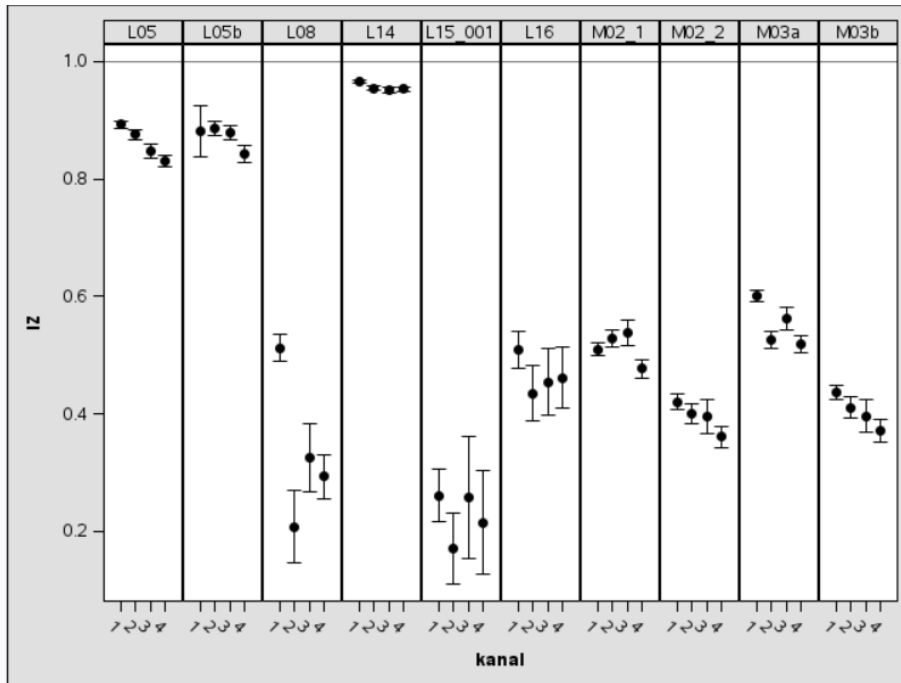
Wskaźniki IZ

Wykres 11. Wskaźniki zgodności IZ dla wybranych zmiennych

Chart 11. Measures of consistency IZ for selected variables

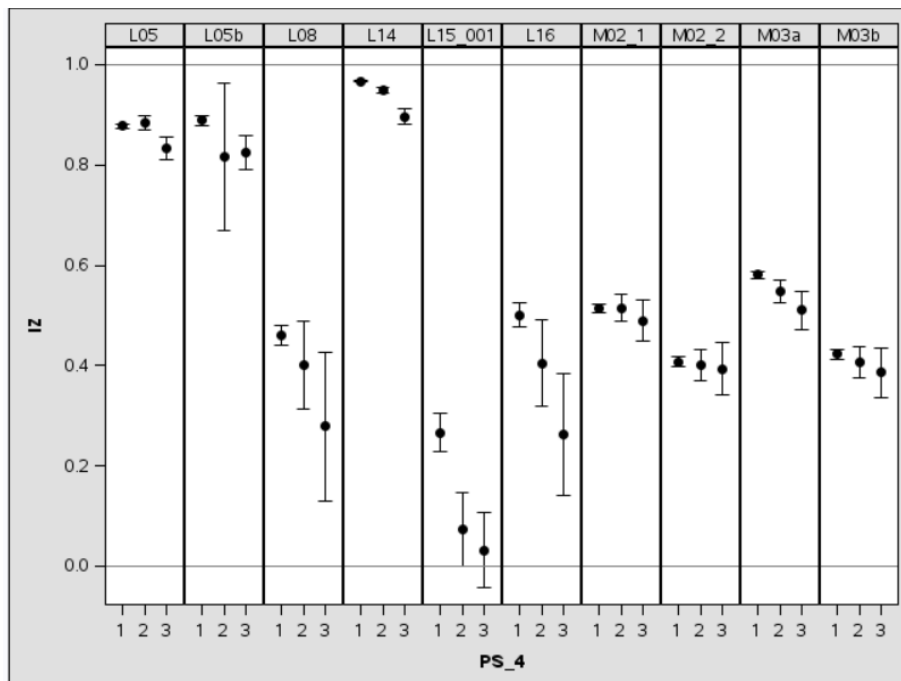


Wykres 12. Wskaźniki zgodności IZ dla wybranych zmiennych według kanałów realizacji spisu zasadniczego
Chart 12. Measures of consistency IZ for selected variables according to modes of contact in census

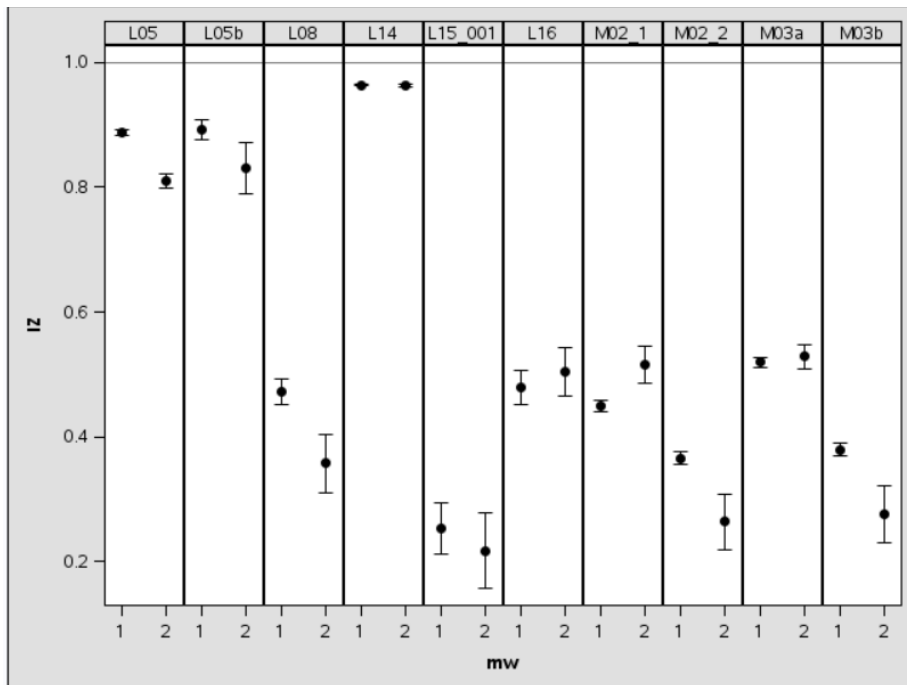


Wykres 13. Wskaźniki zgodności IZ dla wybranych zmiennych według osoby odpowiadającej na pytania spisu zasadniczego

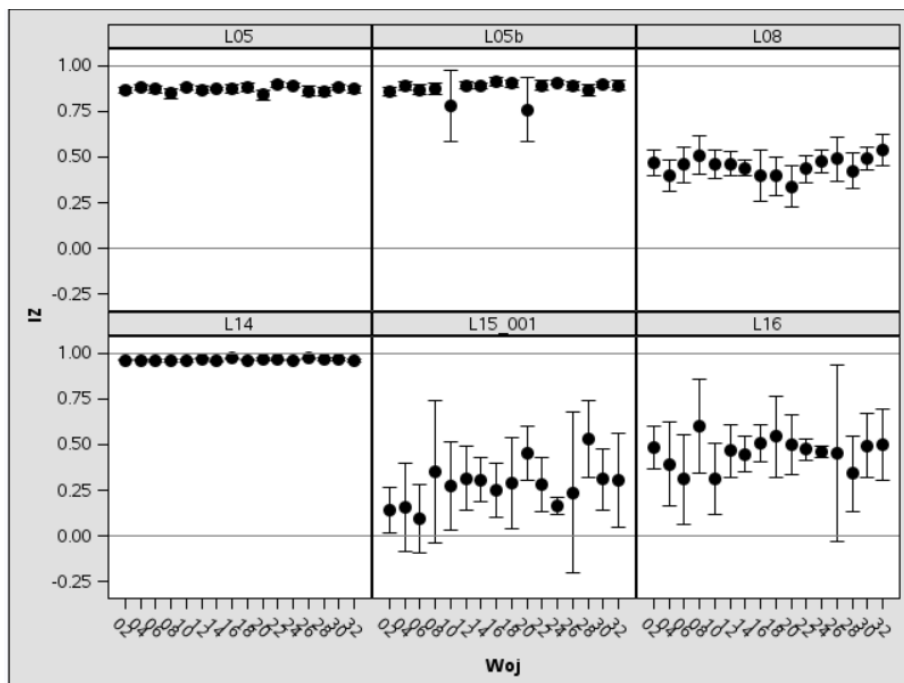
Chart 13. Measures of consistency IZ for selected variables according to the person answering the census questions



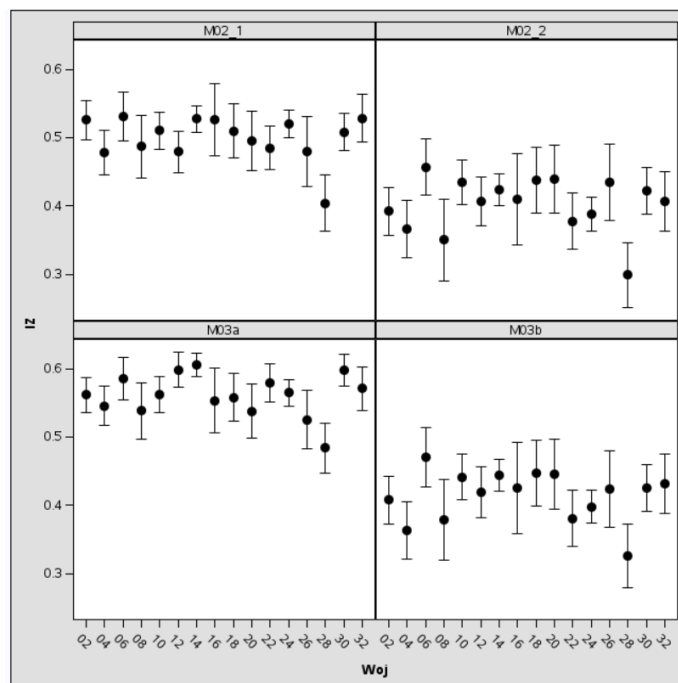
Wykres 14. Wskaźniki zgodności IZ dla wybranych zmiennych według podziału na miasto/wieś
Chart 14. Measures of consistency IZ for selected variables according to urban/rural division



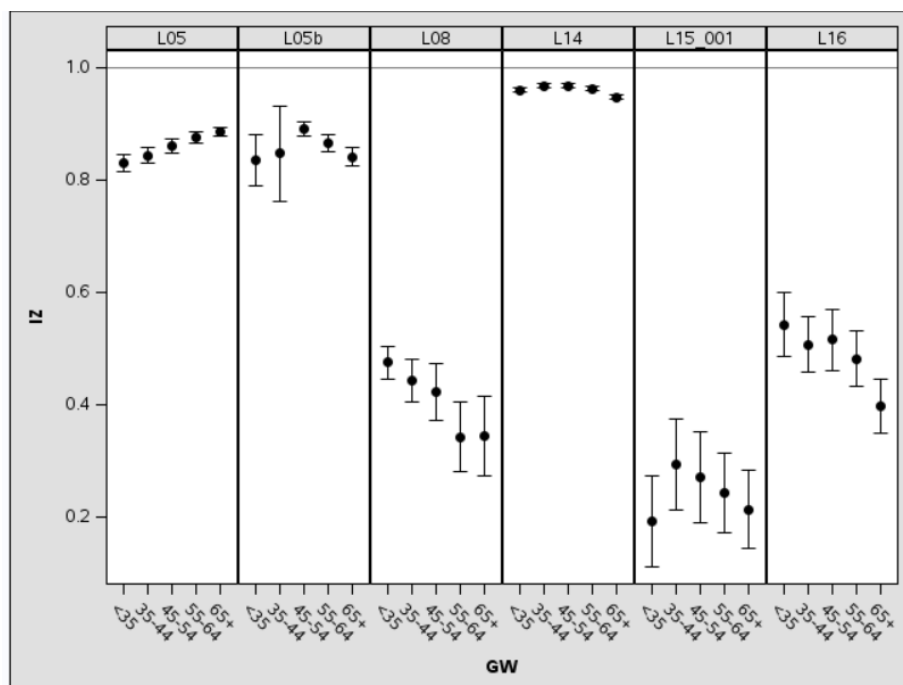
Wykres 15. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących osób według województw
Chart 15. Measures of consistency IZ for selected personal variables by voivodships



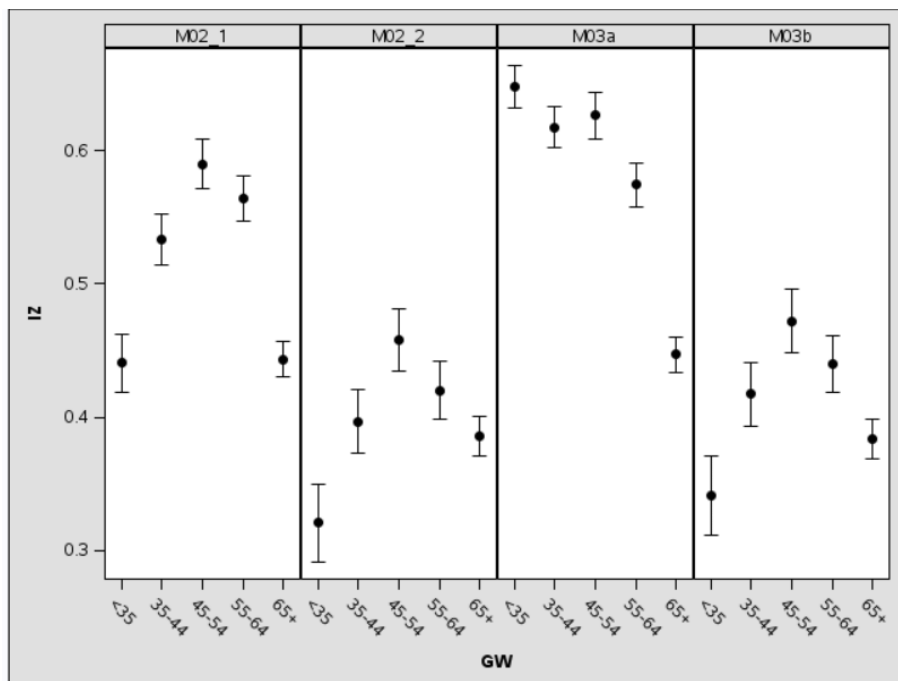
Wykres 16. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących mieszkań według województw
Chart 16. Measures of consistency IZ for selected dwelling variables by voivodships



Wykres 17. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących osób według grup wieku
Chart 17. Measures of consistency IZ for selected personal variables by age groups



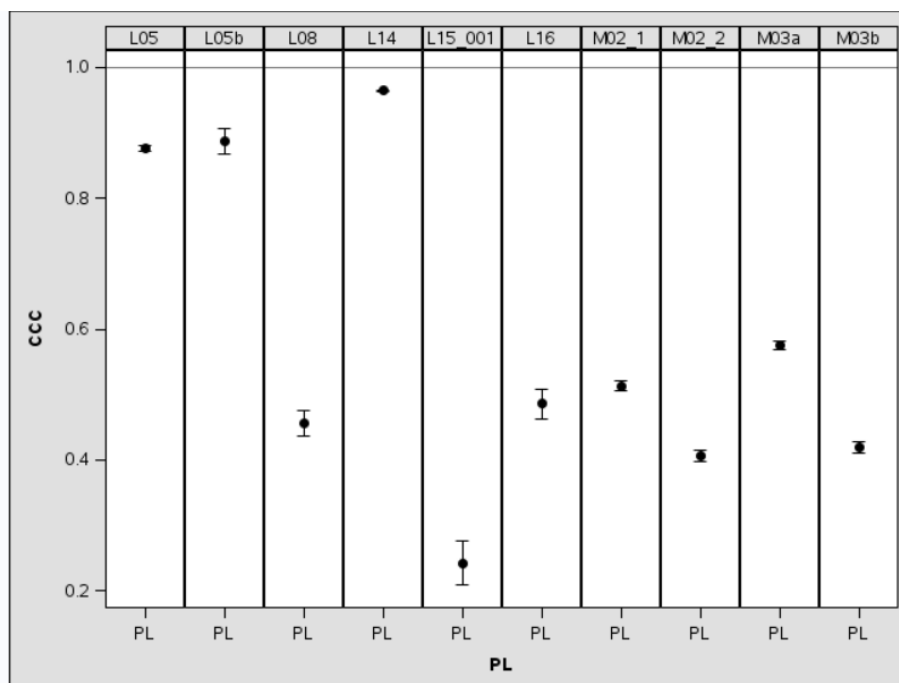
Wykres 18. Wskaźniki zgodności IZ dla wybranych zmiennych dotyczących mieszkań według grup wieku
Chart 18. Measures of consistency IZ for selected dwelling variables by age groups



Wskaźniki CCC

Wartości wskaźników **IZ** oraz **CCC** (a także ocen ich błędów losowych) są bardzo bliskie w sensie numerycznym. Można to tłumaczyć niewielkim wpływem różnicy wartości średnich analizowanych cech, która jest ujęta w definicji wskaźnika **CCC** (co jest konsekwencją tego, że większość analizowanych cech to zmienne binarne, o wartościach 0 lub 1), zatem w dalszym opisie skupiono się na wskaźniku **IZ**. Obliczenia wykazały, że do oceny możliwego błędu związanego z systematycznym obciążeniem uogólnień dla danej cechy przydatne jest wykorzystanie dodatkowych prostych miar typu **BWZ**.

