

Rozprawa doktorska p.t.:

***„Metody statystyczne w poszukiwaniu czynników ryzyka  
występowania wrodzonych wad rozwojowych  
na podstawie danych z Polskiego Rejestru  
Wrodzonych Wad Rozwojowych”***

**Barbara Więckowska**

Katedra i Zakład Informatyki i Statystyki  
Uniwersytet Medyczny w Poznaniu

Poznań, 2011

## **Podziękowania**

*Składam szczególne podziękowania mojemu promotorowi, prof. dr hab. Jerzemu Moczko za wprowadzenie w świat nauki, cenne wskazówki i pomoc w czasie doktoratu.*

*Dziękuję również Przewodniczącej Centralnego Zespołu ds. Polskiego Rejestru Wrodzonych Wad Rozwojowych prof. dr hab. med. Annie Latos-Bieleńskiej, za możliwość współpracy z zespołem Polskiego Rejestru Wrodzonych Wad Rozwojowych, czego efektem jest niniejsza rozprawa doktorska.*

*Serdecznie dziękuję dr Annie Materna-Kiryłuk za wszelką pomoc, cenne wskazówki, wsparcie i dobrą radę.*

*Dziękuję także Radzie Wydziału Lekarskiego II UMP za przeprowadzenie mojego przewodu doktorskiego.*

*Dziękuję również wszystkim pracownikom Zakładu i Katedry Informatyki i Statystyki za wspólnie przeżyte chwile, miłą atmosferę i wsparcie w trudnych chwilach.*

*Pragnę gorąco podziękować mojej Rodzinie, szczególnie Mamie i Tacie za nieustanne wsparcie. Tomaszowi dziękuję za cierpliwość i wiarę w to, że się uda.*

*Projekt został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2011/01/N/NZ7/02689*

## Streszczenie

W pracy porównano różne metody analizy statystycznej służące badaniu wpływu czynników ryzyka na częstość występowania rozszczepu wargi z lub bez rozszczepu podniebienia u dzieci. Oddzielnie analizowano wpływ czynników osobniczych takich jak: kolejności ciąży, liczby poronień samoistnych i oddzielnie wpływ czynników zewnętrznych takich jak: miejsce zamieszkania i czasu porodu.

Dla analizy czynników osobniczych zbudowano różne modele regresji logistycznej. Najefektywniejsze okazały się modele regresji logistycznej, w których w analizie dotyczącej kolejności ciąży występuje wiek matki a w analizie dotyczącej poronień samoistnych występuje kolejność ciąży. Uzyskane wyniki pozwoliły wykazać, że zależność pomiędzy częstością wystąpienia badanej wady a kolejnością ciąży jest wprost proporcjonalna natomiast pomiędzy częstością wystąpienia wady a liczbą wcześniejszych poronień samoistnych odwrotnie proporcjonalna.

Poszukiwanie geograficznych klasterów przestrzennych i czasowo-przestrzennych nie przyniosło spodziewanego efektu. Zastosowane w analizie czynników zewnętrznych: statystyka skanująca Kulldorffa jak i metoda LISA nie przyniosły rezultatu w postaci zidentyfikowanych, istotnych statystycznie klasterów zwiększonej częstości występowania badanej wady wrodzonej. Uzyskane wyniki pozwoliły jednakże na porównanie efektywności tych dwóch metod i wyboru tej lepiej obrazującej postawiony problem medyczny. Tą statystyką okazała się statystyka skanująca Kulldorffa.

## **Abstract**

The aim of this paper was to compare different methods of statistical analyses which are used in the examination of the risks factors that influence the occurrence of left lips with/without cleft palate among children. The influence of individual factors was carried out separately for: the sequence of pregnancy, the number of miscarriages and separately for external factors such as: place of residence, time of labour.

Different logistic regression models were designed to evaluate the influence of internal factors. As a result, the most effective models turned out to be those in which: the sequence of pregnancy analysis contains mother's age and the miscarriages analysis contains the sequence of pregnancy. The obtained results demonstrate that the relationship between the occurrence of the examined defect and the sequence of pregnancy is directly proportional. However, the relationship between the occurrence of the examined defect and the number of previous miscarriages is inversely proportional.

The research of spatial and space-time geographic clusters did not bring the desired effect. The external factors used in the analysis (Kulldorff scanning statistic and LISA method) have not been identified as statistically significant clusters of increased incidence of congenital defects. The obtained results enabled the comparison of the effectiveness of these two methods and the choice of the better one describing the analyzed medical problem (Kulldorff scanning statistics).

## SPIS TREŚCI

<b>1</b>	<b>SPIS OZNACZEŃ I SYMBOLI .....</b>	<b>7</b>
<b>2</b>	<b>WSTĘP.....</b>	<b>8</b>
<b>3</b>	<b>CEL I FORMA ORGANIZACJI PRACY .....</b>	<b>11</b>
<b>4</b>	<b>ISTNIEJĄCY STAN WIEDZY W ZAKRESIE TEMATU BADAŃ .....</b>	<b>13</b>
<b>5</b>	<b>SPOSOBY POZYSKIWANIA DANYCH .....</b>	<b>16</b>
5.1	ZBIERANIE GRUPY BADANEJ.....	16
5.2	DANE GUS DOTYCZĄCE URODZEŃ ŻYWYCH W POPULACJI .....	17
5.3	ZBIERANIE GRUPY KONTROLNEJ.....	17
5.4	MAPY PODZIAŁU ADMINISTRACYJNEGO.....	17
<b>6</b>	<b>ZASTOSOWANE OPROGRAMOWANIE .....</b>	<b>18</b>
<b>7</b>	<b>METODY ANALIZY HISTORII RODNOŚCI .....</b>	<b>19</b>
7.1	DOBÓR GRUPY BADANEJ I GRUPY KONTROLNEJ .....	19
7.1.1	<i>Grupa badana w analizie historii rodności.....</i>	<i>19</i>
7.1.2	<i>Informacje o populacji badanej.....</i>	<i>19</i>
7.1.3	<i>Grupa kontrolna w analizie historii rodności.....</i>	<i>19</i>
7.2	METODY JEDNOWYMIAROWE CZYNNIKÓW ZWIĄZANYCH Z HISTORIĄ RODNOŚCI .....	23
7.2.1	<i>Analiza danych ilościowych i jakościowych o uporządkowanych kategoriach.....</i>	<i>23</i>
7.2.2	<i>Analiza danych dychotomicznych.....</i>	<i>27</i>
7.3	METODY WIELOWYMIAROWE CZYNNIKÓW ZWIĄZANYCH Z HISTORIĄ RODNOŚCI – REGRESJA LOGISTYCZNA. 30	
7.3.1	<i>Założenia analizy regresji logistycznej i warunkowej regresji logistycznej .....</i>	<i>30</i>
7.3.2	<i>Regresja logistyczna i warunkowa regresja logistyczna w analizie historii rodności</i>	<i>34</i>
7.4	WNIOSKI DOTYCZĄCE ANALIZY HISTORII RODNOŚCI.....	46
<b>8</b>	<b>METODY ANALIZY CZYNNIKÓW CZASOWO-PRZESTRZENNYCH .....</b>	<b>47</b>
8.1	DOBÓR GRUPY BADANEJ DLA ZDEFINIOWANEJ POPULACJI .....	49
8.1.1	<i>Grupa badana w analizie czynników czasowo-przestrzennych.....</i>	<i>49</i>
8.1.2	<i>Informacje o populacji badanej.....</i>	<i>49</i>
8.2	DANE PRZESTRZENNE I WIZUALIZACJA.....	50

8.2.1	<i>System informacji geograficznej</i> .....	50
8.2.2	<i>Dane przestrzenne</i> .....	50
8.2.3	<i>Dane opisowe</i> .....	53
8.2.4	<i>Statystyki opisowe</i> .....	53
8.2.5	<i>Sposoby określenia sąsiedztwa</i> .....	58
8.2.6	<i>Autokorelacja przestrzenna</i> .....	62
8.3	<b>PRZEGLĄD METOD BADANIA KLASTERÓW</b> .....	63
8.4	<b>ZAŁOŻENIA TEORII STATYSTYKI GLOBALNEJ MORANA ORAZ WYNIKI ZASTOSOWANIA TEJ METODY</b> .....	65
8.5	<b>LOKALNA ESTYMACJA I WYSZUKIWANIE KLASTERÓW</b> .....	68
8.5.1	<i>Założenia teorii Lokalnych Współczynników Autokorelacji Przestrzennej (LISA)</i> .....	68
8.5.2	<i>Wyniki zastosowania teorii Lokalnych Współczynników Autokorelacji Przestrzennej (LISA)</i> .....	68
8.5.3	<i>Założenia teorii statystyki skanującej Kulldorff'a</i> .....	71
8.5.4	<i>Wyniki zastosowania teorii statystyki skanującej Kulldorff'a</i> .....	74
8.5.5	<i>Porównanie wyników uzyskanych przy pomocy metody LISA i metody Kulldorff'a</i> ..	76
8.6	<b>WNIOSKI DOTYCZĄCE ANALIZY CZYNNIKÓW CZASOWO-PRZESTRZENNYCH</b> .....	85
<b>9</b>	<b>DYSKUSJA</b> .....	<b>86</b>
<b>10</b>	<b>BIBLIOGRAFIA</b> .....	<b>91</b>
<b>11</b>	<b>ZAŁĄCZNIK 1</b> .....	<b>98</b>
<b>12</b>	<b>ZAŁĄCZNIK 2</b> .....	<b>100</b>
<b>13</b>	<b>ZAŁĄCZNIK 3</b> .....	<b>101</b>
<b>14</b>	<b>ZAŁĄCZNIK 4</b> .....	<b>110</b>
<b>15</b>	<b>ZAŁĄCZNIK 5</b> .....	<b>119</b>
<b>16</b>	<b>ZAŁĄCZNIK 6</b> .....	<b>123</b>

---

# 1 SPIS OZNACZEŃ I SYMBOLI

---

CL±P – rozszczep wargi z/bez rozszczepu podniebienia

LR – regresja logistyczna

CLR – warunkowa regresja logistyczna

GIS – geograficzny system informacji

Teryt – kod terytorialny podziału administracyjnego kraju

---

## 2 WSTĘP

---

Analiza środowiskowego uwarunkowania zdrowia populacji to jedna z gałęzi dziedziny nauki, jaką jest epidemiologia. Jednym z jej zasadniczych celów jest poszukiwanie etiologicznych czynników ryzyka rozprzestrzeniania się chorób i inwalidztwa ludzkiego.

Wrodzone wady rozwojowe występują u 2%-4% noworodków i stanowią rosnący problem społeczny i medyczny. Pojęcie wady wrodzonej ([54] Latos-Bieleńska A., Materna-Kiryłuk A. i inni, 1998) jest bardzo szerokie, obejmuje wszelkie zewnętrzne lub wewnętrzne morfologiczne odchylenia od prawidłowego rozwoju niezależnie od etiologii, patogenez, czasu powstania, momentu ustalenia i rozpoznania. Spośród wrodzonych wad rozwojowych o ustalonej etiologii ok. 35% to wady wywołane wyłącznie poprzez czynniki genetyczne, ok. 6.5% przez aberracje chromosomowe a 7,5% wad wrodzonych jest wynikiem mutacji jednogenowych. Około 50% wrodzonych wad rozwojowych o określonej etiologii jest spowodowanych zarówno przez czynniki genetyczne jak i często trudne do ustalenia teratogeny środowiskowe ([52] Latos-Bieleńska A. i Materna-Kiryłuk A., 2006).

Wady wrodzone są ważną przyczyną niepowodzeń rozrodu, zgonów noworodków i niemowląt oraz niepełnosprawności fizycznej lub intelektualnej, która w przypadku wad rozwojowych u dzieci żywo urodzonych, nierzadko może trwać całe życie. Tworzone rejestry wad wrodzonych na całym świecie podejmują działania zmierzające do zmniejszenia społecznych skutków wrodzonych wad rozwojowych. Podjęte przez Światową Organizację Zdrowia (WHO) w 1972 r. starania, by poszczególne rejestry krajowe działały na podobny zasadach tak, by uzyskane dane można było porównać, zaowocowały w 1974 r. ideą powstania rejestru obejmującego kraje Unii Europejskiej. Rezultatem było utworzenie w 1979 r. EUROCAT ([19] Dolk H., 2005). W roku 1997 powstał Polski Rejestr Wrodzonych Wad Rozwojowych (PRWWR) nie jako odwzorowanie innych rejestrów wad, ale ze względu na zastosowane rozwiązania organizacyjne oraz liczne i ważne cele praktyczne jako oryginalne polskie osiągnięcie ([53] Latos-Bieleńska A. i Materna-Kiryłuk A., 2005). W 2001 r. rejestr polski został włączony do EUROCAT stając się częścią systemu monitorowania wad wrodzonych w Europie i na świecie. PRWWR od 2007 r. obejmuje teren całego kraju i monitoruje wszystkie urodzenia w Polsce. W chwili obecnej jest największym rejestrem EUROCAT.

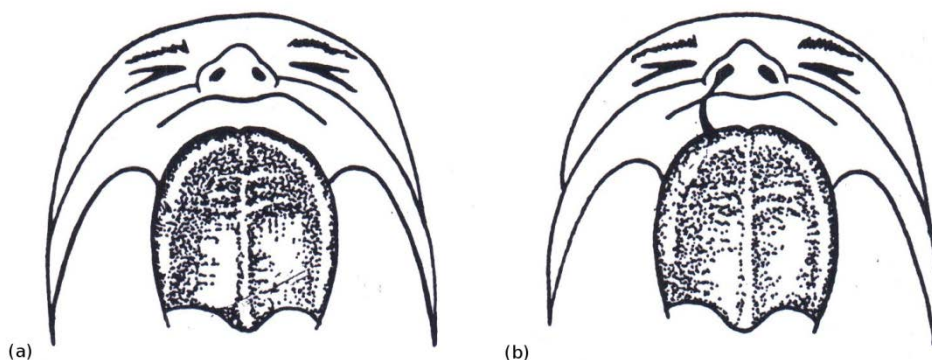
Do PRWWR zgłaszane są wady rozwojowe dzieci przed ukończeniem drugiego roku życia



a także wady rozpoznane prenatalnie i stwierdzone u dzieci martwo urodzonych.

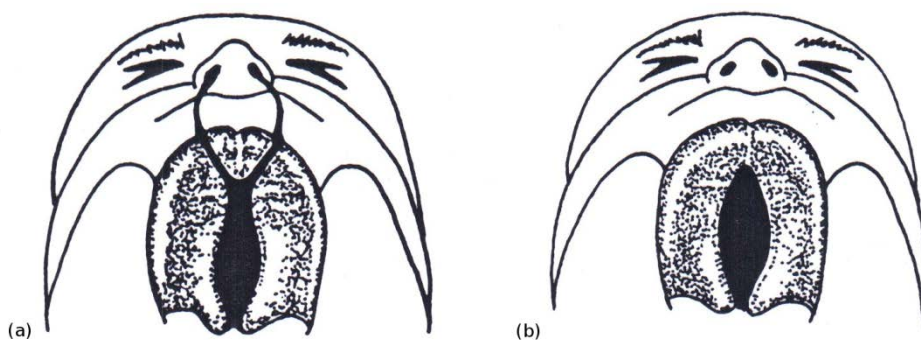
W pracy pod kątem wpływu kolejności ciąży, liczby poronień samoistnych matki oraz jej miejsca zamieszkania a także wielu innych czynników ryzyka występowania wrodzonych wad rozwojowych analizowano **izolowany rozszczep wargi z/bez rozszczepu podniebienia – CL±P** (ang. Cleft Lip with/without cleft palate)

Rozszczep wargi z towarzyszącym mu (lub nie) rozszczepem podniebienia należy do najpoważniejszych, a jednocześnie najczęściej występujących wrodzonych wad rozwojowych części twarzowej czaszki. W Polsce izolowany rozszczep wargi występuje ze średnią częstością 3.1 przypadków na 10 000 żywych urodzeń a rozszczep wargi i podniebienia 5.9 przypadków na 10 000 żywych urodzeń.



Rycina 1 a) brak rozszczepu, b) jednostronny rozszczep wargi

Źródło: Oxford Desk Reference - Clinical Genetics, H.V. Firth, J.A. Hurst, J.G. Hall, 2005



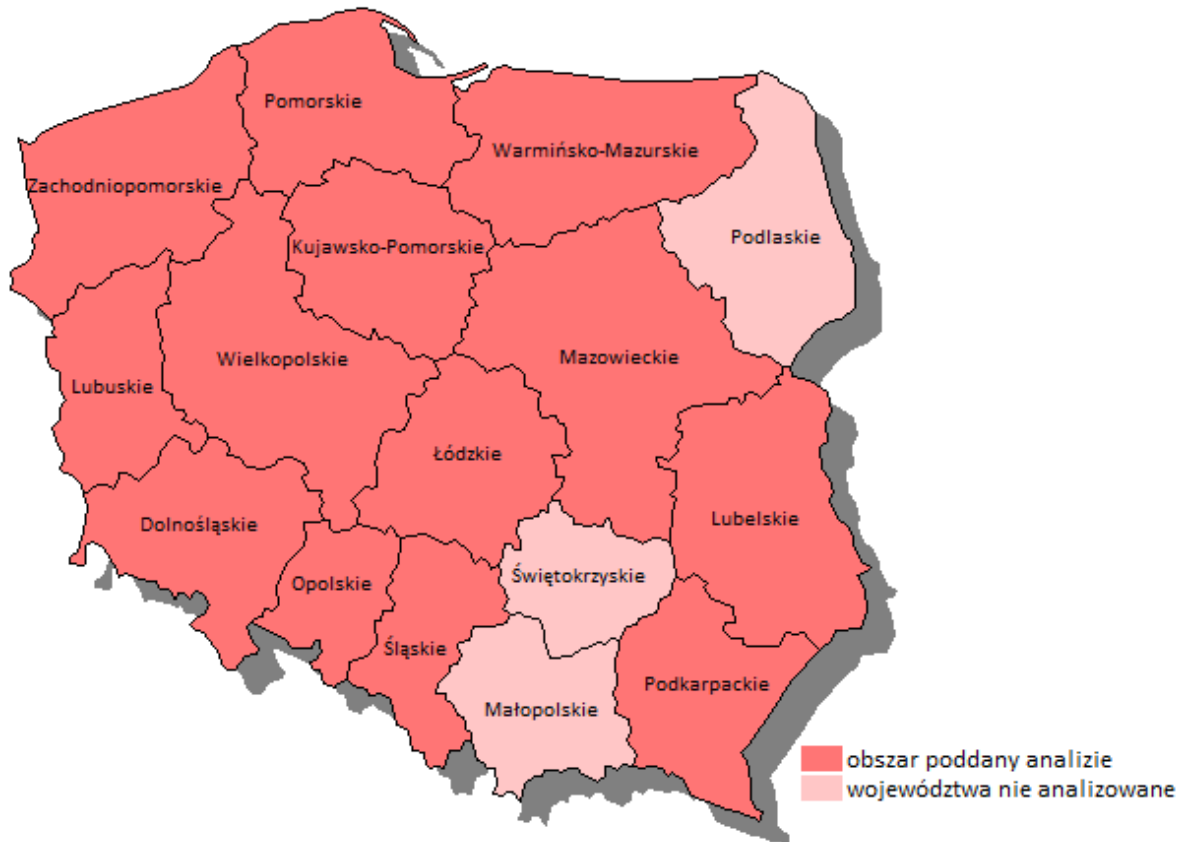
Rycina 2 a) obustronny rozszczep wargi i podniebienia, b) rozszczep podniebienia

Źródło: Oxford Desk Reference - Clinical Genetics, H.V. Firth, J.A. Hurst, J.G. Hall, 2005

Wada ta jest w pełni rozpoznawalna i zgłaszalna a duży udział w jej powstawaniu mają zarówno czynniki genetyczne, jak i znajdujące się w kręgu naszych zainteresowań – czynniki środowiskowe.

Dane do analizy zebrane zostały z całego terenu PRWWR obejmującego w latach 2005-2006 województwa: dolnośląskie, kujawsko-pomorskie, lubelskie, lubuskie, łódzkie,

mazowieckie, opolskie, podkarpackie, pomorskie, śląskie, warmińsko-mazurskie, wielkopolskie i zachodniopomorskie, co stanowi ok. 85% powierzchni Polski (Rycina 3)



Rycina 3 Obszar Polski objęty w latach 2005-2006 przez PRWWR

---

## 3 CEL I FORMA ORGANIZACJI PRACY

---

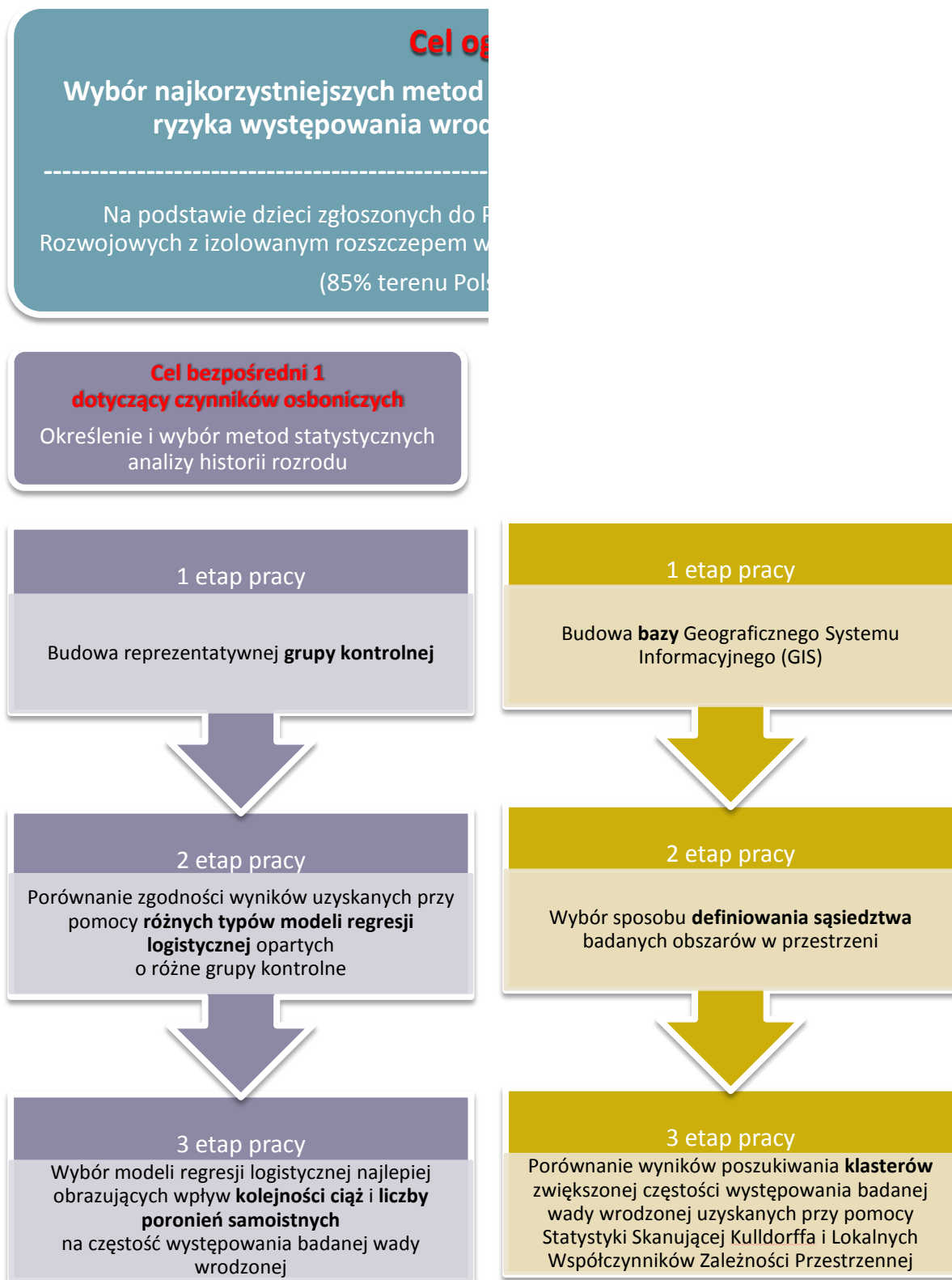
Celem ogólnym pracy jest wybór najkorzystniejszych metod statystycznych takich, które pozwoliłyby w ograniczonym zakresie na poszukiwanie czynników ryzyka występowania wrodzonych wad rozwojowych.

Zakres poszukiwania czynników ryzyka ograniczono do dwóch aspektów tworzących dwa cele bezpośrednie: do analizy wpływu czynników osobniczych (historii rodności matki) i analizy wpływu czynników zewnętrznych działających na matkę (teratogenów środowiskowych związanych w określonej lokalizacją w czasie i przestrzeni) na częstość występowania izolowanego rozszczepu wargi z/bez rozszczepu podniebienia (CL±P).

W związku z dużą różnorodnością metod statystycznych pozwalających na realizację tych celów praca dzieli się na dwa zasadnicze, całkowicie rozdzielne części: **analizę historii rodności** i **analizę czynników czasowo–przestrzennych (analizę klasterów<sup>1</sup>)** (Rycina 4). Realizacja postawionych celów bezpośrednich wiąże się ściśle z kolejnymi etapami pracy.

---

<sup>1</sup> **klaster** według definicji EUROCAT (2003r.) w rozumieniu wad wrodzonych, to skupisko przypadków wystąpienia wad wrodzonych płodu w czasie i / lub przestrzeni które wydaje się być nieprzeciętne



Rycina 4 Cele i forma organizacji pracy

---

## 4 ISTNIEJĄCY STAN WIEDZY W ZAKRESIE TEMATU BADAŃ

---

Niektóre rodzicielskie czynniki są dość dobrze scharakteryzowane jako czynniki ryzyka występowania wielu wrodzonych wad rozwojowych np. wiek matki, palenie papierosów przez matkę czy nie przyjmowanie przez nią kwasu foliowego w pierwszym trymestrze ciąży. Jednakże istnieje kilka elementów historii rodności, takich jak kolejność ciąży i liczba poronień samoistnych, których rola w tworzeniu się tych wad w dalszym ciągu nie jest dobrze rozpoznana. Najpoważniejsze wskazania, że te czynniki mogą być istotne pochodzą z epidemiologicznych badań nad wadą cewy nerwowej ([8] Bianca S. i inni, 2002; [14] Carmi R. i inni, 1994; [64] Rivas F. i inni, 2000; [11] Blanco-Munoz J. i inni, 2006). Kolejność ciąży pojawia się jako potencjalny czynnik ryzyka przy zespole Downa, który może być uzależniony również od wieku matki ([20] Doria-Rose V.P. i inni, 2003; [36] Kallen K., 1997; [15] Castilla E.E. i Paz J.E., 1994). Podobnie jak pochodzenie noworodka z kolejnych ciąż, historia matki dotycząca wcześniejszych poronień może stanowić czynnik podwyższonego ryzyka wady cewy nerwowej ([30] Gardiner A. i inni, 1978) jak również aneuploidów ([9] Bianco K. i inni, 2006). Wszystkie te informacje są w sposób oczywisty potrzebne w poradnictwie prenatalnym. Niestety wiele wcześniejszych badań cierpi z powodu mało-licznych grup i nie wykorzystuje metod wielowymiarowych do oceny wpływu kolejności ciąży i wcześniejszych niepowodzeń rozrodu na występowanie wady wrodzonych ([16] Cengiz B. i inni, 2004).

**W tym badaniu poddano analizie wpływ kolejności ciąży i liczby wcześniejszych poronień samoistnych na częstość występowania rozszczepu wargi z/bez rozszczepu podniebienia (CL±P) w dużym badaniu przypadek-kontrola danych pochodzących z PRWWR. Przybliżone zostały możliwości metody regresji logistycznej, które pozwolą na ocenę niezależnego udziału (wkładu) tych rodzicielskich czynników.**

Zagadnienia dotyczące zwiększenia ryzyka wystąpienia wady wrodzonej pod wpływem zewnętrznych teratogenów związanych z miejscem przebywania a właściwie miejscem zamieszkania matki w czasie ciąży są najczęściej pomijane lub są rozważane w bardzo ograniczonym wymiarze. Dotychczasowe prace dotyczące izolowanego CL±P analizę czynnika geograficznego ograniczają do podziału na miasto/wieś ([10] Bing-Fang Hwang i inni 2008, [57]

Messer L.C. i inni 2010) lub do podziału przestrzeni na inne dwa lub trzy regiony np. północ-południe ([34] Harville E. W. i inni, 2005).

Tymczasem taka forma organizacji badań nie ułatwia identyfikacji teratogenów środowiskowych występujących w konkretnej lokalizacji geograficznej. Szersza wiedza na temat lokalizacji klasterów zwiększonego ryzyka występowania wad izolowanych może stanowić pierwszy etap prac nad identyfikacją, zlokalizowaniem i likwidacją groźnych dla zdrowia zanieczyszczeń środowiska. Znana lokalizacja klasterów pozwoli bowiem na znaczne zawężenie obszaru poszukiwań różnego rodzaju teratogenów środowiskowych do obszaru określonego przez klaster. Wykryty klaster zawęzi obszar poszukiwań teratogenów, przez co obniży koszty ich identyfikacji. Zidentyfikowane w nim teratogeny z kolei pozwolą na prowadzenie ukierunkowanych działań profilaktycznych, jak i na zmniejszenie kosztów zarówno materialnych (związanych z przyszłym leczeniem schorzeń genetycznych uwarunkowanych w dużym stopniu środowiskowo) jak i niemierzalnych (związanych z bólem i cierpieniem rodziców dzieci obarczonych takimi wadami).

Jedną z głównych przyczyn niedostatecznej ilości badań związanych z wykrywaniem klasterów wad wrodzonych, jest późny wspólny rozwój metod ekonometrii i statystyki przestrzennej. Połączenie i wspólne rozwijanie tych metod datuje się na połowę lat siedemdziesiątych XX wieku. Dalszy rozwój metod analizy przestrzennej ściśle związany jest ze znacznym ułatwieniem, jakim stał się powszechnie stosowany komputer i powstanie systemu GIS - niezbędne narzędzie przy obróbce tak dużej ilości danych geograficznych. Chociaż na świecie ta dziedzina wiedzy rozwija się dynamicznie, w Polsce, oprócz tłumaczeń pionierskich prac Klassen'a ([39] Klassen L.H., Paelinck J.H.P. i Wagenaar S.J., 1982; [38] Klassen L.H. i Paelinck J.H.P., 1983) oraz niewielu artykułów i referatów naukowych, ukazały się jedynie dwie publikacje książkowe dotyczące tej problematyki: Ekonometria przestrzenna, ([32] Grabiński T. i inni, 1991) oraz Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN ([42] Kopczevska K., 2007).

Drugą, niemniej ważną przyczyną braku badań nad klasterami występowania wad wrodzonych jest fakt, że podstawą takich badań jest konieczność posiadania kompletnej informacji na temat występowania wad wrodzonych na badanym obszarze. Jedynymi podmiotami dysponującymi tak dokładnymi danymi są duże i długo działające Rejestry Wrodzonych Wad Rozwojowych. Takie rejestry niektórych krajów wchodzących w skład wielkiego rejestru europejskiego EUROCAT oraz rejestr w USA podejmują próby wyszukiwania klasterów dla terytoriów swoich krajów ([www.eurocat-network.eu](http://www.eurocat-network.eu)). PRWWR rozpoczyna

właśnie przygotowania do takiej pracy. Pierwszym etapem pracy związanym z wyszukiwaniem klasterów jest odpowiednie przygotowanie zebranych informacji, tak by tworzyły spójny System Informacji Geograficznej (GIS). Dopiero w następnej kolejności przy pomocy odpowiednich statystyk można skanować przestrzeń i czas w celu identyfikacji takich miejsc, gdzie ryzyko wystąpienia wady jest znacznie wyższe.

**W niniejszej pracy przedstawiono kolejne etapy pracy nad tworzeniem systemu baz informacji potrzebnej w wykrywaniu klasterów. Zostały również przedstawione szerzej i porównane dwie metody wyszukiwania klasterów (LISA i statystyka skanująca Kulldorff'a) na przykładzie rozszczepu wargi z/bez rozszczepu podniebienia (CL±P).**

---

## 5 SPOSOBY POZYSKIWANIA DANYCH

---

### 5.1 ZBIERANIE GRUPY BADANEJ

W pracy wykorzystane zostały dane udostępnione przez Polski Rejestr Wrodzonych Wad Rozwojowych (PRWWR) dotyczące lat 2005 i 2006. Podstawowym źródłem informacji dla rejestru są zgłoszenia wad na zunifikowanych formularzach (załącznik 1). Kwestionariusz zawiera szczegółowe informacje dotyczące czynników ryzyka wystąpienia wady związane z dzieckiem (min. masa urodzeniowa, rok urodzenia, płeć) jak też z rodzicami (min. **miejsce zamieszkania matki**, wykształcenie i wiek rodziców) oraz z **historię rodności** w tym dane dotyczące wcześniejszych ciąży, komplikacji podczas ciąży oraz zakończenia ciąży. Zgłoszenia w formie wypełnionych formularzy a od 2005 roku w formie formularzy elektronicznych, dokonywane są przez lekarzy z placówek pediatrycznych, położniczo-ginekologicznych, dziecięcych poradni specjalistycznych, zakładów genetyki i poradni genetycznych zakładów patomorfologii, praktyk lekarza rodzinnego. wg instrukcji (załącznik 2). Uzyskane w ten sposób informacje są uzupełniane o dane pochodzące od lekarzy sprawdzających karty historii rozwoju noworodka lub inną dokumentację medyczną pod kątem kompletności zgłoszeń do PRWWR co pozwala na identyfikację wad ujawniających się w okresie noworodkowym a nie zgłoszonych do rejestru na formularzach zgłoszeniowych. Zgromadzone informacje poddawane są następnie weryfikacji przez Regionalny Zespół ds. PRWWR na podstawie dokumentacji lekarskiej poradni specjalistycznych historii chorób oddziałów dziecięcych oraz wyników badań cytogenetycznych i patomorfologicznych a następnie wprowadzone do elektronicznej bazy danych. Szczegółowy opis metodologii zbierania danych znajduje się w publikacji [53] Latos-Bieleńska A., Materna-Kiryłuk A., (2005). Podział wad oraz nazewnictwo zostało określone zgodnie z Międzynarodową Statystyczną Klasyfikacją Chorób i Problemów zdrowotnych - rewizja dziesiąta (załącznik 3). Kodowanie wad jest uzupełniane i modyfikowane zgodnie z wytycznymi EUROCAT przez zespół pracowników przeszkolonych i ściśle nadzorowanych przez koordynatora organizacyjnego PRWWR.



## 5.2 DANE GUS DOTYCZĄCE URODZEŃ ŻYWYCH W POPULACJI

Podstawową miarą statystyki opisowej dla epidemiologii wad wrodzonych są **współczynniki częstości** (ang. prevalence) występowania wrodzonych wad rozwojowych dla dzieci żywo urodzonych. Oblicza się je jako stosunek liczby dzieci żywo urodzonych z wrodzoną wadą rozwojową do wszystkich urodzeń żywych w określonym czasie i terenie. Współczynnik jest podawany na 1000 urodzeń żywych. Informacje o wszystkich dzieciach żywo urodzonych są gromadzone przez Główny Urząd Statystyczny (GUS) i udostępniane poprzez strony internetowe jak też publikowane w rocznikach statystycznych. Odpowiednie i bardziej szczegółowe zestawienie takich informacji - niezbędne w niniejszej pracy - zostało uzyskane poprzez złożenie zamówienia.

## 5.3 ZBIERANIE GRUPY KONTROLNEJ

Dane populacyjne dotyczące urodzeń żywych uzyskane z GUS, choć są dla epidemiologii zawsze pierwszym i bardzo cennym źródłem informacji, posiadają bardzo ograniczony zasób zmiennych (potencjalnych czynników ryzyka). Obowiązująca tajemnica statystyczna wprowadzona ustawą z dnia 29 czerwca 1995 r. o statystyce publicznej bardzo ogranicza dostęp do informacji GUS na odpowiednim poziomie szczegółowości. By uzyskać tyle samo informacji o dzieciach zdrowych ile uzyskuje baza PRWWR dla dzieci z wrodzonymi wadami rozwojowymi w 2005 roku, rozpoczęto gromadzenie danych dotyczących grupy kontrolnej (noworodki bez wad rozwojowych). Zgłoszenia odbywają się na tych samych formularzach, na których zgłaszane są wady, również w formie elektronicznej.

## 5.4 MAPY PODZIAŁU ADMINISTRACYJNEGO

Mapy podziału administracyjnego gmin uzyskano z Wojewódzkiego Ośrodka Dokumentacji Geodezyjnej i Kartograficznej w Poznaniu. Stanowią one nie tylko podstawę wizualizacji rozmieszczenia przestrzennego, ale przede wszystkim w połączeniu z informacją o występowaniu wad u dzieci i informacją o liczbie dzieci zdrowych tworzą główną bazę informacji o rozmieszczeniu i koncentracji występowania wad na badanym terenie.

---

## 6 ZASTOSOWANE OPROGRAMOWANIE

---

Wybór oprogramowania wspomagającego przeprowadzenie badań zdeterminowany został przez wybrane metody analizy danych.

Analiza statystyczna przeprowadzona została przy wykorzystaniu programu statystycznego: **PQStat 1.2.4** (*PQStat software*, <http://pqstat.pl/>) oraz programów geostatystycznych: **OpenGeoDa 0.9.8.14** (*GeoDa Center for Geospatial Analysis and Computation*, <http://geodacenter.asu.edu/>), **SatScan 8.2.1** (*Software for the spatial, temporal and space-time scan statistics*, <http://www.satscan.org/>) i **ArcGis 9.2. – ArcMap** (<http://www.esri.com/>).