Dla ilustracji ograniczono się do zamieszczenia wykresu tylko na poziomie kraju, gdyż z powodu bliskości numerycznej wykresy dla wskaźników **CCC** i **IZ** są praktycznie identyczne.

Wykres 19. Wskaźniki zgodności CCC dla wybranych zmiennych**Chart 19. Measures of consistency CCC for selected variables**

Problem braków danych dla cechy L05b

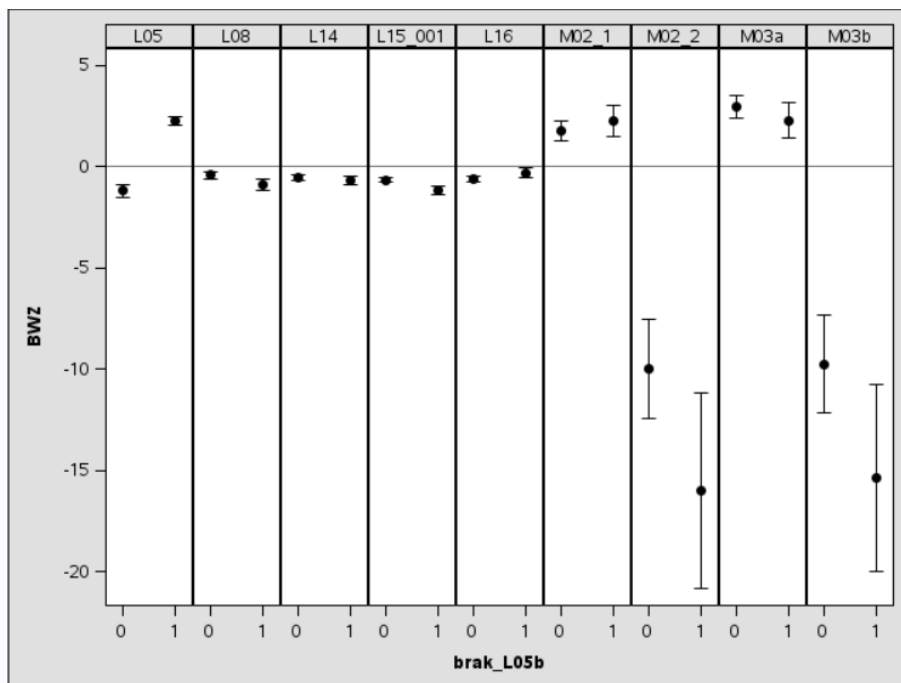
W przypadku zmiennej **L05b**, która dotyczyła istotnej dla rozpoznania badanej populacji liczby osób przebywających pod danym adresem, stwierdzono po połączeniu danych ze spisu zasadniczego i badania kontrolnego bardzo duży odsetek braków dla tej zmiennej w części ze spisu zasadniczego (około 24%). Zjawisko braków dotyczyło sytuacji, gdy w ankiecie podano dla zmiennej **L05** wartość 1 (czyli potwierdzono fakt przebywania również innych osób w mieszkaniu). Aby wyjaśnić czy taki błąd w danych wejściowych wpłynął istotnie na obliczane miary jakości, przeprowadzono obliczenia opisanych wcześniej wskaźników tworząc nową zmienną grupującą **brak_L05b**, o wartości 1 w przypadku występowania braku dla cechy **L05b** oraz z przypisaną wartością 0 w przeciwnym przypadku. Poniżej zamieszczono wykresy dla wybranych do analizy cech oraz specjalnie zdefiniowanego grupowania.

Na podstawie uzyskanych wyników można stwierdzić, że:

1. miary **BWZ**, **IZ** (a także **CCC**, która jak opisano wcześniej daje praktycznie podobne wartości numeryczne) dla większości zmiennych nie wykazują istotnego przesunięcia z powodu występowania braków danych dla cechy **L05b** (poza oczywistą różnicą dla zmiennej **L05** dla podzbioru z brakami w **L05b**),
2. w przypadku zmiennych mieszkaniowych: **M02_2** oraz **M03b** potwierdza się ogólny wniosek o wyraźnym заниzeniu wyników zebranych dla tych kategorii odpowiedzi w spisie zasadniczym w stosunku do badania kontrolnego,
3. występowanie braków danych w zmiennej **L05b** uniemożliwia jednoznaczne wnioskowanie o poziomie błędów dla tej ważnej cechy, jednak nie stanowi przeszkody w analizie ustalonych wskaźników dla pozostałych zmiennych.

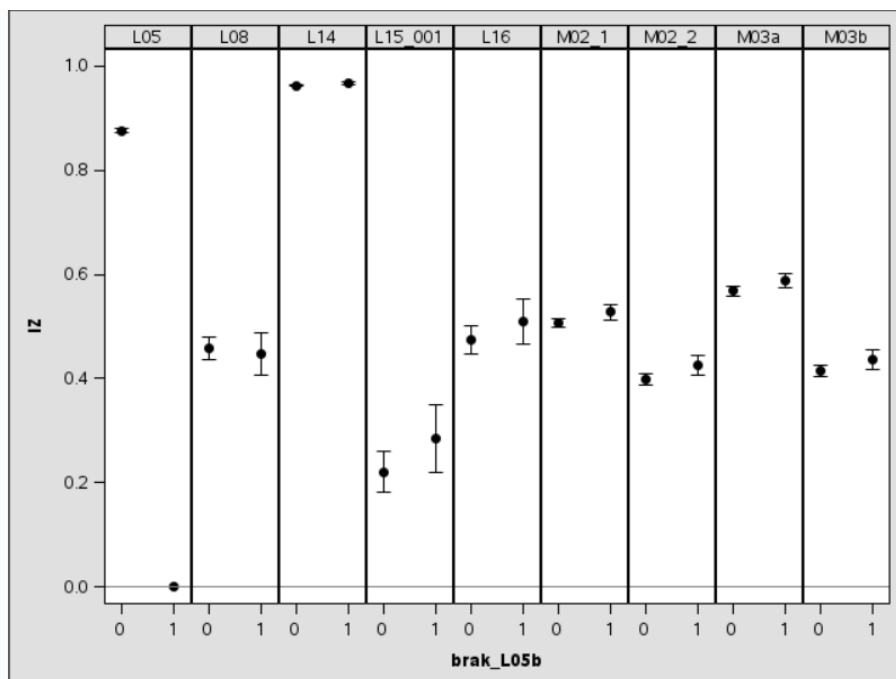
Wykres 20. Wskaźniki zgodności BWZ dla wybranych zmiennych według występowania braków dla cechy L05b (kod 1 na osi x odpowiada podzbiorowi danych z brakami dla cechy L05b, 0 – podzbiorowi bez braków), część 1

Chart 20. Measures of consistency BWZ for selected variables according to the occurrence of missing data for variable L05b (code 1 on the x axis corresponds to the subset with missing data, 0 – the subset without missing data), part 1



Wykres 21. Wskaźniki zgodności IZ dla wybranych zmiennych według występowania braków dla cechy L05b (kod 1 na osi x odpowiada podzbiorowi danych z brakami dla cechy L05b, 0 – podzbiorowi bez braków), część 2

Chart 21. Measures of consistency BWZ for selected variables according to the occurrence of missing data for variable L05b (code 1 on the x axis corresponds to the subset with missing data, 0 – the subset without missing data), part 2



4.2.4. Wnioski

Analiza uzyskanych miar oraz wykresów dla wskaźnika **IZ** prowadzi do następujących ogólnych wniosków (które dotyczą jednocześnie miary **CCC** ze względu na wykazaną empirycznie bliskość numeryczną dla tych dwóch miar):

1. Wartości wskaźników zgodności **IZ** dla analizowanych zmiennych (dla całego kraju) zawarte są w zakresie od 0,13 do 0,96, przy czym najwyższy poziom uzyskano dla cech: **L14**, **L05b**; najniższym poziomem charakteryzują się cechy wynikające z pytania o narodowość tzn. **L15_236**, **L15_037** oraz cechy mieszkaniowe **M02_8** i **M03d**.
2. Wskaźniki **IZ** dla kilku najważniejszych zmiennych liczba osób w mieszkaniu (L05b), poziom wykształcenia (L14), tytuł prawny posiadanego mieszkania (M02) kształtowały się na podobnych poziomach (rzędu 0,8, 0,9 i 0,5 odpowiednio) zarówno w zrealizowanym badaniu kontrolnym po spisie ludności i mieszkań 2011, jak i 2021, co wskazuje na podobną jakość zebranych danych dla tych cech w obu ostatnich spisach ludności i mieszkań.
3. Wskaźniki **IZ** dla cech z dużym poziomem zgodności (bliskim 1) charakteryzują się również mniejszymi wartościami błędów losowych.
4. Dla pewnych kategorii w ramach grupowania błędy są wyraźnie większe (np. kategoria nr 3 cechy PS_4 „Odpowiedzi udzielał ktoś inny” według rodzaju osoby odpowiadającej) – można to uzasadnić wpływem małej wielkości próby dla niektórych podgrup.
5. Na podstawie uzyskanych wyników nie można mówić o występowaniu istotnego zróżnicowania badanego wskaźnika **IZ** ze względu na grupowania według województw dla cech dotyczących osób; natomiast dla cech dotyczących mieszkań zaobserwowano wyraźne obniżenie miary **IZ** dla województwa warmińsko-mazurskiego.

6. W przypadku podziału na miasto/wieś obserwuje się dla większości zmiennych obniżenie stopnia zgodności dla wsi.
7. Dla grupowania według osoby udzielającej odpowiedzi na pytania w spisie zasadniczym widzimy pewną tendencję obniżania się poziomu zgodności w przypadku, gdy odpowiedzi na pytania spisowe nie były udzielane osobiście.
8. Dla grupowania według kanału realizacji spisu nie widać istotnych statystycznie różnic wartości miar zgodności **IZ**, chociaż same wartości wskaźnika są z reguły największe dla samospisu realizowanego przez Internet.
9. Wskaźniki **IZ** w podziale na grupy wieku dla cech dotyczących osób nie wykazują w większości przypadków istotnego zróżnicowania. Jednak dla cech **L08** i **L16** widać tendencję do obniżania się poziomu miary dla kolejnych grup wieku (najniższe wartości dla grupy wieku powyżej 65 lat). Dla cech mieszkaniowych zaobserwowano istotne zróżnicowanie wyników w postaci wzorca (z wyjątkiem cechy **M03a**): istotne wzrosty wskaźnika od najniższej grupy wieku do grupy 45–54 lata (dla której **IZ** osiąga najwyższe wartości), a następnie wyraźny spadek wartości dla kolejnych starszych grup. Z kolei dla zmiennej **M03a** istotnie odstająca i najniższa wartość wystąpiła dla najstarszej grupy wieku.

Podsumowując zebrany materiał empiryczny można sformułować następujące uwagi dotyczące oceny jakości wyników NSP 2021:

1. Przyjęte założenia organizacyjne i metodologiczne w przeprowadzonym badaniu kontrolnym nie dają podstaw do oceny wartości błędów pokrycia.
2. Dla większości kluczowych analizowanych cech uzyskano dobry poziom zgodności (czyli minimalny błąd pomiaru) na poziomie kraju oraz w ramach podstawowych grupowań.
3. Badanie kontrolne wykazało występowanie błędów treści (pomiaru) dla pewnych specyficznych analizowanych cech, w szczególności związanych z trudnością w interpretacji (lub też wrażliwym charakterem) pytań z formularza spisowego przez respondentów oraz związanych z wpływem na poziom tego błędu dodatkowych czynników (np. kanał kontaktu, osoba udzielająca odpowiedzi, podział terytorialny).
4. Uzyskany materiał daje podstawę do ewentualnych dodatkowych analiz w zakresie wskazanych problemów w obrębie zbiorów danych ze spisu zasadniczego dla specyficznych cech.
5. Analizując uzyskane wyniki należy zwrócić uwagę na możliwy dodatkowy czynnik zakłócający związany z efektem „pamięci respondenta”, który wynikał z długiego okresu trwania samego spisu zasadniczego (6 miesięcy) oraz dodatkowego okresu pomiędzy zakończeniem spisu zasadniczego i rozpoczęciem badania kontrolnego (ponad miesiąc).

4.2.5. Conclusions

Analysis of the obtained measures and graphs for the **IZ** indicator leads to the following general conclusions (which also apply to the **CCC** measure due to the empirically demonstrated numerical proximity for the two measures):

1. The values of **IZ** consistency measures for the analyzed variables (for the whole country) are contained in the range from 0.13 to 0.96, with the highest level obtained for the characteristics: **L14**, **L05b**; the lowest levels are for the characteristics resulting from the question on nationality, i.e. **L15_236**, **L15_037**, and the dwelling characteristics **M02_8** and **M03d**.
2. The **IZ** measures for several of the most important variables (e.g. *number of persons in the dwelling, education, legal title of the dwelling owned*) were at similar levels (of the order of 0.8, 0.9 and 0.5, respectively) in both the completed control survey of the 2011 and 2021 population and housing censuses, indicating a similar quality of the collected data for these characteristics in both recent population and housing censuses.
3. The **IZ** measures for characteristics with a high level of agreement (close to 1) are also characterized by smaller random errors.

4. For certain categories within the groups, the errors are clearly larger (e.g. category number 3 of characteristic **PS_4** „Answered by someone else” by type of respondent) – this can be justified by the impact of the small sample size for some subgroups.
5. On the basis of the obtained results, it is not possible to infer of the occurrence of significant differences of the examined **IZ** measure due to grouping by voivodship for the characteristics concerning persons; however, for the characteristics concerning dwellings, a clear reduction of the **IZ** measure for the Warmińsko-mazurskie voivodship was observed.
6. In the case of the urban/rural division, a reduction in the degree of consistency for rural areas is observed for most variables.
7. For grouping by the person answering the questions in the main census, we see a certain tendency of lowering the level of consistency when the census questions were not answered in person.
8. For grouping by census modes of contact, no statistically significant differences in **IZ** consistency measure can be seen, although the measure values themselves are generally highest for self-reports completed via the Internet.
9. **IZ** indicators by age group for person characteristics do not show significant variation in most cases; however, for characteristics **L08** and **L16**, a tendency to decrease the level of the measure for successive age groups (the lowest values for the age group over 65) can be seen; for dwelling characteristics, significant variation in the results in the form of a pattern was observed (except for characteristic **M03a**): significant increases in the indicator from the lowest age group to the 45–54 group (for which the **IZ** reaches the highest values), followed by a clear decrease in values for the next older groups; in turn, for the **M03a** variable, a significant outlier and the lowest value occurred for the oldest age group.

Summarizing the collected empirical material, the following comments can be made on the assessment of the quality of the 2021 Census results:

1. The organizational and methodological assumptions adopted in the control survey carried out do not provide a basis for assessing the value of coverage errors.
2. For most of the key characteristics analyzed, a good level of agreement (i.e. minimum measurement error) was obtained at country level and within the basic groupings.
3. The control survey showed the existence of content (measurement) errors for some specific analyzed characteristics, in particular related to the difficulty of interpretation (or sensitive nature) of the questions from the census form by respondents and related to the influence of additional factors on the level of this error (e.g. contact channel, kind of respondent, geographical distribution).
4. The material obtained provides a basis for possible additional analyses on the problems identified within the main census datasets for specific characteristics.
5. When analyzing the results obtained, attention should be paid to a possible additional distortion factor related to the „respondent memory” effect, which resulted from the long duration of the main census itself (6 months) and the additional period between the end of the main census and the start of the control survey (more than one month).

V. Ocena kompletności wyników NSP 2021 z wykorzystaniem metod analizy demograficznej

V. Results completeness assessment of the 2021 Census using demographic analysis methods

Jednym z głównych celów przeprowadzania narodowych spisów powszechnych jest dostarczenie pełnej informacji o liczbie i strukturach demograficzno-społecznych i ekonomicznych ludności z uwzględnieniem podziału terytorialnego. Spis ludności jest jedynym takim badaniem – ze względu na swój charakter, zakres podmiotowy oraz zasięg – które obejmuje także populację nieuwzględnioną w bieżącej rejestracji (urodzenia, zgonu, migracje).

Każdorazowe przeprowadzenie nowego spisu powszechnego na ogół wiąże się z wdrażaniem lepszych podejść metodologicznych i zastosowaniem nowocześniejszej technologii niż w poprzednim, a co za tym idzie zebraniem danych lepszej jakości. Umożliwia to nie tylko korektę aktualnego stanu ludności, będącej pochodną poprzedniego spisu i prowadzonych w okresach międzyspisywanych bilansów ludności, ale również pozwala na zbadanie, na ile dane opracowane na podstawie bieżącej rejestracji ruchu naturalnego i wędrownego ludności (urodzenia, zgonu i migracje) w latach międzyspisywanych są zbieżne z uzyskanymi wynikami spisu.

Jak już wielokrotnie podkreślano, spis powszechny, jak każde inne badanie, obarczony jest pewnymi błędami. Ze względu na to, że badanie spisowe pełni tak kluczową rolę w budowaniu obrazu sytuacji demograficznej i społeczno-ekonomicznej kraju, niezwykle istotne jest przeprowadzenie kompleksowej analizy mającej na celu ocenę jego jakości.

W 1985 r. Urząd ds. Spisów Powszechnych Stanów Zjednoczonych opublikował dokument opisujący metody i techniki służące ocenie powszechnych spisów ludności i mieszkań (U.S. Bureau of the Census, 1985). Jak wspomniano już w rozdziale III w zamierzeniu dokument ten miał być podstawą przy projektowaniu oraz wdrażaniu programów ewaluacji spisów powszechnych. Skupiono się w nim na dwóch głównych grupach metod: związanych z gospisowym badaniem kontrolnym oraz wykorzystujących metody analizy demograficznej (w szczególności oparte na analizie dwóch lub więcej spisów powszechnych). Również Organizacja Narodów Zjednoczonych (2017) zaleca, aby do oceny jakości spisów stosować metody analizy demograficznej. W niniejszym rozdziale skupiono się na pierwszej grupie metod. Ich podstawą jest porównywanie struktur ludności według płci i wieku, a także wskaźników demograficznych obliczonych na bazie różnych źródeł danych wśród których najważniejszymi są:

- wyniki ocenianego spisu,
- wyniki poprzedniego spisu,
- dane opracowywane w okresie międzyspisywonym, w oparciu o źródła administracyjne (dotyczy zdarzeń demograficznych: urodzenia, zgonu, migracje wewnętrzne, zagraniczne).

Zastosowanie metody polega na przeliczeniu liczby ludności dla lat międzyspisywanych bazując na wynikach nowego, najbardziej aktualnego spisu i porównanie jej z liczbą ludności corocznie opracowywaną na bazie poprzedniego spisu. Ludność taka obliczana jest metodą bilansową – uwzględniającą corocznie zarejestrowane zdarzenia demograficzne (urodzenia, zgonu, migracje) oraz proces starzenia się osób. Warto podkreślić, że od 2010 r. statystyka publiczna opracowuje co roku dwa bilanse ludności: według krajowej oraz międzynarodowej definicji. Różnica między definicjami dotyczy kryteriów w zakresie migracji czasowych uwzględnianych w każdym z bilansów (GUS, 2018). Wszystkie przedstawione wyniki analiz demograficznych w niniejszym rozdziale zostały opracowane dla ludności według definicji krajowej.

Uwzględnienie pojedynczych roczników wieku we wspomnianych analizach struktur demograficznych według wieku pozwala na pogłębione badania w zakresie spójności danych opracowywanych w oparciu o źródła administracyjne (badania statystyczne z zakresu urodzeń, zgonów i migracji) z wynikami ocenianego spisu. Dotyczy to zwłaszcza niespójności w liczbie osób w wieku 0–9 lat, urodzonych już po poprzednim spisie, która jest efektem

bilansowania jedynie danych, dla których głównym źródłem są dane administracyjne. Zestawienie tych liczebności z wynikami obecnego spisu może pomóc odpowiedzieć na pytanie, na ile skuteczne i dokładne było zbieranie danych na temat bieżących zdarzeń demograficznych w zakresie urodzeń, zgonów oraz migracji dzieci.

Dodatkowo, uwzględnienie parametru wieku pozwala na bardziej precyzyjną ocenę wpływu najtrudniejszego komponentu demograficznego jakim są migracje na rozbieżności w strukturach ludności (Paradysz 2010, Gołata 2018). W związku z tym należy wyodrębnić grupy o podwyższonej mobilności z uwagi na wiek, a następnie przeprowadzić osobną, pogłębioną analizę uzyskanych dla nich struktur ludności w oparciu o dane obecnego i poprzedniego spisu oraz dane administracyjne z okresu międzyspisowego. Ponadto, równie ważnym komponentem oceny jakości spisu jest uwzględnienie w analizie osób najstarszych. Jest to grupa szczególna, między innymi z uwagi na fakt, że charakteryzuje się niewielkimi migracjami. Zakładając, że zgony są dobrze rejestrowane to wszelkie rozbieżności dla tej grupy w dużym stopniu są spowodowane różnicami w wynikach poprzedniego i obecnego spisu.

5.1. Metodologia

5.1. Methodology

Dla potrzeb wykonanych analiz posłużono się następującymi pojęciami:

Bilans ludności – metoda szacowania liczby i struktury ludności według płci i wieku w okresach międzyspisowych. Metoda bilansowa opracowywania danych o liczbie i strukturze ludności zakłada przyjęcie wyników poprzedniego spisu powszechnego ludności (NSP 2011) za bazę wyjściową, a następnie uwzględnianie danych o ruchu naturalnym (urodzenia i zgony) oraz migracjach (na pobyt stały i czasowy), dla których źródłem są dane administracyjne.

Ludność bilansowa – liczba ludności w danym roku według definicji krajowej corocznie opracowywana w latach międzyspisowych (tj. w latach 2011–2020) przez GUS na podstawie bilansu ludności, dla którego bazą wyjściową są wyniki NSP 2011 w zakresie liczby i struktury ludności.

Szacunek wsteczny – metoda szacowania liczby i struktury ludności według płci i wieku w okresach międzyspisowych. Metoda szacunku wstecznego zakłada przyjęcie wyników aktualnego spisu powszechnego ludności (NSP 2021) za bazę wyjściową, a następnie odtwarzanie struktury demograficznej w poprzednich latach uwzględniając zarejestrowane dane o ruchu naturalnym (urodzenia i zgony) oraz migracje (na pobyt stały i czasowy) z lat 2011–2020.

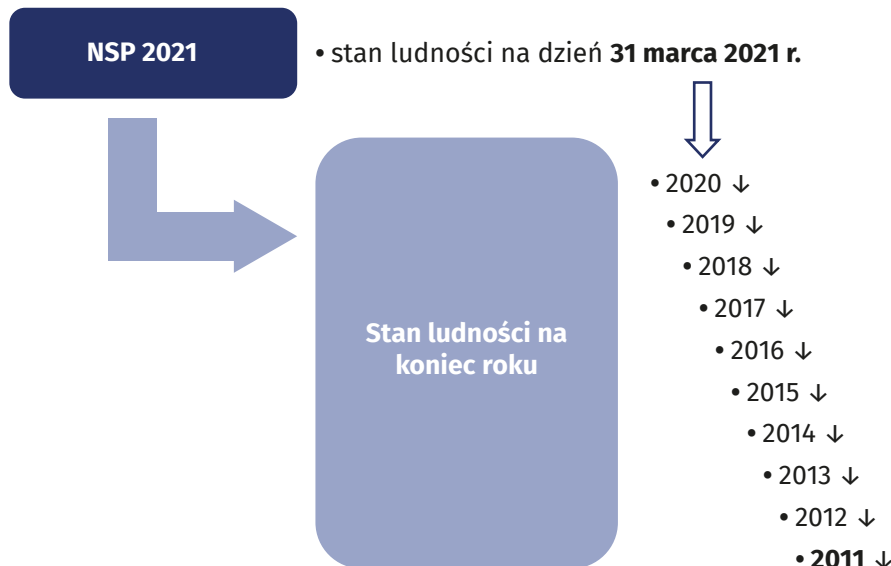
Ludność szacunkowa – liczba ludności opracowana metodą szacunku wstecznego na podstawie wyniku NSP 2021.

Przy wykorzystaniu metod analizy demograficznej obliczono różnice między liczbą ludności otrzymaną na podstawie NSP 2021, a opracowaną na podstawie poprzedniego spisu (NSP 2011) oraz bieżących zdarzeń demograficznych (urodzenia, zgony, migracje). W analizie tej wykorzystano:

- stan ludności na dzień 31 marca 2011 (wynik NSP 2011),
- stan ludności na dzień 31 marca 2021 (wynik NSP 2021),
- stan ludności na koniec roku (tj. 31 grudnia) w latach 2011–2020 (ludność bilansowa),
- bieżąca rejestracja zdarzeń demograficznych w latach 2011–2021 (urodzenia, zgony, migracje).

Przyjmując za punkt wyjścia ludność będącą wynikiem NSP 2021 (w podziale według województw, płci oraz pojedynczych roczników wieku) wykonano obliczenia liczby ludności dla poprzednich lat metodą szacunku wstecznego aż do 31 grudnia 2011 r. Zostało to zilustrowane na poniższym schemacie (Schemat 2).

Schemat 2. Schemat przeprowadzonego szacunku wstecznego
Scheme 2. Scheme of the retrospective estimation



Szacunek wsteczny sporządzono metodą składnikową przy uwzględnieniu zdarzeń demograficznych (urodzenia, zgony, migracje) oraz procesu „odmładzania” ludności według następującego schematu:

Liczba ludności w momencie krytycznego spisu (31 marca 2021 r., wg województw, płci i pojedynczych roczników wieku)

- **urodzenia żywe** (wg województw i płci noworodków),
- + **zgony** (wg województw, płci i wieku zmarłych),
- **zameldowania na pobyt stały** z innych województw i z zagranicy (wg województw, płci i wieku osób),
- + **wymeldowania z pobytu stałego** do innych województw i za granicę (wg województw, płci i wieku osób),
- **zameldowania na pobyt czasowy ponad 3 miesiące** z innych województw (wg województw, płci i wieku osób),
- + **wymeldowania na pobyt czasowy ponad 3 miesiące** do innych województw (wg województw, płci i wieku osób).

W obliczeniach uwzględniono dane dotyczące urodzeń, zgonów oraz migracji w 1 kwartale 2021 r. Zastosowaną metodę przedstawia poniższe równanie:

$$L_{rok}^{\overline{NSP2021}} = \begin{cases} L^{NSP2021} - U_{I kw. 2021} + Z_{I kw. 2021} - IM_{I kw. 2021} + EM_{I kw. 2021} & \text{dla } rok = 2020 \\ L_{rok+1}^{\overline{NSP2021}} - U_{rok+1} + Z_{rok+1} - IM_{rok+1} + EM_{rok+1} & \text{dla } rok \in [2011, 2019], \end{cases}$$

gdzie:

$L_{rok}^{\overline{NSP2021}}$ – ludność na koniec danego roku opracowana metodą szacunku wstecznego,

$L^{NSP2021}$ – ludność na podstawie danych NSP2021,

U_{rok} – urodzenia żywe w danym roku,

$U_{1 kw. 2021}$ – urodzenia żywe w 1 kwartale 2021 r.,

Z_{rok} – zgony w danym roku,

$Z_{1 kw. 2021}$ – zgony w 1 kwartale 2021 r.,

IM_{rok} – imigracje w danym roku,

$IM_{1kw.2021}$ – imigracje w 1 kwartale 2021 r.,

EM_{rok} – emigracje w danym roku,

$EM_{1kw.2021}$ – emigracje w 1 kwartale 2021 r.

Do porównań liczby ludności bilansowej (opracowanej metodą bilansową w okresie międzypisowym, gdzie podstawą były wyniki NSP 2011) z liczbą ludnością otrzymaną z szacunku wstecznego (opracowaną na bazie wyników NSP 2021) zastosowano następujące wzory:

$$R_{rok} = L_{rok}^{\overrightarrow{NSP2011}} - L_{rok}^{\overleftarrow{NSP2021}}$$

$$R_{rok}^{\%} = \frac{L_{rok}^{\overrightarrow{NSP2011}} - L_{rok}^{\overleftarrow{NSP2021}}}{L_{rok}^{\overrightarrow{NSP2011}}} \cdot 100\% = \frac{R_{rok}}{L_{rok}^{\overrightarrow{NSP2011}}} \cdot 100\%$$

gdzie:

$L_{rok}^{\overrightarrow{NSP2011}}$ – ludność na koniec danego roku publikowana przez GUS w okresie międzypisowym (tj. w latach 2011–2020) – ludność bilansowa,

$L_{rok}^{\overleftarrow{NSP2021}}$ – ponownie obliczona ludność na koniec danego roku w okresie międzypisowym na podstawie wyników NSP2021 metodą szacunku wstecznego,

R_{rok} – różnica między liczbą ludności otrzymaną jako bilans dla danego roku i liczbą ludnością otrzymaną z szacunku wstecznego,

$R_{rok}^{\%}$ – różnica R_{rok} wyrażona w wartościach procentowych w odniesieniu do ludności bilansowej.

5.2. Wyniki

5.2. Results

W niniejszym rozdziale analiza oceny jakości danych z NSP 2021 w zakresie liczby i struktury demograficznej ludności została podzielona na kilka etapów. Pierwszy z nich dotyczy przedstawienia najbardziej ogólnych wyników na poziomie krajowym zarówno dla ludności Polski ogółem, jak i z uwzględnieniem płci, a także 5-letnich grup wieku. Następnie przedstawiono wyniki pogłębionej analizy kolejno dla: noworodków i młodszych dzieci, osób w wieku największej mobilności oraz osób najstarszych¹². Kategorie te pełnią kluczową rolę z punktu widzenia analizy spójności danych administracyjnych z wynikami obecnego spisu, ale dodatkowo w przypadku osób najstarszych pozwalają (ze względu na znikome migracje w tej grupie) na bezpośrednie porównanie wyników spisów ludności i mieszkańców z 2011 i 2021 r. Ostatnią część rozdziału stanowi prezentacja oraz omówienie wyników na poziomie wojewódzkim.

5.2.1. Poziom krajowy

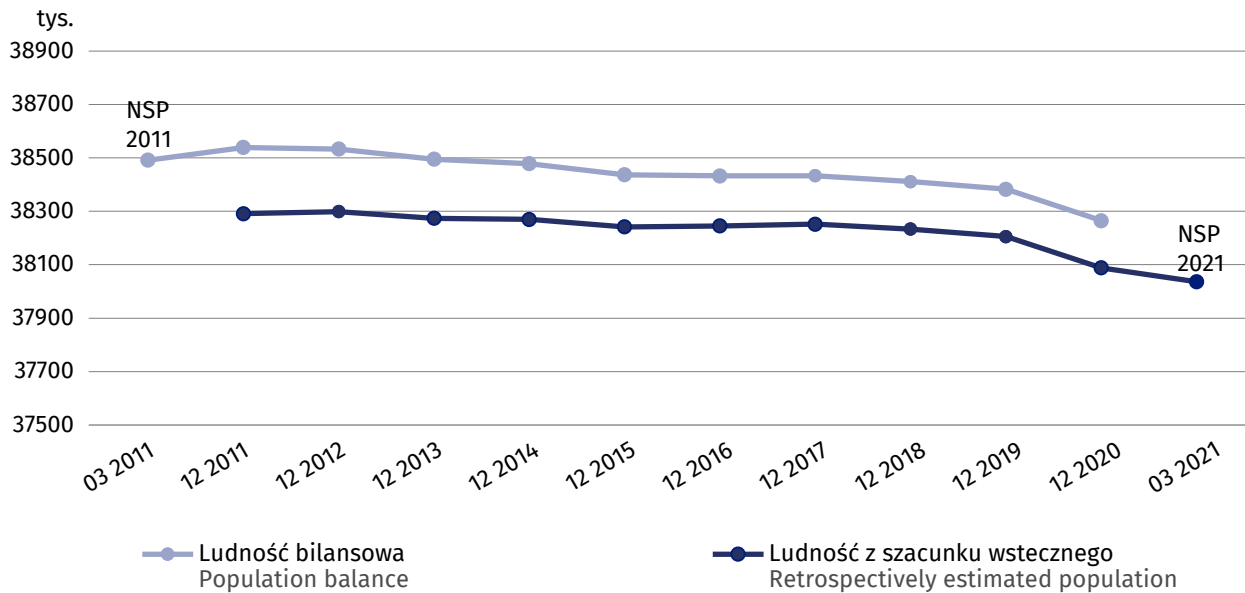
5.2.1. National level

Wykonując szacunek wsteczny, przyjmując za punkt wyjścia wynik NSP 2021, uzyskano na koniec 2020 r. liczbę ludności 38089 tys. Jest ona o 0,46% niższa od liczby ludności bilansowej na koniec 2020 r. (38265 tys.), której bazą wyjściową był wynik NSP 2011 (Wykres 22, Tablica 26). Dla wcześniejszych lat rozbieżność ta staje się coraz większa; na koniec 2011 r. osiąga wartość 0,64% (tj. 247,4 tys.). Wraz z kolejnymi latami po NSP 2011 malał odsetek osób objętych tym badaniem. Wiąże się to z wymieraniem najstarszych roczników wieku (najstarszych w momencie krytycznym NSP 2011), wśród których rozbieżności wyników przeważnie są największe, oraz corocznym uwzględnianiu danych dotyczących urodzeń dzieci po NSP 2011 r.