---

## **7 METODY ANALIZY HISTORII RODNOŚCI**

---

### **7.1 DOBÓR GRUPY BADANEJ I GRUPY KONTROLNEJ**

#### **7.1.1 Grupa badana w analizie historii rodności**

Kryterium włączenia do analizy dzieci z izolowanym CL±P była kompletność informacji dotyczącej kolejności ciąży i liczby poronień samoistnych. Przypadki z mnogimi rozwojowymi wadami wrodzonymi, znanymi zespołami wad, i aberracjami chromosomowymi były wykluczone z analizy. Ze względu na to, że założeniem pracy jest analiza głównie dwóch czynników dotyczących historii rodności: liczby poronień samoistnych i kolejności ciąży na częstość występowania izolowanego CL±P powstały dwie grupy badane. W badaniu wpływu kolejności ciąży wzięły udział wszystkie matki rodzące dziecko z izolowanym CL±P a w badaniu wpływu liczby poronień samoistnych tylko te z ciąży drugiej i kolejnej. Spośród 453 matek dzieci spełniających kryteria 448 matek spełniało warunki włączenia do analizy. Spośród 448 włączonych matek 244 (54.5%) miało przynajmniej jedną wcześniejszą ciążę.

#### **7.1.2 Informacje o populacji badanej**

Informacje dotyczące 628618 urodzeń żywych, o znanej kolejności ciąży i liczby poronień samoistnych w latach 2005, 2006 na terenie Polski objętym PRWWR, pochodzą z danych GUS.

#### **7.1.3 Grupa kontrolna w analizie historii rodności**

W niniejszej pracy kontrolę stanowiły matki dzieci zdrowych (bez wady rozwojowej) urodzonych na terenie PRWWR podczas tego samego okresu (2005 i 2006 roku) co dzieci z grupy badanej. Liczba wszystkich dzieci grupy kontrolnej, nazwanej przez nas grupą „kontrolną pierwotną”, w tym okresie wynosiła 4017.

Bez wielkiej przesady można powiedzieć, że u podstaw statystyki jako nauki leżą dwa problemy: pierwszy to problem wiarygodności wyników analiz dokonanych na próbie z populacji, a drugi to problem uogólnienia tych wyników na populację, z której pobrana była próba. Wiąże się to z reprezentatywnością próby. Ze względu na to, że zbyt duża część matek grupy kontrolnej została wybrana z matek młodych, o wyższym poziomie edukacji oraz zamieszkujących głównie obszar miejski w porównaniu do populacji badanej, podejrzewano, że

nastąpił pewien błąd selekcji (losowania). By mieć pewność, że grupa kontrolna będzie wolna od błędów stronniczości doboru wybrano podgrupę, która lepiej przedstawia rozkład cech populacji. Wybrana grupa kontrolna nazwana przez nas „kontrolą populacyjną”, przedstawiająca miniaturę populacji, z której pochodzą przypadki, wykorzystana została do analizy częstości występowania izolowanego  $CL\pm P$ . Do analizy czynników ryzyka występowania badanej wady wykorzystana została również druga grupa kontrolna nazwana „kontrolą dopasowaną”. Powstała ona również poprzez wylosowanie kontroli spośród całej dużej grupy kontrolnej. Ta druga grupa kontrolna nie była odzwierciedleniem populacji, ale została dobrana do grupy przypadków metodą doboru dopasowanego.

### **7.1.3.1 „Kontrola populacyjna”**

Idealną grupę kontrolną stanowiłyby urodzenia żywe dzieci na terenie PRWWR w latach 2005, 2006. Ze względu na brak informacji dotyczących prokreacji, które stanowią istotny czynnik mogący wpływać na występowanie wrodzonych wad rozwojowych, nie wykorzystaliśmy w sposób bezpośredni danych populacyjnych. W tej sytuacji idealna grupa kontrolna, której potrzebowaliśmy powinna stanowić miniaturę populacji, tzn. powinna to być próba, w której rozkład wszystkich cech odpowiadałby dokładnie rozkładowi tych samych cech w populacji. Taki ideał próby reprezentatywnej jest jednak trudny do osiągnięcia ze względu na fakt, że nie mamy możliwości poznania wszystkich rozkładów populacji. Przy doborze grupy kontrolnej kierowaliśmy się bardziej wyważoną definicją Stephana i McCarthyego ([66] Stephan F.F i McCarthy P.J., 1958, [75] Zieliński A., 2002): „Próba reprezentatywna jest to próba, która dla określonego zbioru zmiennych odwzorowuje populację... [wykonane w jej obrębie] pewne specyficzne analizy... (obliczenia średnich, odchyłeń standardowych itp.) dostarczają wyniki położone w akceptowalnych przedziałach wokół odpowiednich wartości występujących w populacji, z wyjątkiem rzadkich sytuacji, gdy wyniki te wypadają poza granicami tych przedziałów”. Wiedząc, że grupa kontrolna, którą uzyskamy będzie podzbiorem pierwotnej grupy kontrolnej oraz biorąc pod uwagę fakt, że w miarę uwzględniania kolejnych zmiennych wzrastają wymagania w stosunku do liczebności próby szybko osiągając wielkości mało realne, selekcję oparliśmy na kilku podstawowych kryteriach. Informację o rozkładzie cech w populacji, wybranych jako kryteria doboru grupy kontrolnej uzyskaliśmy z GUS. Selekcję kontroli przeprowadziliśmy w oparciu o losowanie warstwowe dla następujących warstw:

- wiek matki ( $\leq 19$  lat, 20-24 lata, 25-29 lat, 30-34 lata, 35-39 lat,  $\geq 40$  lat),

- wykształcenie (podstawowe+podstawowe niepełne, zawodowe, średnie, wyższe),
- obszar miejski/wiejski,
- rok (2005, 2006).

powyższe kryteria wyznaczyły  $6 \cdot 2 \cdot 2 \cdot 4 = 96$  warstw. Losowanie strukturalne, czyli losowanie oddzielne z każdej warstwy odbywało się metodą prostego próbkowania losowego. Przestrzeni próbkowania nie stanowiły dzieci, ale matki dzieci z pierwotnej grupy kontrolnej. Taki sposób wyboru przestrzeni próbkowania umożliwia zakładanie w dalszej części niezależności przypadków, gdyż unikamy w ten sposób wylosowania kilkakrotnie tej samej matki, która mogła urodzić dziecko zarówno w roku 2005 jak i 2006. Wyznaczono niezbędną liczebność próby zgodnie z wzorem:

$$n = \frac{1}{\frac{4d^2}{u_\alpha^2} + \frac{1}{N}} \quad 7.1$$

gdzie:

$d$  – błąd szacunku frakcji,

$N$  – liczebność populacji,

$u_\alpha$  – wartość krytyczna rozkładu normalnego dla zadanego poziomu istotności  $\alpha$ .

Poziom istotności ustalono na  $\alpha=0.05$  a błąd szacunku  $d=0.02$ . Stąd liczebność próby powinna wynieść  $n=2392$ . Ze względu na zbyt małą liczebność w niektórych warstwach przestrzeni losowania nie udało się osiągnąć niezbędnej liczebności próby na poziomie błędu  $d=0.02$  lecz na poziomie 0.021. Łącznie z wszystkich warstw wylosowano 2093 kontroli, z czego dla 2068 (98.8%) znana jest kolejność ciąży i liczba poronień samoistnych i są one zawarte w niniejszej pracy. Spośród 2068 kontroli stanowiących „kontrolę populacyjną” 973 (47.1%) to matki, które miały przynajmniej jedną wcześniejszą ciążę.

W losowaniu wykorzystano generator liczb losowych zawarty w programie Excel 2007.

### 7.1.3.2. „Kontrola dopasowana”

Dopasowanie kontroli do przypadków jest procedurą zwiększającą wydajność badań bez wpływu na siłę asocjacji między cechami ([48] Kupper L.L. i inni, 1981). Może być ono stosowane w badaniach kohort, gdzie w kategoriach dopasowania dobieramy odpowiednią liczbę narażonych. Doboru dokonuje się względem cechy, które później nie będą uwzględniane

w analizie. Grupa badana i kontrolna będzie wówczas zbliżona strukturą wielu cech za wyjątkiem tych, które będą stanowiły czynnik ryzyka występowania badanej wady. U nas dopasowania dokonano ze względu na wiek matki, jako że czynnik ten badany był przez nas wcześniej w pracy ([56] Materna-Kiryłuk A. i inni 2009) i znamy jego wpływ na wadę obecnie przez nas badaną. Przestrzeń próbkowania stanowiła grupa „kontrolna pierwotna”. Grupy „kontrolne dopasowane” uzyskano poprzez dopasowanie w stosunku 1:1, czyli dla każdej matki dziecka z analizowaną wadą, o zadanym wieku wylosowano matkę w tym samym wieku dziecka zdrowego. Wybrano dwie grupy „kontrolne dopasowane”:

- jedna dla wszystkich matek z grupy badanej – grupa ta liczy 442 matki (dla 6 matek nie udało się dokonać dopasowania)
- druga dla matek, które miały przynajmniej jedną wcześniejszą ciążę – grupa ta liczy 236 matek (dla 8 matek nie udało się dokonać dopasowania).

## 7.2 METODY JEDNOWYMIAROWE CZYNNIKÓW ZWIĄZANYCH Z HISTORIĄ RODNOŚCI

Szeroko rozpowszechnione w analizach epidemiologicznych są statystyki oparte o metody jednowymiarowe. W tego rodzaju analizach bada się oddzielnie wpływ każdego z rozważanych czynników ryzyka. Dla danych ilościowych jest to najczęściej test t-Studenta, natomiast dla danych jakościowych są to testy chi-kwadrat dla trendu, chi-kwadrat niezależności wraz z poprawkami. Obok tych testów równie często wyznacza się iloraz szans (*ang. odds ratio, OR*) wraz z przedziałami ufności.

Analiza jednowymiarowa przeprowadzona została dla porównania:

1. grupy badanej z grupą „kontrolną populacyjną”
2. grupy badanej z grupą „kontrolną populacyjną” (matki dzieci z ciąży 2 i dalszej)
3. grupy badanej z grupą „kontrolną dopasowaną”
4. grupy badanej z grupą „kontrolną dopasowaną” (matki dzieci z ciąży 2 i dalszej)

W pracy dokładnie przedstawione i omówione zostały rezultaty **pierwszej wymienionej wyżej analizy jednowymiarowej** kolejności ciąży i liczby przebytych poronień, jak też innych potencjalnych czynników ryzyka analizowanej wady takich jak: masa urodzeniowa, wiek płodowy, wiek matki, wiek ojca, wykształcenie matki, wykształcenie ojca, miejsce zamieszkania matki (miasto, wieś), palenie papierosów przez matkę, infekcje układu oddechowego podczas pierwszego trymestru ciąży, infekcje układu moczowego podczas pierwszego trymestru ciąży, inne infekcje, liczba poronień sztucznych i liczba porodów martwych.

Tabele wynikowe dla kolejnych analiz zostały zawarte w załączniku (Załącznik 4: Tabela A1, A2, A3, A4, A5, A6, A7, A8, A9).

### 7.2.1 Analiza danych ilościowych i jakościowych o uporządkowanych kategoriach

Badanie zależności przedstawiającej trend monotoniczny pozwala lekarzom na stosunkowo łatwe interpretowanie wyników przeprowadzonych analiz. Znaleziony trend obrazujący zwiększające/zmniejszające się ryzyko występowania wady (zmiennej wyjaśnianej) w zależności od wzrostu/spadku wartości potencjalnego czynnika ryzyka (zmiennej wyjaśniającej) daje nieco szersze spojrzenie na dane niż tylko porównanie wartości przeciętnej lub średniej. Nie

mniej występowanie istotnego statystycznie trendu monotonicznego pociąga za sobą zwykle istotną statystycznie różnicę wartości przeciętnych.

W pracy analizę ilościowych czynników ryzyka i czynników jakościowych o uporządkowanych kategoriach wykonano przy pomocy nieparametrycznego testu **chi-kwadrat dla trendu**.

**Tabela 1** Dane ilościowe i jakościowe o uporządkowanych kategoriach w postaci tabeli kontyngencji licznosci oczekiwanych

<b>gradacja czynnika</b> <i>rzyszka (waga: w<sub>i</sub>)</i>	<b>wada</b>	<b>brak wady</b>
1	O <sub>11</sub>	O <sub>12</sub>
2	O <sub>21</sub>	O <sub>22</sub>
3	O <sub>31</sub>	O <sub>32</sub>
4	O <sub>41</sub>	O <sub>42</sub>
...	...	...
r	O <sub>r1</sub>	O <sub>r2</sub>

$$\chi^2 = \frac{\left[ \left( \sum_{i=1}^r w_i \cdot O_{i1} \right) - C_1 \left( \sum_{i=1}^r \frac{w_i \cdot R_i}{n} \right) \right]^2}{\frac{C_1}{n} \left( 1 - \frac{C_1}{n} \right) \left[ \left( \sum_{i=1}^n w_i^2 R_i \right) - n \left( \sum_{i=1}^n \frac{w_i \cdot R_i}{n} \right)^2 \right]} \quad 7.2$$

gdzie:

$R_i$  – sumy wierszy

$C_j$  – sumy kolumn

$O_{i1}$  – wartości obserwowane

Część czynników ryzyka została następnie włączana do analizy wielowymiarowej, jaką jest w pracy regresja logistyczna – czynniki te są w tabeli (*Tabela 2*) zaznaczone.



Tabela 2 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” - zmienne ilościowe i jakościowe o uporządkowanych kategoriach

potencjalne czynniki ryzyka	kontrola populacyjna, N=2068		CL±P, N=448	test chi <sup>2</sup> dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
wiek płodowy w tygodniach #				0.000912
<28	3 (0.1%)	2 (0.4%)	40,0%	
28-31	4 (0.2%)	7 (1.6%)	63.6%	
32-36	101 (4.9%)	27 (6.0%)	21.1%	
37-41	1859 (89.9%)	387 (86.4%)	17.2%	
>=42	73 (3.5%)	11 (2.5%)	13.1%	
bd	28 (1.4%)	14 (3.1%)		
masa urodzeniowa w gramach **				3.815e-010
<1499	4 (0.2%)	5 (1.1%)	55.6%	
1500-1999	12 (0.6%)	14 (3.1%)	53.8%	
2000-2490	60 (2.9%)	26 (5.8%)	30.2%	
2500-2990	263 (12.7%)	86 (19.2%)	24.6%	
3000-3490	804 (38.9%)	151 (33.7%)	15.8%	
3500-3990	661 (32.0%)	124 (27.7%)	15.8%	
4000-4490	216 (10.4%)	30 (6.7%)	12.2%	
>=4500	30 (1.5%)	7 (1.6%)	18.9%	
bd	18 (0.9%)	5 (1.1%)		
wiek matki w latach ☉				0.3083 (NS)
<=19	88 (4.3%)	34 (7.6%)	27.9%	
20-24	482 (23.3%)	108 (24.1%)	18.3%	
25-29	818 (39.6%)	157 (35.0%)	16.1%	
30-34	489 (23.6%)	95 (21.2%)	16.3%	
35-39	157 (7.6%)	41 (9.2%)	20.7%	
>=40	34 (1.6%)	9 (2.0%)	20.9%	
bd	0 (0.0%)	4 (0.9%)		
wiek ojca w latach #				0.667 (NS)
<=19	11 (0.5%)	7 (1.6%)	38.9%	
20-24	245 (11.8%)	67(15.0%)	21.5%	
25-29	728 (35.5%)	128 (28.6%)	15,0%	
30-34	662 (32.0%)	123 (27.5%)	15.7%	
35-39	226 (10.9%)	61 (13.6%)	21.3%	
40-44	97 (4.7%)	24 (5.4%)	19.8%	
45-49	28 (1.4%)	10 (2.2%)	26.3%	
>=50	16 (0.8%)	3 (0.7%)	15.8%	
bd	55 (2.7%)	25 (5.6%)		
wykształcenie matki **				1.222e-011
=< podstawowe	145 (7.0%)	67 (15.0%)	31.6%	
zawodowe	415 (20.1%)	110 (24.6%)	21,0%	
średnie	784 (37.9%)	141 (31.5%)	15.2%	
wyższe	722 (34.9%)	104 (23.2%)	12.6%	
bd	2 (0.1%)	26 (5.8%)		
wykształcenie ojca #				2.076e-010

potencjalne czynniki ryzyka	kontrola populacyjna, N=2068		CL±P, N=448	test chi <sup>2</sup> dla trendu
	licznosc (% z N)	licznosc (% z N)	%dzieci z wadą wg poziomów czynnika	wartosc p
=< podstawowe	69 (3.3%)	38 (8.5%)	35.5%	
zawodowe	677 (32.7%)	162 (36.2%)	19.3%	
średnie	645 (31.2%)	121 (27.0%)	15.8%	
wyższe	588 (28.4%)	71 (15.8%)	10.8%	
bd	89 (4.3%)	56 (12.5%)		
<b>kolejność ciąży ****</b>				<b>1.106e-006</b>
pierworódka	1095 (52.9%)	204 (45.5%)	15.7%	
2 ciąża	644 (31.1%)	129 (28.8%)	16.7%	
3 ciąża	222 (10.7%)	67 (15.0%)	23.2%	
>=4 ciąża	107 (5.2%)	48 (10.7%)	31.0%	
ilość wcześniejszych poronień samoistnych *				0.654 (NS)
0 poronień	1826 (88.3%)	391 (87.3%)	17.6%	
1 poronienie	190 (9.2%)	45 (10.0%)	19.1%	
2 poronienia	44 (2.1%)	11 (2.5%)	20.0%	
3 poronienia	8 (0.4%)	1 (0.2%)	11.1%	

\*\*\*\* zmienna wykorzystana w czterech modelach regresji logistycznej zbudowanych dla kolejności ciąży (prostym, minimalnie dostosowanym, dużym i pełnym)

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla kolejności ciąży (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla kolejności ciąży (pełnym)

⊗ zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla kolejności ciąży (minimalnie dostosowanym)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

# nie uwzględnione w modelach

Dodatkowo dla danych ilościowych ciągłych przedstawiono wyniki testu **t-Studenta** i **t-Studenta z poprawką Cochran'a-Cox'a** (Tabela 3).

Tabela 3 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” - zmienne ilościowe ciągłe

potencjalne czynniki ryzyka	kontrola populacyjna, N=2068		CL±P, N=448		test Fisher'a-Snedecor'a	korekta Cochran'a-Cox'a testu t_Studenta
	średnia (odchlenie st.)	test Lilliefors'a wartość p	średnia (odchlenie st.)	test Lilliefors'a wartość p	wartość p	wartość p
	wiek płodowy w tygodniach	39.2 (1.7)	<0.01	38.9 (2.2)	<0.01	<0.0001
masa urodzeniowa w gramach	3420.0 (508.7)	<0.01	3250.3 (612.9)	<0.01	<0.0001	<0.0001
wiek matki w latach	27.6 (5.0)	<0.01	27.5 (5.6)	<0.01	0.0006	0.5211 (NS)
wiek ojca w latach	30.3 (5.7)	<0.01	30.4 (6.3)	<0.01	0.0039	0.8884 (NS)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

## 7.2.2 Analiza danych dychotomicznych

Porównanie danych dychotomicznych zwykle wykonywane testem chi-kwadrat z poprawką Yates'a lub chi-kwadrat bez tej poprawki można zastąpić odpowiednio przez test dokładny Fisher'a i mniej znany test mid-p. Bazuje ono na danych zebranych w tabeli kontyngencji:

**Tabela 4 Dane dychotomiczne w postaci tabeli kontyngencji  
liczności obserwowanych**

Liczności obserwowane	Wada	Brak wady
Czynnik ryzyka (tak)	<i>a</i>	<i>b</i>
Czynnik ryzyka (nie)	<i>c</i>	<i>d</i>

Test chi-kwadrat (Karl Pearson 1900) dedykowany jest zwykle tabelom o dużych licznościach obserwowanych (według interpretacji [17] Cochran'a W.G., 1952) żadna z liczności oczekiwanych nie może być < 1 oraz nie więcej niż 20% liczebności oczekiwanych może być < 5) a test Fisher'a ([22] Fisher R. A., 1934, [23] Fisher R. A., 1935) tabelom o małych licznościach oczekiwanych. Ten podział wynika głównie ze sposobu wyliczania testu dokładnego Fisher'a, bazującego na silni:

$$P = \frac{\binom{a+c}{a} \binom{b+d}{b}}{\binom{n}{a+b}} = \frac{\frac{(a+c)!}{c!a!} \cdot \frac{(b+d)!}{d!b!}}{\frac{n!}{(n-a-b)!(a+b)!}} \quad 7.3$$

Dokładne prawdopodobieństwo testu Fisher'a jest żmudne w obliczaniach ręcznych, ale też algorytm wyliczania prawdopodobieństwa tego testu jest trudny w implementacji. Zwiększenie mocy obliczeniowej komputerów oraz stale ulepszane algorytmy szybkiego wyliczania wartości prawdopodobieństwa testu Fisher'a na dzień dzisiejszy pozwalają na wyznaczenie jego wartości dla znacznie większych liczności obserwowanych niż jeszcze przed kilku laty. Test ten stał się zatem bardziej uniwersalnym testem od testu chi-kwadrat.

**Test mid-p** jest korektą testu dokładnego Fisher'a. Ta zmodyfikowana wartość poziomu istotności jest rekomendowana przez wielu statystyków ([51] Lancaster H.O., 1961; [2] Anscombe F.J., 1981; [63] Pratt J.W. i Gibbons J.D., 1981; [62] Plackett R.L., 1984; [58] Miettinen O.S., 1985; [59] Miettinen O.S. i Nurminen M., 1985, [6] Barnard G.A. 1989; [5]

Armitage P. i Berry G. 1994; [31] Gardner M.J. i Altman D.G., 2000) jako metoda zmniejszenia konserwatywności testu dokładnego Fisher’a. W rezultacie testem mid-p szybciej odrzucimy hipotezę zerową niż dokładnym testem Fisher’a. Dla dużych prób wartość p otrzymana przy pomocy testu chi-kwadrat z poprawką Yates’a i testu Fisher’a dają zbliżone wyniki, natomiast wartość p testu chi-kwadrat bez korekcji koresponduje z wartością testu mid-p (Tabela 5)

Tabela 5 Testy dla danych dychotomicznych w zależności od stopnia konserwatywności i wymagań dotyczących liczności próby

		Konserwatywność testu	
		Mniej konserwatywny	Bardziej konserwatywny
Wymagana liczność próby	Duża liczność	Test chi-kwadrat	Test chi-kwadrat z poprawką Yates’a na ciągłość ([74] Yates F., 1934)
	Brak wymagań	Test mid-p	Test dokładny Fisher’a

Wartość poziomu istotności testu mid-p wyznaczana jest przez przekształcenie wartości prawdopodobieństwa dla testu dokładnego Fisher’a. Jednostronna wartość p, będąca prawdziwym prawdopodobieństwem ([65] Rothman K.J., 2008), wyznaczana jest z wzoru:

$$p_{I(mid-p)} = p_{I(Fisher)} - 0.5p_{(point)} \quad 7.4$$

gdzie:

$p_{I(Fisher)}$  – wartość jednostronnego prawdopodobieństwa testu dokładnego Fisher’a,

$p_{(point)}$  – wartość prawdopodobieństwa wystąpienia takiego układu wartości w tabeli kontyngencji, jaki zawarty jest w tabeli badanej (wyliczana zgodnie z wzorem (7.3)).

Dwustronna wartość p jest definiowana jako podwojona wartość mniejszego z jednostronnych prawdopodobieństw:  $p_{II(mid-p)} = 2 \cdot p_{I(mid-p)}$ .

Wyniki analizy wykonanej z wykorzystaniem testu mid-p przedstawia tabela (Tabela 6). Część dychotomicznych czynników ryzyka była następnie włączana do analizy wielowymiarowej – czynniki te są w tej tabeli zaznaczone.

Tabela 6 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” - zmienne dichotomiczne

potencjalne czynniki ryzyka	kontrola populacyjna, N=2068		CL±P, N=448	test mid-p
		liczność (% z N)	liczność (% z N)	wartość p
<b>płeć **</b>				<b>0.000105</b>
	męska	1056 (51.1%)	271 (60.5%)	
	żeńska	1012 (48.9%)	172 (38.4%)	
	bd	0 (0.0%)	5 (1.1%)	
miejsce zamieszkania matki *				0.181065 (NS)
	wieś	802 (38.8%)	187 (41.7%)	
	miasto	1266 (61.2%)	256 (57.1%)	
	bd	0 (0.0%)	5 (1.1%)	
<b>palenie tytoniu przez matkę **</b>				<b>7.85E-016</b>
	nie	1747 (85.5%)	325 (72.5%)	
	tak	55 (2.7%)	55 (12.3%)	
	bd	266 (12.9%)	68 (15.2%)	
<b>infekcje ukł. oddech. matki podczas 1<sup>szego</sup> trymestru **</b>				<b>3.10E-007</b>
	nie	1877 (90.8%)	338 (75.4%)	
	tak	36 (1,7%)	27 (6.0%)	
	bd	155 (7.5%)	83 (18.5%)	
<b>inne infekcje matki podczas ciąży **</b>				<b>0.003791</b>
	nie	2020 (97.7%)	419 (93.5%)	
	tak	13 (0.6%)	10 (2.2%)	
	bd	35 (1.7%)	19 (4.2%)	
infekcje ukł. mocz. matki podczas 1 <sup>szego</sup> trymestru *				0.076522 (NS)
	nie	1673 (80.9%)	371 (82.8%)	
	tak	56 (2.7%)	6 (1.3%)	
	bd	339 (16.4%)	71 (15.8%)	
narażenie matki na promieniowanie podczas ciąży *				0.691490 (NS)
	nie	1732 (83.8%)	385 (85.9%)	
	tak	23 (1.1%)	4 (0.9%)	
	bd	313 (15.1%)	59 (13.2%)	
wcześniejsze poronienia sztuczne *				0.693072 (NS)
	nie	2060 (99.6%)	447 (99.8%)	
	tak	3 (0.1%)	1 (0.2%)	
	bd	5 (0.2%)	0 (0.0%)	
wcześniejsze porody martwe *				0.846488 (NS)
	nie	1992 (96.3%)	440 (98.2%)	
	tak	16 (0.8%)	3 (0.7%)	
	bd	60 (2.9%)	5 (1.1%)	

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla kolejności ciąży (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla kolejności ciąży (pełnym)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

## 7.3 METODY WIELOWYMIAROWE CZYNNIKÓW ZWIĄZANYCH Z HISTORIĄ RODNOŚCI – REGRESJA LOGISTYCZNA

### 7.3.1 Założenia analizy regresji logistycznej i warunkowej regresji logistycznej

Analiza wielowymiarowa to zbiór metod statystycznych obejmujących jednocześnie trzy zmienne lub więcej. Klasycznym jej przykładem jest badanie oddziaływania wielu zmiennych - nazywanych **zmiennymi niezależnymi** (lub **wyjaśniającymi**) na jedną zmienną - oznaczaną terminem **zmienna zależna** (lub **wyjaśniana**). Powszechnie stosowane są analizy **liniowej regresji wielokrotnej**, dla której zarówno zmienna niezależna jak i zmienne zależne są wyrażone przynajmniej na skali interwałowej. W badaniach medycznych a w szczególności w badaniach epidemiologicznych znacznie ważniejszą rolę odgrywa **regresja logistyczna**, (*ang. logistic regression, LR*) opracowana przez Berksona w 1944 roku. Jej główną zaletą w stosunku do tradycyjnej regresji liniowej jest osłabienie warunków jej stosowania (analizowane dane nie muszą pochodzić z jednej próby i nie muszą być zapisane przez skalę interwałową).

Zwykle zmienna zależna w regresji logistycznej jest dychotomiczna czyli posiada tylko dwie wartości: posiadanie i brak danej cechy. Istnieje również możliwość wyznaczania regresji logistycznej dla nominalnej zmiennej zależnej o więcej niż dwóch kategoriach - multinominalna regresja logistyczna, oraz dla porządkowej zmiennej zależnej - porządkowa regresja logistyczna. Regresja logistyczna nie tylko osłabia założenia dotyczące skali pomiaru zmiennych, ale dopuszcza do analizy dane pochodzące z wielu prób. W epidemiologii stosuje się ją również często w tzw. badaniach przypadek - kontrola, w których bada się dwie próby: losowane z populacji osób chorych na daną chorobę i z porównawczej populacji osób zdrowych (lub chorych na inną chorobę). Istnieje też możliwość wykonania regresji logistycznej dla grupy kontrolnej dobranej do grupy badanej według wybranych zmiennych w odpowiednim stosunku (1:1 - dla jednego przypadku jedna kontrola, 1:n - dla jednego przypadku n kontroli, m:n - dla m przypadków n kontroli). Regresja taka nosi nazwę **warunkowej regresji logistycznej**, (*ang. conditional logistic regression, CLR*). Zasadniczą zaletą doboru grupy kontrolnej do badanej jest większa precyzja otrzymanych wyników a główną wadą trudność w zebraniu odpowiedniej grupy kontrolnej.

Mimo wielu aspektów różniących regresję liniową i regresję logistyczną obie zakładają,

że zmienne niezależne znajdujące się w modelu nie powinny być redundantne, tzn. każda zmienna powinna wnosić do modelu informacje inne od tych wniesionych przez pozostałe zmienne. Te dwie regresje są również podobne w budowie i w interpretacji.

Model wielokrotnej regresja liniowa wyraża się wzorem:

$$y = a_0 + \sum_{i=1}^m a_i x_i \quad 6.5$$

gdzie:

$y$  - to zmienna zależna,

$x_1, x_2, \dots, x_m$  - to zmienne niezależne,

przy czym zarówno zmienna zależna jak i niezależne zmierzone są przynajmniej na skali interwałowej,

$a_0, a_1, a_2, \dots, a_m$  - współczynniki (nieznane parametry modelu) estymowane metodą najmniejszych kwadratów w oparciu o obserwowane zmienne niezależne oraz o zmienną zależną.

W regresji logistycznej natomiast:

$x_1, x_2, \dots, x_m$  - zmienne niezależne dychotomiczne lub zmienne o równo oddalonych uporządkowanych kategoriach,

$y$  - to zmienna zależna powstała przez przekształcenie zmiennej dychotomicznej  $D \in \{0,1\}$ ,

$a_0, a_1, a_2, \dots, a_m$  - współczynniki modelu LR estymowane metodą największej wiarygodności a modelu CLR warunkową metodą największej wiarygodności w oparciu o obserwowane zmienne niezależne oraz o zmienną zależną.

Proces budowy zmiennej zależnej wyraża następujący schemat:

$$D \in \{0,1\} \stackrel{(I)}{\Rightarrow} P \in \langle 0,1 \rangle \stackrel{(II)}{\Rightarrow} y \in (-\infty, \infty)$$

- (I) Prawdopodobieństwo wystąpienia wady  $P \in \langle 0,1 \rangle$  definiowane jest przez częstość jej występowania i opisane jest zmienną ciągłą ograniczoną do przedziału  $\langle 0,1 \rangle$
- (II) Zastosowanie logistycznej transformacji prawdopodobieństwa umożliwia wyeliminowanie ograniczenia zmiennej ciągłej do przedziału  $\langle 0,1 \rangle$ . Transformacja ta opisywana jest wzorem:

$$P = \frac{e^y}{1 + e^y} \quad 7.6$$

a funkcja odwrotna do niej zwana jest **logitem**

$$y = \ln\left(\frac{P}{1-P}\right) \quad 7.7$$

W rezultacie logistyczny model regresji można opisać wzorem:

$$P = \frac{e^{a_0 + \sum_{i=1}^m a_i x_i}}{1 + e^{a_0 + \sum_{i=1}^m a_i x_i}} \quad 7.8$$

lub

$$a_0 + \sum_{i=1}^m a_i x_i = \ln\left(\frac{P}{1-P}\right) \quad 7.9$$

Współczynniki regresji logistycznej  $a_0, a_1, a_2, \dots, a_m$  są estymowane z pewną dokładnością zwaną błędem  $S_{a_i}$ . Badaną zmienną niezależną można uznać za istotną statystycznie, jeżeli wartość jej współczynnika  $a_i$  jest istotnie różna od 0 ( $H_0: a_i = 0$ ) i jest on wyznaczony z odpowiednio małym błędem zgodnie ze statystyką testu Walda:

$$z^2 = \left(\frac{a_i}{S_{a_i}}\right)^2 \quad 7.10$$

o rozkładzie  $\chi^2$  z jednym stopniem swobody.

W przypadku, gdy zmienna niezależna (np. zmienna  $x_1$ ) posiada więcej niż dwa stany można dokonać jej rozbicia na kilka zmiennych dychotomicznych nazywanych zmiennymi fikcyjnymi (*ang. dummy variable*). Przy czym dla zmiennej o  $k$  kategoriach tworzy się  $k - 1$  zmiennych fikcyjnych:  $d_j$  ( $j \in \{1, 2, \dots, k-1\}$ ), a kategorię dla której nie powstaje zmienna fikcyjna oznacza się jako "bazową". Model regresji logistycznej wyraża się wówczas wzorem:



$$P = \frac{\exp(a_0 + \sum_{j=1}^{k-1} b_j d_j + \sum_{i=2}^m a_i x_i)}{1 + \exp(a_0 + \sum_{j=1}^{k-1} b_j d_j + \sum_{i=2}^m a_i x_i)}. \quad 7.11$$

Zwykle na zmienne fikcyjne rozбивa się zmienną zmierzoną na skali nominalnej, choć można takiego rozбивa dokonać również dla skali silniejszej, niestety traci się wtedy informację o uporządkowaniu kategorii.

Wartość **ilorazu szans (OR)**, wyznaczona tradycyjnie dla danych zebranych w tabelę 2x2 (dwie grupy: badana, kontrolna, dwa stany: narażeni na czynnik ryzyka, nie narażeni na czynnik ryzyka), wyraża szansę wystąpienia wady w grupie narażonej w stosunku do tej szansy w grupy nienarażonej na działanie czynnika ryzyka. Podobnie dzieje się modelu regresji logistycznej:

1. Dla jednej zmiennej niezależnej (np. dla zmiennej  $x_1$ ) o  $k$  kategoriach, rozbitej na  $k-1$  zmiennych fikcyjnych, podaje się  $k-1$  wartości ilorazu szans odnoszących się do poziomu bazowego i obliczanych zgodnie z wzorem:

$$OR_{(d_{j=1}, \text{baza})} = \frac{\exp(a_0 + b_j + \sum_{i=2}^m a_i x_i)}{\exp(a_0 + \sum_{i=2}^m a_i x_i)} = \exp(b_j). \quad 7.12$$

2. Dla zmiennej interwałowej o większej liczbie kategorii (np. dla zmiennej  $x_1$ ) stopniuje się narażenie na czynnik ryzyka. W wyniku czego kategorie zmiennej niezależnej możemy ułożyć od najbardziej do najmniej narażonych na wystąpienie wady. Dla takiej zmiennej podaje się tylko jeden iloraz szans wyliczany zgodnie z wzorem:

$$OR_{x_1=h, x_1=j} = \frac{\exp(a_0 + a_1 h + \sum_{i=2}^m a_i x_i)}{\exp(a_0 + a_1 j + \sum_{i=2}^m a_i x_i)} = \exp(a_1 (h - j)), \quad 7.13$$

gdzie wybrane dwie kategorie zmiennej  $x_1$  to: ( $x_1=h$ ) i ( $x_1=j$ ).

Zazwyczaj dla badanej zmiennej podaje się jednostkowy iloraz szans dla  $h-j=1$ . Oznacza on krotność zmian ryzyka przy wzroście o jednostkę (o jedną kategorię) zmiennej traktowanej jako czynnik ryzyka.

Do analizy **dobroci dopasowania modelu regresji logistycznej** do danych użyto testu Hosmera i Lemeshowa. Najpierw przypadki są sortowane rosnąco według oszacowanego prawdopodobieństwa pojawienia się badanego zjawiska. Wstępnie obserwacje są dzielone na  $G$  grup – zwykle na decyle ( $G=10$ ). Statystykę testu uzyskuje się dla tabel kontyngencji o

$G$  wierszach i dwóch kolumnach (kolumnie licznosci obserwowanych i kolumnie licznosci oczekiwanych) według wzoru:

$$H_L^2 = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g \left(1 - \frac{E_g}{n_g}\right)} \quad 7.14$$

gdzie:

$n_g$  - liczba obserwacji w  $g$ -tej grupie,

$O_g$  - liczba obserwacji w  $g$ -tej grupie posiadających daną cechę ( $D = 1$ ),

$E_g$  - liczba oczekiwanych obserwacji w grupie posiadających daną cechę ( $D = 1$ ).

Statystyka ta ma rozkład  $\chi^2$  z  $(G-2)$  stopniami swobody. Hipoteza zerowa zakłada, że nie ma różnicy między obserwowanymi i oczekiwanymi licznosciami zmiennej zależnej.

### 7.3.2 Regresja logistyczna i warunkowa regresja logistyczna w analizie historii rodności

Budowane modele regresji logistycznej zostały wykorzystane do analizy niezależnego (od innych dostępnych w analizie potencjalnych czynników ryzyka) wpływu

(A) kolejności ciąży,

(B) liczby wcześniejszych poronień samoistnych

na występowanie izolowanego CL±P.

Oddzielny wpływ tych czynników jak i inne najważniejsze potencjalne teratogeny zostały przedstawione szerzej w analizie jednowymiarowej.

#### Zagrozenie współliniowości zmiennych niezależnych

Regresja logistyczna nie nakłada tak wielu ograniczeń na zmienne niezależne jak liniowa regresja wielokrotna, ale przy wyborze zmiennych do modelu należy pamiętać o głównym założeniu regresji, by zmienne niezależne nie były wzajemnie silnie liniowo skorelowane. Włączenie zmiennych skorelowanych zwiększa wariancję estymowanych parametrów powodując zwiększenie budowanego dla nich przedziału ufności, co utrudnia wykrycie badanej zależności. W praktyce trudno jednak wymagać zupełnego braku powiązania. W naszym przypadku czynniki ryzyka związane z ojcem, jak również wiek płodowy zostały wyłączone z

analizy by uniknąć problemu współliniowości i przeparametryzowania modeli (Tabela 7).

Tabela 7 Współczynnik korelacji liniowej Pearsona pomiędzy czynnikami ryzyka izolowanego CL±P (w oparciu o grupę badaną - zmienne ilościowe i jakościowe o uporządkowanych kategoriach)

wsp. korelacji Pearsona	Wiek płodowy <sup>#</sup>	Masa urodzen.	Wiek matki	Wyksz. matki	Wiek ojca <sup>#</sup>	Wyksz. ojca <sup>#</sup>	Kolejn. ciąży	Liczba porodów martwych	Liczba poronień samoist.
Wiek płodowy <sup>#</sup>	1,00	0,64	-0,07	0,09	-0,12	0,14	-0,18	-0,14	0,01
Masa urodzen.	0,64	1,00	-0,05	0,22	-0,10	0,21	-0,21	-0,14	0,00
Wiek matki	-0,07	-0,05	1,00	0,04	0,79	0,04	0,45	0,06	0,01
Wyksz. matki	0,09	0,22	0,04	1,00	-0,09	0,68	-0,30	-0,02	0,04
Wiek ojca <sup>#</sup>	-0,12	-0,10	0,79	-0,09	1,00	0,01	0,40	0,12	-0,02
Wyksz. ojca <sup>#</sup>	0,14	0,21	0,04	0,68	0,01	1,00	-0,29	-0,07	0,00
Kolejn. ciąży	-0,18	-0,21	0,45	-0,30	0,40	-0,29	1,00	0,07	0,21
Liczba porodów martwych	-0,14	-0,14	0,06	-0,02	0,12	-0,07	0,07	1,00	0,01
Liczba poronień samoist.	0,01	0,00	0,01	0,04	-0,02	0,00	0,21	0,01	1,00

# zmienne nie uwzględnione w modelach ze względu na silną zależność liniową z inną zmienną niezależną

*kolor czerwony* – współczynniki korelacji liniowej Pearsona przekraczający wartość 0.3

*pogrubienie czcionki* – istotna statystycznie zależność liniowa pomiędzy zmiennymi ( $p < 0.05$ )

Liczba poronień, kolejność ciąży jak i wiek matki ze względu na współzależność, choć nie tak silną jak matczyne i ojcowskie czynniki ryzyka, najczęściej nie są analizowane wspólnie. Pamiętając jednak, że błąd selekcji zmiennych może wywołać gorsze skutki niż występowanie współzależności, postanowiliśmy uwzględnić w budowanych modelach również te zmienne. Można się bowiem spodziewać, że nie uwzględnianie kolejności ciąży przy badaniu wpływu poronień samoistnych lub wieku matki przy badaniu wpływu kolejności ciąży na występowanie wady może zniekształcić obraz rzeczywistości.

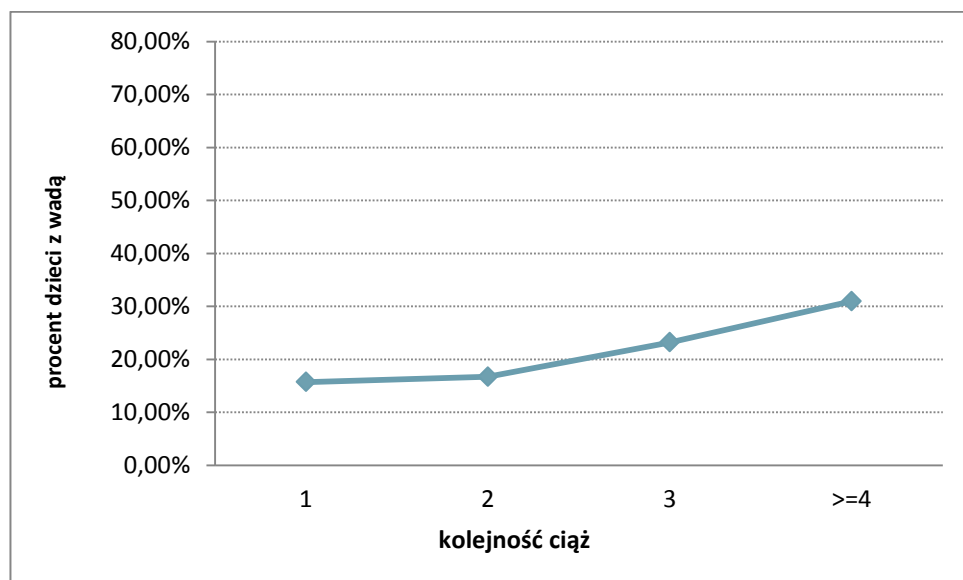
### 7.3.2.1 Modele regresji logistycznej (LR) i warunkowej regresji logistycznej (CLR)

Skonstruowano dwie grupy modeli opierając się na regresji logistycznej (grupa badana i "kontrolna populacyjna") i warunkowej regresji logistycznej (grupa badana i "kontrolna dopasowana"). Dla matek z grupy badanej kod zmiennej dychotomicznej przyjmuje wartość  $D = 1$ , dla matek z grupy kontrolnej  $D = 0$ . Dodatkowo oprócz analizy modeli dla kolejności ciąży i liczby poronień samoistnych zapisanych w skali interwałowej zbudowano modele, w których te zmienne rozbito na zmienne fikcyjne.

Ostatecznie budowane modele porównywano pod kątem istotności kolejności ciąży i liczby poronień samoistnych, by stworzyć pełny obraz wpływu tych dwóch teratogenów na występowanie izolowanego CL±P.

- A. W badaniu kolejności ciąży wzięły udział 448 matki z grupy badanej, 2068 matki z grupy "kontrolnej populacyjnej" oraz 442 z grupy "kontrolnej dopasowanej". By sprawdzić wpływ kolejności ciąży na częstość występowania badanej wady zbudowaliśmy:

- A.A. Proste modele (modele z jedną zmienną niezależną) LR i w CLR (Tabela 8 i Rycina 5):



Rycina 5 Wpływ kolejności ciąży na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną”

- A.A.A. zmienna kolejność ciąży oznaczona przez  $x_1$  zapisana została na skali interwałowej – kolejność ciąży jest liczbą naturalną, dla której najmniejsza możliwa różnica kategorii, to 1[ciąża]

$$P = \frac{e^{a_0 + a_1 x_1}}{1 + e^{a_0 + a_1 x_1}} \quad 6.15$$

A.A.B. zmienna kolejność ciąży została rozłożona na trzy zmienne fikcyjne

$$d_1, d_2, d_3$$

$$P = \frac{e^{a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3}}{1 + e^{a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3}} \quad 6.16$$

gdzie:

$d_1 = 1$ , gdy matka urodziła dziecko z ciąży drugiej,

$d_1 = 0$ , w innym przypadku (z ciąży innej niż druga),

$d_2 = 1$ , gdy matka urodziła dziecko z ciąży trzeciej,

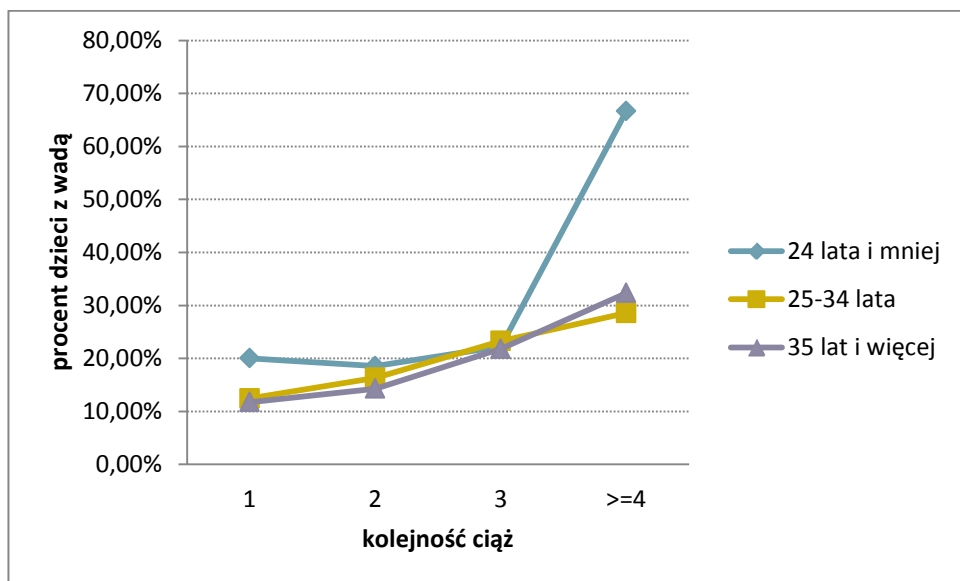
$d_2 = 0$ , w innym przypadku,

$d_3 = 1$ , gdy matka urodziła dziecko z ciąży czwartej lub dalszej,

$d_3 = 0$ , w innym przypadku,

kategorią bazową jest ciąża pierwsza.

A.B. Minimalnie dostosowane modele LR (modele z dwoma zmiennymi niezależnymi), zawierające obok kolejności ciąży również wiek matki (*Tabela 8* i *Rycina 6*). W modelach tych kolejność ciąży rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt A.A.A) oraz w rozbięciu na zmienne fikcyjne (patrz podpunkt A.A.B).



Rycina 6 Wpływ kolejności ciąży (z podziałem na kategorie wieku matki) na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną”

A.C. Duże modele LR i CLR (*Tabela 8*), zawierające obok kolejności ciąży również te zmienne, które w analizie jednowymiarowej różniły istotnie grupę badaną i

kontrolną (Tabela 2 i Tabela 6 dla LR oraz *Tabela 1* i *Tabela 3* w Załączniku 4 dla CLR). W modelach tych kolejność ciąży rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt A.A.A) oraz w rozbiciu na zmienne fikcyjne (patrz podpunkt A.A.B).

A.D. Pełne modele LR i CLR (*Tabela 8*), zawierające obok kolejności ciąży wszystkie te zmienne, które poddano analizie jednowymiarowej (Tabela 2 i Tabela 6 dla LR oraz *Tabela 1* i *Tabela 3* w Załączniku 4 dla CLR), a które nie były zmiennymi nadmiarowymi. W modelach tych kolejność ciąży rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt A.A.A) oraz w rozbiciu na zmienne fikcyjne (patrz podpunkt A.A.B).

Tabela 8 Wpływ kolejności ciąży na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną”

Model	regresja logistyczna					warunkowa regresja logistyczna		
	Kolejność ciąży - skala interwałowa		Kolejność ciąży - zmienne fikcyjne (ciąża 1 = kategoria "bazowa")			Kolejność ciąży - skala interwałowa	Kolejność ciąży - zmienne fikcyjne (ciąża 1 = kategoria "bazowa")	
	OR (95%CI) test Walda	test Hosmera-Lemeshowa	kol. ciąży	OR (95%CI) test Walda	test Hosmera-Lemeshowa	OR (95%CI) test Walda	kol. ciąży	OR (95%CI) test Walda
<b>Prosty model*</b>	1.27 (1.17-1.38) p=5.2x10 <sup>-8</sup>	p=0.3588 (NS)	ciąża 2 ciąża 3 ciąża >=4	1.07 (0.85-1.37), p=0.55 (NS) 1.62 (1.19-2.21), p=0.0024 2.41 (1.66-3.49), p=3.7x10 <sup>-6</sup>	p=1.0000 (NS)	1.38 (1.20-1.59) p= 9.846x10 <sup>-6</sup>	ciąża 2 ciąża 3 ciąża >=4	1.32 (0.94-1.85), p=0.1074 (NS) 1.98 (1.26 -3.13), p=0.0033 3.19 (1.74-5.83), p=0.0002
<b>Minimalnie dostosowany model**</b>	1.40 (1.26-1.54) p=1.3x10 <sup>-10</sup>	p=0.7229 (NS)	ciąża 2 ciąża 3 ciąża >=4	1.22 (0.95-1.58), p=0.1265 (NS) 1.95 (1.39-2.74), p=1.2x10 <sup>-4</sup> 3.37 (2.21-5.15), p=2.2x10 <sup>-8</sup>	p=0.3211 (NS)	X	ciąża 2 ciąża 3 ciąża >=4	X X X
<b>Duży model***</b>	1.22 (1.09-1.36) p=0.0005	p=0.4873 (NS)	ciąża 2 ciąża 3 ciąża >=4	1.03 (0.75-1.40), p=0.8690 (NS) 1.44 (0.97-2.14), p=0.0729 (NS) 2.14 (1.32-3.46), p=0.0019	p=0.5795 (NS)	1.25 (1.01-1.54) p=0.0367	ciąża 2 ciąża 3 ciąża >=4	1.11 (0.69-1.77), p=0.6704(NS) 1.38 (0.74 - 2.56), p=0.3157(NS) 2.44 (1.00-5.95), p=0.0511 (NS)
<b>Pełen model****</b>	1.52 (1.27-1.81) p=3.7x10 <sup>-6</sup>	p=0.7492 (NS)	ciąża 2 ciąża 3 ciąża >=4	1.23 (0.84-1.78), p=0.2853 (NS) 2.26 (1.31-3.90), p=0.0035 4.72 (2.30-9.68), p=2.3x10 <sup>-5</sup>	p=0.08764 (NS)	1.51 (1.15-1.98) p=0.0027	ciąża 2 ciąża 3 ciąża >=4	1.28 (0.79-2.08), p=0.3161 (NS) 2.64 (1.22-5.74), p=0.0139 5.21 (1.77-15.35), p=0.0028

\* model surowy dla kolejności ciąży

\*\* model dla kolejności ciąży dostosowany do wieku matki

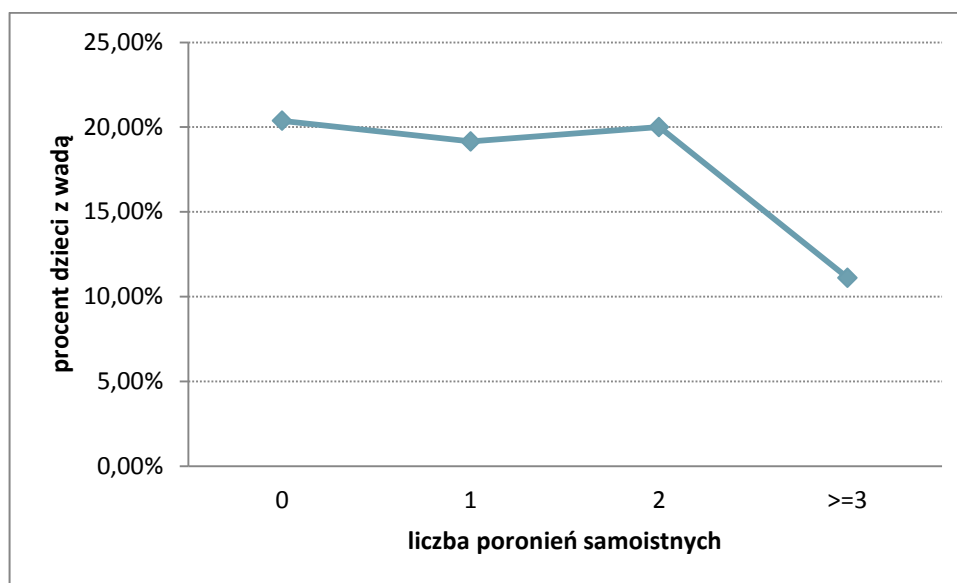
\*\*\* model dla kolejności ciąży dostosowany do tych czynników, które w analizie jednowymiarowej wykazywały istotny wpływ na wadę

\*\*\*\* model dla kolejności ciąży dostosowany do wszystkich czynników, które były poddane analizie jednowymiarowej i nie są nadmiarowe

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

B. W badaniu liczby poronień samoistnych wzięły udział tylko te matki, które miały szansę poronić wcześniejszą ciążę. Takich matek z ciąży drugiej lub dalszej jest 244 w grupie badanej, 973 w grupie "kontrolnej populacyjnej" oraz 236 w grupie "kontrolnej dopasowanej". By sprawdzić wpływ liczby poronień samoistnych na częstość występowania izolowanego CL±P zbudowaliśmy:

B.A. Proste modele LR i CLR (Tabela 9 i Rycina 7):



Rycina 7 Wpływ liczby poronień samoistnych na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciąży drugiej lub dalszej

B.A.A. zmienna liczba poronień samoistnych oznaczona przez  $x_1$  zapisana została na skali interwałowej – liczba poronień jest liczbą naturalną, dla której najmniejsza możliwa różnica kategorii, to 1 [poronienie]

$$P = \frac{e^{a_0 + a_1 x_1}}{1 + e^{a_0 + a_1 x_1}} \quad 7.17$$

B.A.B. zmienna liczba poronień samoistnych została rozłożona na trzy zmienne fikcyjne  $d_1, d_2, d_3$

$$P = \frac{e^{a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3}}{1 + e^{a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3}} \quad 7.18$$

gdzie:

$d_1 = 1$ , gdy matka uprzednio poroniła jedną ciążę,

$d_1 = 0$ , w innym przypadku,

$d_2 = 1$ , gdy matka poroniła uprzednio dwie ciążę,



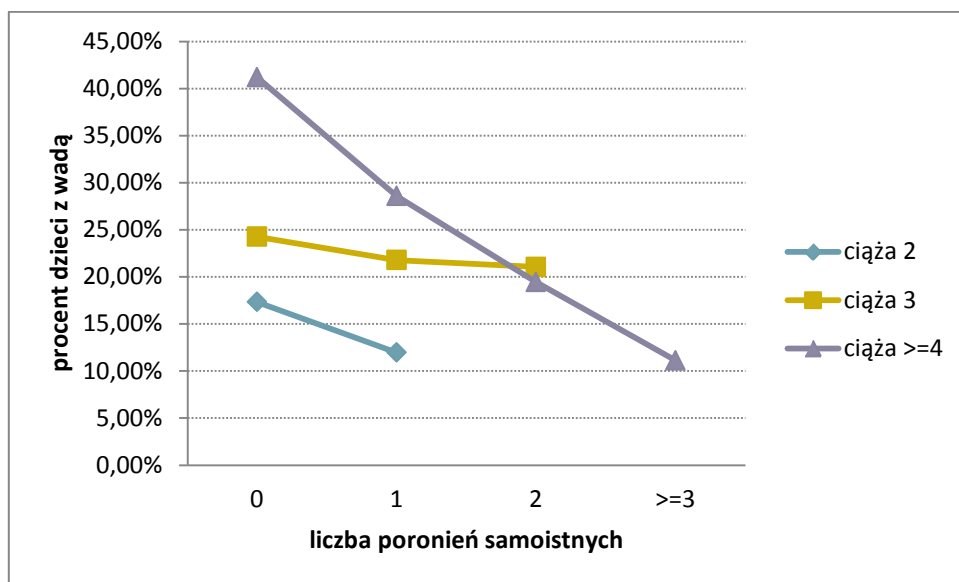
$d_2 = 0$ , w innym przypadku,

$d_3 = 1$ , gdy matka poroniła uprzednio trzy ciąże,

$d_3 = 0$ , w innym przypadku,

kategorią bazową jest brak wcześniejszych poronień.

B.B. Minimalnie dostosowane modele LR i CLR (modele z dwiema zmiennymi niezależnymi), zawierającymi obok liczby poronień samoistnych również kolejność ciąży (*Tabela 9* i *Rycina 8*). W modelach tych liczba poronień samoistnych rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt B.A.A) oraz w rozbiciu na zmienne fikcyjne (patrz podpunkt B.A.B).



Rycina 8 Wpływ liczby poronień samoistnych (z podziałem na kategorie kolejności ciąży) na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciąży drugiej lub dalszej

B.C. Duże modele LR i CLR (*Tabela 9*), zawierające obok liczby poronień samoistnych również te zmienne, które w analizie jednowymiarowej różniły istotnie grupę badaną i kontrolną (Załącznik 4 *Tabela 4* i *Tabela 6* dla LR oraz *Tabela 7* i *Tabela 9* dla CLR). W modelach tych liczba poronień samoistnych rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt B.A.A) oraz w rozbiciu na zmienne fikcyjne (patrz podpunkt B.A.B).

B.D. Pełne modele LR i CLR (*Tabela 9*), zawierające obok liczby poronień samoistnych wszystkie te zmienne, które poddano analizie jednowymiarowej (Załącznik 4 *Tabela 4* i *Tabela 6* dla LR oraz *Tabela 7* i *Tabela 9* dla CLR), a które nie były zmiennymi nadmiarowymi. W modelach tych liczba poronień samoistnych

rozpatrywana była zarówno jako zmienna wyrażona na skali interwałowej (patrz podpunkt B.A.A) oraz w rozbiciu na zmienne fikcyjne (patrz podpunkt B.A.B).

Tabela 9 Wpływ liczby poronień samoistnych na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciąży drugiej lub dalszej

Model	regresja logistyczna					warunkowa regresja logistyczna		
	Liczba poronień - skala interwałowa		Liczba poronień - zmienne fikcyjne (0 poronień = kategoria "bazowa")			Liczba poronień - skala interwałowa	Liczba poronień - zmienne fikcyjne (0 poronień = kategoria "bazowa")	
	OR (95%CI) test Walda	test Hosmera-Lemeshowa	Liczba poronień	OR (95%CI) test Walda	test Hosmera-Lemeshowa	OR (95%CI) test Walda	Liczba poronień	OR (95%CI) test Walda
<b>Prosty model*</b>	0.93 (0.73-1.19) p=0.5776 (NS)	p=0.0000	1 poronieenie 0.93 (0.64-1.33), p=0.6769 (NS) 2 poronienia 0.98 (0.50-1.93), p=0.9472 (NS) 3 poronienia 0.49 (0.06-3.93), p=0.5008 (NS)		p=0.0000	0.92 (0.68-1.25) p=0.5900 (NS)	1 poronieenie 0.92 (0.58-1.47), p=0.7271 (NS) 2 poronienia 1.08 (0.45-2.61), p=0.8574 (NS) 3 poronienia 0.50 (0.45-5.48), p=0.5673 (NS)	
<b>Minimalnie dostosowany model**</b>	0.69 (0.52-0.90) p=0.0068	p=0.6828 (NS)	1 poronieenie 0.75 (0.51-1.09), p=0.1341 (NS) 2 poronienia 0.47 (0.22-1.01), p=0.0523 (NS) 3 poronienia 0.14 (0.02-1.22), p=0.0754 (NS)		p=0.7113 (NS)	0.67 (0.47-0.95) p=0.0249	1 poronieenie 0.69 (0.42-1.14), p=0.1506 (NS) 2 poronienia 0.54 (0.21-1.42), p=0.2126 (NS) 3 poronienia 0.12 (0.01-1.76), p=0.1222 (NS)	
<b>Duży model***</b>	0.63 (0.44-0.90) p=0.0118	p=0.9718(NS)	1 poronieenie 0.76 (0.47-1.23), p=0.2654 (NS) 2 poronienia 0.26 (0.08-0.88), p=0.0299 3 poronienia 0.17 (0.02-1.76), p=0.1390 (NS)		p=0.9574 (NS)	0.51 (0.31-0.83) p=0.0066	1 poronieenie 0.59 (0.31-1.12), p=0.1065 (NS) 2 poronienia 1.20 (0.04-0.90), p=0.0366 3 poronienia 0.13 (0.01-2.53), p=0.1779 (NS)	
<b>Pełen model****</b>	0.56 (0.37-0.85) p=0.0059	p=0.7246 (NS)	1 poronieenie 0.70 (0.41-1.21), p=0.2041 (NS) 2 poronienia 0.22 (0.06-0.83), p=0.0261 3 poronienia 0.10 (0.01-1.20), p=0.0691 (NS)		p=0.5301 (NS)	0.54 (0.31-0.94) p=0.0283	1 poronieenie 0.68 (0.33-1.41), p=0.3023 (NS) 2 poronienia 0.17 (0.03-0.97), p=0.0457 3 poronienia 0.10 (0.00-2.22), p=0.1474 (NS)	

\* model surowy dla liczby poronień samoistnych

\*\* model dla liczby poronień samoistnych dostosowany do kolejności ciąży

\*\*\* model dla liczby poronień samoistnych dostosowany do tych czynników, które w analizie jednowymiarowej wykazywały istotny wpływ na wadę

\*\*\*\* model dla liczby poronień samoistnych dostosowany do wszystkich czynników, które były poddane analizie jednowymiarowej i nie są nadmiarowe

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

Podejrzewając, że to wspólna analiza kolejności ciąży i liczby poronień samoistnych powoduje ujawnienie wpływu poronień samoistnych na występowanie izolowanego CL±P dodatkowo, zbudowaliśmy duży i pełny model regresji usuwając z niego kolejność ciąży (*Tabela 10*).

Tabela 10 Wpływ liczby poronień samoistnych na ryzyko występowania izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciąży drugiej lub dalszej > budowane modele nie zawierają zmiennej: kolejność ciąży

Model	regresja logistyczna					warunkowa regresja logistyczna		
	Liczba poronień - skala interwałowa		Liczba poronień - zmienne fikcyjne (0 poronień = kategoria "bazowa")			Liczba poronień - skala interwałowa	Liczba poronień - zmienne fikcyjne (0 poronień = kategoria "bazowa")	
	OR (95%CI) test Walda	test Hosmera-Lemeshowa	Liczba poronień	OR (95%CI) test Walda	test Hosmera-Lemeshowa	OR (95%CI) test Walda	Liczba poronień	OR (95%CI) test Walda
<b>Duży model***</b>	0.84 (0.60-1.17) p=0.3099 (NS)	p=0.08518 (NS)	1 poronienie 2 poronienia 3 poronienia	0.90 (0.56-1.44), p=0.6475 (NS) 0.59 (0.20-1.72), p=0.3008 (NS) 0.68 (0.07-6.17), p=0.7286 (NS)	p=0.3415 (NS)	0.71 (0.48-1.07) p=0.1024 (NS)	1 poronienie 2 poronienia 3 poronienia	0.75 (0.41-1.38), p=0.3585 (NS) 0.47 (0.13-1.74), p=0.2574 (NS) 0.58 (0.05-7.13), p=0.6726 (NS)
<b>Pełen model****</b>	0.84 (0.59-1.21) p=0.3501 (NS)	p=0.4313 (NS)	1 poronienie 2 poronienia 3 poronienia	0.95 (0.57-1.60), p=0.8556 (NS) 0.50 (0.14-1.76), p=0.2807 (NS) 0.64 (0.06-6.34), p=0.7003 (NS)	p=0.3839 (NS)	0.76 (0.48-1.19) p=0.2287 (NS)	1 poronienie 2 poronienia 3 poronienia	0.85 (0.45-1.76), p=0.7420 (NS) 0.41 (0.09-1.82), p=0.2413 (NS) 0.46 (0.04-5.92), p=0.5499 (NS)

\*\*\* model dla liczby poronień samoistnych dostosowany do tych czynników, które w analizie jednowymiarowej wykazywały istotny wpływ na wagę za wyjątkiem kolejności ciąży

\*\*\*\* model dla liczby poronień samoistnych dostosowany do wszystkich czynników, które były poddane analizie jednowymiarowej i nie są nadmiarowe, za wyjątkiem kolejności ciąży

NS: brak istotności statystycznej - dwustronna wartość  $p \geq 0.05$

## 7.4 WNIOSKI DOTYCZĄCE ANALIZY HISTORII RODNOŚCI

### Wnioski płynące z realizacji pierwszego celu bezpośredniego dotyczącego czynników osobniczych

Określenie i wybór metod statystycznych analizy historii rozrodu

1.

Porównanie zgodności wyników uzyskanych przy pomocy **różnych typów modeli regresji logistycznej** opartych o różne grupy kontrolne:

- Modele regresji logistycznej oparte o **zmienne fikcyjne** są bardziej wrażliwe na małą licznosc poszczególnych kategorii niż modele bez tych zmiennych;
- Wyniki uzyskane dla modeli regresji logistycznej i **warunkowej regresji logistycznej** są zbieżne.

2.

Wybór modeli regresji logistycznej najlepiej obrazujących wpływ **kolejności ciąży** i **liczby poronień samoistnych** na częstość występowania badanej wady wrodzonej:

- Dla liczby **poronień samoistnych** powinno się tworzyć modele, które zawierają również kolejność ciąży. Modele takie wskazują, że im więcej wcześniejszych poronień samoistnych tym mniejsza szansa na urodzenie żywo dziecka z izolowanym CL±P;
- Wraz z **kolejną ciążą** rośnie istotnie ryzyko wystąpienia izolowanego CL±P u dziecka, bez względu na zastosowany w analizie model regresji logistycznej.

---

## 8 METODY ANALIZY CZYNNIKÓW CZASOWO-PRZESTRZENNYCH

---

Podstawową zasadą w geografii jest idea nazywana pierwszym prawem geografii Toblera ([68] Tobler W., 1970): "wszystko jest związane z wszystkim innym, ale rzeczy w pobliżu są bardziej związane niż rzeczy odległe". Jeżeli bowiem coś spowodowało zaistnienie danej własności w jednym miejscu jest również przyczyną istnienia podobnych własności w pobliskich lokalizacjach. Metody analizy czynników czasowo-przestrzennych to domena geografii. Nie mniej prawa, jakimi rządzi się geografia, mają szerokie zastosowanie w epidemiologii.

**Epidemiologia przestrzenna** określona została przez Wallera i Gotwaya ([72] Waller L.A. i Gotway C.A., 2004) jako dynamiczny zbiór teorii i metod analitycznych (metod analizy geostatystycznej), których przedmiotem jest badanie przestrzennych wzorców zachorowalności i śmiertelności. Owe wzorce – nietypowe skupiska (klastery) zwiększonej częstości występowania wad, to podstawowy aspekt zainteresowania epidemiologów. Identyfikacja takich klasterów ułatwia w dalszej kolejności wyszukiwanie działających w nich teratogenów środowiskowych, będących zwykle przyczyną tworzenia się tych skupisk.

W wyszukiwaniu struktur przestrzennych ma zastosowanie wiele metod statystycznych. Niektóre z nich opierają się na danych zapisanych na mapie za pomocą punktu, inne wielokąta a jeszcze inne pola. Najbardziej precyzyjne analizy przestrzenne dotyczą najprostrzej postaci – punktu. W przypadku danych dotyczących ludności przemieszczenie się powoduje duże trudności z przypisaniem osoby do konkretnego punktu na mapie. Jeśli nawet miejsce zamieszkania uznane będzie za najbardziej optymalny punkt, to w przypadku analizy struktur przestrzennych wad wrodzonych zagęszczenie grupy badanej nie musi odzwierciedlać obszarów zwiększonego lub zmniejszonego ryzyka występowania wady wrodzonej. Oczywiście jest bowiem, że zagęszczenie punktów reprezentujących grupę badaną pojawi się w tych obszarach, które są gęściej zaludnione. Aby wyszukiwać klastery wad wrodzonych bazując tylko na grupie badanej, należy zbadać czy otrzymana struktura

przestrzenna zmienia się na przestrzeni krótkiego okresu czasu. Okres ten musiałby być na tyle krótki, by nie miał na niego wpływu migracja ludności, czyli zmieniająca się struktura gęstości zaludnienia. Ponieważ np. wybudowanie nowego osiedla powoduje, że na danym terenie rodzi się więcej dzieci niż wcześniej, a więc zwiększona zostaje szansa na urodzenie dziecka z wadą wrodzoną. Przykład ten pokazuje potrzebę zastosowania grupy odniesienia – jest nią zwykle badana populacja. Mając dane populacyjne dotyczące urodzeń żywych i dane dotyczące występowania wad wrodzonych można wyznaczyć współczynnik częstości (*ang. prevalence*) występowania wady. Tu jednak dane muszą zostać zagregowane. Współczynnik ten jest wówczas wyliczany dla takiej zagregowanej dla wielokątu powierzchni. Zwykle są to dzielnice, całe miejscowości, gminy, województwa itd. Analizy statystyczne można prowadzić na wielokątach, ale można również wyznaczyć jeden charakterystyczny punkt, który może reprezentować wielokąt. W tej sytuacji nie bada się zagęszczenia punktów, ale analizuje przypisaną im wartość prewalencji.

W pracy wykorzystano następujące metody analizy geostatystycznej:

- **statystykę globalną Morana** – do przedstawienia siły i kierunku efektu klasteryzacji,
- **statystykę skanującą Kulldorff'a i Lokalne Współczynniki Autokorelacji Przestrzennej (LISA)** – do identyfikacji klasterów. Porównano wyniki uzyskane przy pomocy tych statystyk.



## 8.1 DOBÓR GRUPY BADANEJ DLA ZDEFINIOWANEJ POPULACJI

Wyszukiwaniem klasterów zwiększonej częstości występowania izolowanego CL±P objęto ok. 85% terenu Polski w roku 2005 i 2006 (*Rycina 3*).

Informacje zbierane przez PRWWR dotyczące dzieci z wadami wrodzonymi pozwalają na przypisanie każdego przypadku do określonego miejsca na mapie poprzez podanie miejsca zamieszkania matki. Z reguły znana jest informacja o gminie oraz konkretna miejscowość. Jednakże Główny Urząd Statystyczny, udostępnia informacje o urodzeniach żywych co najwyżej na poziomie gminy. Jednocześnie nie ma możliwości uzyskania z GUS, poza informacją o gminie, takich danych jak: wiek matki wraz z kolejnością ciąży, wykształceniem, palenie tytoniu przez matkę w czasie ciąży itp., a informacja o dacie urodzenia dziecka ogranicza się do roku urodzenia. Stąd, uwzględniając sposób udostępniania informacji o urodzeniach żywych przez GUS, za najmniejszą możliwą w tej analizie jednostkę podziału geograficznego uznano **gminę**, a za najmniejszą możliwą jednostką podziału czasu uznano **rok**.

### 8.1.1 Grupa badana w analizie czynników czasowo-przestrzennych

Kryterium włączenia do analizy była kompletność informacji dotyczącej gminy zamieszkania i roku porodu. Spośród 448 matek dzieci z izolowanym CL±P 443 matki spełniały warunki włączenia do analizy.

### 8.1.2 Informacje o populacji badanej

Punktem odniesienia przeprowadzanej analizy była populacja licząca 628691 dzieci żywo urodzonych w poszczególnych gminach. Ponieważ teren objęty analizą obejmuje 2558 gmin, to nie było tutaj możliwe zastąpienie danych populacyjnych odpowiednio liczną i reprezentatywną grupą kontrolną.

## 8.2 DANE PRZESTRZENNE I WIZUALIZACJA

### 8.2.1 System informacji geograficznej

Istnieje wiele definicji Geograficznego Systemu Informacji – GIS (ang. Geographic Information System). Ważniejsze z nich można znaleźć w książce Lai ([50] Lai P.C. i inni, 2009). Według definicji Burrough'a ([13] Burrough P. A., 1986), GIS to potężny zestaw narzędzi do zbierania, przechowywania, wyszukiwania, przetwarzania i wyświetlania danych przestrzennych pochodzących z realnego świata.

Baza danych GIS zawiera opisy obiektów geograficznych i zasadniczo składa się z dwóch części, zawierających dwa różne rodzaje danych: **dane przestrzenne i dane opisowe**.

### 8.2.2 Dane przestrzenne

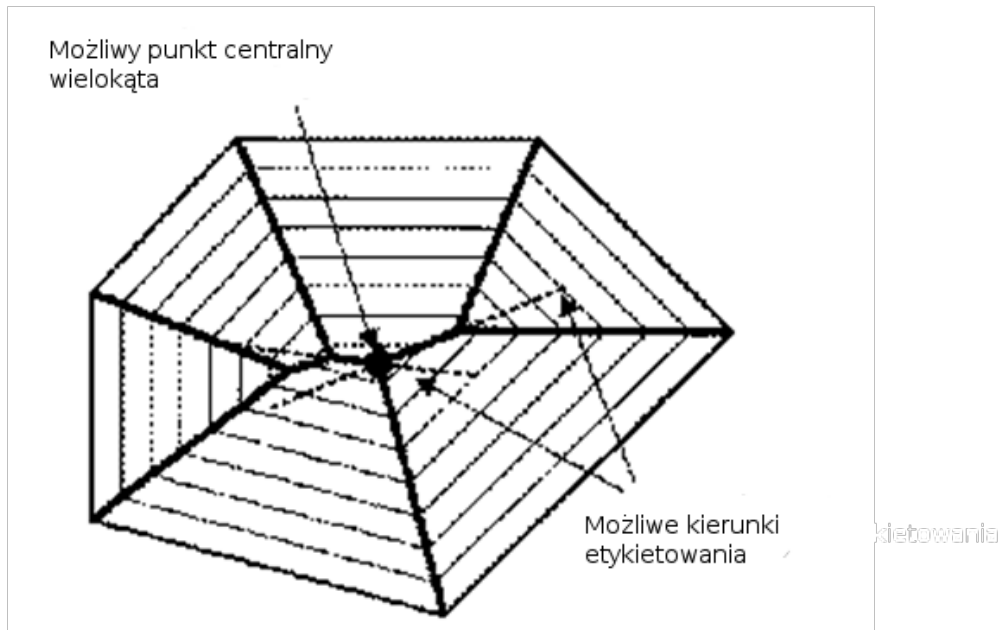
Dane przestrzenne zawierają informacje zarówno o kształcie i lokalizacji bezwzględnej poszczególnych obiektów w wybranym układzie odniesienia, jak również o ich rozmieszczeniu wzajemnym względem innych obiektów (topologia). Obecnie prym wiodzie zapis danych do postaci wektorowej. Powstaje ona najczęściej jako przetworzenie materiału źródłowego, który metodą skanowania zostaje zapisany w formie rastrowej lub został dostarczony w tej formie (zdjęcie satelitarne, zeskanowana mapa).