¹² Punktem wyjścia dla przyjętej metody były dostępne publikacje J. Paradysz (2010) oraz E. Gołata (2018).

Wykres 22. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością obliczoną metodą szacunku wstecznego

Chart 22. Comparison of the population balance (as of 31 December) with the population calculated using the retrospective estimation method



Tablica 26. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością obliczoną metodą szacunku wstecznego

Table 26. Comparison of the population balance (as of 31 December) with the population calculated using the retrospective estimation method

Rok Year	Ludność bilansowa Population balance	Ludność z szacunku wstecznego Retrospectively estimated population	Różnica ^a Difference ^a	
	w tysiącach in thousands		w % in %	
2011	38538,5	38291,1	247,4	0,64
2012	38533,3	38299,2	234,1	0,61
2013	38495,6	38274,0	221,6	0,58
2014	38478,6	38270,1	208,5	0,54
2015	38437,2	38241,6	195,6	0,51
2016	38433,0	38245,4	187,6	0,48
2017	38433,6	38251,7	181,9	0,47

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

Tablica 26. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością obliczoną metodą szacunku wstecznego (dok.)**Table 26. Comparison of the population balance (as of 31 December) with the population calculated using the retrospective estimation method (cont.)**

Rok Year	Ludność bilansowa Population balance	Ludność z szacunku wstecznego Retrospectively estimated population	Różnica ^a Difference ^a	
	w tysiącach in thousands		w % in %	
2018	38411,2	38233,3	177,9	0,46
2019	38382,6	38205,4	177,2	0,46
2020	38265,0	38088,6	176,4	0,46

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

W Tablicy 27 przedstawiono w dwóch wariantach ludność Polski (stan na 31 grudnia) w latach międzypisowych z uwzględnieniem płci. Pierwszy wariant dotyczy ludności bilansowej opartej na wynikach poprzedniego spisu 2011, z kolei drugi obejmuje ludność obliczoną metodą szacunku wstecznego na podstawie wyników najnowszego spisu 2021.

Uzyskane rozbieżności wyrażone w procentach zarówno w przypadku kobiet, jak i mężczyzn kształtują się w poszczególnych latach międzypisowych na podobnym poziomie i są bardzo zbliżone do wyników przedstawionych w Tablicy 26, czyli dla ludności Polski ogółem. Uśredniając rozbieżności na cały okres międzypisowy dla obu płci otrzymano różnice wyników na poziomie około 0,5%. Warto jednak zwrócić uwagę, że przedstawione rozbieżności wyrażone w tysiącach osób są większe dla kobiet niż dla mężczyzn.

Tablica 27. Porównanie ludności bilansowej (stan na 31 grudnia) z ludnością według płci obliczoną metodą szacunku wstecznego**Table 27. Comparison of the population balance (as of 31 December) with the population by sex calculated using the retrospective estimation method**

Rok Year	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a	
			w tysiącach in thousands	w % in %			w tysiącach in thousands	w % in %
2011	18654,6	18533,4	121,2	0,65	19883,9	19757,7	126,2	0,63
2012	18649,3	18534,9	114,4	0,61	19884,0	19764,3	119,7	0,60
2013	18629,5	18521,4	108,1	0,58	19866,1	19752,6	113,5	0,57
2014	18619,8	18518,2	101,6	0,55	19858,8	19751,9	106,9	0,54
2015	18598,0	18503,0	95,0	0,51	19839,2	19738,6	100,6	0,51
2016	18593,2	18502,3	90,9	0,49	19839,8	19743,1	96,7	0,49
2017	18593,2	18505,3	87,9	0,47	19840,4	19746,4	94,0	0,47
2018	18581,9	18496,1	85,8	0,46	19829,3	19737,2	92,1	0,46
2019	18567,1	18481,6	85,5	0,46	19815,5	19723,8	91,7	0,46
2020	18502,2	18417,2	85,1	0,46	19762,8	19671,4	91,4	0,46

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

Analizując wpływ wieku na kształtowanie się różnic między liczbą ludności bilansowej a liczbą ludności z szacunku wstecznego zwraca uwagę fakt, iż jedynie w przypadku najmłodszych grup wieku (0–4 oraz 5–9 lat) ludność z szacunku wstecznego przewyższyła liczbę ludności bilansowej (Tablica 28). Oznacza to, iż spis 2021 wykazał więcej dzieci niż wynikało to z danych, opracowanych na podstawie źródeł administracyjnych (wyniki badań bieżących w zakresie urodzeń, zgonów i migracji).

Największe różnice (powyżej 1%) dotyczą młodych osób w grupach wieku 5–9, 15–24 lata oraz w starszych grupach wieku tj. 55–64 lata i powyżej 80 lat w przypadku mężczyzn i powyżej 85 lat w przypadku kobiet. Z kolei najmniejsze różnice (poniżej 0,5%) wystąpiły dla osób w grupach wieku 10–14 oraz 35–49 lat.

Tablica 28. Porównanie ludności bilansowej (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego z uwzględnieniem płci i 5-letnich grup wieku**Table 28. Comparison of the population balance (as of 31 December, 2020) with the population by sex and 5-years age groups calculated using the retrospective estimation method**

Grupy wieku Age groups	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a	
			w tysiącach in thousands	w % in %			w tysiącach in thousands	w % in %
0– 4 lata	976,9	984,1	-7,1	-0,73	925,3	931,8	-6,5	-0,71
5– 9	982,0	1011,1	-29,2	-2,97	928,5	957,0	-28,5	-3,07
10–14	1059,0	1054,8	4,2	0,39	1006,6	1002,8	3,8	0,38
15–19	920,0	903,4	16,6	1,80	874,4	859,6	14,8	1,69
20–24	1007,1	990,5	16,6	1,65	962,6	952,0	10,6	1,10
25–29	1222,3	1212,7	9,6	0,79	1180,4	1172,6	7,8	0,66
30–34	1431,2	1423,9	7,3	0,51	1388,9	1380,4	8,6	0,62
35–39	1631,4	1631,2	0,2	0,01	1594,3	1587,9	6,4	0,40
40–44	1551,3	1550,9	0,3	0,02	1523,9	1522,4	1,5	0,10
45–49	1351,2	1347,7	3,4	0,25	1342,1	1338,0	4,1	0,30
50–54	1132,7	1125,7	7,0	0,62	1149,3	1140,8	8,6	0,74
55–59	1131,7	1120,1	11,7	1,03	1191,7	1181,0	10,7	0,90
60–64	1260,3	1246,6	13,7	1,09	1419,9	1408,0	12,0	0,84
65–69	1122,8	1112,2	10,6	0,94	1382,8	1372,2	10,5	0,76
70–74	809,9	803,7	6,2	0,76	1107,0	1100,1	7,0	0,63
75–79	390,3	386,8	3,5	0,89	623,2	619,3	3,9	0,63
80–84	292,9	289,6	3,3	1,12	572,9	569,1	3,7	0,65
85–89	160,0	157,2	2,8	1,77	378,3	374,5	3,8	1,01
90 lat i więcej and more	69,3	65,0	4,3	6,24	210,6	202,0	8,5	4,06

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

5.2.1.1. Kompletność szacunków liczby niemowląt

5.2.1.1. Completeness estimation of the infants' number

W celu przeprowadzenia analizy kompletności liczby niemowląt i dzieci spisanych podczas NSP 2021, zastosowano metodę „przetwarzania między urodzeniem a spisem” (Paradysz 2010). Metoda ta polega na porównaniu liczby spisanych niemowląt (0 lat) z liczbą urodzeń żywych, opracowaną na podstawie badań bieżących, opartych na danych administracyjnych w zakresie urodzeń. W wyniku porównania informacji z tych dwóch źródeł wyznacza się odsetek braków lub nadwyżki niemowląt. W publikacji J. Paradysza (2010) zaproponowano przeprowadzenie analizy w dwóch wariantach: bez uwzględnienia umieralności i migracji zagranicznych oraz tylko z uwzględnieniem umieralności. Według E. Gołaty (2018) brak uwzględnienia zgonów niemowląt zmniejsza precyzję, dlatego w niniejszej analizie zdecydowano się uwzględnić także zgony niemowląt.

Powyższą analizę przeprowadzono dla dwóch ostatnich spisów: NSP 2011 oraz NSP 2021. Dodatkowo analizowano kompletność urodzeń w dwóch grupach: pierwsza to najmłodsze niemowlęta, czyli dzieci urodzone między 1 stycznia a momentem krytycznym spisu, a druga analiza obejmuje wszystkie dzieci w wieku 0 lat w momencie krytycznym spisu (przykładowo dla NSP 2021 są to dzieci urodzone między 1 kwietnia 2020 r. a 31 marca 2021 r.).

Odsetek braków lub nadwyżki niemowląt do 3 miesiąca życia (R_0) wyznaczono za pomocą następującego wzoru:

$$R_0 = \frac{L_{0,NSP} - U_{rej} - Z_{0,rej}}{U_{rej} - Z_{0,rej}} \cdot 100\%$$

gdzie:

$L_{0,NSP}$ – liczba spisanych niemowląt w wieku do 3. miesiąca życia,

U_{rej} – liczba niemowląt urodzonych w określonym czasie (w przypadku NSP 2021: 1 stycznia–31 marca 2021),

$Z_{0,rej}$ – liczba zgonów niemowląt w wieku do 3. miesiąca życia.

Odsetek braków wszystkich niemowląt w momencie krytycznym spisu został obliczony w sposób analogiczny do zaprezentowanego powyżej zwiększając w sposób odpowiedni zakres danych.

Przedstawione wyniki (Tablica 29) wskazują na niedoszacowanie w obu spisach liczby niemowląt tj. spisano mniejszą liczbę dzieci w wieku 0 lat w stosunku do liczby niemowląt uwzględnionych w badaniach bieżących z zakresu statystyki urodzeń, dla których źródłem są dane administracyjne. Dodatkowo przedstawione szacunki wskazują na niższe odsetki braków niemowląt w NSP 2021 w porównaniu do NSP 2011.

Zgodnie z wynikami badań bieżących z zakresu urodzeń oraz zgonów liczba dzieci w wieku 0 lat (dzieci urodzone między 1 kwietnia 2020 r. a 31 marca 2021 r.) w momencie krytycznym NSP 2021 powinna wynieść 348119 osób, podczas gdy faktycznie spisanych zostało o 7412 niemowląt mniej. Stanowi to 2,12% wszystkich niemowląt, które powinien objąć spis. Dla porównania w NSP 2011 różnica ta była znacznie większa i wynosiła 12851 (3,18%). Największe braki można stwierdzić dla niemowląt urodzonych w trzech pierwszych miesiącach danego roku, dla NSP 2011 odsetek brakujących dzieci wyniósł 15,55%, a dla NSP 2021 już tylko 3,41%. W NSP 2021 nie odnotowano znaczących różnic według płci – odsetki braków były niemal identyczne dla obu płci (w momencie krytycznym NSP 2021: dla chłopców – 2,10%, dla dziewczynek – 2,15%).

Tablica 29. Analiza kompletności liczby niemowląt dla NSP 2011 oraz NSP 2021**Table 29. Completeness analysis of the infants' number in the 2011 Census and the 2021 Census**

Płeć Sex	Liczba urodzeń Number of births	Liczba zmarłych niemowląt Number of infant's deaths	Liczba niemowląt, którą powinien objąć spis Number of infants that the census should be recorded	Liczba spisanych niemowląt Number of infants recorded	Różnica w % ^a Difference in % ^a
	w tysiącach in thousands				
Zdarzenia demograficzne dotyczące niemowląt urodzonych w 3 pierwszych miesiącach 2011 r.					
Demographic events concerning infants born in the first 3 months of 2011					
Ogółem Total	95,6	0,3	95,2	80,4	-14,8
Chłopcy Boys	49,4	0,2	49,3	41,3	-8,0
Dziewczeta Girls	46,1	0,2	46,0	39,1	-6,8
Zdarzenia demograficzne dotyczące dzieci w wieku 0 lat w momencie krytycznym NSP 2011					
Demographic events for children aged 0 at the critical moment of the 2011 Census					
Ogółem Total	404,6	1,9	402,8	389,9	-12,9
Chłopcy Boys	209,7	1,0	208,7	200,6	-8,1
Dziewczeta Girls	194,9	0,9	194,1	189,3	-4,8
Zdarzenia demograficzne dotyczące niemowląt urodzonych w 3 pierwszych miesiącach 2021 r.					
Demographic events concerning infants born in the first 3 months of 2021					
Ogółem Total	83,2	0,2	83,0	0,25	1342,1
Chłopcy Boys	42,7	0,1	42,5	0,62	1149,3
Dziewczeta Girls	40,5	0,1	40,4	1,03	1191,7
Zdarzenia demograficzne dotyczące dzieci w wieku 0 lat w momencie krytycznym NSP 2021					
Demographic events for children aged 0 at the critical moment of the 2021 Census					
Ogółem Total	349,3	1,2	348,1	340,7	-7,4
Chłopcy Boys	179,2	0,7	178,5	174,8	-3,8
Dziewczeta Girls	170,1	0,5	169,6	165,9	-3,7

a Różnica między liczbą dzieci spisanych a liczbą dzieci, które powinien objąć spis. Różnica ta została również wyrażona w wartościach procentowych w odniesieniu do liczby dzieci, którą powinien objąć spis.

a Difference between the number of children comes from the Census and the number of children. This difference was also expressed as a percentage value in relation to the number of children

5.2.1.2. Kompletność szacunków liczby dzieci w wieku 0–9 lat

5.2.1.2. Completeness estimation of the number of children aged 0–9 years

W tej części analizy zastosowano analogiczny algorytm jak w przypadku niemowląt tj. „przetrwania między urodzeniem a spisem” (Paradysz 2010). Zdecydowano jednak o uwzględnieniu, obok urodzeń i zgonów, również salda migracji zagranicznych na pobyt stały. Liczbę dzieci na moment krytyczny spisu (31 marca 2021) obliczono przy pomocy poniższego równania:

$$L_{rok,x} = \begin{cases} U_{rok} - Z_{rok,x} + Migr_{rok,x} & \text{dla } x = 0 \\ L_{rok-1,x-1} - Z_{rok,x} + Migr_{rok,x} & \text{dla } x > 0, \end{cases}$$

gdzie:

Lrok,x – ludność w wieku x lat w danym roku,

Urok – liczba urodzeń w danym roku,

Zrok,x – liczba zgonów w wieku x lat w danym roku,

Migrrok,x – saldo migracji zagranicznych w wieku x lat w danym roku.

Warto odnotować, że w analizowanej grupie 0–9 lat jedynie dla dzieci w wieku 0 oraz 2 lat różnice były ujemne, co oznacza mniejszą liczbę dzieci w wieku 0 oraz 2 lat, otrzymaną na podstawie wyników NSP 2021 (Tablica 30). Należy jednak zauważyć, iż w przypadku dzieci dwuletnich uzyskane różnice były bliskie zeru, co oznacza, że w ich przypadku spójność danych jest niemal całkowita.

W pozostałych analizowanych rocznikach wieku stwierdzono nadwyżkę dzieci spisanych w porównaniu do oszacowanych na podstawie badań bieżących z zakresu ruchu naturalnego (urodzenia i zgonu) oraz migracji. Największa nadwyżka wystąpiła wśród dzieci w wieku 9 lat (3,77% ogółem) oraz 6 lat (3,62% ogółem).

Rozbieżności szacunków liczby dzieci w poszczególnych rocznikach wieku były podobne dla chłopców i dziewczynek (Tablica 30). Największe różnica według płci dotyczy dziewięciolatków (0,58 p. proc.).