Tworząc topologiczny model wektorowy uzyskujemy informacje dotyczące między innymi sąsiedztwa i odległości obiektów niezbędne do analizy klastrow. Obiektami przestrzennymi poddanymi niniejszej analizie są gminy identyfikowane jednoznacznie w oparciu o kod terytorialny - teryt. Siedmiocyfrowy kod **teryt** podziału administracyjnego dla kraju jest trójczłonowy, a jego poszczególne człony oznaczają:

- I człon - dwucyfrowy symbol województwa
- II człon - dwucyfrowy symbol powiatu
- III człon - trzycyfrowy symbol gminy, w którym: trzecia cyfra stanowi symbol rodzaju jednostki czyli rodzaj gminy

Położenie geograficzne gmin definiowane jest przez granice ich obszaru – poprzez **wielokąty** (Rycina 10). W formie wielokątów zapisanych w postaci wektorowej w skali 1:250000 zapisane są pozyskane mapy. Dodatkowo dla celów dalszych analiz położenie gminy przedstawiono również za pomocą **centroidy** wyznaczonej na podstawie wielokątu (Rycina

10). Wyliczona centroida jest punktem znajdującym się wewnątrz wielokąta i reprezentuje geometryczne uściślenie intuicyjnego „środką” (Rycina 9)

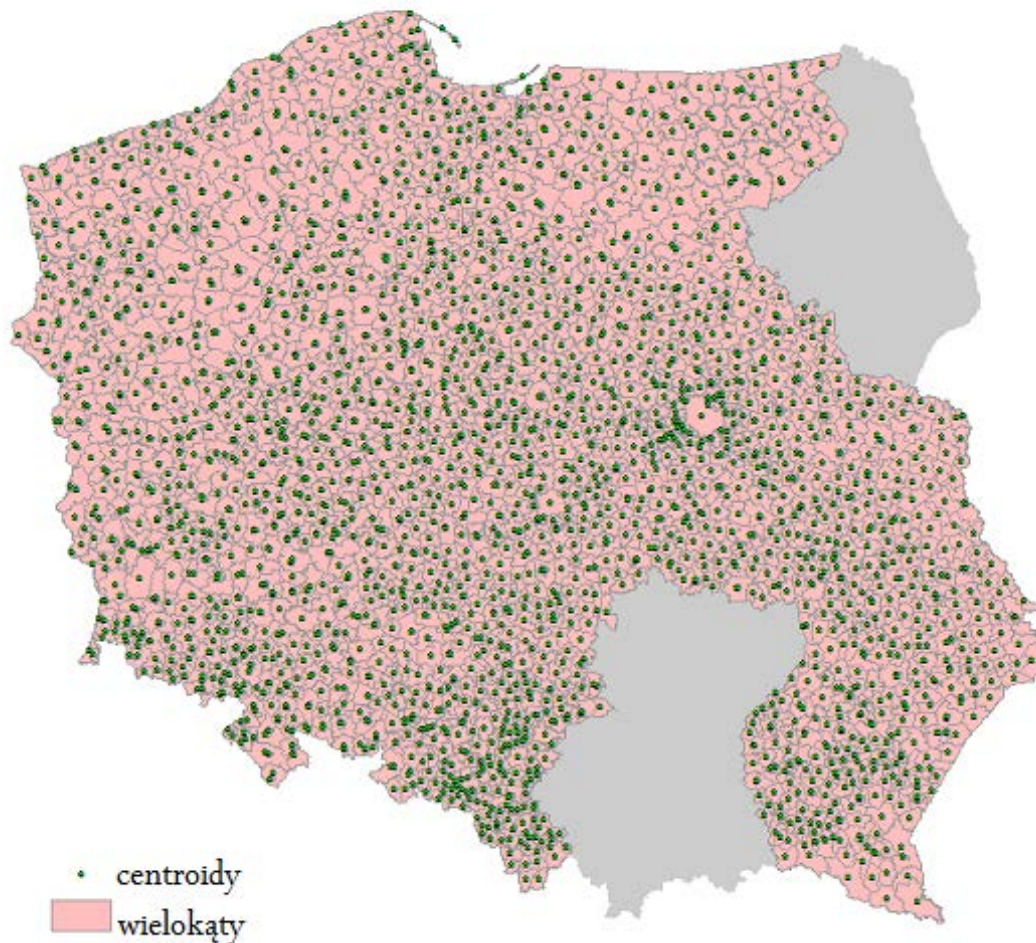


**Rycina 9 Przykładowa centroida wyznaczony na podstawie wielokąta**  
Źródło: Geographic Information Analysis, David O’Sullivan, David J. Unwin 2010

By różnice w gęstości zaludnienia w gminach miejsko-wiejskich miały odzwierciedlenie w tworzonych centroidach rozbito je na część miejską i część wiejską gminy. Stąd gminy dzielone są w zależności od wielkości populacji zgodnie z kodem teryt (ostatnia cyfra jego III członu) na:

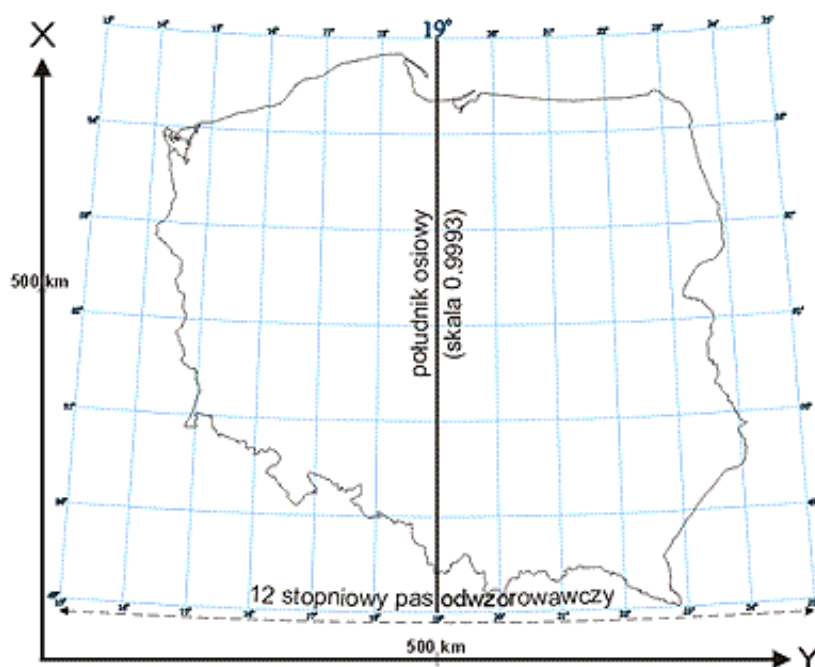
- gminy miejskie,
- gminy wiejskie,
- miasto w gminie miejsko-wiejskiej,
- obszar wiejski w gminie miejsko-wiejskiej.

Dla analizowanych województw stanowią one łącznie 2558 oddzielnych obszarów – nazywanych w dalszym ciągu gminami.



**Rycina 10. 2558 gminy Polski objętych analizą: jako wielokąty i jako punkty - centroidy wyliczone na podstawie wielokątów**

Rzeczywiste położenie obiektu w przestrzeni wyrażone w postaci współrzędnych geograficznych zostało przeliczone do jego reprezentacji na płaszczyźnie mapy. Dla celów niniejszej analizy przekształcenia dokonano w oparciu o układ 1992 (Rycina 11).



Rycina 11 Układ „1992” – wykorzystuje jednostrefowe odwzorowanie Gaussa-Krugera elipsoidy GRS-80 (WGS-84) z południkiem osiowym 19 stopni.

### 8.2.3 Dane opisowe

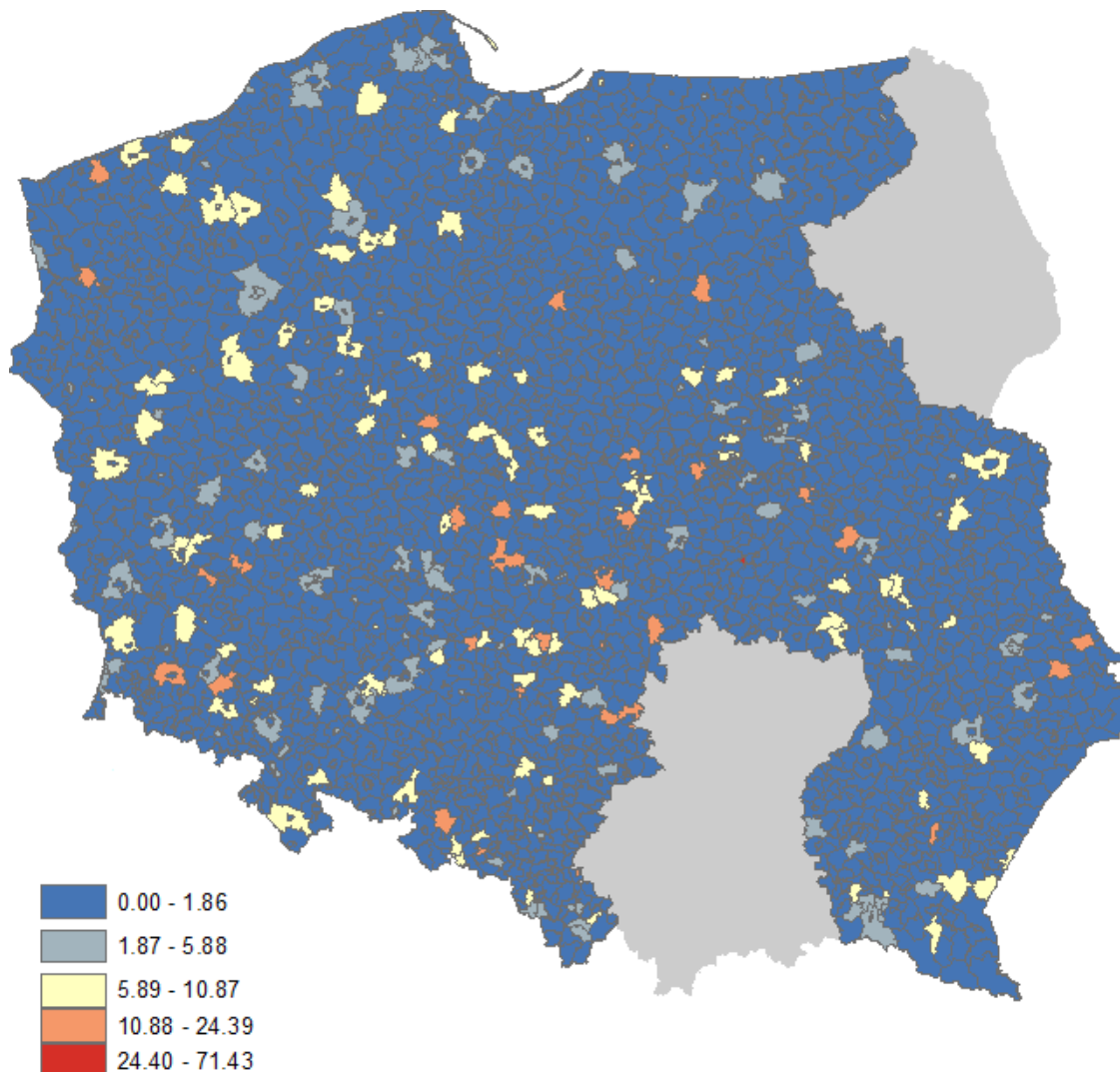
Dane opisowe zapisane są w postaci relacyjnej bazy danych i stanowią dodatkowe informacje o poszczególnych obiektach geograficznych. Dla potrzeb niniejszej analizy odpowiednie dane PRWWR oraz UŻ zostały przekształcone tworząc relacyjną bazę danych (MySQL). Zbudowana baza złożona jest z tabel zawierających: dane o ilości urodzeń dzieci wg poszczególnych grup wad oraz dane o ilości urodzeń żywych. Informacje te zapisane są w rozbiciu na rok i gminę urodzenia dziecka. W bazie 2558 gmin identyfikowanych jest jednoznacznie w oparciu o kody teryt.

### 8.2.4 Statystyki opisowe

Podstawowe statystyki opisowe wad wrodzonych to współczynniki częstości występowania tych wad w określonym roku. Ze względu, na to że najmniejszą jednostką administracyjną położenia geograficznego dzieci w zbudowanej bazie jest gmina, częstość występowania wady wrodzonej wyznaczono dla jej obszaru.

Do przedstawienia współczynników na płaszczyźnie mapy używa się **kartogramu**. By zwiększyć czytelność przedstawianych map nie wyrysowuje się dokładnej wartości współczynników dla każdej gminy, ale dzieli się je na klasy. Istnieje wiele metod dzielenia

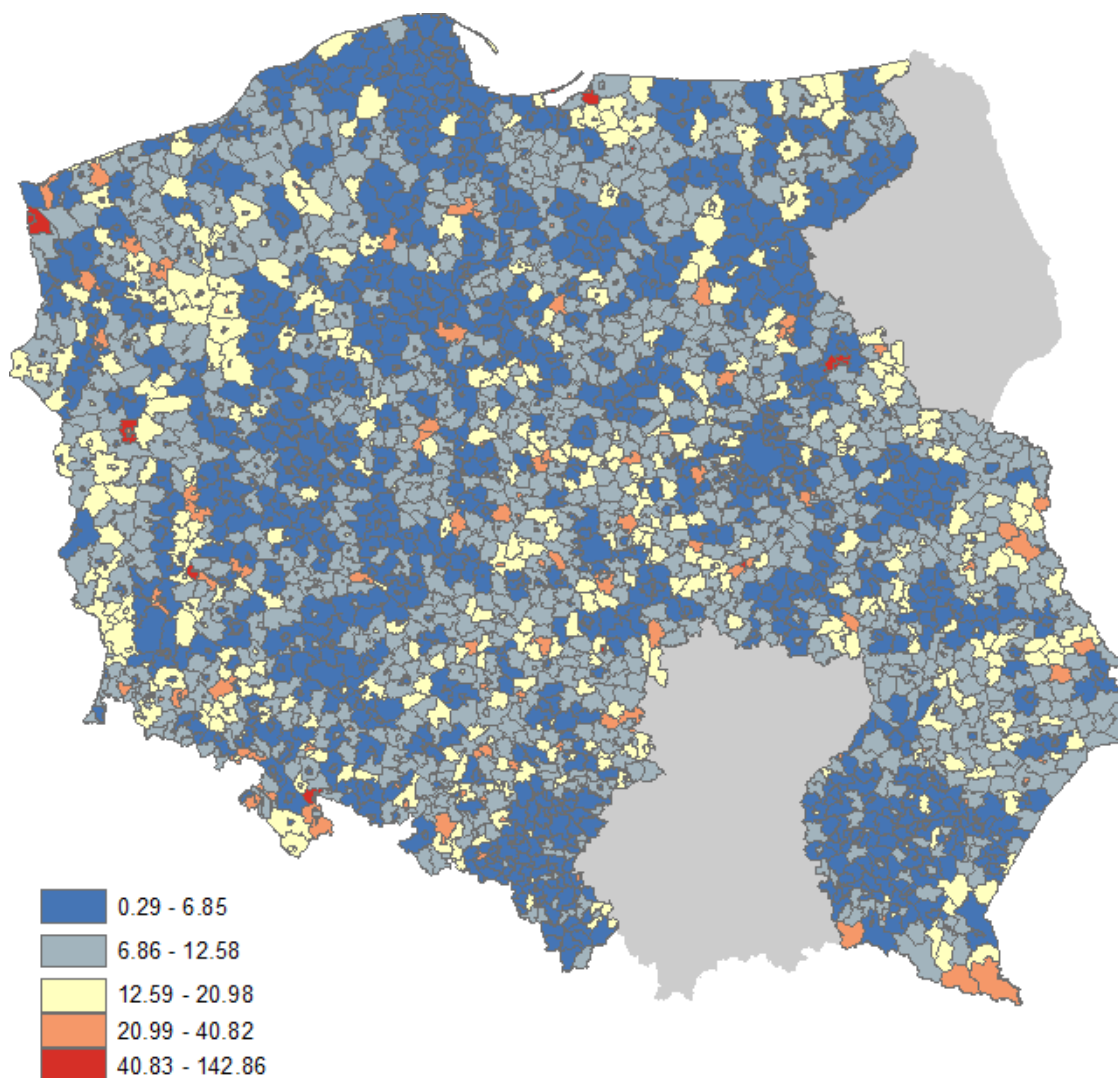
współczynników na klasy ([50] Poh-Chin Lai i inni, 2009). Zwykle tworzy się 5 lub 7 klas. W pracy podziału dokonano przydzielając gminy na podstawie ich współczynnika na 5 klas bazując na naturalnym ułożeniu grup w rozkładzie danych (Rycina 12 – dla roku 2005 i 2006 łącznie, Załącznik 5: Rycina A1 i Rycina A2). W tym celu użyto statystycznej formuły naturalnego podziału – Jenks' Natural Breaks ([35] Jenks G.F., 1967) tworzącej podgrupy w uporządkowanym zbiorze danych w taki sposób by zminimalizować wariancję wartości wewnątrz grup.



**Rycina 12** Surowe współczynniki występowania izolowanego CL±P w latach 2005, 2006 w badanych województwach, przedstawione według naturalnego podziału Jenks'a

Bardzo często, szczególnie w epidemiologii, spotyka się zbiory danych, w których liczności grupy badanej dla wielu obszarów są bardzo małe, wiele z nich to 0. Stąd wyliczane współczynniki są zbyt wrażliwe (niestabilne) na małe zmiany danych. Prosty sposób poradzenia sobie w takiego typu przypadkach, zaproponowany przez Cressi i jego współpracowników ([18] Cressi N.A., 1993), polega na dodaniu 1 do liczności dzieci z wadą w

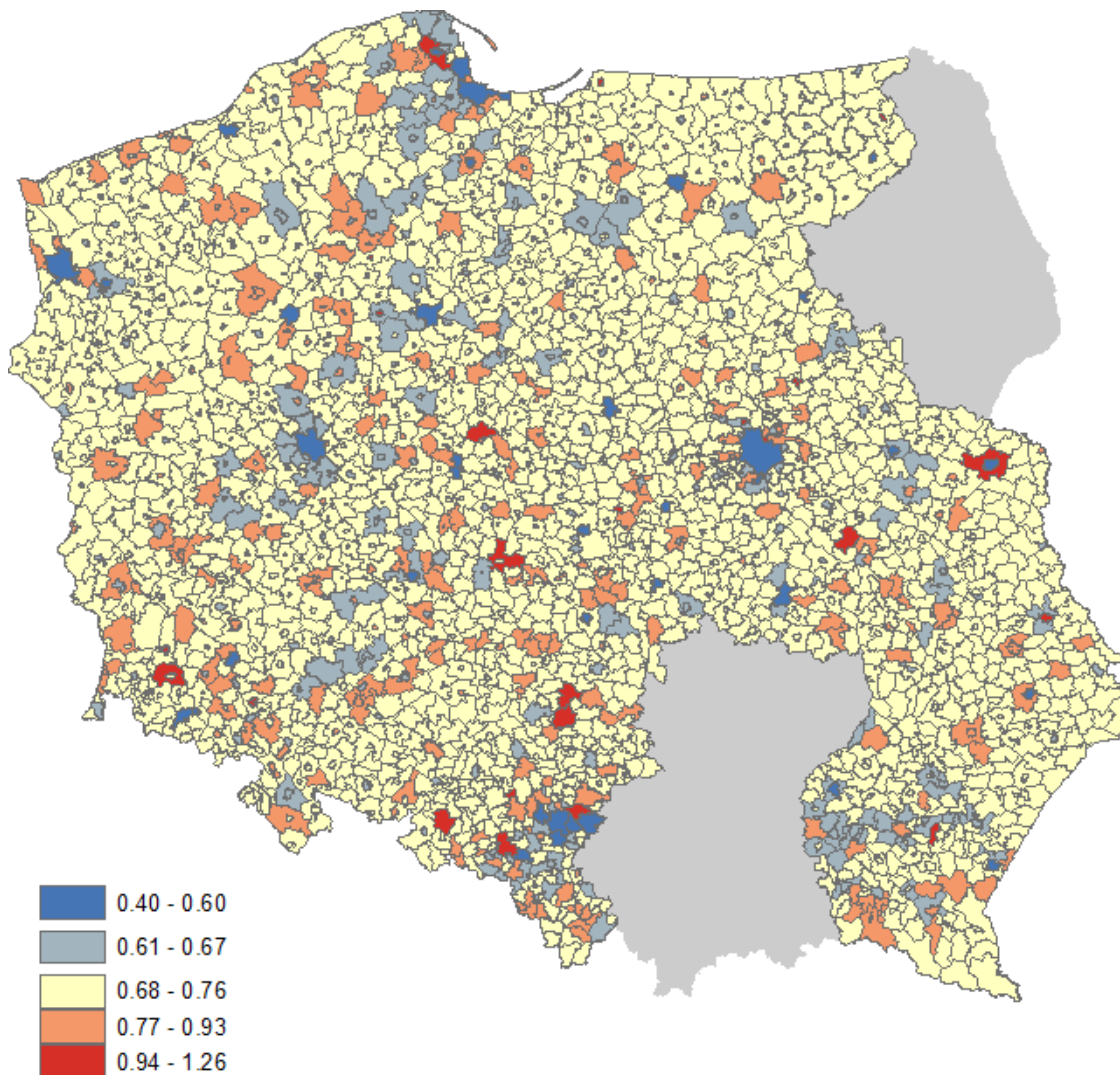
każdej gminie przed wyliczeniem współczynników częstości występowania wad (Rycina 13 – dla roku 2005 i 2006 łącznie, Załącznik 5: Rycina A3 i Rycina A4). Ten zabieg pozwala wprowadzić rozróżnienie pomiędzy obszarami określanymi poprzednio jako 0, ale problem niestabilności pozostaje.



**Rycina 13** Surowe współczynniki występowania izolowanego CL±P w latach 2005, 2006 w badanych województwach – zmodyfikowane zgodnie z metodą Cressi, przedstawione według naturalnego podziału Jenks'a

Inną alternatywą jest zastosowanie empirycznego wygładzenia Bayesa (ang. Empirical Bayes – EBS) – poprzez zastosowanie tej metody rozpiętość surowych współczynników kurczy się i staje bardziej zbliżona do średniej. Metoda ta opiera się na wyliczonej średniej ważonej z różnicy pomiędzy surowymi współczynnikami dla każdej gminy i średniej całego obszaru. Wagi są proporcjonalne do badanej populacji w każdej gminie. Na skutek tego redukcji ulega również niestabilność. Małe gminy (o małej populacji), mające tendencję do znacznej zmiany

wyświetlanych współczynników a duże do niewielkiej zmiany, stabilizują swoje wartości zbliżając się do średniej (Rycina 14).

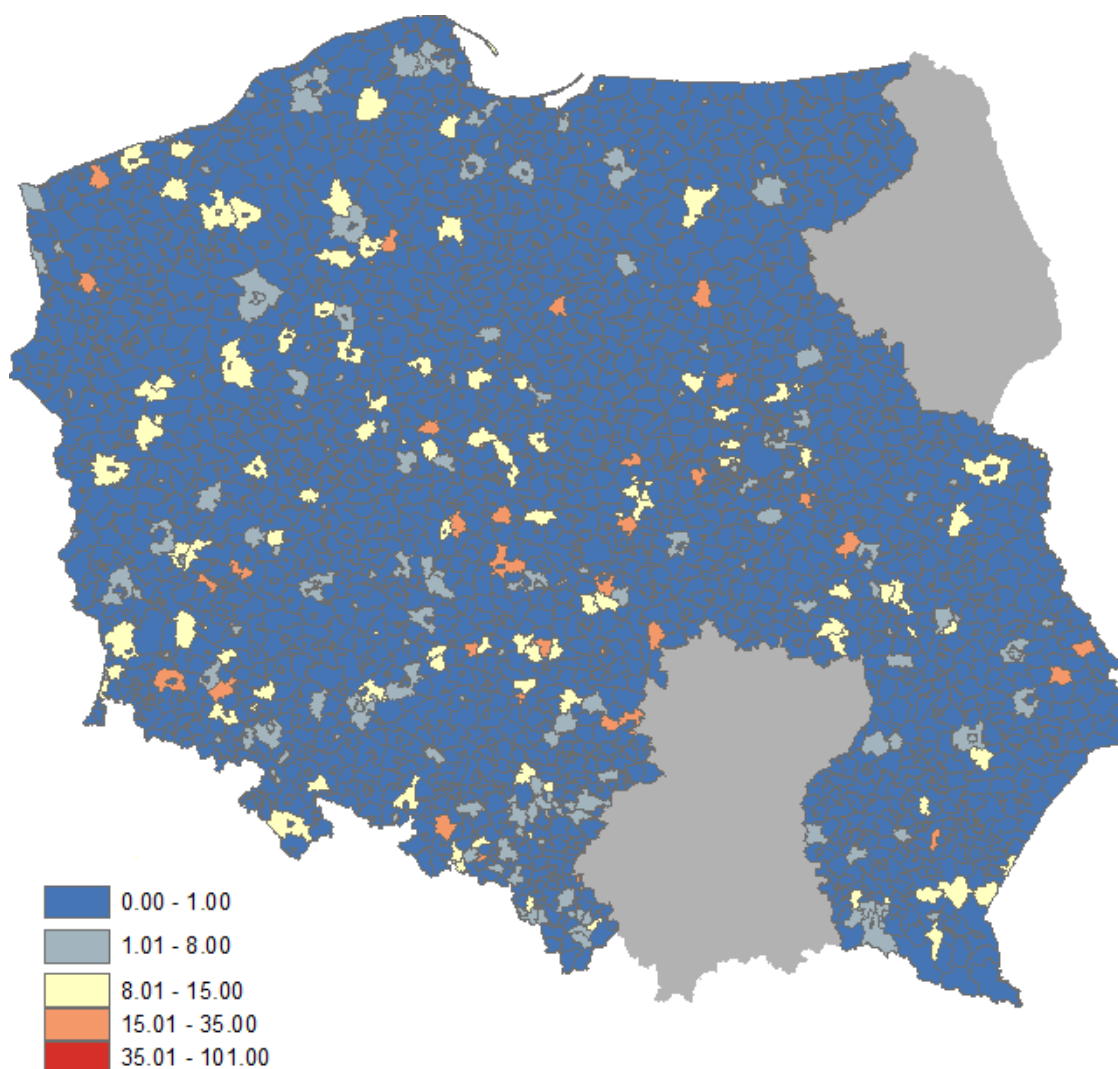


**Rycina 14 Wyglądzone współczynniki (EBS) występowania izolowanego CL±P w latach 2005, 2006 w badanych województwach, przedstawione według naturalnego podziału Jenks'a**

Matematyczne szczegóły i dalsze uwagi dotyczące wygładzania współczynników metodą Bayesa można znaleźć w pracach Anselina oraz Boscoe ([4] Anselin L. i inni, 2004; [12] Boscoe F. P. i inni, 2003).

Mapy nadmiernego ryzyka (ang. Excess Risk Maps) – przedstawiają natomiast zwiększone lub zmniejszone w stosunku do średniej ryzyko wystąpienia wady. Kolor niebieski reprezentuje gminy w których współczynnik jest mniejszy niż średni współczynnik wyliczony dla całego obszaru (iloraz tych współczynników  $<1$ ), czerwony reprezentuje gminy w których współczynnik jest większy niż średni współczynnik wyliczony dla całego obszaru (iloraz tych współczynników  $>1$ ) (Rycina 15).





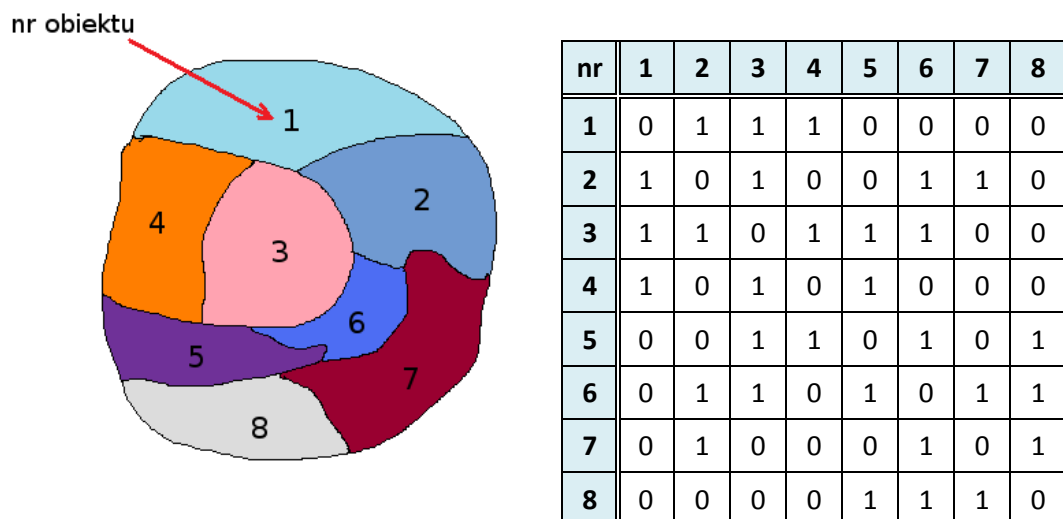
**Rycina 15** Mapa nadmiernego ryzyka występowania izolowanego CL±P w latach 2005, 2006 w badanych województwach, przedstawiona według naturalnego podziału Jenks'a

Zaproponowane sposoby przedstawiania badanego zjawiska na płaszczyźnie mapy rozwiązują inny ważny problem w epidemiologii jakim jest ochrona danych osobowych. Mapy muszą być tak zbudowane, by nie można było zidentyfikować konkretnych osób na ich podstawie. Stąd raczej nie używa się punktów do oznaczenia lokalizacji miejsca zamieszkania poszczególnych pacjentów. Przedstawione rozwiązanie polega głównie na przedstawieniu współczynników częstości występowania wady dla zagregowanych jednostek, ale stosuje się również inne metody maskowania ([33] Hampton K. H. i inni, 2010).

## 8.2.5 Sposoby określenia sąsiedztwa

Wybór sposobu rozumienia sąsiedztwa determinuje wyniki analiz. Jest jednym z pierwszych etapów statystyki przestrzennej. Przestrzenną strukturę sąsiedztwa można opisać za pomocą macierzy (lub grafu) sąsiedztwa a także macierzy wag. Macierze te można tworzyć na dwa sposoby: korzystając z kryterium wspólnej granicy lub wyznaczając współrzędne środków regionów i obliczając odległości pomiędzy nimi.

- Macierz sąsiedztwa** jest tablicą kwadratową o zerowych wartościach na diagonalu, gdzie sąsiedztwo pomiędzy elementem  $i$  oraz  $j$  jest oznaczane wartością binarną. Według **kryterium sąsiedztwa wspólnej granicy** wartość 0 oznacza brak sąsiedztwa pomiędzy elementami tzn. obszary te nie mają wspólnej granicy (wspólnego odcinka lub wspólnego punktu), wartość 1 występowanie sąsiedztwa pomiędzy elementami tzn. obszary mają wspólną granicę. Wspólna granica zwyczajowo rozumiana jest jako wspólny odcinek o niezerowej długości (tzn. odcinek dłuższy niż 1 punkt) – jest to sąsiedztwo typu **Rook**, lub jako dowolny odcinek (również o zerowej długości, czyli punkt) – jest to sąsiedztwo typu **Queen**. W podobny sposób można stworzyć macierz sąsiedztwa w oparciu o **kryterium odległości** tzn. macierz sąsiadów w promieniu  $d$  km. Wartość 0 oznacza że elementy macierzy są oddalone od siebie o więcej niż  $d$  km a wartość 1 o odległość mniejszą bądź równą  $d$  km. Przy czym odległość definiowana jest przez wybraną metrykę.



Rycina 16 Przykład macierzy sąsiedztwa wspólnej granicy 8 wielokątów – obiektów na jakie została podzielona mapa

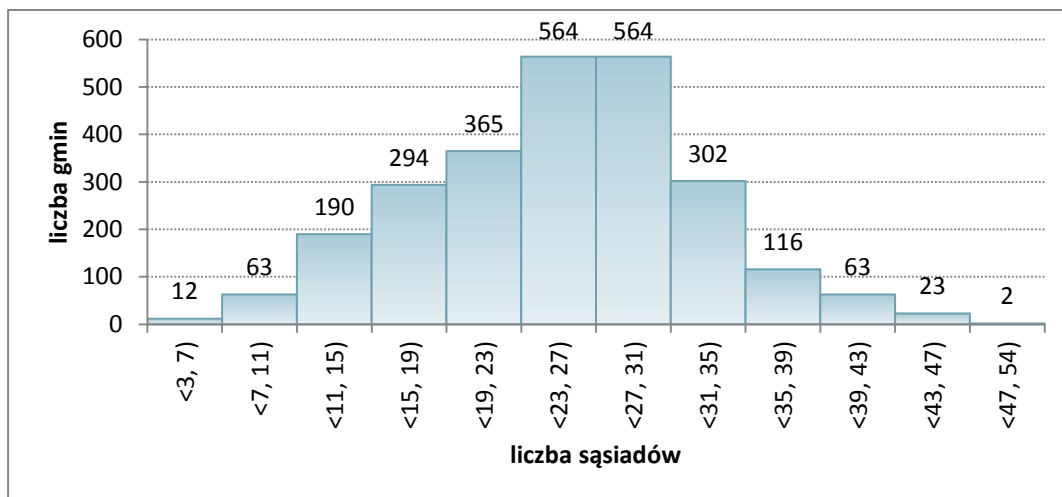
- **Macierz wag  $W$**  to macierz, która niesie więcej informacji niż tylko istnienie sąsiedztwa lub jego brak. Może zawierać informacje o odległości sąsiadujących obiektów, lub stopniu sąsiedztwa (sąsiad bezpośredni, sąsiad sąsiada,... ), czy też długości wspólnej granicy między sąsiadami. Najczęściej jednak uzyskuje się ją poprzez standaryzację rzędami do jedynki macierzy sąsiedztwa, czyli dla każdego wiersza  $i$ :

$$\sum_{j=1} w_{ij} = 1, \quad 7.1$$

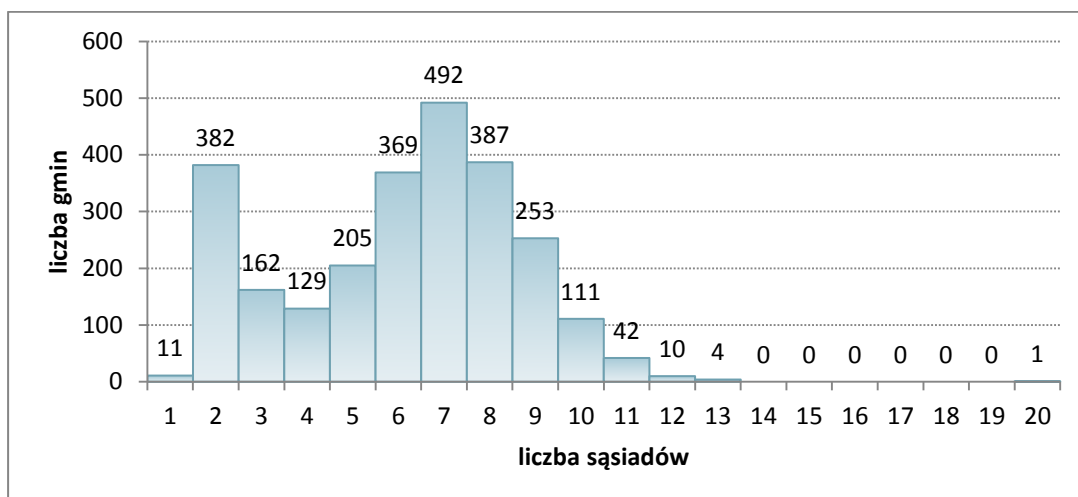
gdzie element  $w_{ij} = \frac{1}{n}$  gdy region ma  $n$  sąsiadów.

W niniejszej pracy za kryterium sąsiedztwa przyjęto kryterium odległości. Zdecydowano się na:

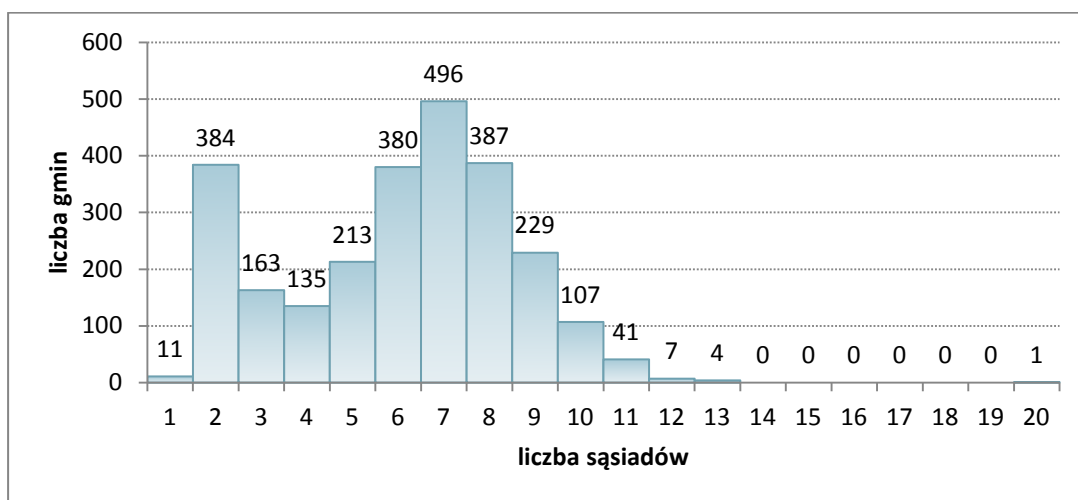
- Uznanie za sąsiadujące względem zadanej gminy, te gminy których centroidy znajdują się wewnątrz okręgu o promieniu 30 km. Przy czym odległość wyrażona jest poprzez metrykę euklidesową. Tak przyjęte założenie (wykorzystywane dalej przez analizę Morana i LISA) oznacza, że będziemy sprawdzać, czy gminy są podobne pod względem częstości występowania badanej wady na przestrzeni 30 kilometrowego obszaru. W promieniu 30 km ilość sąsiadów każdej gminy jest dużo większa niż w przypadku metody Queen i Rook (Rycina 17, Rycina 18, Rycina 19) . Sposób dokonania podziału gmin miejsko wiejskich na miasto i obszar wiejski, w przypadku sąsiedztwa typu Queen i sąsiedztwa typu Rook, powodował występowanie zbyt wielu gmin mających bardzo małą liczbę sąsiadów (blisko 400 gmin posiadało wówczas mniej niż trzech sąsiadów).



Rycina 17 Histogram liczby sąsiadów odległych o mniej niż 30 km



Rycina 18 Histogram liczby sąsiadów – sąsiedztwo typu Queen



Rycina 19 Histogram liczby sąsiadów – sąsiedztwo typu Rook

- Wyliczenie macierzy wag standaryzowanej rzędami do jedynki (macierz ta wykorzystywana jest przez statystykę globalną Morana i LISA).
- Ograniczenie wielkości wyszukiwanych klasterów do okręgu o promieniu długości 30 km (ograniczenie to wykorzystuje statystyka skanująca Kulldorff'a).

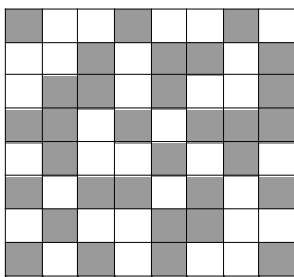
Są to optymalne założenia spełniające oczekiwania co do wielkości klasterów a także czasu prowadzenia analizy.

## 8.2.6 Autokorelacja przestrzenna

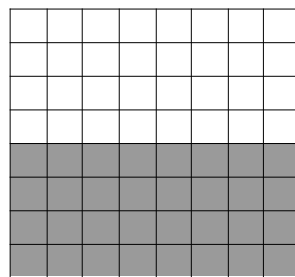
Autokorelacja przestrzenna oznacza, że bliskie geograficznie obserwacje są bardziej podobne do siebie niż te odległe. Zjawisko to powoduje tworzenie się klasterów przestrzennych o różnych kształtach. Skrajnym przypadkiem autokorelacji przestrzennej są zmiany skokowe określane mianem przestrzennej heterogeniczności (niestabilności procesów). Przykładem takim mogą być klasterzy wysokich wartości w miastach i niskich wartości poza miastami ([24] Florax R. i Nijkamp P., 2003)

Autokorelacja przestrzenna może nie występować – mówimy wówczas o **przestrzennej losowości**. Uzyskany rozkład przestrzenny jest tak samo prawdopodobny jak w każdy inny rozkład (Rycina 20a)). Gdy wartości sąsiednie są sobie podobne, to możemy mówić o występowaniu **autokorelacji dodatniej** (Rycina 20b)). **Ujemna autokorelacja** występuje wówczas, gdy obszary sąsiednie są bardziej różne niż mogłoby to wynikać z rozkładu losowego (Rycina 20c)).

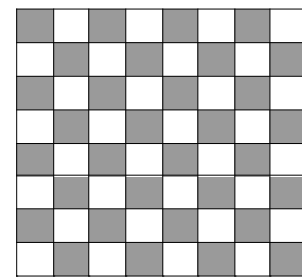
a) losowość przestrzenna



b) autokorelacja dodatnia



c) autokorelacja ujemna



Rycina 20 Przykładowy rozkład przestrzenny

## 8.3 PRZEGLĄD METOD BADANIA KLASTERÓW

Fotheringham ([25] Fotheringham A.S., 1997, [26] Fotheringham A.S., 2000) wyróżnia cztery główne grupy metod służące badaniu zależności przestrzennej: 1) Automaty Analizy Geograficznej (GAM), 2) Lokalne Współczynniki Autokorelacji Przestrzennej (LISA), 3) Geograficzna Regresja Ważona, 4) Matematyczne modele przepływów. O ile grupa trzecia i czwarta służą modelowaniu lokalnej zależności przestrzennej, o tyle pierwsza i druga są metodami pozwalającymi na identyfikację klasterów. GAM jest zbiorem metod przeznaczonych do analizy danych punktowych, zaś LISA dla danych zagregowanych. W tej pracy przedstawione będzie zastosowanie statystyki skanującej Kulldorff'a, która z jednej strony nawiązuje do GAM, a z drugiej strony przy pewnych założeniach może zostać wykorzystana również do danych zagregowanych. Klustery zidentyfikowane na podstawie statystyki Kulldorff'a porównane zostaną z wynikami uzyskanymi metodą LISA.

Geographical Analysis Machine (GAM) została zaproponowana przez Openshawa ([61] Openshaw S. i inni, 1987) i jest wykorzystywana dla analizy danych punktowych. Wyznacza ona zbiory okręgów różnej wielkości, o środkach w punktach będących węzłami regularnej siatki (zazwyczaj złożonej z kwadratów) pokrywającej analizowany obszar. Rzeczywista liczba przypadków zachorowań na wybraną chorobę wewnątrz każdego okręgu jest następnie porównywana z wartością oczekiwaną przy założeniu, że liczba zachorowań ma rozkład Poissona. Turnbull ([70] Turnbull B.W. i inni, 1990) opracował test pozwalający na badanie lokalizacji i istotności klastrów. Ten test polega na utworzeniu okien w kształcie okręgu zawierających z góry ustaloną liczbę zagrożonych jednostek. Besag i Newell rozwinęli metodę GAM dla sytuacji, gdy badacz ustala liczbę  $k$  będącą oczekiwaną wielkością klastra. Zazwyczaj  $k$  waha się pomiędzy 2 a 10. Metoda została opisana w pracach Besaga i Newella ([7] Besag J. i Newell J., 1991) oraz Alexander i Cuzick ([1] Alexander F.E. i Cuzick J., 1992). Przestrzenna statystyka skanująca Kulldorff'a ([47] Kulldorff M. i Nagarwalla N. 1995, [43] Kulldorff M., 1997) nawiązuje zarówno do GAM Openshawa jak i procedury opracowanej przez Turnbulla. Umożliwia ona wykrycie klastrów nie tylko w kształcie koła, ale również elipsy. Po raz pierwszy została zastosowana w pracy Kulldorff'a i Nagarwalli (1995) do wyznaczenia klasterów zachorowań na białaczkę w stanie Nowy Jork. Przegląd prac prowadzonych w oparciu o przestrzenną statystykę skanującą zawarty został w książce: *Spatial Analysis in Epidemiology* ([67] Stevenson M. i inni, 2008). Kulldorff zaproponował i rozwinął również

czasowo-przestrzenną wersję statystyki skanującej ([45] Kulldorff M. i inni, 1998a; [44] Kulldorff M., 2001; [46] Kulldorff M. i inni, 2005).

Statystyki  $I_i$  Morana oraz  $C_i$  Geary'ego ([3] Anselin L., 1995) wchodzące w skład lokalnych współczynników zależności przestrzennej LISA (ang. Local Indicators of Spatial Association) są lokalnymi wariantami globalnych statystyk przestrzennych i działają na danych zagregowanych. Pozwalają one na wykrycie lokalnej zależności (autokorelacji) przestrzennej, która ujawnia się w postaci klasterów o podobnych wartościach analizowanej zmiennej, oraz kontrastujących względem swojego otoczenia wartości odstających.



## 8.4 ZAŁOŻENIA TEORII STATYSTYKI GLOBALNEJ MORANA ORAZ WYNIKI ZASTOSOWANIA TEJ METODY

Przed przystąpieniem do identyfikacji klasterów, współczynniki częstości występowania izolowanego CL±P dla wszystkich gmin analizowano przy użyciu **globalnej statystyki Morana I**. Statystyka ta jest wykorzystywana do testowania globalnej autokorelacji przestrzennej. Statystkę globalną Morana traktuje się jak współczynnik korelacji i interpretuje podobnie jak współczynnik korelacji liniowej Pearsona. Współczynnik ten nazywany zwykle współczynnikiem autokorelacji zaproponowany został w roku 1950 przez Morana ([60] Moran P.A.P., 1950). Wskazuje on siłę i kierunek przestrzennego efektu aglomeracji i wyraża się wzorem:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad 8.2$$

gdzie:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2$$

$z_i, z_j$  – to wartość atrybutów dla porównywanych obiektów (współczynniki częstości występowania izolowanego CL±P dla poszczególnych gmin),

$\bar{z}$  – to średnia wartość atrybutów dla wszystkich obiektów (wartość oczekiwana częstości występowania izolowanego CL±P dla całego analizowanego obszaru),

$w_{ij}$  – element przestrzennej macierzy wag  $W$  (macierz wag jest standaryzowana rzędami do jedynki),

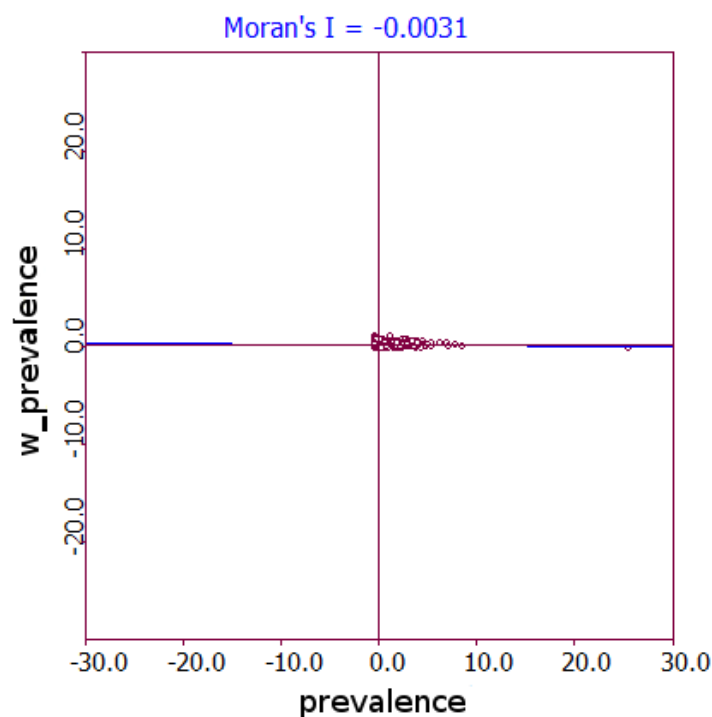
$n$  – liczba obiektów (liczba gmin).

Wartość dodatnia oznacza występowanie klasterów podobnych wartości, wartość 0 brak autokorelacji, a wartość ujemna to występowanie tzw. hot spots czyli zdecydowanie różnych wartości w obszarach sąsiedzkich.

Statystyka I Morana ma dla dużych licznosci rozkład normalny, choć zwykle rozkład tej statystyki otrzymuje się poprzez symulację Monte Carlo. Hipoteza zerowa w tym teście zakłada brak autokorelacji przestrzennej ( $H_0: I = 0$ ).

Wartość statystyki globalnej Morana dla przeprowadzonej analizy wynosi  $I=-0.0031$  a wartość  $p=0.3430$ , co świadczy o braku autokorelacji, tzn. tworzony obszar nie wykazuje tendencji do tworzenia klasterów zwiększonej bądź zmniejszonej częstości występowania izolowanego CL±P. Wyniki te uzyskano wykorzystując macierz wag powstałą w oparciu o macierz sąsiedztwa gmin w promieniu 30 km.

Wykres punktowy statystyki I Morana wykorzystywany jest do przedstawienia związków przestrzennych (klasterów) obserwacji nietypowych (ang. outliers) a także niestabilności przestrzennej ([3] Anselin, 1995). Na osi odciętych wykresu znajduje się badana zmienna standaryzowana (współczynnik częstości występowania izolowanego CL±P), na osi rzędnych zmienna opóźniona przestrzennie (ang. spatial lag). Opóźnienie przestrzenne jest średnią ważoną z wartości zmiennej (wartości współczynników częstości występowania izolowanego CL±P) w gminach sąsiadujących według wybranej macierzy wag. Współczynnik nachylenia linii regresji to wartość globalnej statystyki I Morana.



Rycina 21 Wykres punktowy statystyki I Morana surowych współczynników występowania izolowanego CL±P w latach 2005, 2006 w badanych województwach

Interpretacja wykresu Morana rozpoczyna się od analizy poszczególnych ćwiartek wykresu co można zapisać w postaci tabeli w następujący sposób:

**Tabela 11 Interpretacja 4 ćwiartek wykresu Morana**

	niskie wartości w rejonach sąsiadujących	wysokie wartości w rejonach sąsiadujących
wysokie wartości w rejonie badanym	<b>wysokie-niskie</b> ujemna autokorelacja	<b>wysokie-wysokie</b> dodatnia autokorelacja
niskie wartości w rejonie badanym	<b>niskie-niskie</b> dodatnia autokorelacja	<b>niskie-wysokie</b> ujemna autokorelacja

W przeprowadzonej analizie nachylenie prostej jest bliskie 0 (prosta pokrywa się z osią odciętych). Nie można omawiać tu rodzaju autokorelacji, ponieważ analizowany obszar nie wykazuje tendencji do tworzenia istotnych statystycznie klasterów zwiększonej lub zmniejszonej częstości występowania izolowanego  $CL \pm P$ . Można się spodziewać, że wyszukiwanie klasterów nie przyniesie oczekiwanych rezultatów, a więc nie będzie można wiązać występowania tej wady z teratogenem występującym w określonej lokalizacji geograficznej.

W analizie Morana wzbogaconej o informację dotyczącą roku urodzenia dziecka, współczynniki dla danej gminy w roku 2005 (oś X) odnoszone są do współczynników gmin sąsiednich w roku 2006 (oś Y). Wyniki tej analizy (Załącznik 6), podobnie jak analizy przestrzennej (Rycina 21) świadczą o braku autokorelacji.

## 8.5 LOKALNA ESTYMACJA I WYSZUKIWANIE KLASTERÓW

### 8.5.1 Założenia teorii Lokalnych Współczynników Autokorelacji Przestrzennej (LISA)

Jednym z testów należących do Lokalnych Współczynników Zależności Przestrzennej (LISA) jest **statystyka lokalna Morana**  $I_i$ . Bada ona, czy region jest otoczony przez sąsiadów o podobnych czy różnych wartościach współczynnika. Wartość statystyki tego testu wyznaczana jest dla każdego obiektu (dla każdej gminy) i definiowana jest analogicznie jak jej globalny odpowiednik:

$$I_i = \frac{(z_i - \bar{z}) \sum_{j=1}^n w_{ij} (z_j - \bar{z})}{\frac{\sum_{j=1}^n (z_j - \bar{z})^2}{n}}$$

8.3

Wartość dodatnia oznacza, że gmina jest otoczona przez gminy o podobnej do niej wartości badanego współczynnika, wartość ujemna oznacza, że gmina jest otoczona przez gminy o znacząco różnych wartościach.

Rozkład statystyki lokalnej Morana otrzymuje się poprzez symulację Monte Carlo. Hipoteza zerowa w tym teście zakłada brak autokorelacji przestrzennej ( $H_0: I_i = 0$ ).

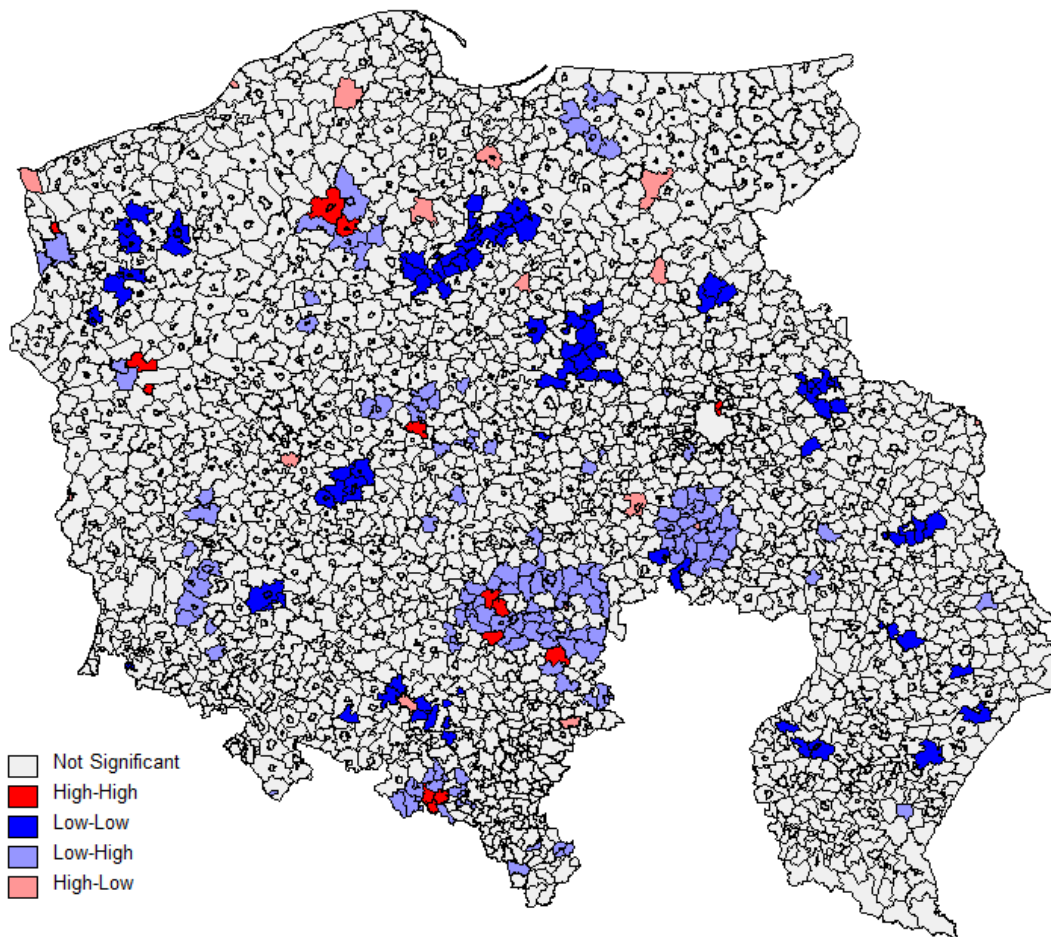
Ze względu na to, że lokalna statystyka Morana może być skorelowana dla sąsiadujących gmin (kilka gmin jest rozważanych kilkakrotnie – przy wyznaczaniu kolejnych wartości  $I_i$ ) poziom istotności powinien być skorygowany o średnią liczbę sąsiadów ([3] Anselin, 1995).

Wykorzystywana jest tu poprawka Bonferroniego  $\alpha_1 = \frac{\alpha}{k}$  lub Sidaka  $\alpha_1 = 1 - (1 - \alpha)^{1/k}$ , gdzie  $k$  jest średnią liczbą sąsiadów.

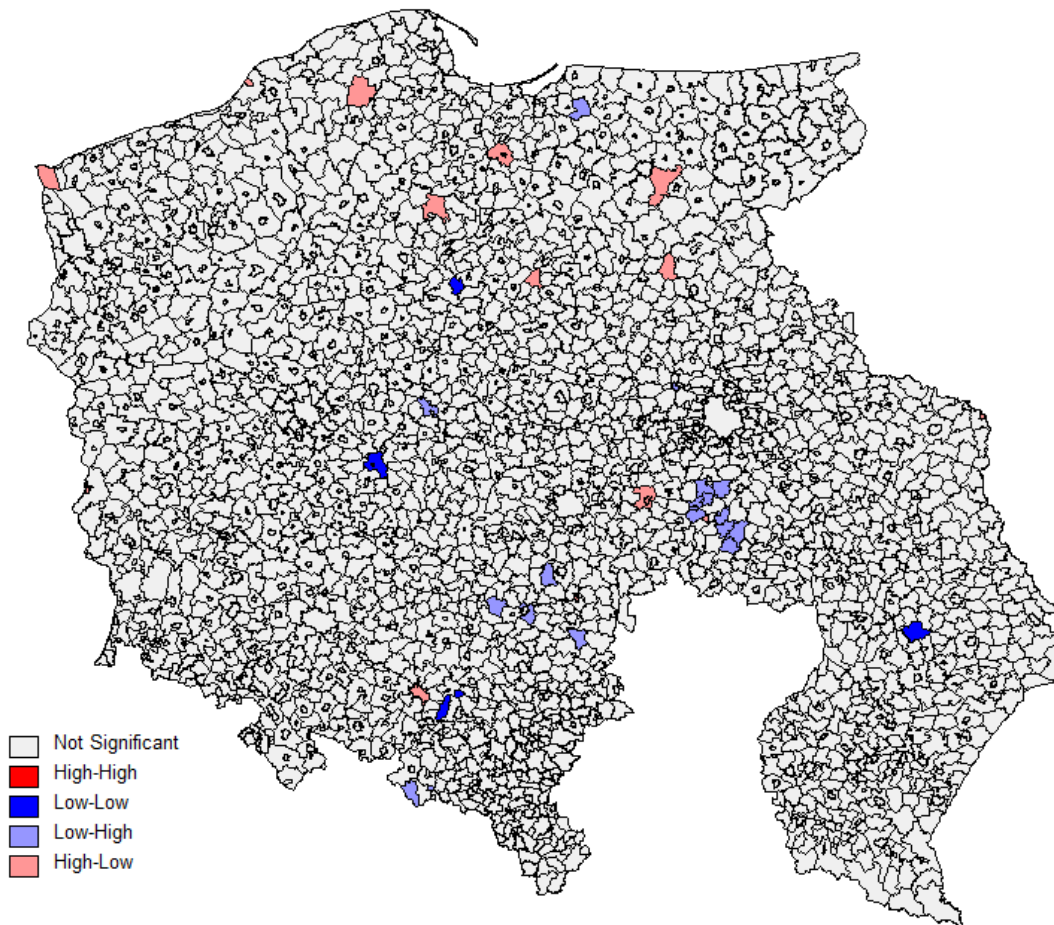
### 8.5.2 Wyniki zastosowania teorii Lokalnych Współczynników Autokorelacji Przestrzennej (LISA)

Kombinacja informacji z wykresu punktowego Morana i z istotności statystyk LISA przedstawia na mapie tzw. reżimy przestrzenne (Rycina 22, Rycina 23). Interesująca z punktu widzenia niniejszego badania jest dodatnia autokorelacja, a zatem klaster wyższych wartości współczynników. Są to te gminy, które mają wysoką wartość współczynnika w badanej gminie i gminach uznanych za sąsiadujące oznaczone na mapie jako wysokie-wysokie (ang. High-High) – kolor czerwony. Kolor ten występuje na mapie (Rycina 22), gdzie zaznaczone są

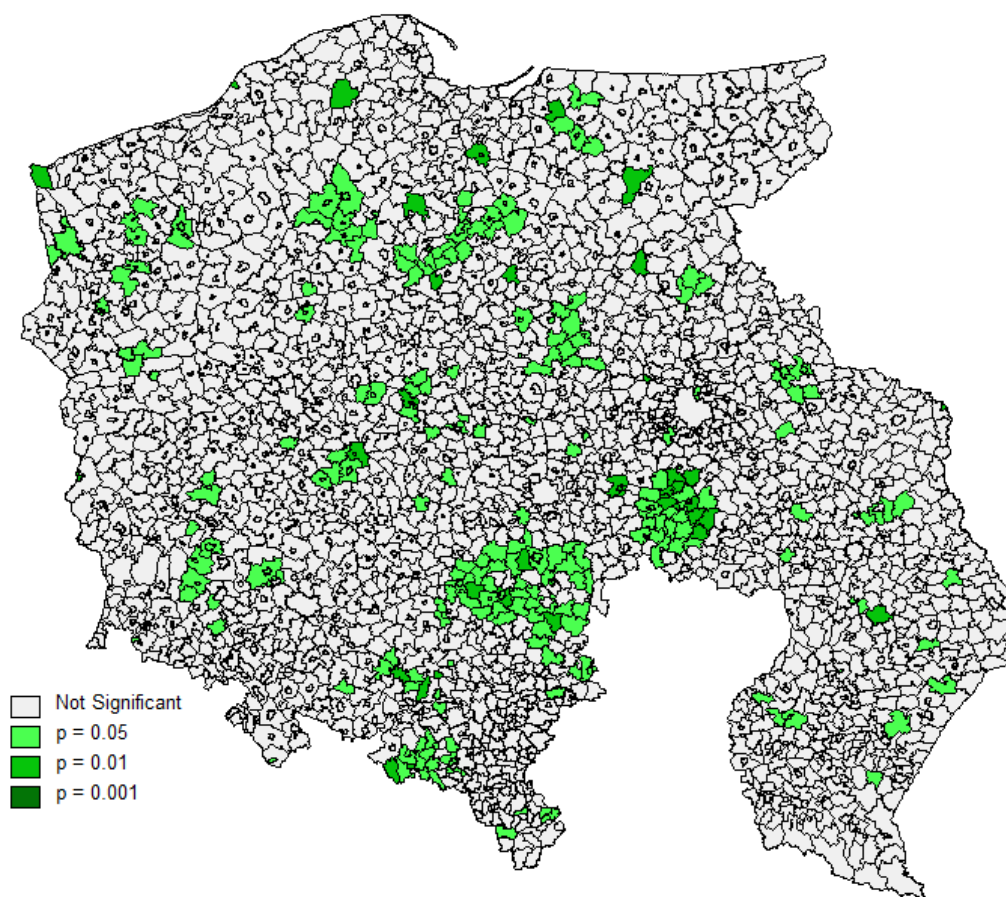
obszary istotne na poziomie  $\alpha=0.05$ , ale nie ma go już na mapie (Rycina 23), gdzie zaznaczone są obszary istotne na poziomie  $\alpha=0.01$ . Rycina 24 przedstawia dokładny obraz istotności statystycznej przynależności gmin do ćwiartek wykresu punktowego Morana. Przy czym wyznaczony poziom istotności według poprawki Bonferroniego i Sidaka dla średniej liczby gmin sąsiadujących równej 24 wynosi  $\alpha_1 = 0.0021$ . Zatem wyznaczone skupiska nie stanowią klasterów istotnych statystycznie.



Rycina 22 Przynależność gmin do ćwiartek wykresu punktowego Morana na poziomie istotności 0.05



Rycina 23 Przynależność gmin do ćwiartek wykresu punktowego Morana na poziomie istotności 0.01



Rycina 24 Poziom istotności dla mapy przynależności gmin do ćwiartek wykresu punktowego Morana

Podobnie, w wyniku zastosowania analizy przestrzennej wzbogaconej o informację dotyczącą roku urodzenia dziecka brak jest istotnych statystycznie klasterów ([Załącznik 6](#)).

### 8.5.3 Założenia teorii statystyki skanującej Kulldorff'a

Statystyki skanujące stanowią odrębną gałąź metod przeznaczonych do wyszukiwania klasterów. Wyszukiwanie klasterów metodą opracowaną przez Kulldorff'a dokonuje się poprzez zastosowanie okna skanującego. Okno to:

- przedział – dla klasterów czasowych,
- okrąg lub elipsa – dla klasterów przestrzennych,
- walec – dla klasterów czasowo-przestrzennych.

Używa się różnego rozmiaru okna skanującego wyznaczanego przy pomocy metryki euklidesowej. Dla każdej pozycji i rozmiaru takiego okna wyliczane są liczności obserwowane

i liczności oczekiwane wewnątrz i na zewnątrz okna oraz skanowana jest przestrzeń w celu znalezienia najbardziej prawdopodobnego klasteru. Z pozoru proces skanowania można określić jako nieskończony, ale nakładając na niego dwa warunki ograniczono go do skończonej ilości kroków. Pierwszym warunkiem jest ograniczenie ilości punktów zaczepienia okna skanującego. Centrum budowania okna może być każda centroida lub inny ograniczony zbiór punktów przestrzeni w określonej jednostce czasu. Drugim warunkiem jest ograniczenie wielkości okna skanującego poprzez podanie wielkości populacji do jakiej można zwiększać okno lub zakresu przestrzennego jaki może pokrywać okno. Zakres ten jest wówczas definiowany przez długość promienia określonego w metrach. Podobne ograniczenie dotyczy również jednostki czasu w klasterach czasowo-przestrzennych. Ograniczenia te dają możliwość ingerowania w wielkość i usytuowanie okien skanujących przez co pozwalają na dowolne sterowanie procesem skanującym. Pozwalają na wyszukiwanie klasterów niezogniskowanych jak i badanie zasięgu klasteru wokół znanego wcześniej punktu zwiększonego ryzyka tzw. ogniska.

Badanie istotności statystycznej okna skanującego określa **test ilorazu wiarygodności**.

Testujemy  $H_0 : p = q$  przeciwko  $H_1 : p > q$ ,

gdzie:

$p$  - prawdopodobieństwo wystąpienia wady wewnątrz okna skanującego  $Z$ ,

$q$  - prawdopodobieństwo wystąpienia wady na zewnątrz okna skanującego  $Z$ .

Budowa testu rozpoczyna się od wyznaczenia funkcji wiarygodności  $L(Z, p, q)$  dla ustalonego okna  $Z$ . Zebrane dane populacyjne oraz dotyczące wad wrodzonych pozwalają na to, by funkcję wiarygodności oprzeć o model Poissona.

Funkcja wiarygodności dla okna  $Z$  ma postać:

$$L(Z, p, q) = \left( \frac{c}{E(c)} \right)^c \left( \frac{C - c}{C - E(c)} \right)^{C - c} I() \quad 7.4$$

gdzie:

$c$  – liczba przypadków (liczba dzieci z wadą wrodzoną) wewnątrz okna skanującego,

$C$  – całkowita liczba przypadków,

$E(c)$  – licznosc oczekiwana wewnątrz okna skanującego dostosowana,



$C-E(c)$  – liczność oczekiwana na zewnątrz okna skanującego,

$I()$  – funkcja indykatorowa wynosi 1 gdy liczba przypadków wewnątrz okna jest większa niż liczność oczekiwana lub 0 w przeciwnym przypadku.

Procedurę wyznaczania funkcji wiarygodności powtarza się dla każdej pozycji i rozmiaru okna skanującego.

Kolejnym krokiem jest wyznaczenie maksimum tej funkcji:

$$L(\bar{Z}) \stackrel{\text{def}}{=} \sup_{p>q} L(\bar{Z}, p, q) \quad 7.5$$

Najbardziej prawdopodobnym klasterem (ang. the most likely cluster)  $\bar{Z}$  jest to okno skanujące, dla którego maksimum funkcji wiarygodności jest większe niż dla każdego innego okna:

$$\bar{Z} = \left\{ Z : L(\bar{Z}) > L(Z) \quad \forall Z \right\} \quad 7.6$$

Jest to klaster o najmniejszym prawdopodobieństwie przypadkowego pojawienia się.

Jeśli przez  $L_0$  oznaczymy:

$$L_0 \stackrel{\text{def}}{=} \sup_{p=q} L(\bar{Z}, p, q)$$

wówczas ilorazem wiarygodności dla najbardziej prawdopodobnego klasteru będzie wyrażenie:

$$LR = \frac{L(\bar{Z})}{L_0} \quad 7.7$$

zwane testem ilorazu wiarygodności. Rozkład uzyskanej w ten sposób statystyki testowej okazuje się trudny do uzyskania w postaci analitycznej, dlatego też uzyskuje go się tzw. metodą Monte Carlo. W tym celu generowane zostają różne ustawienia losowe danych; następnie porównuje się maksimum funkcji wiarygodności dla danych rzeczywistych z maksimum funkcji wiarygodności dla różnych ustawień danych fikcyjnych. Dla 9999 powtórzeń (wygenerowanych ustawień) test jest istotny na poziomie 0.05, jeśli wartość statystyki dla rzeczywistego ustawienia danych znajduje się pomiędzy 500 najwyższymi wartościami statystyki testowej pochodzącej z tych powtórzeń.

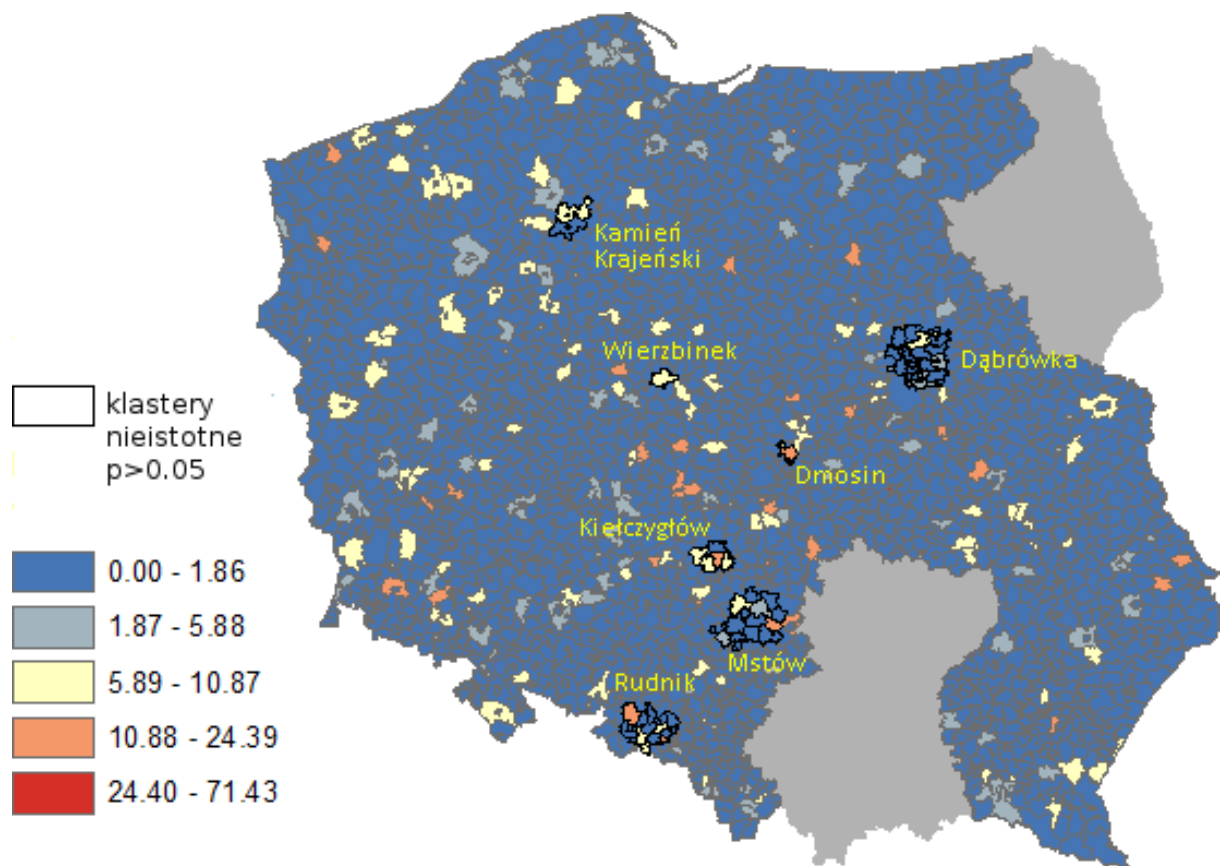
Po wyznaczeniu najbardziej wyróżniającego się klastera poszukuje się kolejnych klasterów. W tym celu powyższa procedura jest ponawiana, przy czym z analizowanego obszaru wyłącza się poprzednio wykryty klaster.

## 8.5.4 Wyniki zastosowania teorii statystyki skanującej Kulldorff'a

### 8.5.4.1.1 Analiza klasterów przestrzennych

Okno skanujące zostało zdefiniowane jako okrąg. Podstawowym założeniem dotyczącym wyszukiwania klasterów przestrzennych było ograniczenie wielkości okna skanującego tak, by analiza oparta była na podobnych założeniach jak statystyka Morana wykorzystująca w sposób bezpośredni macierz wag. Promień do jakiego może się rozszerzać okno skanujące określono więc na 30 km.

Zgodnie z oczekiwaniami, żaden z wykrytych „prawdopodobnych klasterów” nie okazał się być istotnym statystycznie (Rycina 25, Tabela 12). Podobnie, w wyniku zastosowania analizy czasowo-przestrzennej brak jest istotnych statystycznie klasterów (Załącznik 6).