Tablica 30. Analiza kompletności liczby dzieci w wieku 0–9 lat spisanych w NSP 2021

Table 30. Completeness analysis of the number of children aged 0–9 enumerated in the 2021 Census

Wiek Age	Liczba dzieci spisanych w NSP 2021 Number of children enumerated in the 2021 Census			Liczba dzieci oszacowana na podstawie badań bieżących Number of children based on current surveys			Różnica ^a Difference ^a					
	ogółem total	chłopcy boys	dziewczeta girls	ogółem total	chłopcy boys	dziewczeta girls	ogółem total	chłopcy boys	dziewczeta girls	ogółem total	chłopcy boys	dziewczeta girls
	w tysiącach in thousands						w % in %					
0 lat	340,7	174,8	165,9	349,0	179,0	170,0	-8,3	-4,2	-4,1	-2,39	-2,36	-2,42
1	371,7	190,9	180,8	370,2	190,1	180,1	1,5	0,7	0,8	0,40	0,37	0,43
2	386,1	198,7	187,4	386,2	198,7	187,5	-0,1	0,0	-0,1	-0,03	-0,01	-0,04

a Różnica między liczbą dzieci spisanych a liczbą dzieci oszacowaną na podstawie badań bieżących z zakresu ruchu naturalnego i migracji. Różnica ta została również wyrażona w wartościach procentowych w odniesieniu do liczby dzieci oszacowanej na podstawie danych administracyjnych.

a Difference between the number of children obtained from the census and the number of children based on current surveys of vital statistics and migration. This difference was also expressed as a percentage value in relation to the number of children based on the administrative data

Tablica 30. Analiza kompletności liczby dzieci w wieku 0–9 lat spisanych w NSP 2021 (dok.)**Table 30. Completeness analysis of the number of children aged 0–9 enumerated in the 2021 Census (cont.)**

Wiek Age	Liczba dzieci spisanych w NSP 2021 Number of children enumerated in the 2021 Census			Liczba dzieci oszacowana na podstawie badań bieżących Number of children based on current surveys			Różnica ^a Difference ^a					
	ogółem total	chłopcy boys	dziewczęta girls	ogółem total	chłopcy boys	dziewczęta girls	ogółem total	chłopcy boys	dziewczęta girls	ogółem total	chłopcy boys	dziewczęta girls
	w tysiącach in thousands									w % in %		
3	406,7	209,0	197,7	401,5	206,2	195,3	5,2	2,8	2,4	1,29	1,37	1,21
4	401,2	205,7	195,4	391,7	201,0	190,7	9,4	4,7	4,8	2,41	2,33	2,49
5	388,3	199,3	188,9	376,9	193,5	183,4	11,4	5,9	5,5	3,02	3,04	3,00
6	390,8	200,8	190,0	377,1	193,9	183,3	13,7	6,9	6,7	3,62	3,58	3,67
7	384,4	197,3	187,1	373,6	192,0	181,7	10,7	5,3	5,4	2,87	2,76	2,99
8	396,9	204,3	192,7	384,3	197,5	186,8	12,6	6,7	5,9	3,29	3,40	3,17
9 lat	404,6	207,6	197,1	389,9	200,6	189,4	14,7	7,0	7,7	3,77	3,49	4,07

a Różnica między liczbą dzieci spisanych a liczbą dzieci oszacowaną na podstawie badań bieżących z zakresu ruchu naturalnego i migracji. Różnica ta została również wyrażona w wartościach procentowych w odniesieniu do liczby dzieci oszacowanej na podstawie danych administracyjnych.

a Difference between the number of children obtained from the census and the number of children based on current surveys of vital statistics and migration. This difference was also expressed as a percentage value in relation to the number of children based on the administrative data

5.2.1.3. Kompletność szacunków liczby osób w wieku podwyższonej mobilności

5.2.1.3. Completeness estimation of the number of persons at the age with the higher mobility

Kolejnym etapem analizy było zbadanie spójności między wynikami w zakresie liczby ludności bilansowej (otrzymanej na podstawie NSP 2011), a ludności z szacunku wstecznego (otrzymanego na podstawie NSP 2021) dla osób dorosłych o szczególnie podwyższonej mobilności, czyli w grupie wieku 25–40 lat. Przeanalizowano salda migracji zagranicznych na pobyt stały dla pojedynczych roczników wieku dla obu płci łącznie. Analiza wykazała, że osoby w wieku 25–40 lat stanowiły w 2020 r. aż 33% ogółu migracji na pobyt stały, opracowanych na podstawie badań bieżących, dla których źródłem danych pozostają dane administracyjne.

W Tablicy 31 przedstawiono wyniki dla pojedynczych roczników wieku z uwzględnieniem płci (według stanu na dzień 31 grudnia 2020 r.). Największe rozbieżności dotyczą osób w wieku 25 i 26 lat. W przypadku mężczyzn różnica ta jest rzędu 1%, co przekłada się na blisko 5 tys. osób więcej w ludności bilansowej dla obu roczników wieku łącznie. W przypadku kobiet nadwyżka ta jest mniejsza i wynosi ok. 0,81%, co przekłada się w sumie na ponad 2,5 tys. osób. Warto odnotować, że również dla kobiet w wieku 33 lat wspomniana nadwyżka jest na podobnym poziomie tj. rzędu 0,81%.

Należy podkreślić, że dla pozostałych roczników wieku otrzymano dużą zgodność (różnica poniżej 1%) między wynikiem szacunku wstecznego a ludnością bilansową.

Tablica 31. Porównanie ludności bilansowej w wieku 25–40 lat według płci (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego**Table 31. Comparison of the population balance aged 25–40 by sex (as of 31 December, 2020) with the population calculated using the retrospective estimation method**

Grupy wieku Age groups	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a	
			w tysiącach in thousands	w % in %			w tysiącach in thousands	w % in %
25 lat	220,55	218,04	2,51	1,14	212,98	211,26	1,73	0,81
26	234,08	231,74	2,34	1,00	227,13	225,30	1,83	0,81
27	244,55	242,50	2,05	0,84	235,71	234,09	1,61	0,68
28	253,70	252,12	1,58	0,62	245,15	243,77	1,37	0,56
29	269,42	268,27	1,15	0,43	259,46	258,16	1,30	0,50
30	275,22	274,22	1,00	0,36	265,90	264,84	1,06	0,40
31	277,41	276,14	1,27	0,46	268,31	266,80	1,51	0,56
32	285,11	283,22	1,89	0,66	277,44	275,48	1,96	0,70
33	290,96	289,02	1,95	0,67	282,00	279,72	2,28	0,81
34	302,53	301,32	1,20	0,40	295,29	293,53	1,76	0,60
35	320,08	319,86	0,22	0,07	312,26	310,75	1,51	0,48
36	330,16	330,13	0,03	0,01	321,71	320,25	1,46	0,45
37	338,68	338,73	-0,05	-0,01	332,12	330,69	1,44	0,43
38	328,89	328,92	-0,03	-0,01	320,58	319,34	1,25	0,39
39	313,60	313,54	0,06	0,02	307,64	306,91	0,73	0,24
40 lat	318,60	318,64	-0,04	-0,01	313,29	312,88	0,41	0,13

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

5.2.1.4. Kompletność szacunków dla osób w wieku 90 lat i więcej

5.2.1.4. Completeness estimation for the persons aged 90 years and more

Następnym etapem oceny jakości spisu było zbadanie zgodności między ludnością bilansową, a uzyskaną w wyniku przeprowadzenia szacunku wstecznego dla osób najstarszych (powyżej 90 lat). Osoby w tym wieku charakteryzują się najmniejszą skłonnością do migracji.

Dla każdego z analizowanych roczników wieku przedstawiony wynik wskazuje na większą liczbę ludności bilansowej, niż tę, którą uzyskano z szacunku wstecznego (Tablica 32). Wraz z wiekiem rosną również różnice wyrażone w procentach, są one znacznie mniejsze dla kobiet niż dla mężczyzn. Różnice wyrażone w liczbach absolutnych w grupie wieku 90 lat i więcej nie przekraczają 1000 osób, z wyłączeniem kobiet w grupie wieku 100 i więcej lat.

Tablica 32. Porównanie ludności bilansowej (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego dla osób w wieku 90 lat i więcej według płci i wieku

Table 32. Comparison of the population balance (as of December 31, 2020) with the population calculated using the retrospective estimation for the persons aged 90 years and more by sex and age

Grupy wieku Age groups	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a	
			w liczbach bezwzględnych in absolute numbers	w % in %			w liczbach bezwzględnych in absolute numbers	w % in %
90 lat	19046	18413	633	3,32	51126	50152	974	1,91
91	14150	13581	569	4,02	40211	39285	926	2,30
92	10757	10297	460	4,28	32134	31258	876	2,73
93	7517	7155	362	4,82	23986	23302	684	2,85
94	5239	4946	293	5,59	18332	17683	649	3,54
95	4069	3748	321	7,89	14646	13970	676	4,62
96	2701	2404	297	11,00	9860	9369	491	4,98
97	2037	1750	287	14,09	7171	6688	483	6,74
98	1389	1134	255	18,36	4896	4441	455	9,29
99	835	661	174	20,84	2913	2594	319	10,95
100 lat i więcej and more	1571	896	675	42,97	5311	3300	2011	37,87

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

5.2.2. Poziom wojewódzki 5.2.2. Voivodship level

Jak zaprezentowano wcześniej, w skali całego kraju na koniec 2020 r. rozbieżności pomiędzy ludnością bilansową a ludnością z szacunku wstecznego kształtują się na poziomie ok. 0,46%, co przekłada się na nieco ponad 176,4 tys. osób na korzyść ludności bilansowej.

W niniejszym rozdziale zaprezentowano wyniki oceny kompletności w ujęciu regionalnym.

Analiza wykazała, że rozbieżności na poziomie województw kształtują się na poziomie od kilkunastu do kilkudziesięciu tysięcy osób. Zgodnie z przyjętą metodologią obliczania różnic, wynik dodatni oznacza większą liczbę ludności bilansowej. Należy zwrócić uwagę, że dla poszczególnych województw przedstawione różnice dla obu płci mają ten sam kierunek (niedoszacowanie lub przeszacowanie) i zbliżoną wartość (Tablica 33). Ze względu na brak zróżnicowania kolejne analizy wykonano łącznie dla kobiet i mężczyzn (Mapa 1).

Tablica 33. Porównanie ludności bilansowej według płci i województw (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego

Table 33. Comparison of the population balance by sex and voivodships (as of December 31 2020) with the population based on the retrospective estimation method

Województwo Voivodship	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro- spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro- spectively estimated population	różnica ^a difference ^a	
			w tysiącach in thousands	w % in %			w tysiącach in thousands	w % in %
Dolnośląskie	1389,4	1398,4	-9,0	-0,65	1501,9	1509,9	-8,1	-0,54
Kujawsko- pomorskie	999,1	983,5	15,5	1,56	1062,9	1048,0	14,8	1,40
Lubelskie	1015,2	996,0	19,3	1,90	1080,0	1061,0	19,1	1,77
Lubuskie	489,8	483,0	6,9	1,40	517,3	510,4	7,0	1,35
Łódzkie	1161,5	1151,2	10,3	0,89	1276,5	1264,6	11,9	0,93
Małopolskie	1654,9	1665,5	-10,6	-0,64	1755,5	1767,2	-11,7	-0,67
Mazowieckie	2595,7	2642,4	-46,7	-1,80	2829,4	2875,3	-45,9	-1,62
Opolskie	472,0	462,4	9,7	2,05	504,7	493,5	11,3	2,23
Podkarpackie	1038,4	1026,3	12,2	1,17	1082,8	1069,9	12,9	1,19
Podlaskie	571,5	563,0	8,5	1,49	601,8	593,6	8,2	1,36
Pomorskie	1141,7	1147,0	-5,3	-0,46	1204,9	1211,2	-6,3	-0,52
Śląskie	2163,4	2125,1	38,4	1,77	2328,9	2287,0	41,9	1,80
Świętokrzyskie	596,7	584,3	12,4	2,08	627,9	615,3	12,6	2,01
Warmińsko- mazurskie	692,5	677,2	15,3	2,21	724,0	708,4	15,6	2,15

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

Tablica 33. Porównanie ludności bilansowej według płci i województw (stan na 31 grudnia 2020 r.) z ludnością obliczoną metodą szacunku wstecznego (dok.)**Table 33. Comparison of the population balance by sex and voivodships (as of December 31 2020) with the population based on the retrospective estimation method (cont.)**

Województwo Voivodship	Mężczyźni Males				Kobiety Females			
	ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a		ludność bilansowa population balance	ludność z szacunku wstecznego retro-spectively estimated population	różnica ^a difference ^a	
			w tysiącach in thousands	w % in %			w tysiącach in thousands	w % in %
Wielkopolskie	1700,4	1705,0	-4,6	-0,27	1796,0	1802,0	-6,0	-0,33
Zachodniopomorskie	819,8	807,0	12,8	1,57	868,2	854,1	14,1	1,63

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

Najmniejsze rozbieżności¹³ pomiędzy liczbą ludności na koniec 2020 r. stwierdzono w województwach wielkopolskim (-0,30%), pomorskim (-0,49%), dolnośląskim (-0,59%) i małopolskim (-0,65%). Z kolei największe różnice występują w województwach warmińsko-mazurskim (2,18%), opolskim (2,14%) oraz świętokrzyskim (2,04%).

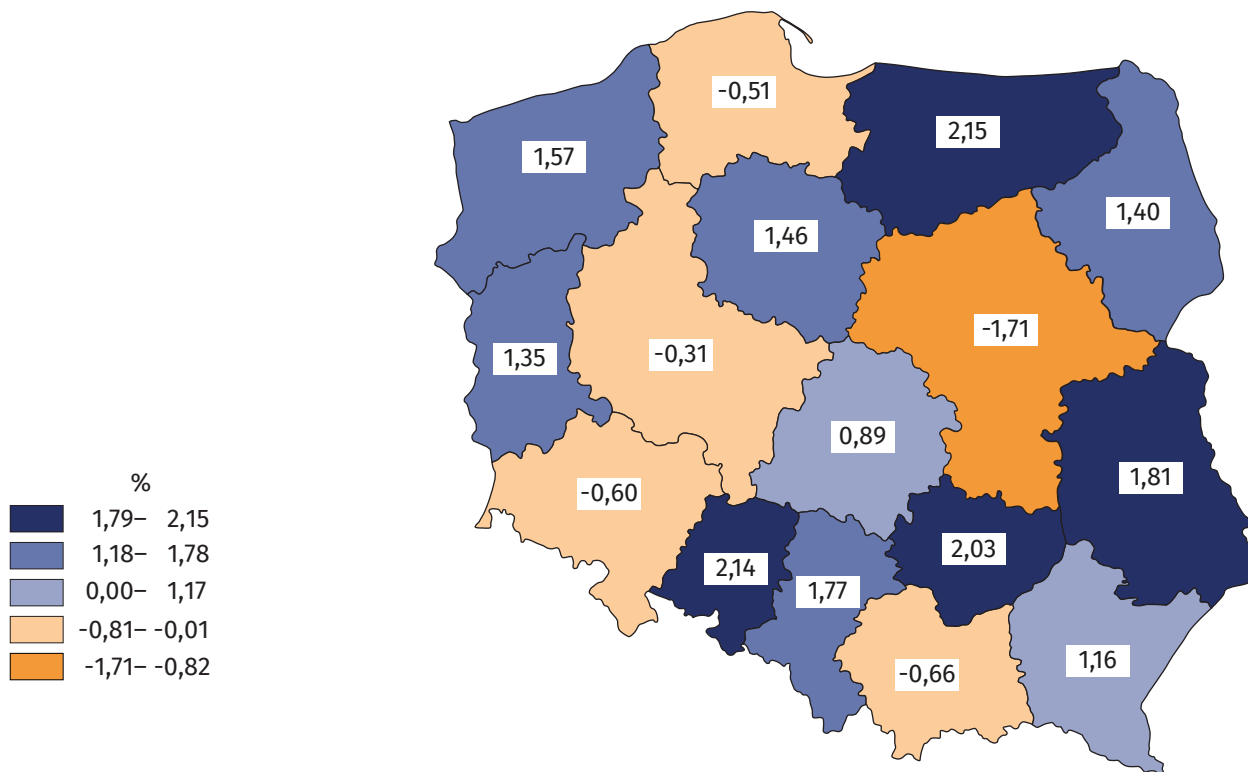
Warto odnotować, iż we wszystkich województwach, w których obecny spis wykazał większą liczbę ludności (tj. różnice ujemne), znajdują się główne aglomeracje miejskie (Wrocław, Kraków, Warszawa, Trójmiasto i Poznań), jednak z wyjątkiem województwa mazowieckiego różnica nie przekracza 1%. Są to województwa, które charakteryzują się wysokimi imigracjami, głównie w celach zarobkowych i często bez formalnego udokumentowania, co nie jest odzwierciedlone w wynikach badań bieżących z zakresu migracji zagranicznych i wewnętrznych. W pozostałych województwach, charakteryzujących się niższymi wskaźnikami ekonomicznymi¹⁴ (np. współczynnik aktywności zawodowej, wskaźnik zatrudnienia, stopa bezrobocia, PKB na 1 mieszkańca), ludność bilansowa przewyższyła tę, jaka została opracowana na podstawie spisu NSP 2021.

13 Analiza wielkości rozbieżności dla poszczególnych województw została przygotowana w oparciu o wartości bezwzględne uzyskanych różnic. Im bliżej zera znajduje się wartość bezwzględna danej różnicy, tym mniejsza jest rozbieżność.

14 Regiony Polski 2022, <https://stat.gov.pl/obszary-tematyczne/inne-opracowania/miasta-województwa/regiony-polski-2022,6,16.html> (dostęp 15 marca 2023 r.)