Rycina 25 Najbardziej prawdopodobne, ale nie istotne statystycznie klastery wyznaczone poprzez statystykę skanującą Kulldorff'a oraz surowe współczynniki występowania izolowanego CLP w latach 2005, 2006 w badanych województwach, przedstawione według naturalnego podziału Jenks'a

Tabela 12 Wyniki analizy skanującej Kulldorff'a

	Najbardziej prawdopodobny klaster	Kolejne prawdopodobne klaster					
	okrąg 1	okrąg 2	okrąg 3	okrąg 4	okrąg 5	okrąg 6	okrąg 7
<b>Gmina środka okręgu</b>	Dąbrówka	Kiełczygłów	Kamień Krajeński	Wierzbinek	Rudnik	Dmosin	Mstów
<b>Województwo środka okręgu</b>	Mazowieckie	Łódzkie	Kujawsko-Pomorskie	Wielkopolskie	Śląskie	Łódzkie	Śląskie
<b>Środek okręgu wsp. X</b>	922553.4	775559.7	667504	738954.8	727272.9	827208.9	801879.9
<b>Środek okręgu wsp. Y</b>	5832887.4	5684262.8	5934625.5	5817182	5560426.8	5762194.5	5638929.2
<b>Długość promienia (w metrach)</b>	22213.96	11420.41	13050.51	7028.53	16572.53	5301.59	21075.5
<b>Liczność - populacja</b>	3537	256	299	147	1116	161	2965
<b>Obserwowana liczba przypadków</b>	17	4	4	3	7	3	12
<b>Oczekiwana liczba przypadków</b>	4.98	0.36	0.42	0.21	1.57	0.23	4.18
<b>Współczynnik częstości na 1000</b>	4.806333051	15.625	13.37793	20.40816	6.272401	18.63354	4.047218
<b>Relatywne ryzyko</b>	3.51	11.18	9.57	14.62	4.51	13.31	2.92
<b>Statystyka LR</b>	9.007723	5.999308	5.437293	5.244226	5.058321	4.98122	4.908645
<b>Wartość p</b>	0.110 (NS)	0.818 (NS)	0.954 (NS)	0.977 (NS)	0.990 (NS)	0.993 (NS)	0.995 (NS)

NS: brak istotności statystycznej -- wartość p &gt;= 0.05

### 8.5.5 Porównanie wyników uzyskanych przy pomocy metody LISA i metody Kulldorff'a

Podstawowym a zarazem najbardziej ogólnym rezultatem analizy przestrzennej jest wynik Statystyki globalnej Morana. Zgodnie z jego wskazaniem badany obszar nie wykazuje tendencji do tworzenia klasterów – brak autokorelacji. Potwierdzeniem tego wyniku są rezultaty zarówno analizy skanującej Kulldorff'a jak i analizy LISA skorygowanej o poprawkę Bonferroniego (ewentualnie Sidaka) – wyznaczone klastery są tu nieistotne statystycznie. Podjęta analiza porównawcza lokalnych metod wyszukiwania klasterów obejmuje zatem takie obszary, które jedynie wyróżniają się podwyższonym współczynnikiem występowania izolowanego  $CL \pm P$ . Miejsca te nazywane będą w dalszej części pracy **L-skupiskami** i **K-skupiskami** odpowiednio do stosowanej do ich wyszukania metody.

Dokonałiśmy podziału gmin na te, które przynależą do skupisk i te, które są poza nimi (Tabela 13).

Tabela 13 Przynależność gmin do L-skupisk i do K-skupisk

Liczba gmin	Poza K-skupiskami	K-skupiska	Razem
Poza L-skupiskami	a = 2318	b = 9	2327
L-skupiska	c = 173	d = 58	231
Razem	2491	67	2558

Można zauważyć, że do skupisk wyznaczonych metodą LISA przynależy 231 gmin, podczas gdy do skupisk wyznaczonych metodą Kulldorff'a tylko 67. Różnica jest tak duża, że względu na przyjęte założenia i sposób realizacji zadania przez obie metody. Tutaj przyjęty promień okręgu w obu analizach wynosił 30km. W metodzie Kulldorff'a okno rozszerza się w poszukiwaniu skupiska tak długo, aż osiągnie zadany promień, następnie wybrany zostaje taki rozmiar okna, który najdokładniej obejmuje skupisko. W analizie LISA zawsze wykorzystywany jest ten z góry zadany promień okręgu, zdefiniowany poprzez sposób rozumienia sąsiedztwa.

Wyznaczona zgodność (wartości na głównej przekątnej w tabeli Tabela 13 – oznaczone jako a i d) jest duża, 2376 spośród 2558 gmin (92.89%) zostało zakwalifikowanych tak samo przez obie metody. Współczynnik Kappa Cohen'a wprowadza korekcje tej wartości o szanse na zgodność (o zgodność, która pojawia się dla przypadkowego rozłożenia danych w tabeli). Zgodność wyrażona współczynnikiem Kappa jest istotna statystycznie  $p < 0.000001$ , choć jest mniejsza niż ta nie skorygowana o szansę na zgodność:  $Kappa(95\%CI) = 36.34\%(29.40\%, 43.29\%)$ .

Wszystkich skupisk wyznaczonych metodą LISA jest 16. Metoda ta pozwala na ich zazębianie się, stąd niektóre gminy należą do kilku skupisk. Zazębiające się skupiska połączono tworząc grupy jednorodne. Wyznaczono w ten sposób 7 spójnych, niepokrywających się nowych skupisk (Tabela 14):

Tabela 14 Przynależność gmin do L-skupisk z wyszczególnieniem zazębiających się skupisk i skupisk połączonych w grupy jednorodne

		zazębiające się skupiska wyznaczone metodą LISA															Liczba gmin	
		SK1	SK2	SK3	SK4	SK5	SK6	SK7	SK8	SK9	SK10	SK11	SK12	SK13	SK14	SK15		SK16
L-skupiska = skupiska połączone w grupy jednorodne	Poza L-skupiskami	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2327
	L-supisko 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13
	L-supisko 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6
		0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	6
	L-supisko 3	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	7
		0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	4
		0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	12
	L-supisko 4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	30
	L-supisko 1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	40
	L-supisko 2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5
		0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
		0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	8
		0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	11
		0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2
		0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	14
		0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	22
		0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
	L-supisko 5	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	2
		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
		1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	26
<b>Razem</b>																	2558	

0 – gminy nie należące do danego skupiska

1 – gminy należące do danego skupiska

Poszukując kolejnego skupiska metoda Kulldorff'a wyklucza z analizy gminy, które znalazły się w poprzednim skupisku. Stąd skupiska wyznaczone metodą statystyki skanującej Kulldorff'a nie zazębiają się. Jest ich również 7 (Tabela 15):

Tabela 15 Przynależność gmin do K-skupisk z ich wyszczególnieniem

K-skupiska	Liczba gmin
Poza K-skupiskami	2491
K-skupisko 1	23
K-skupisko 2	5
K-skupisko 3	5
K-skupisko 4	2
K-skupisko 5	15
K-skupisko 6	2
K-skupisko 7	15
<b>Razem</b>	<b>2558</b>

Porównanie zasięgu skupisk wskazuje, że obie metody wyznaczyły podobny obszar (Tabela 16). L-skupiska są większe i zawierają w sobie K-skupiska (sześć z siedmiu K-skupisk zawiera się w L-skupiskach).

Do wyznaczonych skupisk przynależy 96 dzieci z izolowanym CL±P. Z czego 42, to dzieci, które znalazły się w L-skupiskach i K-skupiskach, 46 tylko w L-skupiskach, 8 tylko w K-skupiskach (Tabela 17).

Tabela 16 Przynależność gmin do poszczególnych L-skupisk i do K-skupisk

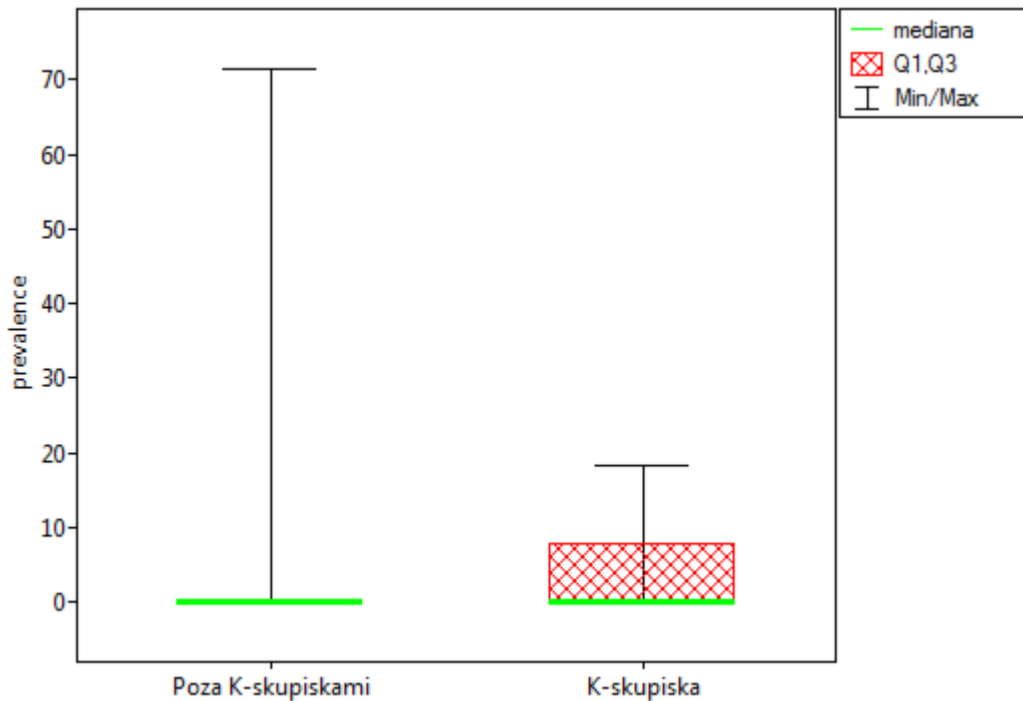
Liczba gmin	Poza K-skupiskami	K-skupisko 1	K-skupisko 2	K-skupisko 3	K-skupisko 4	K-skupisko 5	K-skupisko 6	K-skupisko 7	Razem
Poza L-skupiskami	2318	4	0	0	0	0	2	3	2327
L-skupisko 1	21	19	0	0	0	0	0	0	40
L-skupisko 2	50	0	5	0	0	0	0	12	67
L-skupisko 3	18	0	0	5	0	0	0	0	23
L-skupisko 4	28	0	0	0	2	0	0	0	30
L-skupisko 5	22	0	0	0	0	15	0	0	37
L-skupisko 6	13	0	0	0	0	0	0	0	13
L-skupisko 7	21	0	0	0	0	0	0	0	21
Razem	2491	23	5	5	2	15	2	15	2558

Tabela 17 Przynależność dzieci z izolowanym CL±P do poszczególnych L-skupisk i K-skupisk w latach 2005 i 2006

Liczność-grupa badana	Poza K-skupiskami	K-skupisko 1	K-skupisko 2	K-skupisko 3	K-skupisko 4	K-skupisko 5	K-skupisko 6	K-skupisko 7	Razem
Poza L-skupiskami	347	4	0	0	0	0	3	1	355
L-skupisko 1	18	13	0	0	0	0	0	0	31
L-skupisko 2	5	0	4	0	0	0	0	11	20
L-skupisko 3	3	0	0	4	0	0	0	0	7
L-skupisko 4	4	0	0	0	3	0	0	0	7
L-skupisko 5	6	0	0	0	0	7	0	0	13
L-skupisko 6	4	0	0	0	0	0	0	0	4
L-skupisko 7	6	0	0	0	0	0	0	0	6
Razem	393	17	4	4	3	7	3	12	443

Zgodność oszacowana współczynnikiem Kappa jest istotna statystycznie  $p < 0.000001$  i wynosi:  
Kappa(95%CI) = 48.49%(38.81%, 58.18%),

Porównanie współczynnika częstości występowania izolowanego CL±P pomiędzy obszarami objętymi przez skupisko i znajdującymi się poza nim wskazuje, że obie metody wskazały na właściwe usytuowanie skupisk (Rycina 26, Rycina 27). W obu przypadkach współczynnik częstości występowania izolowanego CL±P w skupiskach jest istotnie statystycznie wyższy niż poza nimi (średnia rang gmin w skupiskach jest wyższa niż poza skupiskami).



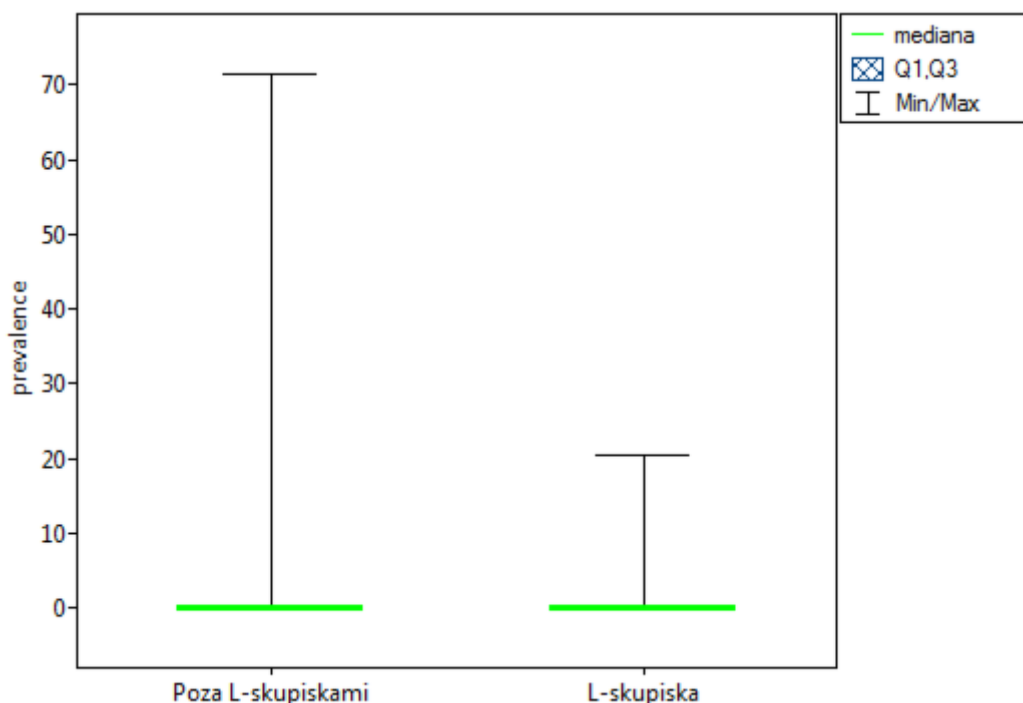
Rycina 26 Rozkład współczynników częstości występowania izolowanego CL±P wewnątrz i na zewnątrz K-skupisk w latach 2005 i 2006.

Test Mann'a-Whitney'a, statystyka z poprawką na ciągłość i na rangi wiązane:  $Z = 8.6971776$ ,  $p < 0.000001$ ,

K-skupiska: średnia rang = 1731.738806,

Poza K-skupiskami: średnia rang = 1267.33621.





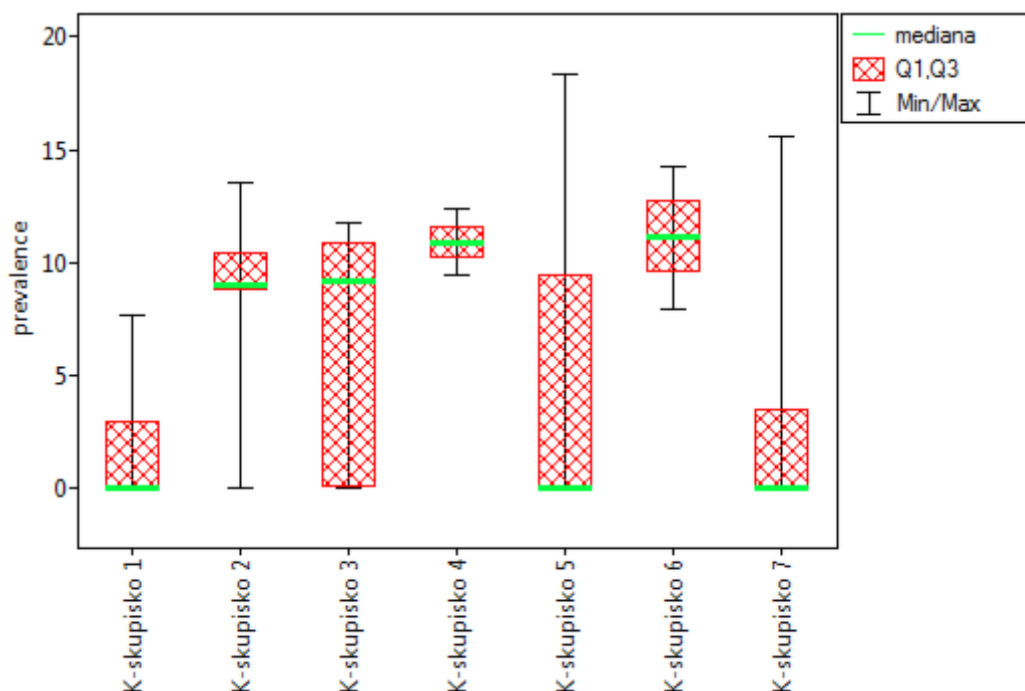
Rycina 27 Rozkład współczynników częstości występowania izolowanego CL±P wewnątrz i na zewnątrz L-skupisk w latach 2005 i 2006.

Test Mann'a-Whitney'a, statystyka z poprawką na ciągłość i na rangi wiązane:  $Z = 6.0219823$ ,  $p < 0.000001$ ,

L-skupiska: średnia rang = 1442.493506,

Poza L-skupiskami: średnia rang = 1263.319725.

W przypadku analizy Kulldorff'a, według wskazań testu Kruskal'a-Wallis'a, współczynniki częstości występowania izolowanego CL±P różnią się nieznacznie pomiędzy skupiskami – statystyka  $H = 16.1854547$ ,  $p = 0.0127924$  (Rycina 28). Różnica jednak jest tak niewielka, że test Post-Hoc Dunn'a nie lokalizuje różnic między parami skupisk (Tabela 18).



Rycina 28 Rozkład współczynników częstości występowania izolowanego CL±P w K-skupiskach w latach 2005 i 2006.

Test Kruskal'a-Wallis'a, statystyka  $H = 16.1854547$ ,  $p = 0.0127924$ ,

K-skupisko 1: średnia rang = 28.2826087,

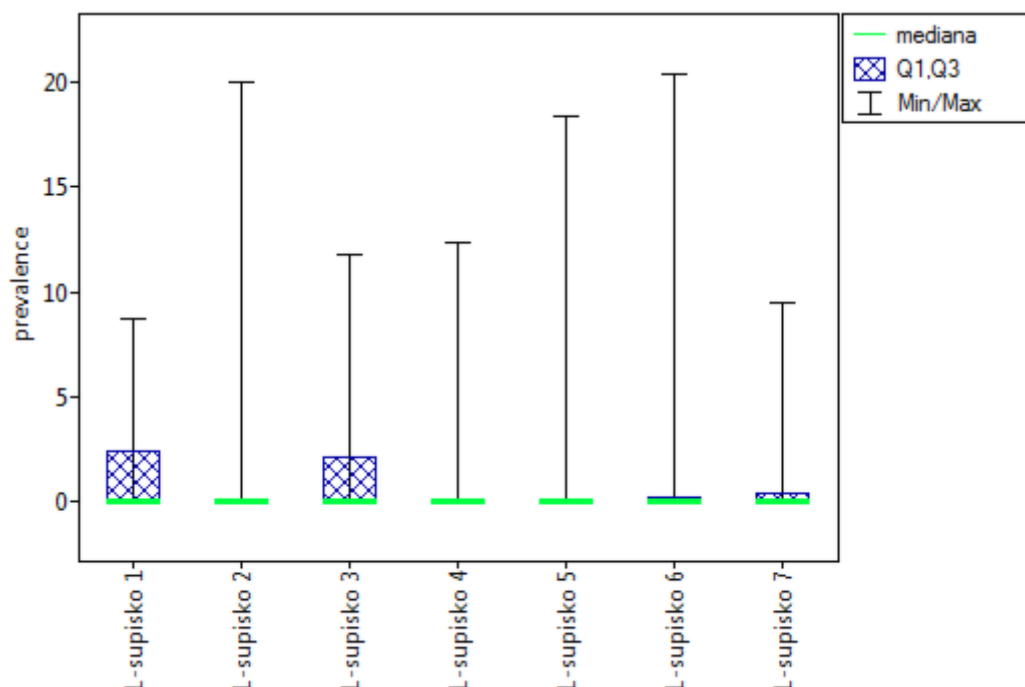
K-skupisko 2: średnia rang = 49.5,

K-skupisko 3: średnia rang = 42.6,  
 K-skupisko 4: średnia rang = 59.5,  
 K-skupisko 5: średnia rang = 33.3666667,  
 K-skupisko 6: średnia rang = 58,  
 K-skupisko 7: średnia rang = 28.7666667.

Tabela 18 Wyniki testu Post-Hoc Dunn'a dla różnic w rozkładach współczynników częstości izolowanego CL±P w K-skupiskach w latach 2005 i 2006

	K-skupisko 1	K-skupisko 2	K-skupisko 3	K-skupisko 4	K-skupisko 5	K-skupisko 6	K-skupisko 7
<b>K-skupisko 1</b>	Różnica śr. rang	21.22	14.32	31.22	5.08	29.72	0.48
	NIR	29.21	29.21	43.64	19.65	43.64	19.65
	Statystyka Z	2.21	1.49	2.17	0.79	2.07	0.07
	Wartość p	0.57	1	0.63	1	0.81	1
<b>K-skupisko 2</b>	21.22	Różnica śr. rang	6.90	10.00	16.13	8.50	20.73
	29.21	NIR	37.44	49.53	30.57	49.53	30.57
	2.21	Statystyka Z	0.56	0.61	1.60	0.52	2.06
	0.5739	Wartość p	1	1	1	1	0.8263
<b>K-skupisko 3</b>	14.32	6.90	Różnica śr. rang	16.90	9.23	15.40	13.83
	29.21	37.44	NIR	49.53	30.57	49.53	30.57
	1.49	0.56	Statystyka Z	1.04	0.92	0.94	1.37
	1	1	Wartość p	1	1	1	1
<b>K-skupisko 4</b>	31.22	10.00	16.90	Różnica śr. rang	26.13	1.50	30.73
	43.64	49.53	49.53	NIR	44.56	59.20	44.56
	2.17	0.61	1.04	Statystyka Z	1.78	0.08	2.10
	0.6250	1	1	Wartość p	1	1	0.7590
<b>K-skupisko 5</b>	5.08	16.13	9.23	26.13	Różnica śr. rang	24.63	4.60
	19.65	30.57	30.57	44.56	NIR	44.56	21.62
	0.79	1.60	0.92	1.78	Statystyka Z	1.68	0.65
	1	1	1	1	Wartość p	1	1
<b>K-skupisko 6</b>	29.72	8.50	15.40	1.50	24.63	Różnica śr. rang	29.23
	43.64	49.53	49.53	59.20	44.56	NIR	44.56
	2.07	0.52	0.94	0.08	1.68	Statystyka Z	1.99
	0.8099	1	1	1	1	Wartość p	0.9714
<b>K-skupisko 7</b>	0.48	20.73	13.83	30.73	4.60	29.23	Różnica śr. rang
	19.65	30.57	30.57	44.56	21.62	44.56	NIR
	0.07	2.06	1.37	2.10	0.65	1.99	Statystyka Z
	1	0.8263	1	0.7590	1	0.9714	Wartość p

Podobnie skupiska wykryte metodą LISA nie różnią się istotnie statystycznie pod względem współczynników częstości występowania izolowanego CL±P – statystyka testu Kruskal'a-Wallis'a  $H=2.7655124$ ,  $p=0.8376494$  (Rycina 29).



Rycina 29 Rozkład współczynników częstości występowania izolowanego CL±P w L-skupiskach w latach 2005 i 2006

Test Kruskal'a-Wallis'a, statystyka  $H=2.7655124$ ,  $p=0.8376494$ ,

L-skupisko 1: średnia rang = 125.925,

L-skupisko 2: średnia rang = 111.0149254,

L-skupisko 3: średnia rang = 118.9130435,

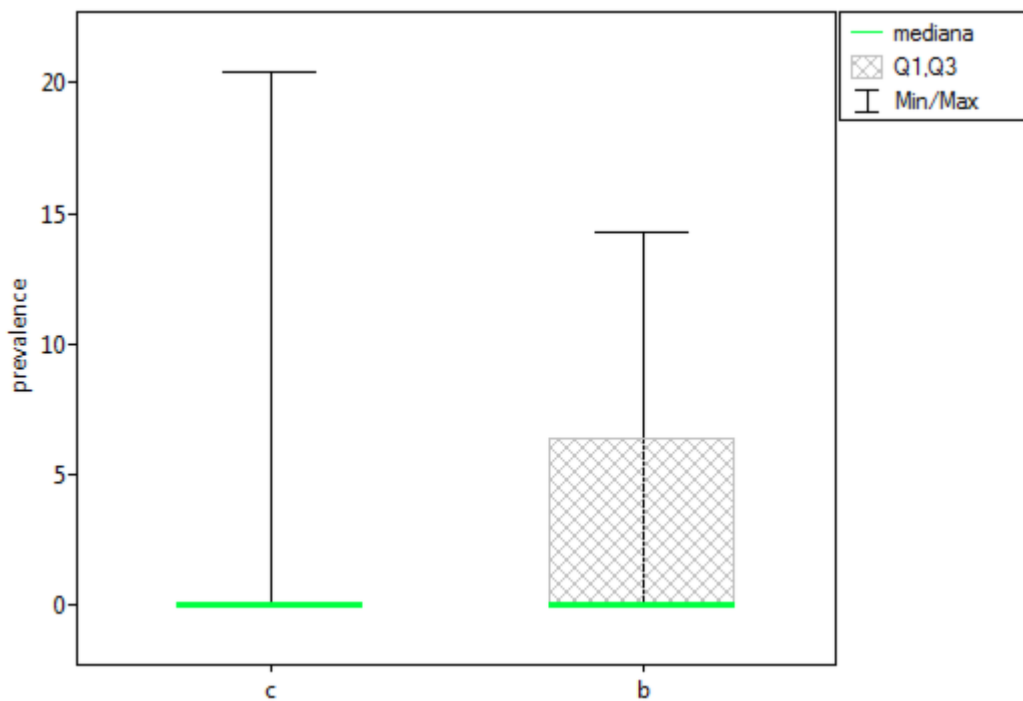
L-skupisko 4: średnia rang = 111.8666667,

L-skupisko 5: średnia rang = 112.6081081,

L-skupisko 6: średnia rang = 119.9615385,

L-skupisko 7: średnia rang = 119.2380952.

Poszukiwanie różnic pomiędzy skupiskami zakreślonymi metodą Kulldorff'a i LISA opiera się na współczynnikach częstości wyznaczonych dla gmin, które przynależą do skupiska tylko według jednej z porównywanych metod (Tabela 13: liczba tych gmin oznaczona została jako  $b$  i  $c$ ). Porównanie współczynników częstości występowania izolowanego CL±P pomiędzy tymi gminami wskazuje na występowanie różnic: statystyka testu Mann'a-Whitney'a  $Z=2.1228984$ ,  $p=0.0337624$  (Rycina 30). Współczynnik ten jest wyższy dla gmin oznaczonych jako  $b$  (należących do K-skupisk ale leżących poza L-skupiskami) niż dla gmin oznaczonych jako  $c$  (należących do L-skupisk ale leżących poza K-skupiskami).



Rycina 30 Rozkład współczynników częstości występowania izolowanego CL±P wewnątrz gmin należących do skupiska wyznaczonego przy pomocy tylko jednej metody, w latach 2005 i 2006.

Test Mann'a-Whitney'a, statystyka z poprawką na ciągłość i na rangi wiązane:  $Z = 2.1228984$ ,  $p = 0.0337624$ ,

a – gminy przynależące do L-skupisk a nie przynależące do K-skupisk, średnia rang = 90.196532,

b – gminy przynależące do K-skupisk a nie przynależące do L-skupisk, średnia rang = 116.555556.

## 8.6 WNIOSKI DOTYCZĄCE ANALIZY CZYNNIKÓW CZASOWO-PRZESTRZENNYCH

### Wnioski płynące z realizacji celu bezpośredniego 2 dotyczącego czynników zewnętrznych

Określenie i wybór metod statystycznych analizy czynników czasowo - przestrzennych

1.

Wybór sposobu **definiowania sąsiedztwa** badanych obszarów w przestrzeni:

- By spełnić oczekiwania co do
  - wielkości klasterów,
  - czasu prowadzenia analiz,
  - możliwość późniejszego porównywania metody LISA i Kulldorff'a,uznano za sąsiadujące względem zadanej gminy, te gminy których centroidy znajdują się wewnątrz okręgu o promieniu 30 km. Przyjęto metrykę euklidesową.

2.

Porównanie wyników poszukiwania **klasterów** zwiększonej częstości występowania badanej wady wrodzonej uzyskanych przy pomocy Statystyki Skanującej Kulldorff'a i Lokalnych Współczynników Zależności Przestrzennej (LISA):

- Metoda LISA i metoda Kulldorff'a zgodnie wskazują lokalizację skupisk;
- Obie metody, wykryte skupiska zwiększonego ryzyka wystąpienia izolowanego CL±P uznają za nie istotne statystycznie.;
- Współczynniki częstości, w obu metodach, są istotnie wyższe w obszarach zakreślonych przez klaster niż poza nimi;
- Nie zlokalizowano istotnych różnic we współczynniku częstości występowania izolowanego CL±P pomiędzy poszczególnymi skupiskami, choć metoda Kulldorff'a (działająca na licznosciach a nie na współczynnikach) wyznaczyła skupiska o nieco większym zróżnicowaniu współczynnika niż metoda LISA.
- Skupiska wyznaczone metodą LISA są większe niż te wyznaczone metodą Kulldorff'a. Jest to wynik innego sposobu rozumienia zadanego rozmiaru poszukiwanego klasteru:
  - metoda Kulldorff'a pozwala na znalezienie optymalnej wielkości klasteru mniejszej lub równej zadanemu rozmiarowi,
  - metoda LISA szuka klasterów tylko o zadanej wielkości.

---

## 9 DYSKUSJA

---

Praca została podzielona na dwie zupełnie odrębne części: **analizę historii rodności** (głównie kolejność ciąży i liczba poronień samoistnych) i **analizę klasterów** (miejsce zamieszkania matki i czas porodu). W każdej części wykorzystane są inne metody statystyczne, ale również inne grupy odniesienia. W analizie historii rodności zastosowana została metoda regresji logistycznej i warunkowej regresji logistycznej w oparciu o kilka różnych grup kontrolnych. W analizie klasterów użyto metody LISA i statystyki skanującej Kulldorff'a w oparciu o dane populacyjne. Wydaje się, że najlepiej byłoby połączyć te dwie części pracy i rozwiązać postawiony problem jedną z zastosowanych metod – tzn. wzbogacić modele regresji logistycznej o informacje dotyczące geograficznego położenia danych, lub metody statystyki Kulldorff'a czy LISA o historię rodności. W ostatnich latach metody te zaczęły się rozwijać w tym właśnie kierunku. Modele regresji logistycznej mogą wykorzystywać geograficzne rozłożenie danych – jako jeden z czynników ryzyka. Istnieje specjalny rodzaj regresji – Geograficzna Regresja Ważona, która bierze pod uwagę macierz wag przestrzennych. Metoda ta, mimo że pozwala na użycie dychotomicznej zmiennej zależnej, używana jest jednak w nieco innym kontekście. Stosuje się ją do modelowania zależności i trendów przestrzennych a nie klasterów ([28] Fotheringham A.S. i inni 2002; [27] Fotheringham A.S. i inni 2006). Podobnie, metody analizy klasterów można wzbogacić o inne czynniki ryzyka – opis i przykłady wykorzystania takich analiz można znaleźć w pracach: Kleinmana oraz Klassena ([41] Kleinman K. i inni, 2005; [37] Klassen A. i inni 2005). Metody te koncentrują się jednak głównie na danych przestrzennych, inne czynniki ryzyka traktowane są tu jako pomocnicze współzmiennne (ang. covariates). Wydaje się więc, że połączenie tych dwóch części pracy byłoby trudne. Okazuje się, że jest to niemożliwe ze względu na brak odpowiednich danych. Główną przyczyną rozdzenia analiz jest brak możliwości uzyskania odpowiednich danych populacyjnych. GUS ze względu na tajemnicę statystyczną zabraniającą podawania danych w taki sposób by można było zidentyfikować osobę, nie udostępnia (a w niektórych przypadkach nie zbiera) danych na odpowiednim poziomie szczegółowości – potrzebnym do analizy wielu czynników ryzyka wad wrodzonych. Grupa kontrolna ten odpowiedni poziom szczegółowości posiada, ale nie jest odpowiednio liczna

do analizy klasterów i nie była stratyfikowana w stosunku do tak małych jednostek położenia przestrzennego jakimi są gminy.

Nie udało się przeprowadzić jednego spójnego badania czynników ryzyka izolowanego CL±P. Ten brak spójności rekompensuje jednak szczegółowość i dokładność przeprowadzonych analiz. Dodatkowym atutem tej pracy jest wybór takich czynników ryzyka (kolejności ciąży, liczby poronień samoistnych, miejsca zamieszkania matki i czasu porodu), które zwykle nie są wnikliwie analizowane i które wciąż nie są dobrze opisane. Dostępna literatura wskazująca, że te czynniki mogą być istotne dotyczy głównie ciężkich wad takich jak: wady cewy nerwowej, wady serca, jak też wad stosunkowo częstszych jak zespół Downa.

Bazując na danych PRWWR wskazano silne dowody na występowanie zależności pomiędzy kolejnością ciąży a wzrostem ryzyka wystąpienia rozszczepu wargi. To podwyższone ryzyko jest niezależne od wieku matki, wcześniejszych niepowodzeń rozrodu i innych potencjalnych czynników zidentyfikowanych w naszym badaniu. Taką samą zależność wykryto dla wad serca, nieco słabszą dla agenezji nerek i rozszczepu podniebienia ([55] Materna-Kiryłuk A., Więckowska B. i inni, 2011). W pracach nad wadą cewy nerwowej opisano podobną zależność ([73] Whiteman D. i inni, 2000; [11] Blanco-Munoz J. i inni, 2006), ale są też prace wskazujące na zwiększone ryzyko dla matek, które rodziły wielokrotnie oraz tych będących w ciąży po raz pierwszy ([29] Frey L. i Hauser W.A., 2003). Na brak wpływu kolejności ciąży na występowanie wady cewy nerwowej wskazano w pracy Cengiz ([16] Cengiz B. i inni, 2004).

W wyniku analizy pewnej podgrupy matek rodzących dzieci z ciąży drugiej lub dalszej, możemy wesprzeć hipotezę, że historia wcześniejszych niepowodzeń rozrodu zmniejsza ryzyko wystąpienia izolowanego CL±P. Wynik ten potwierdza wcześniejsze doniesienia dotyczące badań nad wadą cewy nerwowej ([49] Kurinczuk J.J. i Clarke M., 1993). Jednakże, na brak bezpośredniego wpływu poronień samoistnych wskazują prace Carmi oraz Cengiz ([14] Carmi R. i inni., 1994; [16] Cengiz B. i inni, 2004), przy czym w pierwszej z nich grupę kontrolną stanowiły dzieci z innymi wadami a druga praca bazuje na bardzo mało licznej próbie. Podjęte próby wykazania, na podstawie PRWWR, wpływu poronień samoistnych na występowanie wad serca, agenezji nerek i rozszczepu podniebienia również nie wskazały na występowanie takiej zależności ([55] Materna-Kiryłuk A., Więckowska B. i inni, 2011).

Natomiast praca Karen Todoroff ([69] Todoroff K. i Shaw G.M., 2000) wskazuje, że nie tylko wystąpienie wcześniejszych poronień, ale czas od ostatniego poronienia może mieć wpływ na ryzyko wystąpienia wady cewy nerwowej. Tak samo Eduardo Villamor ([71] Villamor E. i inni, 2008) wskazuje na istotny wpływ czasu jaki upłynął od ostatniej ciąży na występowanie rozszczepu podniebienia (CP). Niestety zebrane przez PRWWR dane nie pozwalają określić czasu od ostatniej ciąży i nie było tu możliwe zweryfikowanie tej tezy. W temacie tym nie ma jednak zgodności. Odwrotną zależność opisuje David Whiteman ([73] Whiteman D. i inni, 2000). Te odmienne rezultaty przeprowadzonych przez różnych badaczy analiz mogą być konsekwencją niepełnej zgłaszalności poronień. Poronienia bowiem we wczesnym okresie ciąży mogą się odbyć bez wiedzy matki.

Warto podkreślić, iż istotne zmniejszenie ryzyka wraz ze wzrostem liczby uprzednich poronień samoistnych obserwowaliśmy dla modeli regresji logistycznej opartych o poronienia samoistne wyrażone na skali interwałowej, ale takie modele, w których występuje również kolejność ciąży. Po jej usunięciu z modelu efekt ten zanika. Dodatkowo, ponieważ zależność częstości występowania wady od liczby poronień samoistnych i kolejności ciąży przybiera kształt trendu liniowego, modele regresji logistycznej uwzględniające ten typ zależności dają efekty, bardziej odpowiadające temu co obserwujemy w próbie niż modele oparte o zmienne fikcyjne ([40] Kleinbaum D.G. i inni, 2002). W badaniach ze zmiennymi fikcyjnymi iloraz szans nie informuje o stałym wzroście lub spadku ryzyka pomiędzy poszczególnymi kategoriami zmiennej a podaje szansę wystąpienia wady wyliczoną w stosunku do grupy odniesienia. W pracy dodatkowo, oprócz modeli regresji logistycznej równoległe, ale w oparciu o inną grupę kontrolną budowano modele warunkowej regresji logistycznej. Pomimo, że w wyniku tej bliźniaczej analizy nie uzyskaliśmy identycznych wyników, to jej rezultaty wsparły uprzednią analizę i potwierdziły główne hipotezy.