Mapa 1. Różnica procentowa między ludnością bilansową (stan na 31 grudnia 2020 r.) a ludnością szacunkową (stan na 31 grudnia 2020 r.) według województw

Map 1. Relative difference between the population balance (as of 31 December 2020) and the estimated population (as of 31 December 2020) by voivodships



Uwzględnienie w analizie dodatkowo 15-letnich grup wieku (Tablica 34) umożliwiło bardziej szczegółowe przedstawienie rozbieżności dla województw.

Największym zróżnicowaniem charakteryzują się grupy wieku 15–29 (od -6,17% w województwie mazowieckim do 8,40% w świętokrzyskim) oraz 90 lat i więcej (od -0,39% w województwie wielkopolskim do 17,35% w opolskim). Z kolei najbardziej wyrównane są grupy wieku 45–59 (od -0,37% w województwie mazowieckim do 2,30% w województwie opolskim) oraz 60–74 lata (od 0,20% w województwie świętokrzyskim do 2,17% w opolskim).

Grupa wieku 15–29 lat warunkuje w największym stopniu całkowite rozbieżności dla poszczególnych województw – zarówno pod względem kierunku (niedoszacowanie/przeszacowanie), jak i wartości. We wszystkich województwach stwierdzone dla tej grupy wieku rozbieżności w wartościach absolutnych są największe spośród wszystkich pozostałych grup wieku. Warto zwrócić uwagę, że w województwach o najkorzystniejszych wartościach wskaźników ekonomicznych, rozbieżności te są ujemne, co oznacza większą liczbę osób spisanych w NSP 2021 niż wynikało to z prowadzonych corocznie bilansów ludności i wyników poprzedniego spisu (NSP 2011).

Na szczególną uwagę zasługuje najmłodsza grupa wieku (0–14 lat). We wszystkich województwach, oprócz śląskiego, stwierdzono większą liczbę dzieci otrzymaną na podstawie spisu NSP 2021 względem bilansu ludności.

Obecny spis wskazuje na mniejszą liczbę osób powyżej 60 lat we wszystkich województwach w porównaniu do bilansu ludności wg stanu na 31 grudnia 2020 r. W największym stopniu dotyczy to województwa śląskiego.

Tablica 34. Różnica w tysiącach osób oraz różnica procentowa^a między ludnością bilansową a ludnością obliczoną metodą szacunku wstecznego z uwzględnieniem 15-letnich grup wieku oraz województw
Table 34. The absolute (in thousands of persons) and the relative difference^a between the population balance and the population developed using the retrospective estimation method by 15-years age groups and voivodships

Województwo Voivodship	Grupy wieku Age groups						
	0–14 lat	15–29	30–44	45–59	60–74	75–89	90 lat i więcej and more
W tysiącach In thousands							
Dolnośląskie	-3,2	-18,0	-5,8	1,7	6,0	1,5	0,9
Kujawsko- pomorskie	-3,6	19,2	7,4	3,3	3,1	0,6	0,4
Lubelskie	-5,7	23,2	13,9	4,3	1,5	0,5	0,5
Lubuskie	-0,6	10,6	1,3	0,7	1,2	0,4	0,2
Łódzkie	-0,7	12,5	5,4	1,8	1,8	0,7	0,6
Małopolskie	-10,0	-20,2	-0,3	2,9	3,9	0,9	0,6
Mazowieckie	-13,0	-51,6	-35,5	-3,7	6,8	2,3	1,9
Opolskie	-2,1	8,1	2,6	4,7	4,0	2,4	1,3
Podkarpackie	-11,3	22,2	7,6	3,9	2,0	0,4	0,3
Podlaskie	-5,1	11,4	4,4	2,9	1,9	0,6	0,5
Pomorskie	-1,4	-11,2	-5,0	1,5	3,3	0,8	0,4
Śląskie	1,8	25,4	12,1	11,9	16,3	8,5	4,3
Świętokrzyskie	-3,0	17,0	7,6	2,4	0,5	0,2	0,4
Warmińsko- mazurskie	-0,8	20,3	6,2	3,0	1,6	0,4	0,1
Wielkopolskie	-4,6	-7,5	-1,3	0,9	2,0	-0,1	-0,1
Zachodniopo- morskie	0,1	14,8	3,5	3,3	4,1	0,9	0,3
W % In %							
Dolnośląskie	-0,76	-4,18	-0,82	0,31	1,03	0,85	3,96
Kujawsko- pomorskie	-1,16	5,58	1,53	0,82	0,80	0,50	3,09
Lubelskie	-1,86	6,46	2,90	1,08	0,39	0,37	3,15
Lubuskie	-0,41	6,50	0,54	0,37	0,61	0,70	3,54

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

Tablica 34. Różnica w tysiącach osób oraz różnica procentowa^a między ludnością bilansową a ludnością obliczoną metodą szacunku wstecznego z uwzględnieniem 15-letnich grup wieku oraz województw (dok.)

Table 34. The absolute (in thousands of persons) and the relative difference^a between the population balance and the population developed using the retrospective estimation method by 15-years age groups and voivodships (cont.)

Województwo Voivodship	Grupy wieku Age groups						
	0–14 lat	15–29	30–44	45–59	60–74	75–89	90 lat i więcej and more
Łódzkie	-0,21	3,30	0,97	0,39	0,37	0,44	3,04
Małopolskie	-1,82	-3,51	-0,03	0,44	0,68	0,40	2,67
Mazowieckie	-1,45	-6,17	-2,67	-0,37	0,71	0,66	4,19
Opolskie	-1,61	5,34	1,15	2,30	2,17	3,45	17,35
Podkarpackie	-3,53	5,87	1,52	0,95	0,56	0,30	2,01
Podlaskie	-2,93	5,73	1,64	1,26	0,90	0,73	4,74
Pomorskie	-0,34	-2,90	-0,89	0,34	0,78	0,64	2,68
Śląskie	0,27	3,78	1,15	1,34	1,86	2,67	13,88
Świętokrzyskie	-1,81	8,40	2,73	1,01	0,20	0,24	3,93
Warmińsko- mazurskie	-0,37	8,34	1,88	1,09	0,60	0,48	1,61
Wielkopolskie	-0,79	-1,30	-0,15	0,13	0,33	-0,03	-0,39
Zachodniopo- morskie	0,03	5,47	0,88	1,02	1,20	0,91	2,66

a Różnica między ludnością bilansową a ludnością z szacunku wstecznego wyrażona zarówno w wartościach absolutnych, jak i w wartościach procentowych względem ludności bilansowej.

a Difference between the population balance and retrospectively estimated population expressed in absolute value and percentages value in relation to the population balance.

5.2.3. Podsumowanie

1. W celu oceny jakości spisu, zgodnie z dokumentem Organizacji Narodów Zjednoczonych (2017) „Principles and Recommendations for Population and Housing Censuses” zastosowano jedną z zalecanych metod, tj. analizę demograficzną polegającą na porównywaniu odpowiednich struktur ludności według płci i wieku obliczonych na bazie różnych źródeł danych. Zgodnie z tym dla określonego roku obliczone zostały dwie struktury demograficzne: metodą bilansową na podstawie wyników poprzedniego spisu (NSP 2011) oraz metodą szacunku wstecznego na podstawie aktualnego spisu (NSP 2021).
2. Różnice między liczbą ludności Polski w latach 2011–2020 kształtują się średnio na poziomie 0,5%, co przekłada się na ok. 200 tys. osób. Należy podkreślić, że obserwowane rozbieżności pomiędzy wynikami spisów nie zależą od płci, zarówno dla kobiet, jak i dla mężczyzn kształtują się one na zbliżonym poziomie w całym okresie międzyspisowym.
3. Inna sytuacja obserwowana jest przy uwzględnieniu wieku, stąd też pełni on zasadniczą rolę przy ocenie jakości analizowanych danych. Ocenie poddano liczbę dzieci w przedziale wieku od 0 do 9 lat (według sta-

nu na 31 grudnia 2020 r.), czyli urodzone już po spisie 2011. Ich liczebność na koniec 2020 r. jest wyłącznie wynikiem bilansu stanu i struktury ludności, opracowywanym na podstawie bieżących udokumentowanych zdarzeń demograficznych (urodzeń, zgonów i migracji). W spisie 2021 stwierdzono większą liczebność dzieci w tej grupie wieku (o 71,3 tys.) niż wynikało to jedynie z badań bieżących w zakresie urodzeń, zgonów oraz migracji na pobyt stały opracowywanych w okresie międzyspisowym. Warto podkreślić, że osiągnięto wyraźnie większą spójność w grupie wieku 0-9 lat niż w poprzedniej edycji spisu z 2011 r., zwłaszcza w przypadku dzieci do 4 roku życia. Obecnie stwierdzone rozbieżności w tej grupie wieku (0-4 lat) są o blisko 15% mniejsze niż analogiczne (na koniec 2010 r.) występujące w spisie z 2011 r.

4. Ważnym komponentem oceny jakości danych jest wyodrębnienie grupy osób charakteryzujących się największą mobilnością. Analiza częstości migracji wykazała, że w 2020 r. były to osoby w grupie wieku 25-40 lat. Dla osób w grupie wieku od 30 do 39 lat uzyskano pięciokrotnie większą spójność w porównaniu do poprzedniego spisu. Na koniec 2010 r. wskaźnik rozbieżności pomiędzy wynikami spisu (NSP 2011) a ludnością bilansową (na podstawie NSP 2011) kształtował się na poziomie 1,5%, a na koniec 2020 r. wskaźnik analogicznych rozbieżności dla ocenianego spisu (NSP 2021) obniżył się do poziomu 0,3%.
5. Liczebności osób w wieku 25 i 26 lat – uzyskane na podstawie zarówno obecnego, jak i poprzedniego spisu – są znacznie niższe niż wskazują na to wyniki bilansu ludności. Warto odnotować, że obecnie obserwowane dla nich rozbieżności – na poziomie 1% – są dwukrotnie mniejsze niż w poprzedniej edycji spisu. Oznacza to poprawę spójności danych badań bieżących w zakresie urodzeń, zgonów oraz migracji na pobyt stały z wynikami spisu dla osób w tych rocznikach wieku.
6. Bezpośrednie porównanie wyników dwóch spisów przeprowadzono również na podstawie analizy demograficznej osób najstarszych (90 lat i więcej). Pozwala to na pominięcie w znacznym stopniu najstabilniej udokumentowanego komponentu zmiany demograficznej jakim są migracje, np. w 2020 r. emigracja na pobyt stały wyniosła zaledwie 12 osób, a imigracja 13 dla osób najstarszych w wieku 90 lat i więcej. Rezultaty analizy wskazują na większą liczbę osób w analizowanej grupie wieku w ludności bilansowej w porównaniu do wielkości obliczonych w szacunku wstecznym. Wyniki NSP 2021 umożliwiły skorygowanie tych wartości o kilkaset osób przypadających na poszczególne roczniki wieku.
7. Przeprowadzona analiza demograficzna na poziomie województw wskazuje na duże zróżnicowanie rozbieżności pomiędzy ludnością bilansową a ludnością z szacunku wstecznego, które kształtują się od kilkunastu do kilkudziesięciu tysięcy osób. Przy czym, tak jak na poziomie krajowym również i w tym przypadku rozbieżności nie zależą od płci.
8. We wszystkich województwach, w których obecny spis (NSP 2021) wykazał większą liczbę ludności znajdują się główne aglomeracje miejskie (Wrocław, Kraków, Warszawa, Trójmiasto i Poznań). W pozostałych województwach liczba ludności bilansowej przewyższyła liczbę ludności opracowaną na podstawie obecnego spisu NSP 2021.

5.2.4. Summary

1. To assess the quality of the census, in accordance with the United Nations (2017) "Principles and Recommendations for Population and Housing Censuses", one of the recommended methods was used. Therefore, demographic analysis based on comparing the corresponding population structures by sex and age calculated from various data sources. For this purpose, two demographic structures were computed for a given year: one using the balance method based on the results of the 2011 Census and second using the retrospective estimation method based on the 2021 Census.
2. The differences between the population of Poland in the years 2011-2020 are on average at the level of 0.5%, which amounts to approximately 200 thousand persons. It should be emphasized that the observed discrepancies between the results of the censuses do not depend on sex, both for women and men they are at a similar level throughout the entire period between the censuses.
3. On the other hand, a different situation is observed when age is taken into account. Therefore, it plays a fundamental role in assessing the quality of census. Particularly noteworthy are children aged 0 to 9 (as

of December 31 2020), which were born after the 2011 Census. Their number at the end of 2020 is the result of the population balance based on current statistics (births, deaths and migration). In the 2021 Census, the number of children aged 0–9 was higher (by 71.3 thousand) compared to the corresponding current statistics on births, deaths and migration for permanent stay developed in the period between the censuses. It is worth emphasising that a clearly greater coherence was achieved in the 0–9 age group than in the previous edition of the 2011 Census, especially in the case of children up to 4 years of age. The discrepancies in this age group (0–4 years old) are nearly 15% smaller than the corresponding ones (at the end of 2010) in the 2011 Census.

4. An important component of assessing the quality of the data is the identification of a group of persons characterized by the greatest mobility. The analysis of migration frequency showed that in 2020 there were persons in the age group of 25 to 40. For those aged 30 to 39 consistency of results was five times higher than in the previous census. At the end of 2010, the discrepancies between the results of the 2011 Census and the population balance (based on the 2002 Census) amounted to 1.5%. At the end of 2020, the corresponding discrepancies indicator for the 2021 Census decreased to 0.3%.
5. The number of persons aged 25 and 26, developed on the basis of both the 2011 and 2021 Census, is much smaller than indicated by the results of the population balance. It is worth noting that the discrepancies currently observed for them – at the level of 1% – are two times lower than in the previous edition of the census (i.e. the 2011 Census). This means an improvement in the coherence of the current statistics on births, deaths and migration for permanent stay with census results for persons in these ages.
6. A direct comparison of the results of the two censuses was also carried out on the basis of a demographic analysis of the oldest persons (aged 90 and over). This makes it possible to omit to a large extent the least documented component of demographic change, which is migration. For example, in 2020, the number of emigration for permanent stay amounted to only 12, and immigration to 13 for the persons aged 90 and over. The results of the analysis indicate a greater number of elderly persons in the population balance than in the retrospective estimate. The outcome of the 2021 Census made it possible to correct these values by several hundred persons for single ages.
7. The demographic analysis performed at the level of voivodships indicates a large diversity of discrepancies between the population balance and the population computed using retrospective estimation, which range from a few to several thousand persons. At the same time, as at the national level, the discrepancies are not correlated with sex.
8. In all voivodships where the 2021 Census recorded a larger population, there are major urban agglomerations (Wrocław, Kraków, Warszawa, Trójmiasto i Poznań). The remaining voivodships are characterized by the opposite situation i.e. the number from the population balance exceeded the number of population obtained on the basis of the 2021 Census.

VI. Ocena jakości wyników spisu w zakresie mieszkań

VI. Quality assessment of the Census 2021's results on housing

Kolejnym głównym celem realizacji spisów powszechnych jest monitorowanie wielkości oraz stanu zasobów mieszkaniowych pozostających do dyspozycji ludności na terenie kraju. Jakkolwiek, w ramach corocznych programów badań statycznych statystyki publicznej pozyskiwane są informacje na temat mieszkań, czy budynków z mieszkaniami, które umożliwiają obserwację zmian zachodzących w zasobach mieszkaniowych, to spisy powszechne jako badania obejmujące swoim zakresem całą badaną populację stanowią referencyjne źródło informacji wykorzystywane do aktualizacji danych opracowywanych w ramach bieżących badań statystyki publicznej.

W przypadku zagadnień związanych z tematyką mieszkań oraz budynków, w których są one zlokalizowane, informacje wynikowe ze spisu wykorzystywane są m.in. przy tworzeniu bilansów zasobów mieszkaniowych. Metodyka ustalania informacji dotyczących stanu liczebnego i wielkości mieszkań, czy też wyposażenia mieszkań w urządzenia i instalacje sanitarno-techniczne dla kolejnych lat sprawozdawczych zakłada, że na moment krytyczny spisu (tj. w przypadku NSP 2021 31 marca 2021 r.) przyjmowane są informacje wynikowe uzyskane w ramach spisów powszechnych i stanowią one jednocześnie punkt wyjścia do naliczania informacji według stanu na koniec kolejnych lat sprawozdawczych. Dane na temat zmian zachodzących w zasobach mieszkaniowych, tj. przyrostach i ubytkach mieszkań w okresach międzyspisowych, pochodzą natomiast z badań statystycznych statystyki publicznej. Należy także podkreślić, że spisy powszechne, w ramach których obserwacją objęte są zarówno mieszkania, jak osoby je zamieszkujące, a dane zbierane dla tych populacji odnoszą się do tego samego momentu referencyjnego, stanowią najbardziej dokładne źródło informacji umożliwiające monitorowanie warunków mieszkaniowych ludności.