Analiza miejsca zamieszkania matki i czasu porodu pozwoliła na zlokalizowanie nieistotnych statystycznie skupisk zwiększonej częstości występowania izolowanego CL±P. Wynik ten może sugerować nieznaczny wpływ teratogenów środowiskowych na tworzenie się tej wady. Trudno znaleźć potwierdzenie tej hipotezy w dostępnej literaturze. Jednak nie dlatego, że uzyskano odmienne wyniki, ale dlatego, że wada ta nie była analizowana tak szczegółowo pod kątem geograficznego występowania. Wcześniejsze analizy dotyczące izolowanego CL±P pomijają zupełnie analizę czynnika geograficznego, czasami ograniczają się



do podziału na miasto/wieś ([10] Bing-Fang Hwang i inni 2008; [57] Messer L.C. i inni, 2010) lub do podziału przestrzeni na inne dwa lub trzy regiony. Taką analizę przeprowadził Harville dla populacji norweskiej ([34] Harville E.W. i inni, 2005). Znalazł on różnice w częstości występowania CL±P pomiędzy regionem północnym, wschodnim i środkowo-zachodnim. Podobnie analizowano wadę cewy nerwowej w Belfaście ([21] Elwood J.H. i Elwood J.M., 1984), dzieląc miasto na dzielnice. Tu jednak nie udało się znaleźć istotnych statystycznie różnic. Informacje dotyczące wpływu lokalizacji przestrzennej miejsca zamieszkania matki na częstość występowania wady cewy nerwowej można znaleźć w pracy Freya ([29] Frey L. i Hauser W.A., 2003). Jest to praca przeglądowa dotycząca etiologii tejże wady. Frey wskazuje, że w Europie kontynentalnej nie znaleziono dowodów na istotny wpływ geograficznej lokalizacji zamieszkania matki, ale już na terytorium Wielkiej Brytanii stwierdzono tzw. gradientową zależność, w Kanadzie różnice dotyczą wschodniej i zachodniej jej części. Klaster wykryto natomiast na terenie Chin i w Indiach.

W niniejszej pracy nie znaleziono istotnych statystycznie klasterów a jedynie skupiska o podwyższonej częstości występowania CL±P. Na ich podstawie udało się porównać zastosowane metody statystyczne. Stwierdzono, że przestrzenna lokalizacja wykrytych L-skupisk i K-skupisk jest zbliżona. Można więc uznać, że wyniki uzyskane obiema metodami są zgodne. Pojawiające się w bardziej szczegółowej analizie różnice wynikają z faktu, że w metodzie Kulldorff'a sposób poszukiwania klasteru pozwala na znalezienie optymalnej jego wielkości podczas gdy w metodzie LISA ta wielkość jest zadawana z góry. Brak elastyczności okna L-skupiska w niniejszej analizie powoduje, że obejmuje ono również te gminy, które ze względu na niski współczynnik nie musiałyby znajdować się w skupisku. Stąd, pomimo zgodności w lokalizacji przestrzennej, różny zasięg (wielkość) skupisk wyznaczonych przez porównywane metody powoduje różnice we współczynniku częstości występowania izolowanego CL±P pomiędzy nimi.

Poszukując klasterów badacz zwykle nie potrafi dokładnie sprecyzować ich rozmiaru. Jest to tym trudniejsze im większy obszar jest przeszukiwany i im więcej różnego rozmiaru okien może się w nim znaleźć. Pamiętać należy, że każdy rzeczywisty klaster, występujący w ramach badanego obszaru, może mieć zupełnie inny rozmiar. Badaczowi trudno jest wskazać ten jeden rozmiar, który miałby być odpowiednim dla każdego klasteru, w niektórych

przypadkach może być to niemożliwe. Dużo łatwiej jest wskazać maksymalny, mający sens z punktu widzenia badanego zjawiska, promień do którego może się rozszerzać okno. Dlatego też z punktu widzenia praktyki stosowania, metoda Kulldorff'a okazała się być metodą bardziej uniwersalną i precyzyjną. Wykrycie istotnego statystycznie klasteru jest początkiem długiej i kosztownej procedury poprzedzającej wszczęcie alarmu związanego z wykrytym klasterem (ang. cluster investigation protocols<sup>2</sup>), dlatego ważne jest aby jego istnienie było potwierdzone przez więcej niż jedną metodę statystyczną. W naszym przypadku globalna statystyka Morana wskazała na brak efektu klasteryzacji, podobnie jak lokalna statystyka Kulldorff'a. Analiza LISA wyznaczyła klaster, które jednak, po uwzględnieniu poprawki Bonferroniego (Sidaka), okazały się być nieistotne statystycznie. Nie ma zatem przesłanek pozwalających wsząć ową procedurę. Nie mniej wszystkie dostępne w bazie informacje dotyczące dzieci z izolowanym CL±P , które znalazły się w zlokalizowanych skupiskach zostały jeszcze raz uważnie przeanalizowane przez wykwalifikowanych pracowników Polskiego Rejestru Wrodzonych Wad Rozwojowych. Nie znaleziono w nich niepokojących przesłanek świadczących o powiązaniu tych przypadków z zaistniałym w obrębie tych skupisk teratogenem środowiskowym.

---

<sup>2</sup> <http://www.eurocat-network.eu/>

---

## 10 BIBLIOGRAFIA

---

1. **Alexander F.E., Cuzick J.**, *Methods for the assessment of disease cluster*, In: Elliott P., Cuzick J., English D., Stern R. (Eds), *Geographical and Environmental Epidemiology: Methods for Small – Area Studies*, Oxford University Press, Oxford, **1992**.
2. **Anscombe F.J.**, *Computing in Statistical Science through APL*, Springer-Verlag, New York, **1981**.
3. **Anselin L.**, *Local indicators of spatial association – LISA*, *Geographical Analysis*, **1995**, 27, 93-115.
4. **Anselin L., Kim Y.-W., and Syabri I.**, *Web-based analytical tools for the exploration of spatial data*, *Journal of Geographical Systems*, **2004b**, 6, 197–218.
5. **Armitage P., Berry G.**, *Statistical Methods in Medical Research (3rd edition)*, Wiley-Blackwell, Oxford, **1994**.
6. **Barnard G.A.**, *On alleged gains in power from lower p-values*, *Statistics in Medicine*, **1989**, 8, 1469-1477.
7. **Besag J., Nowell J.**, *The detection of clusters in rare diseases*, *Journal of the Royal Statistical Society, Series A*, **1991**, 154, 143-155.
8. **Bianca S., Bianca M., Bonaffini F., Ettore G.**, *The role of maternal reproductive history in the aetiology of neural tube defects*, *Medical Hypotheses*, **2002**, 58, 113-114.
9. **Bianco K., Caughey A.B., Shaffer B.L., Davis R., Norton M.E.**, *History of miscarriage and increased incidence of fetal aneuploidy in subsequent pregnancy*, *Obstetrics & Gynecology*, **2006**, 107, 1098-1102.
10. **Bing-Fang Hwang, Jouni J.K. Jaakkola,** *Ozone and Other Air Pollutans and the risk of Oral Cleft*, *Environmental Health Perspectives*, **2008**, 116(10), 1411-1415.
11. **Blanco-Muñoz J., Lacasaña M., Borja-Aburto V.H.,** *Maternal miscarriage history and risk of anencephaly*, *Paediatric and Perinatal Epidemiology*, **2006**, 20(3), 210–218.

12. **Boscoe F.P., McLaughlin C., Schymura M.J., Kielb C.L.**, *Visualization of the spatial scan statistic using nested circles*, Health and Place, **2003**, 9, 273-277.
13. **Burrough P.A.**, *Principles of Geographic Information Systems for Land Assessment*, Oxford University Press, New York, **1986**.
14. **Carmi R., Gohar J., Meizner I., Katz M.**, *Spontaneous abortion-high risk factor for neural tube defects in subsequent pregnancy*, American Journal of Medical Genetics, **1994**, 51, 93-97.
15. **Castilla E.E., Paz J.E.**, *Parity and Down's syndrome*, Lancet, **1994**, 344, 1645-1646.
16. **Cengiz B., Söylemez F., Oztürk E., Cavdar A.O.**, *Serum zinc, selenium, copper, and lead levels in women with second-trimester induced abortion resulting from neural tube defects: a preliminary study.*, Biological Trace Element Research, **2004**, 97(3), 225-35.
17. **Cochran W. G.**, *The chi-square goodness-of-fit test*, Annals of Mathematical Statistics, **1952**, 23(3), 15-345.
18. **Cressi N. A.**, *Statistics for Spatial Data (rev.edition)*, Wiley, Hoboken, **1993**.
19. **Dolk H.** , *EUROCAT: 25 years of European surveillance of congenital anomalie*, Archives of disease in childhood. Fetal and neonatal edition, **2005**, 90, 355-358.
20. **Doria-Rose V.P., Kim H.S., Augustine E.T., Edwards K.L.**, *Parity and the risk of Down's syndrome*, American Journal of Epidemiology, **2003**, 158, 503-508.
21. **Elwood J.H., Elwood J.M.**, *Investigation of area differences in the prevalence at birth of anencephalus in Belfast*, International Jurnal of Epidemiology , **1984**, 13(1), 45-52.
22. **Fisher R.A.**, *Statistical methods for research workers (5th edition)*, Oliver and Boyd, Edinburgh, **1934**.
23. **Fisher R.A.**, *The logic of inductive inference*, Journal of the Royal Statistical Society Series A, **1935**, 98, 39-54.

- 24. Florax R., Nijkamp P.,** *Misspecification in Linear Spatial Regression Models*, Tinbergen Institute Discussion Paper, **2003**, TI 2003-081/3.
- 25. Fotheringham A.S.,** *Trends in Quantitative Methods I: Stressing the Local*, *Progress in Human Geography*, **1997**, 21, 88-96.
- 26. Fotheringham A.S.,** *Context-dependent spatial analysis: A role for GIS?*, *Journal of Geographical Systems*, **2000**, 2, 71-76.
- 27. Fotheringham A.S. ,** *Quantification, Evidence and Positivism*, In: *Approaches to Human Geography eds. Valentine G. and Aitken S*, Sage, London, **2006**.
- 28. Fotheringham A.S., Brunson C., Charlton M.E. ,** *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, Chichester, **2002**.
- 29. Frey L. and Hauser W. A.,** *Epidemiology of Neural Tube Defects*, *Epilepsia*, **2003**, 44 (3 Suppl), 4-13.
- 30. Gardiner A., Clarke C., Cowen J., Finn R., McKendrick O.,** *Spontaneous abortion and fetal abnormality in subsequent pregnancy*, *British Medical Journal*, **1978**, 1, 1016-1018.
- 31. Gardner M.J., Altman D.G. ,** *Statistics with Confidence - Confidence Intervals and Statistical Guidelines (2nd edition)* , British Medical Journal Books, London, **2000**.
- 32. Grabiński T., Ludwiczak B., Malina A., Zeliaś A.,** *Ekonometria przestrzenna*, PWE, Warszawa, **1991**.
- 33. Hampton K.H., Fitch M.K., Allshouse W.B., Doherty I.A., Gesink D.C., Leone P.A., Serre M.L., and Miller W.C.,** *Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking*, *American Journal of Epidemiology*, **2010**, 172 (9), 1062-1069.
- 34. Harville E. W., Wilcox A. J., Lie R. T., Vindenes H., Åbyholm F.,** *Cleft Lip and Palate versus Cleft Lip Only: Are They Distinct Deffects?*, *American Journal of Epidemiology*, **2005**, 162 (5), 448-453.

- 35. Jenks G.F.**, *The Data Model Concept in Statistical Mapping*, International Yearbook of Cartography , **1967**, 7, 186-190.
- 36. Kallen K .**, *Parity and Down syndrome*, American Journal of Medical Genetics, **1997**, 70, 196-201.
- 37. Klassen A., Kulldorff M., Curriero F. ,** *Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors*, International Journal of Health Geographics, **2005**, 4(1).
- 38. Klassen L.H., Paelinck J.H.P.**, *Ekonometria przestrzenna*, PWN, Warszawa, **1983**.
- 39. Klassen L.H., Paelinck J.H.P., Wagenaar S.J.**, *Systemy przestrzenne*, PWN, Warszawa, **1982**.
- 40. Kleinbaum D.G., Klein M., Pryor E.R.**, *Logistic Regression: A Self-Learning Text, (2nd edition)*, Springer Science+Business Media Inc., New York, **2002**.
- 41. Kleinman K., Abrams A., Kulldorff M., Platt R.**, *A model-adjusted space-time scan statistic with an application to syndromic surveillance*, Epidemiology and Infection, **2005**, 133, 409-419.
- 42. Kopczewska K.**, *Ekonometria i Statystyka Przestrzenna*, CeDeWu, Warszawa, **2007**.
- 43. Kulldorff M.**, *A spatial scan statistic*, Communications in Statistics: Theory and Methods, **1997**, 26, 1481-1496.
- 44. Kulldorff M.**, *Prospective time periodic geographical disease surveillance using a scan statistic*, Journal of the Royal Statistical Society, Series A, **2001**, 164, 61-72.
- 45. Kulldorff M., Athas W. F., Feuer E. J., Miller B. A., Key C. R.**, *Evaluating cluster alarms. A space-time scan statistic and brain cancer in Los Alamos, New Mexico*, American Journal of Public Health, **1998a**, 88, 1377-1380.
- 46. Kulldorff M., Heffernan R., Hartman J., Assuncao R., Mostashari F.**, *A space-time permutation scan statistic for disease outbreak detection*, PLoS Medicine 2, **2005**, 2(3), e59.

- 47. Kulldorff M., Nagarwalla N.,** *Spatial disease clusters: Detection and inference*, *Statistics in Medicine*, **1995**, 14, 799-819.
- 48. Kupper L.L., Karon J.M., Kleinbaum D.G., i inni,** *Matching in epidemiological studies: validity and efficiency considerations*, *Biometrics*, **1981**, 37, 271-92.
- 49. Kurinczuk J.J., Clarke M.,** *A case-control study to investigate the role of recent spontaneous abortion in the aetiology of neural tube defects*, *Paediatric and Perinatal Epidemiology*, **1993**, 7(2), 167-76.
- 50. Lai P.C., So F.M., Chan K.W.,** *Spatial epidemiological approaches in disease mapping and analysis*, CRC Press, Boca Raton, **2009**.
- 51. Lancaster H.O.,** *Significance tests in discrete distributions*, *Journal of the American Statistical Association*, **1961**, 56, 223-234.
- 52. Latos-Bieleńska A., Materna-Kiryluk A.,** *Wrodzone wady rozwojowe w Polsce w latach 2003-2004. Dane z Polskiego Rejestru Wrodzonych Wad Rozwojowych*, red. A. Latos-Bieleńska, A. Materna-Kiryluk, OWN, Poznań, **2006**.
- 53. Latos-Bieleńska A., Materna-Kiryluk A.,** *Polish Registry of Congenital Malformation – aims and organization of the Registry monitoring 300 000 births a year*, *Journal of Applied Genetics*, **2005**, 46(4), 341-348.
- 54. Latos-Bieleńska A., Materna-Kiryluk A. i inni,** *Polski Rejestr Wrodzonych Wad rozwojowych*, red. Latos-Bieleńska A., OWN, Poznań, **1998**.
- 55. Materna-Kiryluk A., Więckowska B., Wiśniewska K., Borszewska-Kornacka M.K., Godula-Stuglik U., Limon J., Rusin J., Sawulicka-Oleszczuk H., Szwatkiewicz-Warowicka E., Walczak M.,** *Maternal reproductive history and the risk of isolated congenital malformations*, *Paediatric and Perinatal Epidemiology*, **2011**, 25(2), 135-143.
- 56. Materna-Kiryluk A., Wiśniewska K., Badura-Stronka M., Mejnartowicz J., Więckowska B., Balcar-Boroń A., Czerwionka-Szaflarska M., Gajewska E., Godula-Stuglik U., Krawczyński M., Limon J., Rusin J., Sawulicka-Oleszczuk H., Szwatkiewicz-**

- Warowicka E., Walczak M., Latos-Bieleńska A.,** *Parental age as a risk factor for isolated congenital malformations in a Polish population, Paediatric and Perinatal Epidemiology, 2009, 23 (1), 29-40.*
- 57. Messer L. C., Luben T.J., Mendola P., Carozza S. E., Horel S. A., Langlois P. H.,** *Urban-Rural Residence the Occurrence of Cleft Lip and Cleft Palate in Texas, 1999-2000, Annals of Epidemiology, 2010, 20, 32-39.*
- 58. Miettinen O.S.,** *Theoretical Epidemiology: Principles of Occurrence Research in Medicine,* John Wiley and Sons, New York, **1985.**
- 59. Miettinen O.S., Nurminen M.,** *Comparative analysis of two rates, Statistics In Medicine, 1985, 4, 213-226.*
- 60. Moran P.A.P.,** *Notes on continuous stochastic phenomena, Biometrika, 1950, 37, 17-23.*
- 61. Openshaw S., Charlton M. E., Wymer C., Craft A.,** *A mark I geographical analysis machine for the automated analysis of point data sets, International Journal of Geographical Information System, 1987, 1, 335-358.*
- 62. Plackett R.L.,** *Discussion of Yates' "Tests of significance for 2x2 contingency tables", Journal of Royal Statistical Society Series A, 1984, 147, 426-463.*
- 63. Pratt J.W., Gibbons J.D.,** *Concepts of Nonparametric Theory,* Springer-Verlag, New York, **1981.**
- 64. Rivas F., Davalos I.P., Olivares N., Davalos N.O., Perez-Medina R., Gomez-Partida G., Chakraborty R.,** *Reproductive history in mothers of children with neural tube defects, Gynecol Obstet Invest, 2000, 49, 255-260.*
- 65. Rothman K.J., Greenland S., Lash T.L.,** *Modern Epidemiology (3rd edition),* Lippincott Williams and Wilkins, Philadelphia, **2008.**
- 66. Stephan F.F., McCarthy P.J.,** *Sampling options: an analysis of survey procedure,* Wiley, New York, **1958.**



- 67. Stevenson M., Stevens K.B., Rogers D.J., Clements A.C.A., Pfeiffer D.U., Robinson T.P.,** *Spatial Analysis in Epidemiology*, Oxford University press, Oxford, **2008**.
- 68. Tobler W.,** *A computer movie simulating urban growth in the Detroit region*, *Economic Geography*, **1970**, 46, 234-240.
- 69. Todoroff K., Shaw G.M.,** *Prior Spontaneous Abortion, Prior Elective Termination, Interpregnancy Interval, and Risk of Neural Tube Defects.*, *American Journal of Epidemiology*, **2000**, 151 (5), 505-511.
- 70. Turnbull B.W., Iwano E.J., Burnett W.S., Howe H.L., Clark L.C.,** *Monitoring for clusters of disease: application to leukemia incidence in upstate New York*, *American Journal of Epidemiology*, **1990**, 132(1 Suppl), 136–143.
- 71. Villamor E., Sparén P., Cnattingius S.,** *Risk of oral clefts in relation to prepregnancy weight change and interpregnancy interval*, *American Journal of Epidemiology*, **2008**, 167(11), 1305-11.
- 72. Waller L.A., Gotway C.A.,** *Applied Spatial Statistics for Public Health Data*, Wiley, Hoboken, **2004**.
- 73. Whiteman D., Murphy M., Hey K., O'Donnell M., Goldacre M.,** *Reproductive Factors, Subfertility, and Risk of Neural Tube Defects: A CaseControl Study Based on the Oxford Record Linkage Study Register*, *American Journal of Epidemiology*, **2000**, 152(9), 823-828.
- 74. Yates F.,** *Contingency tables involving small numbers and the chi-square test*, *Journal of the Royal Statistical Society*, **1934**, 1(2), 17-235.
- 75. Zieliński A.,** *Dobór próby w badaniach epidemiologicznych*, *Przegląd Epidemiologiczny*, **2002**, 56, 489-98.

# 11 ZAŁĄCZNIK 1

## ZGŁOSZENIE WADY U DZIECKA W WIEKU 0-2 LAT DO POLSKIEGO REJESTRU WRODZONYCH WAD ROZWOJOWYCH

### UWAGA !!!

Rejestr dotyczy wad u dzieci żywo i martwo urodzonych oraz wad wykrytych prenatalnie. Fakt zgłoszenia dziecka do Rejestru należy natychmiast odnotować na 7 stronie Książeczki Zdrowia Dziecka w punkcie "Testy przesiewowe", w rubryce "inne". Jeśli zgłoszenia dokonuje lekarz położnik lub neonatolog, fakt ten należy także odnotować w Historii Rozwoju Noworodka na stronie 2 w rubryce "Kontrola, badania".

Zgłoszenie proszę przelać bezpośrednio na adres:

POLSKI REJESTR WRODZONYCH WAD ROZWOJOWYCH

Katedra i Zakład Genetyki Medycznej UM

ul. Grunwaldzka 55 paw. 15, 60-352 POZNAŃ

tel.: (0-61) 854 73 45, 854 73 49; fax: (0-61) 854 73 48, strona internetowa: [www.rejestrwad.pl](http://www.rejestrwad.pl)

Przewodnicząca Zespołu ds. PRWWR: *Prof.dr hab.med. Anna Latos-Bieleńska*

Koordinator Organizacyjny: *Dr med. Anna Materna-Küryluk*

W sprawach dotyczących Rejestru można także kontaktować się z Wojewódzkim Zespołem ds. PRWWR dla woj. wielkopolskiego i południowej części lubuskiego:

Przewodniczący Wojewódzkiego Zespołu: *Prof.dr hab. med. Marian Krawczyński*

Klinika Gastroenterologii Dziecięcej i Chorób Metabolicznych UM, ul. Szpitalna 27/33, 60-572 POZNAŃ

**WAŻNE:** przy wypełnianiu formularza zwrócić uwagę, czy żaden punkt nie został pominięty

(\*) właściwe pokreślić

ID pacjenta

1. Nazwisko i imię dziecka:
2. Data urodzenia dziecka: --;      3. Płeć dziecka: męska/żeńska/nieokreślona/nieznana(\*);
4. Imiona rodziców: .....
5. Dokładne miejsce zamieszkania (matki): .....  
powiat ..... województwo: .....
6. Dziecko z ciąży: pojedynczej – bliźniaczej – trójacznej – innej ..... (\*)  
W przypadku ciąży mnogiej, którym kolejnym dzieckiem jest dziecko zgłaszane: I – II – III – .....  
Zapłodnienie drogą in vitro? NIE - TAK: .....
7. Dziecko z kolejnej ciąży: , porodu:       8. Wiek płodowy przy urodzeniu:  t.c;
9. Masa ciała urodzeniowa: g      10. Noworodek urodzony: żywo – martwo – nie wiadomo
11. Wiek dziecka (lub wiek ciążowy) w dniu rozpoznania wady: .....  
ewent. data rozpoznania --
12. Czy wada została rozpoznana prenatalnie? NIE – TAK: w  tygodniu ciąży;
13. Metoda diagnostyki prenatalnej: USG płodu – w których tygodniach ciąży .....  
– amniocenteza – inne techniki – jakie? .....
14. Badanie kariotypu dziecka/płodu: wykonano i wynik znany – wykonano i wynik nieznany – nie wykonywano – badanie nieudane – nie wiadomo;  
wynik badania kariotypu: .....
15. Poprzednie ciążę (poza dzieckiem zgłaszanym): liczba dzieci żyjących ; liczba dzieci zmarłych ;  
liczba poronień samoistnych ; liczba poronień sztucznych ; liczba porodów martwych
16. Przebieg ciąży: prawidłowy – nieprawidłowy – nie wiadomo (\*)  
krwawienie z dróg rodnych (tydz. ciąży?) : .....  
infekcje w czasie ciąży (jakie i tydz. ciąży?) : .....  
cukrzyca, padaczka, inne choroby matki (jakie?) : .....  
małowodzie, wielowodzie, inna patologia ciąży (jaka?): .....  
leki (tydz. ciąży, jakie): .....  
alkohol, używki (tydz. ciąży, jakie, dawka?): .....  
badania rtg w ciąży(jakie, tydz. ciąży): .....  
kwas foliowy – dawka i okres przyjmowania: .....  
toksoplazmoza (czy wykonano badanie, w którym tyg. ciąży?): .....

17. Wada/y: rodzaj stwierdzonej wady rozwojowej (wad) – dokładny opis. Jeśli wykonano badania diagnostyczne potwierdzające wadę, podać jakie.	Kod ICD 10(#)

18. Czy nastąpił zgon dziecka ze zgłaszaną wadą rozwojową: *TAK – NIE*. Jeśli tak podać wiek, w jakim wystąpił zgon ..... i ewentualnie datę zgonu   -  -
19. Czy wykonano badanie anatomopatologiczne: tak i wynik znany – tak i wynik nieznan - nie wykonano – płód zmacerowany – nie wiadomo (\*). Opis badania anatomopatologicznego  
.....  
.....  
.....
20. Wiek matki przy urodzeniu dziecka    lat (ewent. także data urodzenia:   -  -  );
21. Wykształcenie matki: niepełne podstawowe – podstawowe – zasadnicze zawodowe – średnie zawodowe – średnie ogólnokształcące – pomaturalne – wyższe (\*)
22. Zawód wykonywany matki: ..... 23. Ekspozycja na szkodliwe czynniki środowiskowe (w miejscu pracy i inne) stałe.....  
w I trymestrze ciąży.....  
Czy u matki wykonano przypominające szczepienie przeciw różyczce po 20 roku życia: *TAK – NIE – NIE WIADOMO*
24. Wiek ojca przy urodzeniu dziecka    lat (ewent. także data urodzenia:   -  -  );
25. Wykształcenie ojca: niepełne podstawowe – podstawowe – zasadnicze zawodowe – średnie zawodowe – średnie ogólnokształcące – pomaturalne – wyższe (\*)
26. Zawód wykonywany ojca ..... 27. Ekspozycja na szkodliwe czynniki środowiskowe (w miejscu pracy i inne, np. alkohol); jakie:.....  
.....
28. Czy rodzice są spokrewnieni: *TAK – NIE – NIE WIADOMO* (*RODZAJ POKREWIEŃSTWA:.....*)
29. Czy w rodzinie występowały przypadki takiej samej wady lub zespołu wad: *TAK – NIE*.  
Jeśli tak, to u kogo: *RODZICE – RODZENSTWO – INNI KREWNI – jacy?* .....
30. Czy w rodzinie występowały inne wady rozwojowe lub choroby uwarunkowane genetycznie – u kogo i jakie?:  
.....

ID zgłoszenia \_\_\_\_\_

Data zgłoszenia wady:

pieczętka jednostki zgłaszającej

pieczętka lekarza zgłaszającego

Zgadzam się na zgłoszenie wady u dziecka do Polskiego Rejestru Wrodzonych Wad Rozwojowych. Administratorem danych jest Katedra i Zakład Genetyki Medycznej AM w Poznaniu (adres w nagłówku karty). Mają Państwo prawo do wglądu, poprawiania i uzupełniania danych osobowych. (Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych) UWAGA: w/w dane służą do badań nad epidemiologią i etiologią wad rozwojowych na terenie Polski i są objęte tajemnicą lekarską.

Data

Podpis rodzica

---

# 12 ZAŁĄCZNIK 2

---

## INSTRUKACJA

**Wypełnione formularze zgłoszeń należy przysyłać bezpośrednio na adres:**

POLSKI REJESTR WRODZONYCH WAD ROZWOJOWYCH  
Katedra i Zakład Genetyki Medycznej  
Akademii Medycznej w Poznaniu  
Skrytka pocztowa 79  
60-955 Poznań 37

### **WYTYCZNE DOTYCZĄCE WYPEŁNIANIA KARTY ZGŁOSZENIA WADY ROZWOJOWEJ DO POLSKIEGO REJESTRU WRODZONYCH WAD ROZWOJOWYCH**

**Zgłoszenie dotyczy wad u dzieci w wieku od urodzenia do ukończenia 24 miesiąca życia, a także wad u dzieci martwo urodzonych i wad wykrytych prenatalnie.**

Formularz zgłoszeniowy powinien wypełniać lekarz oddziału: noworodkowego, dziecięcego, położniczo-ginekologicznego; poradni: rejonowej-dziecięcej, specjalistycznej.

W celu uniknięcia dublowania zgłoszenia wady (u dzieci urodzonych żywo), lekarz który pierwszy stwierdził wadę rozwojową u danego dziecka i wypisał formularz wpisuje rodzaj zgłoszonej wady oraz datę zgłoszenia do Rejestru w Książeczce Zdrowia Dziecka na 3 stronie, w rubryce " Kwalifikacja do grupy dyspanseryjnej " lub na 7 stronie nowej Książeczki Zdrowia Dziecka w punkcie " Testy przesiewowe ", w rubryce " inne ". Jeśli zgłoszenia dokonuje lekarz położnik lub neonatolog, fakt ten należy także odnotować w Historii Rozwoju Noworodka na 2 stronie, w rubryce " Kontrole, badania ".

Jednostka zgłaszająca wadę wrodzoną wysyła na bieżąco wypełnione formularze bezpośrednio do Zespołu d/s Polskiego Rejestru Wad Rozwojowych ( adres na formularzu zgłoszenia ).

#### **Technika wypełniania formularza**

1. Formularz zgłoszeniowy wypełnia lekarz stwierdzający wadę.
2. W przypadku, gdy dziecko pochodzi z ciąży mnogiej należy wypełnić osobny formularz dla każdego noworodka z wadą rozwojową.
3. W przypadku wątpliwości, czy daną nieprawidłowość kwalifikować jako wadę, należy zawsze wypełnić formularz.
4. Opis wady należy podać w języku polskim. Należy dokładnie opisać każdą stwierdzoną wadę w punkcie 17 z podaniem narządu i lokalizacji ( w załączeniu lista wad podlegających zgłoszeniu).
5. Informacje dotyczące badania patomorfologicznego (punkt 19 ) wypełnić zarówno dla noworodków żywo urodzonych, które zmarły jak i martwych urodzeń.
6. Formularz powinien być podpisany przez lekarza zgłaszającego wadę oraz zawierać jego imienną pieczęć. Na formularzu winna się również znajdować czytelna pieczęć jednostki zgłaszającej.

## 13 ZAŁĄCZNIK 3

### Międzynarodowa Statystyczna Klasyfikacja Chorób i Problemów zdrowotnych (10)

<i>Grupa</i>	<i>Wady</i>
Poza kategorią Q ICD 10	I42.4 zwłóknienie sprężyste wsierdza (fibroelastoza)
	K07.0 znaczne nieprawidłowości wymiaru szczęk
	K07.1 nieprawidłowości stosunku łuków zębowych
	K07.6 zaburzenia stawów skroniowo-żuchwowych
	K07.9 nieprawidłowość zębowo-twarzowa, nie określona
	K00.6 zaburzenia w wyrzynaniu zębów
Q00-07 wady układu nerwowego	Q00.0 bezmózgowie
	Q00.2 rozszczep potylicy (iniencephalia)
	Q01.1 przepuklina mózgowa nosowo-czołowa
	Q01.2 przepuklina mózgowa potyliczna
	Q01.8 przepuklina mózgowa w innej lokalizacji
	Q01.9 przepuklina mózgowa, nie określona
	Q02 małogłowie
	Q03.0 wady rozwojowe wodociągu mózgu Sylwiusza
	Q03.1 zarośnięcie otworów Magendiego i Luschki
	Q03.9 wodogłowie wrodzone, nie określone
	Q04.0 wwr spoidła wielkiego
	Q04.2 przodomózgowie jednokomorowe (holoprosencephalia)
	Q04.3 inne wady mózgu z ubytkiem tkanek
	Q04.6 wrodzone torbiele mózgu
	Q04.9 wwr mózgu, nie określone
	Q05.0 rozszczep kręgosłupa szyjnego ze współistn. wodogłowiem
	Q05.1 rozszczep kręgosłupa piersiowego ze współistn. wodogłowiem
	Q05.2 rozszczep kręgosłupa lędźwiowego ze współistn. wodogłowiem
	Q05.3 rozszczep kręgosłupa krzyżowego ze współistn. wodogłowiem
	Q05.4 nieokreślony rozszczep kręgosłupa ze współistn. wodogłowiem
	Q05.5 rozszczep kręgosłupa szyjnego bez wodogłowia
	Q05.6 rozszczep kręgosłupa piersiowego bez wodogłowia
	Q05.7 rozszczep kręgosłupa lędźwiowego bez wodogłowia
	Q05.8 rozszczep kręgosłupa krzyżowego bez wodogłowia
	Q05.9 rozszczep kręgosłupa, nie określony
	Q07.0 zespół Arnoldda-Chiariego
	Q07.9 wwr układu nerwowego, nie określone
	Q06.4 wodordzenie
Q10-18 wady oka, ucha, twarzy i szyi	Q10.3 inne wwr powiek
	Q11.1 inne bezocze
	Q11.2 małocze
	Q12.0 wrodzona zaćma
	Q13.0 szczelina tęczówki

<i>Grupa</i>	<i>Wady</i>
	Q13.1 brak tęczówki
	Q13.2 inne wwr rogówki
	Q13.3 wrodzone zmętnienie rogówki
	Q14.1 wwr siatkówki
	Q14.3 wwr naczyńówki
	Q14.8 inne wwr tylnego odcinka oka
	Q14.9 wwr tylnego odcinka oka, nie określona
	Q15.0 jaskra wrodzona
	Q15.9 wwr oka, nie określona
	Q16.0 wrodzony brak małżowiny (usznej)
	Q16.1 wrodzony brak, zarośnięcie i zwężenie przewodu słuchowego (zewn.)
	Q17.0 małżowina uszna dodatkowa
	Q17.2 małżowina uszna mała
	Q17.9 wwr ucha, nie określona
	Q18.0 zatoka, przetoka i torbiel szczeliny skrzelowej
	Q18.2 inne wwr szczeliny skrzelowej
	Q18.3 taśma szyjna (skrzydlik szyjny)
	Q18.4 duża szpara ustna (makrostomia)
	Q18.8 inne określone wady rozwojowe twarzy i szyi
Q20-28 wady serca	Q20.0 wspólny pień tętniczy
	Q20.1 odejście obu dużych naczyń z prawej komory
	Q20.2 odejście obu dużych naczyń z lewej komory
	Q20.3 nieprawidłowe połączenia komorowo-tętnicze
	Q20.4 odejście obu dużych naczyń serca z jednej komory
	Q20.5 nieprawidłowe połączenie przeds.-komorowe
	Q20.8 inne wwr jam i połączeń sercowych
	Q21.0 ubytek przegrody międzykomorowej
	Q21.1 ubytek przegrody międzyprzedsionkowej
	Q21.2 ubytek przegrody przedsionkowo-komorowej
	Q21.3 tetralogia Fallota
	Q21.8 inna wwr przegród serca
	Q21.9 wrodzona wada przegrody serca, nie określona
	Q22.0 zarośnięcie zastawki pnia płucnego
	Q22.1 wrodzone zwężenie zastawki pnia płucnego
	Q22.2 wrodzona niedomykliwość zastawki pnia płucnego
	Q22.3 inne wwr zastawki pnia płucnego
	Q22.4 wrodzone zwężenie zastawki trójdzielnej
	Q22.5 choroba Ebsteina
	Q22.6 zespół niedorozwoju prawej części serca
	Q22.8 inne wwr zastawki trójdzielnej
	Q22.9 wwr zastawki trójdzielnej, nie określona
	Q23.0 wrodzone zwężenie ujścia tętniczego lewego
	Q23.1 wrodzona niedomykalność zastawki aorty
	Q23.2 wrodzone zwężenie zastawki dwudzielnej
	Q23.3 wrodzona niedomykalność zastawki dwudzielnej
	Q23.4 zespół niedorozwoju lewej części serca
	Q23.8 inne wwr zastawki: aorty i dwudzielnej
	Q23.9 wrodzona wada zastawki: aorty i dwudzielnej, nie określona
	Q24.0 prawostronne położenie serca
	Q24.2 serce trójprzedsionkowe

<i>Grupa</i>	<i>Wady</i>
	Q24.5 wada rozwojowa naczyń wieńcowych
	Q24.6 wrodzony blok serca
	Q24.8 inne określone wrodzone wady rozwojowe serca
	Q25.0 drożny przewód tętniczy
	Q25.1 zwężenie cieśni aorty
	Q25.2 zarośnięcie aorty
	Q25.3 zwężenie aorty
	Q25.4 inne wwr aorty
	Q25.5 zarośnięcie pnia płucnego
	Q25.6 zwężenie pnia płucnego
	Q25.7 inne wwr pnia płucnego
	Q25.8 inne wwr dużych tętnic
	Q25.9 wwr dużych tętnic, nie określona
	Q26.0 wrodzone zwężenie żyły głównej
	Q26.1 przetrwała płodowa lewa górna żyła główna
	Q26.4 nieprawidłowe ujście żył płucnych, nie określone
	Q26.8 inne wwr dużych żył układowych
	Q28.3 inne wady rozwojowe naczyń mózgowych
Q30-34 wady układu oddechowego	Q30.0 zarośnięcie nozdrzy
	Q30.1 agenezja lub niedorozwój nosa
	Q30.8 inne wwr nosa
	Q31.2 niedorozwój krtani
	Q31.4 wrodzony świst krtaniowy
	Q31.8 inne wwr krtani
	Q31.9 wwr krtani, nie określona
	Q32.0 wrodzone rozmiękanie tchawicy
	Q32.1 inne wwr tchawicy
	Q32.4 inne wwr oskrzeli
	Q33.0 wrodzona torbielowatość płuc
	Q33.1 dodatkowy płat płuca
	Q33.3 niewytworzenie płuca
	Q33.6 niedorozwój lub dysplazja płuca
	Q33.8 inne wwr płuc
	Q33.9 wwr płuca, nie określona
	Q34.9 wwr ukł. oddechowego, nie określona
Q35-37 rozszczepy wargi i/lub podniebienia	Q35.0 rozszczep podniebienia twardego, obustronny
	Q35.1 rozszczep podniebienia twardego, jednostronny
	Q35.2 rozszczep podniebienia miękkiego, obustronny
	Q35.3 rozszczep podniebienia miękkiego, jednostronny
	Q35.4 rozszczep podniebienia twardego i miękkiego, obustronny
	Q35.5 rozszczep podniebienia twardego i miękkiego, jednostronny
	Q35.6 rozszczep podniebienia pośrodkowy
	Q35.8 rozszczep podniebienia nieokreślony, obustronny
	Q35.9 rozszczep podniebienia nieokreślony, jednostronny
	Q36.0 rozszczep wargi, obustronny
	Q36.1 rozszczep wargi pośrodkowy
	Q36.9 rozszczep wargi, jednostronny
	Q37.0 rozszczep podniebienia twardego wraz z rozszczepem wargi, obustronny