Ze względu na kluczową rolę, jaką spisy powszechne odgrywają przy opracowywaniu informacji charakteryzujących zasoby mieszkaniowe oraz warunki mieszkaniowe ludności, wyniki te wymagają szczegółowej oceny ich jakości. Powinna ona uwzględniać zarówno poszczególne komponenty oceny jakości badań, jak i kolejne etapy realizacji badania spisowego, tj. począwszy od budowy wykazów jednostek statystycznych objętych badaniem spisowym (patrz rozdział pierwszy opracowania), poprzez przetwarzanie danych pochodzących z różnych źródeł informacji, procesy naliczania danych, aż po ocenę uzyskanych informacji wynikowych.

Biorąc pod uwagę obecny etap przetwarzania danych spisowych, ocena jakości uzyskanych wyników spisu prezentowana w ramach niniejszego rozdziału polega na analizie spójności wyników dotyczących bilansów zasobów mieszkaniowych opracowanych dla lat 2011–2020 według stanu na 31 grudnia każdego z ww. lat sprawozdawczych przy zastosowaniu jako punkt wyjściowy do przeliczeń wyników pochodzących z dwóch ostatnich edycji spisu, tj. NSP 2011 oraz NSP 2021. Ze względu na dostępność ostatecznych danych spisowych w zakresie tematyki mieszkań i budynków analizy dotyczą informacji na temat liczby mieszkań dla lat 2011–2020 dla Polski ogółem, a w przypadku 2020 r. również na poziomie województw.

6.1. Metodologia

6.1. Methodology

Tablica 35. Oznaczenia

Table 35. Indications

$L_M^{\overline{NSP_{2011}}}$ rok	Liczba mieszkań według stanu na koniec danego roku opracowywana corocznie przez GUS w ramach bilansu zasobów mieszkaniowych na podstawie wyniku NSP 2011
$L_M^{\overline{NSP_{2021}}}$ rok	Liczba mieszkań według stanu na koniec danego roku opracowana metodą szacunku wstecznego na podstawie wyniku NSP 2021.

Tablica 35. Oznaczenia (dok.)
Table 35. Indications (cont.)

R_{L_M}	Różnica między liczbą mieszkań $L_M^{\overleftarrow{NSP}_{2011}}$, a $L_M^{\overleftarrow{NSP}_{2021}}$.
$R_{L_M}^{\%}$	Różnica R_{L_M} wyrażona w % względem liczby mieszkań $L_M^{\overleftarrow{NSP}_{2011}}$.

W przypadku zasobów mieszkaniowych analiza spójności polega na porównaniu informacji na temat liczby mieszkań według stanu na 31 grudnia dla lat 2011–2020 opracowanych w ramach bilansu zasobów mieszkaniowych, w których jako punkt wyjścia wykorzystano wyniki NSP 2011, z szeregiem danych naliczonym przy zastosowaniu metody szacunku wstecznego bazującego na danych NSP 2021.

Analiza składa się z dwóch etapów. W pierwszej kolejności naliczono szereg danych dla lat 2011–2020 z wykorzystaniem metody szacunku wstecznego. W drugim etapie dokonano porównania danych uzyskanych w ramach szeregów czasowych naliczonych przy różnych podstawach spisu.

Metoda szacunku wstecznego polega na opracowaniu dla kolejnych lat 2011–2020 informacji odnośnie do liczby mieszkań przy wykorzystaniu jako punkt wyjścia informacji wynikowych pochodzących z NSP 2021. Procedura ta uwzględniła zjawiska zachodzące w zasobach mieszkaniowych, tj. przyrosty i ubytki mieszkań.

Do naliczeń wykorzystano dane na temat liczby mieszkań pochodzące z następujących źródeł informacji:

- ostateczne informacje wynikowe NSP 2021 według stanu na 31 marca 2021 r.,
- dla lat 2011–2020 mieszkania dodane oraz usunięte z NOBC w oparciu o informacje wykazane w ramach Wykazu budynków mieszkalnych i niemieszkalnych oraz obiektów zbiorowego zakwaterowania przekazanych do użytku przygotowywany dla aktualizacji Krajowego rejestru urzędowego podziału terytorialnego kraju (TERYT) dla lat 2011–2020 (badanie 1.47.04 Wydane pozwolenia na budowę i zgłoszenia z projektem budowlanym budowy obiektów budowlanych oraz efekty działalności budowlanej) oraz weryfikację adresów mieszkań przeprowadzoną przez gminy oraz urzędy statystyczne w ramach prac przygotowawczych do NSP 2021 – ustalone jako różnice pomiędzy stanami NOBC na koniec kolejnych lat sprawozdawczych,
- mieszkania oddane do użytku oraz zubytkowane w I kwartale 2021 r. i przyjęte w NOBC jako przyrost lub ubytek na podstawie danych zebranych w ramach formularza B-07 Sprawozdanie o budynkach mieszkalnych i mieszkaniach w budynkach niemieszkalnych oddanych do użytkowania oraz Wykazu budynków mieszkalnych i niemieszkalnych oraz obiektów zbiorowego zakwaterowania przekazanych do użytku przygotowywany dla aktualizacji Krajowego rejestru urzędowego podziału terytorialnego kraju (TERYT) (badanie 1.47.04 Wydane pozwolenia na budowę i zgłoszenia z projektem budowlanym budowy obiektów budowlanych oraz efekty działalności budowlanej).

Algorytm naliczania wstecznego szacunku przedstawiają poniższe równania:

$$L_M^{\overleftarrow{NSP}_{rok}} = \begin{cases} L_M^{\overleftarrow{NSP}_{2021}} - P^{L_M}_{I kw. 2021} + U^{L_M}_{I kw. 2021} & \text{dla } rok = 2020 \\ L_M^{\overleftarrow{NSP}_{rok+1}} - P^{L_M}_{rok+1} + U^{L_M}_{rok+1} & \text{dla } rok \in [2011; 2019] \end{cases}$$

gdzie:

- $L_M^{\overleftarrow{NSP}_{2011}}$ – liczba mieszkań na koniec danego roku opracowana metodą szacunku wstecznego,
 $L_M^{\overleftarrow{NSP}_{2021}}$ – liczba mieszkań na podstawie danych NSP2021,
 $P^{L_M}_{I kw. 2021}$ – przyrost liczby mieszkań w I kwartale 2021,
 $U^{L_M}_{I kw. 2021}$ – ubytek liczby mieszkań w I kwartale 2021,
 $P^{L_M}_{rok+1}$ – przyrost liczby mieszkań w roku t+1,

$U_{rok+1}^{L_M}$ – ubytek liczby mieszkań w roku t+1.

W kolejnym etapie dokonano porównań wyników uzyskanych metodą szacunku wstecznego z informacjami, które zostały opracowane przez GUS w ramach corocznych bilansów zasobów mieszkaniowych za lata 2011–2020 przy podstawie danych spisowych NSP 2011.

Do porównań danych na temat liczby mieszkań opracowanej na bazie wyników NSP 2011 z tymi samymi danymi uzyskanymi w ramach szacunku wstecznego (tj. opracowanych na bazie wyników NSP 2021) zastosowano następujące wzory:

$$R_{L_M_{rok}} = L_M_{rok}^{\overline{NSP2011}} - L_M_{rok}^{\overline{NSP2021}}$$

$$R_{L_M_{rok}}^{\%} = \frac{L_M_{rok}^{\overline{NSP2011}} - L_M_{rok}^{\overline{NSP2021}}}{L_M_{rok}^{\overline{NSP2011}}} \cdot 100\% = \frac{R_{L_M_{rok}}}{L_M_{rok}^{\overline{NSP2011}}} \cdot 100\%$$

gdzie:

$L_M_{rok}^{\overline{NSP2011}}$ – liczba mieszkań na koniec danego roku publikowana przez GUS w okresie międzyspisowym (tj. w latach 2011–2020) i szacowana w ramach corocznych bilansów zasobów mieszkaniowych na podstawie wyników spisu NSP 2011,

$L_M_{rok}^{\overline{NSP2021}}$ – liczba mieszkań na koniec danego roku w okresie międzyspisowym (tj. w latach 2011–2020) szacowana na podstawie wyników spisu NSP 2021 metodą szacunku wstecznego,

$R_{L_M_{rok}}$ – różnica między liczbami mieszkań opracowanymi na bazie wyników NSP2011 i NSP2021,

$R_{L_M_{rok}}^{\%}$ – różnica $R_{L_M_{rok}}$ wyrażona w wartościach procentowych w odniesieniu do tradycyjnie szacowanej liczby mieszkań w okresie międzyspisowym, czyli $L_M_{rok}^{\overline{NSP2011}}$.

Uzyskane różnice określają rozbieżności w danych na temat liczby mieszkań według stanu na 31 grudnia 2021 r. dla poszczególnych lat okresu 2011–2020 obliczonymi na bazie wyników NSP 2011 i NSP 2021. Dodatkowo wartości wskazują, że liczba mieszkań opracowana w ramach corocznych bilansów zasobów mieszkaniowych przy wykorzystaniu jako podstawę NSP 2011 jest wyższa niż ta opracowana metodą szacunku wstecznego. Z kolei, ujemne wartości wskaźnika świadczą o tym, że liczba mieszkań opracowana w ramach corocznych bilansów zasobów mieszkaniowych przy wykorzystaniu jako podstawy wyników NSP 2011 jest niższa niż ta opracowana metodą szacunku wstecznego.

6.2. Wyniki

6.2. Results

Ocena jakości wyników NSP 2021 w zakresie mieszkań koncentruje się na analizie spójności wyników w zakresie liczby mieszkań opracowanych przy wykorzystaniu wyników uzyskanych w ramach kolejnych spisów powszechnych, tj. NSP 2011 i NSP 2021. W pierwszym etapie porównane zostały informacje na temat liczby mieszkań dla Polski ogółem według stanu na 31 grudnia dla poszczególnych lat okresu międzyspisowego. Drugi etap porównań obejmuje natomiast analizę danych na temat liczby mieszkań na poziomie województw według stanu na 31 grudnia 2020 r.

Przyjmując za podstawę do naliczeń szacunku wstecznego wyniki NSP 2021 wyestymowano, że na koniec 2020 r. w Polsce zlokalizowanych było 15190,7 tys. mieszkań, tj. o ok. 1,2% więcej niż w populacji mieszkań ustalonych przy wykorzystaniu wyników NSP 2011.

Największą rozbieżność pomiędzy danymi odnotowano w 2019 r., w przypadku którego liczba mieszkań ustalona za pomocą metody szacunku wstecznego wyniosła 15153,0 tys. i była o prawie 2,3% wyższa niż ta opracowana w ramach bilansu zasobów mieszkaniowych przy podstawie NSP 2011. Należy zauważyć, że informacje opraco-

wane według stanu na 31 grudnia 2019 r. przy podstawie NSP 2021 uwzględniają wyniki przeprowadzonej na przetomie lat 2020 i 2021 weryfikacji adresów mieszkań przeprowadzonej przez urzędy gmin oraz jednostki TERYT w wojewódzkich urzędach statystycznych. Na skutek weryfikacji do NOBC dopisano bowiem ok. 215 tys. adresów mieszkań, a prawie 260 tys. zostało wykluczonych z zasobów mieszkaniowych. Zmiany te zostały uwzględnione w systemie NOBC według stanu na 31 grudnia 2020r. W kolejnych latach międzypisowych różnica pomiędzy danymi wahała się od 1,5% w 2018 r. do 2,1% w 2011 r. z tendencją rosnącą wraz ze zwiększającym się odstępem czasu od NSP 2021.

Tablica 36. Liczba mieszkań (stan na 31 grudnia) opracowana na bazie wyników NSP 2011 oraz metodą szacunku wstecznego przy podstawie NSP 2021 dla Polski ogółem dla lat 2011–2020

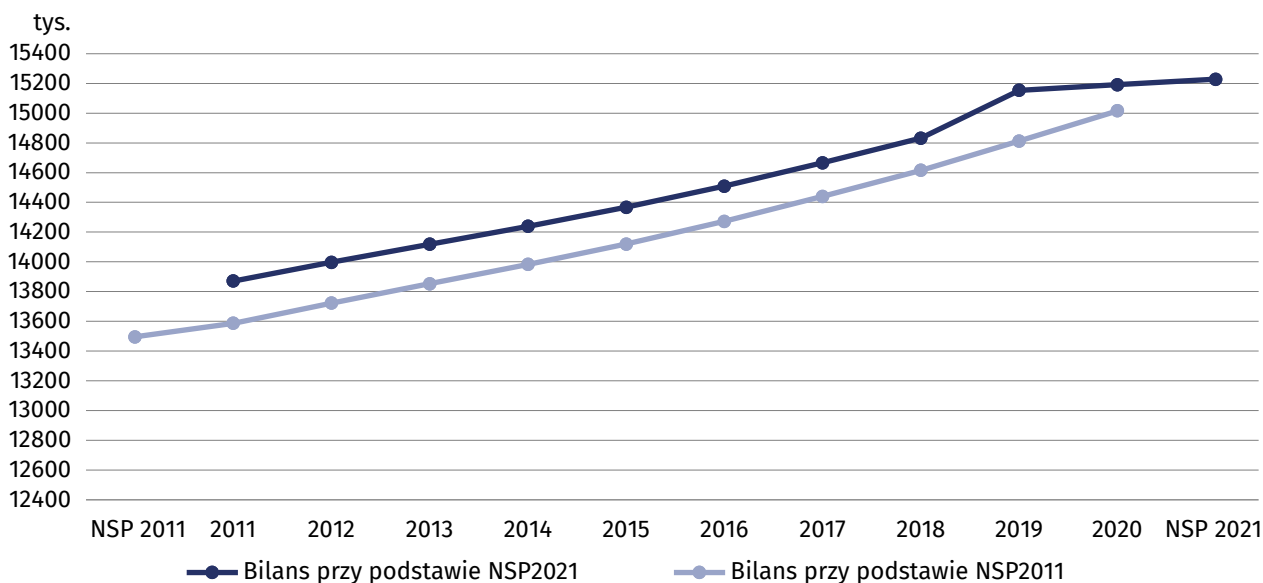
Table 36. Number of dwellings (as of 31 December) using the results of the 2011 Census and calculated using the retrospective estimation method on the basis of the 2021 Census in Poland for years 2011–2021

Rok Year	Liczba mieszkań na podstawie szacunku wstecznego, w tys. Number of dwellings from the retrospective estimation, in thousands	Liczba mieszkań na podstawie bilansu zasobów mieszkaniowych, w tys. Number of dwellings based on balance of housing stocks, in thousands	Różnica, w tys. Difference, in thousands	Różnica, w % Difference, in %
2011	13871,8	13587,4	-284,3	-2,093
2012	13997,6	13722,8	-274,8	-2,002
2013	14118,2	13852,9	-265,3	-1,915
2014	14239,3	13983,0	-256,3	-1,833
2015	14366,7	14119,5	-247,3	-1,751
2016	14508,8	14272,0	-236,8	-1,659
2017	14666,6	14439,8	-226,8	-1,571
2018	14831,7	14615,1	-216,6	-1,482
2019	15153,0	14812,8	-340,3	-2,297
2020	15190,7	15015,3	-175,4	-1,168

Doświadczenia uzyskane w ramach NSP 2021, w tym w szczególności te związane z weryfikacją adresów mieszkań przeprowadzoną w ramach prac przygotowawczych do spisu, wskazują, że rozbieżności pomiędzy informacjami dotyczącymi liczby mieszkań opracowanymi w oparciu o kolejne edycje spisów stanowią odzwierciedlenie różnic występujących pomiędzy stanem prawnym, a stanem faktycznym istniejącym w terenie. Dla lat międzypisowych informacje na temat bilansów zasobów mieszkaniowych ustalane są w oparciu o dane dotyczące mieszkań oddawanych do użytku bądź ubytkowanych, które przekazywane są statystyce publicznej przez powiatowych inspektorów nadzoru budowlanego na podstawie zgłoszeń inwestorów, czy właścicieli nieruchomości. Należy jednak uwzględnić, że nie wszystkie budynki mieszkalne, pomimo ich zamieszkiwania, są formalnie oddawane do użytku, czy też budynki nienadające się do zamieszkiwania oznaczane jako wykluczone z zasobów mieszkaniowych. Informacje na ten temat są natomiast weryfikowane podczas spisów powszechnych, które z założenia obejmują swoim zakresem całą populację mieszkań i umożliwiają jednoczesne pozyskanie informacji o adresach zamieszkiwania ludności. Takie podejście do badania zasobów mieszkaniowych pozwala na ustalenie informacji jak najbardziej zbliżonych do rzeczywistości.

Wykres 23. Porównanie liczby mieszkań (stan na 31 grudnia) otrzymanej na bazie wyników NSP 2011 z liczbą mieszkań obliczoną metodą szacunku wstecznego przy podstawie NSP 2021

Chart 23. Comparison of the number of dwellings (as of 31 December) using the results of the 2011 Census with the number of dwellings calculated using the retrospective estimation method on the basis of the 2021 Census



Podczas gdy na poziomie kraju różnica pomiędzy liczbą mieszkań według stanu na 31 grudnia 2020 r. ustaloną w oparciu o wyniki NSP 2021 była wyższa o prawie 1,2% od wyniku opracowanego przy podstawie danych NSP 2011, to w przekroju wojewódzkim wskaźnik ten charakteryzował się zróżnicowaniem i to zarówno co do jego wartości, jak i kierunku zmian. Liczba mieszkań zgodnie z szacunkiem wstecznym wyniosła wtedy 15190,7 tys. wobec 15015,3 tys. mieszkań jako zasoby opracowane przy podstawie NSP 2011.

Tablica 37. Liczba mieszkań (stan na 31 grudnia 2020 r.) opracowana na bazie wyników NSP 2011 oraz metodą szacunku wstecznego przy podstawie NSP 2021 według województw

Table 37. Number of dwellings (as of 31 December 2020) obtained on the basis of the 2011 Census and calculated using the retrospective estimation method based on the 2021 Census at the voivodship level

Województwo Voivodship	Liczba mieszkań na podstawie szacunku wstecznego, w tys. Number of dwellings from the retrospective estimation, in thousands	Liczba mieszkań na podstawie bilansu zasobów mieszkaniowych, w tys. Number of dwellings based on balance of housing stocks, in thousands	Różnica, w tys. Difference, in thousands	Różnica, w % Difference, in %
Polska ogółem	15015,3	15190,7	-175,4	-1,168
Dolnośląskie	1228,4	1256,8	-28,5	-2,317
Kujawsko-pomorskie	768,3	776,7	-8,4	-1,096
Lubelskie	793,8	786,5	7,3	0,915
Lubuskie	385,9	390,8	-4,8	-1,254

Tablica 37. Liczba mieszkań (stan na 31 grudnia 2020 r.) opracowana na bazie wyników NSP 2011 oraz metodą szacunku wstecznego przy podstawie NSP 2021 według województw (dok.)

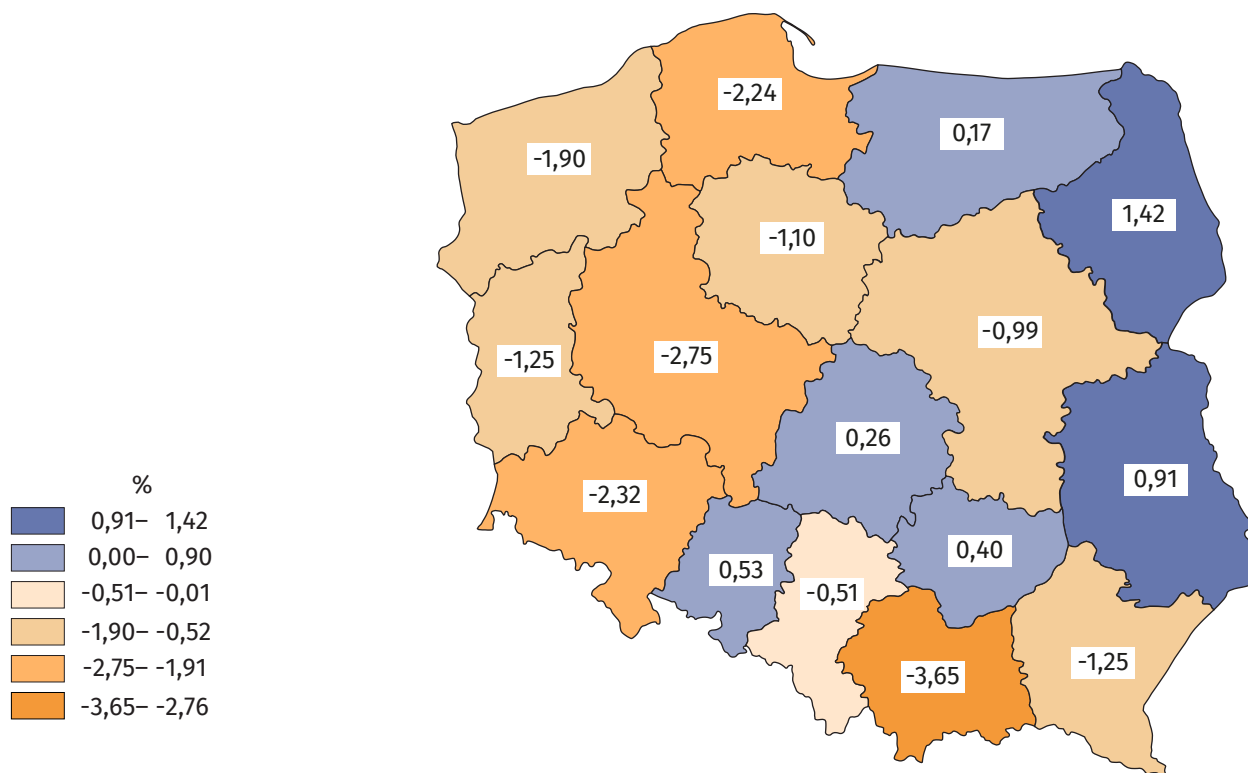
Table 37. Number of dwellings (as of 31 December 2020) obtained on the basis of the 2021 Census and calculated using the retrospective estimation method based on the 2021 Census at the voivodship level (cont.)

Województwo Voivodship	Liczba mieszkań na podstawie szacunku wstecznego, w tys. Number of dwellings from the retrospective estimation, in thousands	Liczba mieszkań na podstawie bilansu zasobów mieszkaniowych, w tys. Number of dwellings based on balance of housing stocks, in thousands	Różnica, w tys. Difference, in thousands	Różnica, w % Difference, in %
Łódzkie	1042,9	1040,1	2,7	0,260
Małopolskie	1239,0	1284,2	-45,2	-3,648
Mazowieckie	2388,2	2411,8	-23,7	-0,991
Opolskie	362,3	360,4	1,9	0,527
Podkarpackie	687,4	696,0	-8,6	-1,249
Podlaskie	466,4	459,8	6,6	1,421
Pomorskie	914,1	934,6	-20,5	-2,243
Śląskie	1803,8	1812,9	-9,1	-0,506
Świętokrzyskie	454,1	452,3	1,8	0,404
Warmińsko-mazurskie	531,4	530,5	0,9	0,167
Wielkopolskie	1272,3	1307,3	-35,0	-2,749
Zachodniopomorskie	677,1	690,0	-12,8	-1,897

W przeważającej liczbie województw liczba mieszkań według stanu na 31 grudnia 2020 r. opracowana z wykorzystaniem metody bilansu wstecznego w oparciu o wyniki NSP 2021 była, podobnie jak w przypadku Polski ogółem, wyższa niż ustalona w ramach bilansu zasobów mieszkaniowych sporządzanych w oparciu o wyniki NSP 2011.

Mapa 2. Różnica procentowa między liczbą mieszkań opracowaną na podstawie NSP 2011, a szacunkową liczbą mieszkań opracowaną przy podstawie NSP 2021 według województw według stanu na 31 grudnia 2020 r.

Map 2. Relative difference between the number of dwellings based on the 2011 Census and the estimated number of dwellings based on the 2021 Census by voivodships as of 31 December 2020



Najmniejsze rozbieżności pomiędzy liczbą mieszkań na koniec 2020 r. zidentyfikowano w województwach warmińsko-mazurskim (0,17%), łódzkim (0,26%), czy świętokrzyskim (0,40%) i jednocześnie były to województwa, w przypadku których liczba mieszkań ustalona w oparciu o NSP 2021 była nieco niższa niż ta opracowana przy wykorzystaniu NSP 2011.

Największe różnice pomiędzy danymi na temat liczby mieszkań w odniesieniu do informacji udostępnionych w ramach bilansu zasobów mieszkaniowych przy podstawie NSP 2011 odnotowano natomiast w województwie małopolskim (-3,65%), wielkopolskim (-2,75%), dolnośląskim (-3,32%), czy pomorskim (-2,24%). Liczba mieszkań w tych województwach ustalona przy podstawie NSP 2021 była wyższa niż ta naliczona w ramach bilansów zasobów mieszkaniowych wykorzystujących jako punkt wyjścia wyniki NSP 2011.

6.3. Podsumowanie

1. W celu oceny jakości spisu w zakresie mieszkań dokonano porównania informacji na temat liczby mieszkań opracowanych dla lat 2011–2020 według stanu na 31 grudnia danego roku w ramach corocznych bilansów zasobów mieszkaniowych z wykorzystaniem jako podstawę wyniki NSP 2011 z informacjami naliczonymi w ramach tzw. bilansu wstecznego, dla którego punktem wyjścia są wyniki NSP 2021.
2. Różnice występujące pomiędzy danymi dotyczącymi liczby mieszkań ustalonymi w oparciu o wyniki NSP 2011 i NSP 2021 odzwierciedlają rozbieżności odnotowywane pomiędzy stanem formalnym a stanem faktycznym

zjawisk identyfikowanych w zasobach mieszkaniowych, tj. oddawania mieszkań do użytku oraz ich wykluczenia z zasobów mieszkaniowych.

3. W latach 2011–2020 liczba mieszkań zlokalizowanych na terenie Polski opracowana na podstawie wyników NSP 2021 była przeciętnie o 1,62% wyższa niż ta ustalona w oparciu o NSP 2011. Przekłada się to na ok. 252 tys. mieszkań. W przypadku 2020 r. rozbieżność w liczbie mieszkań wyniosła 175 tys., co oznacza, że po uwzględnieniu wyników NSP 2021 liczba mieszkań była o ok. 1,2% większa niż w populacji mieszkań ustalonych przy wykorzystaniu wyników NSP 2011. Wraz ze zwiększającą się rozpiętością czasu pomiędzy spisami różnica w liczbie mieszkań była coraz wyższa.
4. Różnice w liczbie mieszkań uzyskanych w ramach przygotowywania corocznych bilansów zasobów mieszkaniowych przy podstawie NSP 2011 oraz bilansu wstecznego przy podstawie NSP 2021 były zróżnicowane terytorialnie zarówno pod względem znaku, jak i ich wielkości.
5. Według stanu na 31 grudnia 2020 r. w przeważającej liczbie województw, podobnie jak w przypadku Polski ogółem, liczba mieszkań ustalona w ramach bilansu wstecznego przy podstawie NSP 2021 była wyższa niż wyniki uzyskane w ramach bilansu zasobów mieszkaniowych w oparciu o NSP 2011. Odwrotną zależność odnotowano natomiast w województwach: lubelskim, łódzkim, opolskim, podlaskim, świętokrzyskim i warmińsko-mazurskim.

6.3. Summary

1. In order to assess the quality of the housing census, information on the number of dwellings compiled for the years 2011–2020 as part of the annual housing stock balances, based on the results of the 2011 Census, was compared with information calculated under the retrospective estimation, for which the starting point is the results of the 2021 Census.
2. The differences between the data on the number of dwellings determined on the basis of the results of the 2011 Census and the 2021 Census reflect discrepancies recorded between the formal and actual status of the phenomena identified in the housing stocks, i.e. the completed dwellings and the exclusion of dwellings from the housing stocks.
3. In the years 2011–2020, the number of dwellings located in Poland based on the results of the 2021 Census was on average 1.62% higher per year than that determined on the basis of the 2011 Census. This concerns approx. 252 thousand dwellings. In the case of 2020, the discrepancy in the number of dwellings amounted to 175,000, which means that after taking into account the results of the 2021 Census, the number of dwellings was approx. 1.2% higher than in the population of dwellings determined using the results of the 2011 Census. The greater time span between the censuses, the higher difference in the number of dwellings was recorded.
4. Differences in the number of dwellings obtained as part of the preparation of annual housing stock balances based on the 2011 Census as well as the retrospective estimation based on the 2021 Census were territorially differentiated both in terms of the sign as well as their size.
5. As of 31 December, 2020 in the majority of voivodships, as in the case of Poland in general, the number of dwellings compiled under the retrospective estimation based on the 2021 Census was higher than the results obtained from balance of housing stocks based on the 2011 Census. An inverse relationship was recorded in the following voivodships: lubelskie, łódzkie, opolskie, podlaskie, świętokrzyskie i warmińsko-mazurskie.

Bibliografia

Bibliography

- Baffour, B., Valente, P. (2012). An evaluation of census quality, *Statistical Journal of the IAOS*, Vol. 28, pp. 121–135. https://www.researchgate.net/publication/281286440_An_evaluation_of_census_quality
- Bank Danych Lokalnych GUS, <https://bdl.stat.gov.pl/bdl/start>
- Baza Demografia GUS, <https://demografia.stat.gov.pl/bazademografia/Tables.aspx>
- Bethlehem, J. (2009). The rise of survey sampling.
- Beręsewicz, M., Gołata, E. (2019). Przyszłość badań statystycznych wobec potrzeb społecznych i gospodarczych na przykładzie spisu ludności. W: Gorynia, M. (red.), *Ewolucja nauk ekonomicznych: jedność a różnorodność, relacje do innych nauk, problemy klasyfikacyjne* (s. 173-194). Warszawa: PAN.
- Bracha, Cz. (1996). *Teoretyczne podstawy metody reprezentacyjnej*, Warszawa: PWN.
- Brown, J. J., Beaujouan, E. (2013). Review of the Rolling Census Approach: and other survey-based options.
- Chipperfield, J., Preston, J. (2007). Efficient bootstrap for business surveys, *Survey Methodology* (2007), Vol. 33, No. 2, pp. 167–172.
- D'Orazio, M., Di Zio, M., Scanu, M. (2006). *Statistical matching: Theory and practice*. John Wiley & Sons.
- Dygaszewicz, J. (2009). Spisy powszechne XXI wieku. *Wiadomości Statystyczne. The Polish Statistician*, 54(06), s. 13–27.
- Eurostat, Working Group, (2003) *Methodological documents – definition of quality in statistics*, <https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf/> (dostęp: 15 marca 2023)
- Gołata, E. (2018). *Koniec ery tradycyjnych spisów ludności*, Poznań: Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Hung-Mo Lin, Hae-Young Kim, Williamson John M. and Virginia M. Lesser (2012). Estimating agreement coefficients from sample survey data, *Survey Methodology*, Vol. 38, No. 1, pp. 63–72.
- Kordos, J. (1987), *Dokładność danych w badaniach społecznych*, Biblioteka *Wiadomości Statystycznych*, t. 35, Warszawa: GUS.
- Kordos, J. (2007). Some Aspects of Post-Enumeration Surveys in Poland. *Statistics in Transition – new series*, Vol. 8(3), pp. 563–576.
- Krywult-Albańska, M. (2013). Powszechne spisy ludności na przykładzie wybranych państw. *Aspekty metodologiczne. Politeja – Pismo Wydziału Studiów Międzynarodowych i Politycznych Uniwersytetu Jagiellońskiego*, 10(26), s. 275–290.
- Nordholt, E. S., Netherlands, S. (2015). The Dutch census 2011. *Statistika*, Vol. 95(1), pp. 86–92.
- Paradysz, J. (2010) *Konieczność estymacji pośredniej w spisach powszechnych*, Poznań: Uniwersytet Ekonomiczny w Poznaniu.
- Paradysz, J. (2009). Spisy jako źródło informacji o warunkach życia ludności w Polsce. *Wiadomości Statystyczne. The Polish Statistician*, 54(07), s. 1–9.

Roszka, W. (2013). *Statystyczna integracja danych w badaniach społeczno-ekonomicznych*. Poznań: Uniwersytet Ekonomiczny w Poznaniu.

Schenkel, J. F., Zhang, L. C. (2022). Adjusting misclassification using a second classifier with an external validation sample. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Shepherd, B. E., Han, K., Chen, T., Bian, A., Pugh, S., Duda, S. N., Lumley, T., Heerman, W.J., Shaw, P. A., (2022). Multi-wave validation sampling for error-prone electronic health records. *Biometrics*.

Szabłowski, P. J., Wesołowski, J., Wieczorkowski, R., (1996). Indeks zgodności jako miara jakości danych (na podstawie wyników spisu kontrolnego do Mikrospisu 1995). *Wiadomości Statystyczne*, nr 4, s. 43–49.

Szreder, M. (2010). *Metody i techniki sondażowych badań opinii*. Polskie Wydawnictwo Ekonomiczne.

United Nations. (2017). Principles and recommendations for population and housing censuses https://unstats.un.org/unsd/publication/seriesM/Series_M67Rev3en.pdf

Valente, P. (2010). Census taking in Europe: how are populations counted in 2010?. *Population Societies*, Vol. 467(5), pp. 1–4.

Rozporządzenie (WE) Parlamentu Europejskiego i Rady nr 763/2008 z dnia 9 lipca 2008 r. w sprawie spisów powszechnych ludności i mieszkań (Dz. Urz. UE L 218 z 13 sierpnia 2008), <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32008R0763&from=IT>

Rozporządzenie wykonawcze Komisji (UE) 2017/881 z dnia 23 maja 2017 r. w sprawie wykonania rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 763/2008 w sprawie spisów powszechnych ludności i mieszkań w odniesieniu do ustaleń dotyczących raportów jakości i ich struktury oraz formatu technicznego przekazywania danych zmieniające rozporządzenie (UE) nr 1151/2010, <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A32017R0881>