<i>Grupa</i>	<i>Wady</i>
	Q37.1 rozszczep podniebienia twardego wraz z rozszczepem wargi, jednostronny
	Q37.2 rozszczep podniebienia miękkiego wraz z rozszczepem wargi, obustronny
	Q37.3 rozszczep podniebienia miękkiego wraz z rozszczepem wargi, jednostronny
	Q37.4 rozszczep podniebienia miękkiego, twardego i wargi, obustronny
	Q37.5 rozszczep podniebienia miękkiego, twardego i wargi, jednostronny
	Q37.8 rozszczep podniebienia wraz z rozszczepem wargi nie określony, obustronny
	Q37.9 rozszczep podniebienia wraz z rozszczepem wargi nie określony, jednostronny
Q38-45 wady przewodu pokarmowego	Q38.2 język olbrzymi
	Q38.3 inne wwr języka
	Q38.5 wwr podniebienia niesklasyfikowane gdzie indziej
	Q38.6 inne wwr jamy ustnej
	Q39.0 zarośnięcie przełyku bez przetoki
	Q39.1 zarośnięcie przełyku z przetoką tchawiczo-przełykową
	Q39.2 wrodzona przetoka tchawiczo-przełykowa bez zarośnięcia
	Q40.0 wrodzone przerostowe zwężenie odźwiernika
	Q40.1 wrodzona przepuklina rozworu przełykowego przepony
	Q40.2 inne określone wady rozwojowe żołądka
	Q41.0 wrodzony brak, zarośnięcie lub zwężenie dwunastnicy
	Q41.1 wrodzony brak, zarośnięcie lub niedorozwój j. czczego
	Q41.2 wrodzony brak, zarośnięcie lub niedorozwój j. krętego
	Q41.9 wrodzony brak, zarośnięcie lub zwężenie nie określonej części j. cienkiego
	Q42.8 wrodzony brak, zarośnięcie lub zwężenie innych części j. grubego
	Q42.0 wrodzony brak, zarośnięcie lub zwężenie odbytnicy z przetoką
	Q42.1 wrodzony brak, zarośnięcie lub zwężenie odbytnicy bez przetoki
	Q42.2 wrodzony brak, zarośnięcie lub zwężenie odbytu z przetoką
	Q42.3 wrodzony brak, zarośnięcie lub zwężenie odbytu bez przetoki
	Q43.0 uchyłek Meckela
	Q43.1 choroba Hirschsprunga
	Q43.3 wwr umocowania jelit
	Q43.5 odbył przemieszczony
	Q43.6 wrodzona przetoka odbytnicy i odbytu
	Q43.7 przetrwały stek (kloaka)
	Q44.4 torbiel przewodu żółciowego wspólnego
	Q44.5 inne wwr przewodów żółciowych
	Q44.6 torbielowatość wątroby
	Q44.7 inne wwr wątroby
	Q45.1 trzustka pierścieniowata
	Q45.8 inne określone wady rozwojowe układu pokarmowego
	Q45.9 wrodzona wada układu pokarmowego, nie określona



<i>Grupa</i>	<i>Wady</i>
Q50-56 wady narządów płciowych	Q50.0 wrodzony brak jajnika
	Q50.1 torbiel rozwojowa jajnika
	Q50.6 inne wwr jajowodu lub więzadła szerokiego macicy
	Q51.4 macica jednoróżna
	Q52.2 wrodzona przetoka odbytniczo-pochwowa
	Q52.5 zarośnięcie warg sromowych
	Q52.6 wwr łechtaczki
	Q52.7 inne wwr sromu
	Q52.9 wrodzona wada narządów płciowych żeńskich, nie określona
	Q54.1 spodziewstwo prąciowe
	Q54.2 spodziewstwo prąciowo-mosznowe
	Q54.3 spodziewstwo kroczone
	Q54.8 spodziewstwo, inne
	Q54.9 spodziewstwo, nie określone
	Q55.0 brak lub agenezja jądra
	Q55.1 niedorozwój jąder i moszny
	Q55.2 inne wwr jąder lub moszny
	Q55.5 wrodzony brak lub aplazja prącia
	Q55.6 inne wwr prącia
	Q55.8 inne określone wwr narządów płciowych męskich
	Q55.9 wwr narządów płciowych męskich, nie określona
	Q56.2 obojnactwo rzekome żeńskie, niesklasyfikowane gdzie indziej
	Q56.4 płeć niezdeterminowana, nie określona
Q60-64 wady układu moczowego	Q60.0 agenezja nerki, jednostronna
	Q60.1 agenezja nerki, obustronna
	Q60.2 agenezja nerki, nie określona
	Q60.3 niedorozwój nerki, jednostronny
	Q60.4 niedorozwój nerki, obustronny
	Q60.5 niedorozwój nerki, nie określony
	Q60.6 Poterra, sekwencja
	Q61.0 wrodzona torbiel nerki (pojedyncza)
	Q61.1 wielotorbielowatość nerek, typu dziecięcego
	Q61.3 wielotorbielowatość nerek, nie określona
	Q61.4 dysplazja nerek
	Q61.8 inne torbielowate choroby nerek
	Q61.9 choroba torbielowata nerek, nie określona
	Q62.0 wodonercze wrodzone
	Q62.1 zarośnięcie lub zwężenie moczowodu
	Q62.2 wrodzony moczowód olbrzymi (poszerzenie moczowodu)
	Q62.4 niewykształcenie moczowodu (brak)
	Q62.5 zdwojenie moczowodu
	Q62.7 wrodzony odpływ pęcherzowo-moczowodowo-nerkowy
	Q63.0 nerka dodatkowa
	Q63.1 nerka płatowata, podkowiasta lub złączenie nerek
	Q63.2 nerka przemieszczona (ektopiczna)
	Q63.3 nerka olbrzymia lub hyperplastyczna
	Q63.8 inna określona wada rozwojowa nerki
	Q63.9 wwr nerki, nie określona

<i>Grupa</i>	<i>Wady</i>
	Q64.0 wierzchniactwo
	Q64.1 wynicowanie pęcherza moczowego
	Q64.2 wrodzona zastawka cewki tylnej
	Q64.3 inne zarośnięcie lub zwężenie cewki i szyi pęcherza moczowego
	Q64.4 wada rozwojowa przewodu omocznego
	Q64.5 wrodzony brak pęcherza lub cewki moczowej
	Q64.7 inne wwr pęcherza i cewki moczowej
	Q64.9 wwr ukł moczowego, nie określona
Q65-79 wady ukł. mięśniowo-szkieletowego	Q65.0 wrodzone zwichnięcie st. biodrowego, jednostronne
	Q65.1 wrodzone zwichnięcie st. biodrowego, obustronne
	Q65.2 wrodzone zwichnięcie st. Biodrowego, nie określone
	Q65.3 wrodzone nadwichnięcie st. biodrowego, jednostronne
	Q65.8 inne wrodzone zniekształcenia stawu biodrowego
	Q65.9 wrodzone zniekształcenie stawu biodrowego, nie określone
	Q66.0 stopa końsko-szpotawa, wrodzona
	Q66.2 szpotawość śródstopia wrodzona
	Q66.3 inne wrodzone zniekształcenia szpotawe stopy
	Q66.4 stopa piętowo-koślawa wrodzona
	Q66.5 wrodzona stopa płaska
	Q66.6 inne wrodzone zniekształcenia koslawe stopy
	Q66.8 inne wrodzone zniekształcenia stopy
	Q67.0 asymetria twarzy
	Q67.2 długogłowie (dolichocephalia)
	Q67.4 inne wrodzone zniekształcenia czaszki, twarzy i szczęk
	Q67.5 wrodzone zniekształcenie kręgosłupa
	Q67.6 klatka piersiowa szewska (lejkowata)
	Q67.7 klatka piersiowa kurza
	Q67.8 klatka piersiowa dzwonowata
	Q68.0 wrodzone zniekształcenie mięśnia most.-oboj.-sutkowego
	Q68.2 wrodzone zniekształcenie kolana
	Q68.4 wrodzone wygięcie piszczeli lub strzałki
	Q69.0 dodatkowy(e) palec(ce) rąk
	Q69.1 dodatkowy(e) kciuk(i)
	Q69.2 dodatkowy(e) palec(ce) stóp
	Q69.9 palce dodatkowe, nie określone
	Q70.0 palce rąk splecione (z kościorostem)
	Q70.1 palce rąk zrośnięte (bez kościorostu i bno)
	Q70.3 zrost palców stóp (bez kościorostu)
	Q70.4 palce dodatkowe ze zrostem palców (polisyndaktylia)
	Q70.9 zrost palców, nie określony
	Q71.0 wrodzony brak całkowity kończyn(y) górnych(ej)
	Q71.1 wrodzony brak ramienia i przedramienia z zachowaną dłonią
	Q71.2 wrodzony brak przedramienia i dłoni
	Q71.3 wrodzony brak dłoni i palców
	Q71.4 podłużne zniekształcenie zmniejszające kości promieniowej
	Q71.5 podłużne zniekształcenie zmniejszające kości łokciowej
	Q71.6 ręka przypominająca szczyłce homara
	Q71.8 inne zniekształcenia zmniejszające kończyn górnych

<i>Grupa</i>	<i>Wady</i>
	Q71.9 zniekształcenie zmniejszające kończyny górnej, nie określone
	Q72.0 wrodzony brak całkowity kończyn(y) dolnych(ej)
	Q72.1 wrodzony brak uda i podudzia z zachowaną stopą
	Q72.2 wrodzony brak podudzia i stopy
	Q72.3 wrodzony brak stopy i palców
	Q72.4 podłużne zniekształcenie zmniejszające kości udowej
	Q72.6 podłużne zniekształcenie zmniejszające kości strzałkowej
	Q72.7 rozszczep stopy
	Q72.8 inne zniekształcenia zmniejszające kończy(y) dolnych(ej)
	Q72.9 zniekształcenie zmniejszające kończyny dolnej, nie określone
	Q73.8 inne zmniejszające zniekształcenia kończyn(y), nie określonych(ej)
	Q74.0 inne wwr kończyn(y) górnych(ej), z wyłączeniem obręczy barkowej
	Q74.1 wwr stawu kolanowego
	Q74.2 inne wwr kończyn(y) dolnych(ej), z wyłączeniem obręczy miednicznej
	Q74.3 wrodzona sztywność wielostawowa
	Q74.8 inne określone wwr kończyn(y)
	Q74.9 nieokreślona wwr kończyn(y)
	Q75.0 przedwczesne skostnienie szwów czaszkowych
	Q75.1 dyzostoza czaszkowo-twarzowa (Crouzona, zespół)
	Q75.3 wielkogłowie
	Q75.8 inne określone wwr kości czaszki i twarzy
	Q75.9 wwr kości czaszki i twarzy, nie określona
	Q76.0 utajony rozszczep kręgosłupa
	Q76.1 Klippel-Feila, zespół
	Q76.3 wrodzone skrzywienie boczne kręgosłupa spowodowane wadami kości
	Q76.4 inne wwr kręgosłupa nie związane ze skrzywieniem bocznym
	Q76.6 inne wwr żeber
	Q76.7 wwr mostka
	Q76.8 inne wwr klatki piersiowej kostnej
	Q76.9 wwr klatki piersiowej kostnej, nie określona
	Q77.0 brak tkanki chrzęstnej (achondrogenезja)
	Q77.1 karłowatość, postać śmiertelna (thanatophoric dwarfism)
	Q77.3 chondrodysplazja punktowa (chondrodysplasia punctata)
	Q77.4 achondroplazja
	Q77.8 inne osteochondrodysplazje z upośledzeniem wzrostu kości długich i kręgosłupa
	Q77.9 osteochondrodysplazja z upośledzeniem wzrostu kości długich i kręgosłupa, nie określone
	Q78.0 k+C435ostnienie niedoskonałe+C467+C275
	Q78.6 mnogie wyrośla kostne wrodzone
	Q78.9 osteochondrodysplazja, nie określona
	Q79.1 inne wwr przepony
	Q79.2 przepuklina pępowinowa
	Q79.3 wytrzewienie (gastroschisis)
	Q79.4 zespół suszonej śliwki (zespół wiotkiego brzucha)

<i>Grupa</i>	<i>Wady</i>
	Q79.5 inne wwr ściany jamy brzusznej
	Q79.8 inne wwr ukł. mięśniowo-szkieletowego
	Q79.9 wwr ukł. mięśniowo-szkieletowego, nie określona
Q80-85 wady powłok ciała	Q80.2 rybia łuska, blaszkowata (collodion baby)
	Q80.9 wrodzona rybia łuska, nie określona
	Q81.2 pęcherzowe oddzielanie naskórka postać dystroficzna
	Q81.9 pęcherzowe oddzielanie naskórka, nie określone
	Q82.3 nietrzymanie barwnika
	Q82.8 inne określone wady rozwojowe skóry
	Q84.0 wrodzone łysienie
	Q84.8 inne określone wwr powłok ciała
	Q84.9 wwr powłok ciała, nie określona
	Q85.0 nerwiakowłókniakowatość niezłośliwa (ch. von Recklinhausena)
	Q85.1 stwardnienie guzowate (ch. Bourneville'a)
	Q85.8 inne fakomatozy niesklasyfikowane gdzie indziej
	Q85.9 fakomatoza, nie określona
Q86-87 inne zespoły wad (wady mnogie)	P35.0 wrodzony zespół różyczkowy
	P35.1 wrodzone zakażenie wirusem cytomegalii
	P37.1 wrodzona toksoplazmoza
	Q86.0 płodowy zespół alkoholowy
	Q86.8 inne zespoły wwr spowodowane przez znane czynniki zew.
	Q87.0 zespoły wwr dotyczące głównie wyglądu twarzy
	Q87.2 zespoły wwr dotyczące głównie kończyn
	Q87.4 Marfana, zespół
	Q87.2 zespoły wwr dotyczące głównie kończyn
	Q87.0 zespoły wwr dotyczące głównie wyglądu twarzy
	Q87.1 zespoły wwr związanych głównie z niskim wzrostem
	Q87.3 zespoły wwr związane z wczesnym nadmiernym rozwojem
	Q89.7 mnogie wwr niesklasyfikowane gdzie indziej
Q89 inne lub niesklas. wady wrodzone	Q89.0 wwr śledziony
	Q89.1 wwr nadnerczy
	Q89.3 ułożenie odwrotne
	Q89.4 bliźnięta zrosnięte (zrosłaki)
	Q89.9 wrodzona wada nieokreślona
Q90-99 aberracje chromosomowe, niesklas. gdzie indziej	Q90.0 trisomia 21, mejotyczna nierozdzielność
	Q90.1 trisomia 21, mitotyczna nierozdzielność (mozaika)
	Q90.2 trisomia 21, translokacja
	Q90.9 Downa, zespół, nie określony
	Q91.0 trisomia 18, mejotyczna nierozdzielność
	Q91.2 trisomia 18, translokacja
	Q91.3 Edwardsa, zespół, nie określony
	Q91.4 trisomia 13, mejotyczna nierozdzielność
	Q91.5 trisomia 13, mitotyczna nierozdzielność (mozaika)
	Q91.7 Patau, zespół, nie określony

<i>Grupa</i>	<i>Wady</i>
	Q92.1 trisomia całego chromosomu, mozaika (mitotyczna nierozdzielcz)
	Q92.3 trisomia mała (duplikacja mniej niż całego ramienia chromosomu)
	Q92.5 duplikacje współistniejące z innymi złożonymi przemieszczeniami
	Q92.6 dodatkowy chromosom markerowy
	Q92.7 triploidia i poliploidia
	Q92.8 inne określone całkowite i częściowe trisomie autosomów
	Q93.2 chromosom przemieszczony z ringiem lub dicentryczny
	Q93.3 delecja krótkiego ramienia chromosomu 4 (zesp. Wolffa-Hirschorna)
	Q93.4 delecja krótkiego ramienia chromosomu 5 (zesp. cri-du-chat)
	Q93.5 inne delecje części chromosomów
	Q95.3 zrównoważone przemieszcz. chromosomów płciowych/autosomów u nieprawidłowego osobnika
	Q96.0 kariotyp 45,X
	Q96.4 mozaika 45,X/inne linie komórkowe z nieprawidłowym chromosomem płciowym
	Q96.9 Turnera, zespół, nie określony
	Q98.0 Klinefeltera, zespół, kariotyp 47,XXY
	Q98.4 Klinefeltera, zespół, nie określony
	Q99.8 inne określone aberracje chromosomowe
	Q99.9 aberracja chromosomowa, nie określona
	Q97.2 mozaika, linie komór. z różnymi liczbami chrom. X

wwr - wrodzona wada rozwojowa lub wrodzone wady rozwojowe  
bno - bliżej nie określona

# 14 ZAŁĄCZNIK 4

Tabela A 1 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną” - zmienne ilościowe i jakościowe o uporządkowanych kategoriach

potencjalne czynniki ryzyka	kontrola dopasowana, N=442		CL±P, N=442	test $\chi^2$ dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
wiek płodowy w tygodniach #				0.3759 (NS)
	<28 2 (0.5%)	2 (0.5%)	50.0%	
	28-31 0 (0.0%)	7 (1.6%)	100.0%	
	32-36 33 (7.5%)	26 (5.9%)	44.1%	
	37-41 388 (87.8%)	382 (86.4%)	49.6%	
	>=42 14 (3.2%)	11 (2.5%)	44.0%	
	bd 5 (1.1%)	14 (3.2%)		
masa urodzeniowa w gramach **				1.55e-005
	<1499 1 (0.2%)	5 (1.1%)	83.3%	
	1500-1999 5 (1.1%)	14 (3.2%)	73.7%	
	2000-2490 10 (2.3%)	25 (5.7%)	71.4%	
	2500-2990 59 (13.3%)	85 (19.2%)	59.0%	
	3000-3490 173 (39.1%)	148 (33.5%)	46.1%	
	3500-3990 134 (30.3%)	123 (27.8%)	47.9%	
	4000-4490 51 (11.5%)	30 (6.8%)	37.0%	
	>=4500 7 (1.6%)	7 (1.6%)	50.0%	
	bd 2 (0.5%)	5 (1.1%)		
wiek matki w latach \$				
	<=19	-	-	
	20-24	-	-	
	25-29	-	-	
	30-34	-	-	
	35-39	-	-	
	>=40	-	-	
	bd	-	-	
wiek ojca w latach #				0.7304 (NS)
	<=19 4 (0.9%)	7 (1.6%)	63.6%	
	20-24 52 (11.8%)	66 (14.9%)	55.9%	
	25-29 151 (34.2%)	128 (29.0%)	45.9%	
	30-34 134 (30.3%)	123 (27.8%)	47.9%	
	35-39 65 (14.7%)	60 (13.6%)	48.0%	
	40-44 24 (5.4%)	24 (5.4%)	50.0%	
	45-49 5 (1.1%)	9 (2.0%)	64.3%	
	>=50 3 (0.7%)	3 (0.7%)	50.0%	
	bd 4 (0.9%)	22 (5.0%)		
wykształcenie matki **				5.193e-005
	=< podstawowe 43 (9.7%)	66 (14.9%)	60.6%	
	zawodowe 85 (19.2%)	106 (24.0%)	55.5%	
	średnie 162 (36.7%)	141 (31.9%)	46.5%	

	wyższe	151 (34.2%)	104 (23.5%)	40.8%	
	bd	1 (0.2%)	25 (5.7%)		
<b>wykształcenie ojca #</b>					<b>4.212e-005</b>
	=< podstawowe	23 (5.2%)	37 (8.4%)	61.7%	
	zawodowe	149 (33.7%)	160 (36.2%)	51.8%	
	średnie	131 (29.6%)	120 (27.1%)	47.8%	
	wyższe	127 (28.7%)	71 (16.1%)	35.9%	
	bd	12 (2.7%)	54 (12.2%)		
<b>kolejność ciąży ***</b>					<b>0.0003845</b>
	<b>pierworódka</b>	<b>240 (54.3%)</b>	<b>203 (45.9%)</b>	<b>45.8%</b>	
	<b>2 ciąża</b>	<b>130 (29.4%)</b>	<b>126 (28.5%)</b>	<b>49.2%</b>	
	<b>3 ciąża</b>	<b>47 (10.6%)</b>	<b>65 (14.7%)</b>	<b>58.0%</b>	
	<b>&gt;=4 ciąża</b>	<b>25 (5.7%)</b>	<b>48 (10.9%)</b>	<b>65.8%</b>	
ilość wcześniejszych poronień samoistnych *					0.5211 (NS)
	0 poronień	384 (86.9%)	386 (87.3%)	50.1%	
	1 poronienie	41 (9.3%)	44 (10.0%)	51.8%	
	2 poronienia	14 (3.2%)	11 (2.5%)	44.0%	
	3 poronienia	3 (0.7%)	1 (0.2%)	25.0%	

\*\*\* zmienna wykorzystana w trzech modelach regresji logistycznej zbudowanych dla kolejności ciąży (prostym, dużym i pełnym)

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla kolejności ciąży (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla kolejności ciąży (pełnym)

\$ zmienna tworząca dopasowanie grupy badanej i kontrolnej

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

# nie uwzględnione w modelach

**Tabela A 2 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną” - zmienne ilościowe ciągle**

potencjalne czynniki ryzyka	kontrola populacyjna, N=442		CL±P, N=442		test Fishera-Snedecora	korekta Cochrana-Coxa testu t_Studenta
	średnia (odchlenie st.)	test Lillieforsa wartość p	średnia (odchlenie st.)	test Lillieforsa wartość p	wartość p	wartość p
wiek płodowy w tygodniach	39.0 (1.9)	<0.01	38.9 (2.2)	<0.01	0.0050	0.3689 (NS)
<b>masa urodzeniowa w gramach</b>	3406.9 (541.3)	<0.01	3251.4 (615.1)	<0.01	0.0076	<b>0.0001</b>
wiek matki w latach	27.4 (5.5)	<0.01	27.4 (5.5)	<0.01	x	x
wiek ojca w latach	30.5 (5.8)	<0.01	30.3 (6.3)	<0.01	0.0936	0.7601* (NS)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

\*wartość p testu test t-Studenta

Tabela A 3 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną”- zmienne dichotomiczne

potencjalne czynniki ryzyka	kontrola dopasowana, N=442		CL±P, N=442	test mid-p
		liczność (% z N)	liczność (% z N)	wartość p
pleć **				0.000454
	męska	218 (49.3%)	267 (60.4%)	
	żeńską	224 (50.7%)	170 (38.5%)	
	bd	0 (0.0%)	5 (1.1%)	
miejsce zamieszkania matki *				0.897237(NS)
	wieś	187 (42.3%)	183 (41.4%)	
	miasto	255 (57.7%)	254 (57.5%)	
	bd	0 (0.0%)	5 (1.1%)	
palenie tytoniu przez matkę **				5.42E-009
	nie	356 (80.5%)	322 (72.9%)	
	tak	10 (2.3%)	54 (12.2%)	
	bd	76 (17.2%)	66 (14.9%)	
infekcje ukł. oddech. matki podczas 1 <sup>szego</sup> trymestru **				0.000668
	nie	411 (93.0%)	334 (75.6%)	
	tak	9 (2.0%)	26 (5.9%)	
	bd	22 (5.0%)	82 (18.6%)	
inne infekcje matki podczas ciąży **				0.018753
	nie	435 (98.4%)	413 (93.4%)	
	tak	2 (0.5%)	10 (2.3%)	
	bd	5 (1.1%)	19 (4.3%)	
infekcje ukł. mocz. matki podczas 1 <sup>szego</sup> trymestru **				0.016498
	nie	342 (77.4%)	366 (82.2%)	
	tak	17 (3.8%)	6 (1.4%)	
	bd	83 (18.8%)	70 (15.8%)	
narażenie matki na promieniowanie podczas ciąży *				0.941474(NS)
	nie	361 (85.5%)	381 (90.3%)	
	tak	4 (0.9%)	4 (0.9%)	
	bd	77 (18.2%)	57 (13.5%)	
wcześniejsze poronienia sztuczne *				0.500566(NS)
	nie	441 (99.8%)	441 (99.8%)	
	tak	0 (0.0%)	1 (0.2%)	
	bd	1 (0.2%)	0 (0.0%)	
wcześniejsze porody martwe *				0.993527(NS)
	nie	431 (97.5%)	434 (98.2%)	
	tak	3 (0.7%)	3 (0.7%)	
	bd	8 (1.8%)	5 (1.1%)	

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla kolejności ciąży (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanych dla kolejności ciąży (pełnym)

NS: brak istotności statystycznej -- dwustronna wartość-p >=0.05



Tabela A 4 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciążą drugiej lub dalszej - zmienne ilościowe i jakościowe o uporządkowanych kategoriach

potencjalne czynniki ryzyka	kontrola dopasowana, N=973		CL±P, N=244	test $\chi^2$ dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
wiek płodowy w tygodniach #				0.001497
	<28 1 (0.1%)	2 (0.8%)	66.7%	
	28-31 4 (0.4%)	5 (2.0%)	55.6%	
	32-36 48 (4.9%)	17 (7.0%)	26.2%	
	37-41 883 (90.8%)	210 (86.1%)	19.2%	
	>=42 20 (2.1%)	4 (1.6%)	16.7%	
	bd 17 (1.7%)	6 (2.5%)		
masa urodzeniowa w gramach **				1.776e-007
	<1499 4 (0.4%)	4 (1.6%)	50.0%	
	1500-1999 7 (0.7%)	10 (4.1%)	58.8%	
	2000-2490 23 (2.4%)	14 (5.7%)	37.8%	
	2500-2990 134 (13.8%)	44 (18.0%)	24.7%	
	3000-3490 345 (35.5%)	81 (33.2%)	19.0%	
	3500-3990 316 (32.5%)	68 (27.9%)	17.7%	
	4000-4490 123 (12.6%)	15 (6.1%)	10.9%	
	>=4500 13 (1.3%)	4 (1.6%)	23.5%	
	bd 8 (0.8%)	4 (1.6%)		
wiek matki w latach ⊗				0.9811 (NS)
	<=19 3 (0.3%)	3 (1.2%)	50.0%	
	20-24 343 (35.3%)	90 (36.9%)	20.8%	
	25-29 120 (12.3%)	27 (11.1%)	18.4%	
	30-34 346 (35.6%)	74 (30.3%)	17.6%	
	35-39 133 (13.7%)	38 (15.6%)	22.2%	
	>=40 28 (2.9%)	8 (3.3%)	22.2%	
	bd 0 (0.0%)	4 (1.6%)		
wiek ojca w latach #				0.3654 (NS)
	<=19 1 (0.1%)	1 (0.4%)	50.0%	
	20-24 49 (5.0%)	15 (6.1%)	23.4%	
	25-29 242 (24.9%)	55 (22.5%)	18.5%	
	30-34 391 (40.2%)	81 (33.2%)	17.2%	
	35-39 167 (17.2%)	51 (20.9%)	23.4%	
	40-44 74 (7.6%)	19 (7.8%)	20.4%	
	45-49 21 (2.2%)	8 (3.3%)	27.6%	
	>=50 7 (0.7%)	2 (0.8%)	22.2%	
	bd 21 (2.2%)	12 (4.9%)		
wykształcenie matki **				9.302e-007
	=< podstawowe 63 (6.5%)	36 (14.8%)	36.4%	
	zawodowe 237 (24.4%)	66 (27.0%)	21.8%	
	średnie 347 (35.7%)	72 (29.5%)	17.2%	
	wyższe 325 (33.4%)	52 (21.3%)	13.8%	
	bd 1 (0.1%)	18 (7.4%)		

potencjalne czynniki ryzyka	kontrola dopasowana, N=973		CL±P, N=244	test $\chi^2$ dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
<b>wykształcenie ojca #</b>				<b>7.943e-006</b>
=< podstawowe	33 (3.4%)	24 (9.8%)	42.1%	
zawodowe	356 (36.6%)	90 (36.9%)	20.2%	
średnie	278 (28.6%)	68 (27.9%)	19.7%	
wyższe	271 (27.9%)	37 (15.2%)	12.0%	
bd	35 (3.6%)	25 (10.2%)		
<b>kolejność ciąży ***</b>				<b>1.544e-005</b>
2 ciąży	644 (66.2%)	129 (52.9%)	16.7%	
3 ciąży	222 (22.8%)	67 (27.5%)	23.2%	
>=4 ciąży	107 (11.0%)	48 (19.7%)	31.0%	
<b>ilość wcześniejszych poronień samoistnych ****</b>				<b>0.5776 (NS)</b>
<b>0 poronień</b>	<b>731 (75.1%)</b>	<b>187 (76.6%)</b>	<b>20.4%</b>	
<b>1 poronienie</b>	<b>190 (19.5%)</b>	<b>45 (18.4%)</b>	<b>19.1%</b>	
<b>2 poronienia</b>	<b>44 (4.5%)</b>	<b>11 (4.5%)</b>	<b>20.0%</b>	
<b>3 poronienia</b>	<b>8 (0.8%)</b>	<b>1 (0.4%)</b>	<b>11.1%</b>	

\*\*\*\* zmienna wykorzystana w czterech modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (prostym, minimalnie dostosowanym, dużym i pełnym)

\*\*\* zmienna wykorzystana w trzech modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (minimalnie dostosowanym, dużym i pełnym)

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla liczby poronień samoistnych (pełnym)

⊗ zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla liczby poronień samoistnych (minimalnie dostosowanym)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

# nie uwzględnione w modelach

**Tabela A 5 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciążą drugiej lub dalszej - zmienne ilościowe ciągłe**

potencjalne czynniki ryzyka	kontrola populacyjna, N=973		CL±P, N=244		test Fishera-Snedecora	korekta Cochrana-Coxa testu t_Studenta
	średnia (odchlenie st.)	test Lillieforsa wartość p	średnia (odchlenie st.)	test Lillieforsa wartość p	wartość p	wartość p
wiek płodowy w tygodniach	39.0 (1.7)	<0.01	38.6 (2.5)	<0.01	0,000000	<b>0.014215</b>
masa urodzeniowa w gramach	3439.8 (540.6)	<0.01	3225.4 (647.2)	<0.01	0,000246	<b>0.000003</b>
wiek matki w latach	30.0 (4.7)	<0.01	30.0 ( 5.1)	<0.01	0,056809	0.875784 * (NS)
wiek ojca w latach	32.4 (5.5)	<0.01	32.6 (6.1)	<0.01	0,043596	0.524942 (NS)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

\*wartość p testu test t-Studenta

Tabela A 6 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną populacyjną” matek rodzących dziecko z ciąży drugiej lub dalszej - zmienne dichotomiczne

potencjalne czynniki ryzyka	kontrola dopasowana, N=973		CL±P, N=244	test mid-p
	liczność (% z N)		liczność (% z N)	wartość p
pleć **				0.046098
	męska	510 (52.4%)	143 (58.6%)	
	żeńską	463 (47.6%)	97 (39.8%)	
	bd	0 (0.0%)	4 (1.6%)	
miejsce zamieszkania matki *				0.919556(NS)
	wieś	421 (43.3%)	106 (43.4%)	
	miasto	552 (56.7%)	137 (56.1%)	
	bd	0 (0.0%)	1 (0.4%)	
palenie tytoniu przez matkę **				8.61E-008
	nie	804 (82.6%)	178 (73.0%)	
	tak	32 (3.3%)	31 (12.7%)	
	bd	137 (14.1%)	35 (14.3%)	
infekcje ukł. oddech. matki podczas 1 <sup>szego</sup> trymestru **				0.000007
	nie	866 (89.0%)	180 (73.8%)	
	tak	13 (1.3%)	16 (6.6%)	
	bd	94 (9.7%)	48 (19.7%)	
inne infekcje matki podczas ciąży **				0.012129
	nie	945 (97.1%)	226 (92.6%)	
	tak	7 (0.7%)	7 (2.9%)	
	bd	21 (2.2%)	11 (4.5%)	
infekcje ukł. mocz. matki podczas 1 <sup>szego</sup> trymestru *				0.535907(NS)
	nie	812 (83.5%)	206 (84.4%)	
	tak	18 (1.8%)	3 (1.2%)	
	bd	143 (14.7%)	35 (14.3%)	
narażenie matki na promieniowanie podczas ciąży *				0.397314(NS)
	nie	796 (81.8%)	207 (84.8%)	
	tak	15 (1.5%)	2 (0.8%)	
	bd	162 (16.6%)	35 (14.3%)	
wcześniejsze poronienia sztuczne *				0.776184(NS)
	nie	965 (99.2%)	243 (99.6%)	
	tak	3 (0.3%)	1 (0.4%)	
	bd	5 (0.5%)	(0.0%)	
wcześniejsze porody martwe *				0.632692(NS)
	nie	897 (92.2%)	236 (96.7%)	
	tak	16 (1.6%)	3 (1.2%)	
	bd	60 (6.2%)	5 (2.0%)	

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla liczby poronień samoistnych (pełnym)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

Tabela A 7 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną” matek rodzących dziecko z ciąży drugiej lub dalszej - zmienne ilościowe i jakościowe o uporządkowanych kategoriach

potencjalne czynniki ryzyka	kontrola dopasowana, N=236		CL±P, N=236	test $\chi^2$ dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
wiek płodowy w tygodniach #				0.02753
	<28 0 (0.0%)	2 (0.8%)	100.0%	
	28-31 1 (0.4%)	5 (2.1%)	83.3%	
	32-36 12 (5.1%)	16 (6.8%)	57.1%	
	37-41 215 (91.1%)	203 (86.0%)	48.6%	
	>=42 5 (2.1%)	4 (1.7%)	44.4%	
	bd 3 (1.3%)	6 (2.5%)		
masa urodzeniowa w gramach **				3.738e-005
	<1499 0 (0.0%)	4 (1.7%)	100.0%	
	1500-1999 1 (0.4%)	9 (3.8%)	90.0%	
	2000-2490 6 (2.5%)	13 (5.5%)	68.4%	
	2500-2990 31 (13.1%)	44 (18.6%)	58.7%	
	3000-3490 82 (34.7%)	77 (32.6%)	48.4%	
	3500-3990 82 (34.7%)	66 (28.0%)	44.6%	
	4000-4490 30 (12.7%)	15 (6.4%)	33.3%	
	>=4500 2 (0.8%)	4 (1.7%)	66.7%	
	bd 2 (0.8%)	4 (1.7%)		
wiek matki w latach \$				
	<=19	-	-	
	20-24	-	-	
	25-29	-	-	
	30-34	-	-	
	35-39	-	-	
	>=40	-	-	
	bd	-	-	
wiek ojca w latach #				0.683 (NS)
	<=19 0 (0.0%)	1 (0.4%)	100.0%	
	20-24 13 (5.5%)	13 (5.5%)	50.0%	
	25-29 58 (24.6%)	55 (23.3%)	48.7%	
	30-34 81 (34.3%)	81 (34.3%)	50.0%	
	35-39 50 (21.2%)	50 (21.2%)	50.0%	
	40-44 24 (10.2%)	19 (8.1%)	44.2%	
	45-49 3 (1.3%)	7 (3.0%)	70.0%	
	>=50 0 (0.0%)	2 (0.8%)	100.0%	
	bd 7 (3.0%)	8 (3.4%)		
wykształcenie matki **				0.0005965
	=< podstawowe 12 (5.1%)	34 (14.4%)	73.9%	
	zawodowe 62 (26.3%)	61 (25.8%)	49.6%	
	średnie 86 (36.4%)	72 (30.5%)	45.6%	
	wyższe 76 (32.2%)	52 (22.0%)	40.6%	
	bd 0 (0.0%)	17 (7.2%)		

potencjalne czynniki ryzyka	kontrola dopasowana, N=236		CL±P, N=236	test $\chi^2$ dla trendu
	liczność (% z N)	liczność (% z N)	%dzieci z wadą wg poziomów czynnika	wartość p
<b>wykształcenie ojca #</b>				<b>0.0002507</b>
=< podstawowe	8 (3.4%)	23 (9.7%)	74.2%	
zawodowe	82 (34.7%)	87 (36.9%)	51.5%	
średnie	69 (29.2%)	67 (28.4%)	49.3%	
wyższe	68(28.8%)	37 (15.7%)	35.2%	
bd	9 (3.8%)	22 (9.3%)		
<b>kolejność ciąży ***</b>				<b>0.0003658</b>
2 ciąży	159 (67.4%)	123 (52.1%)	43.6%	
3 ciąży	51 (21.6%)	65 (27.5%)	56.0%	
>=4 ciąży	26 (11.0%)	48 (20.3%)	64.9%	
<b>ilość wcześniejszych poronień samoistnych ****</b>				<b>0.587 (NS)</b>
<b>0 poronień</b>	<b>177 (75.0%)</b>	<b>181 (76.7%)</b>	<b>50.6%</b>	
<b>1 poronienie</b>	<b>46 (19.5%)</b>	<b>43 (18.2%)</b>	<b>48.3%</b>	
<b>2 poronienia</b>	<b>10 (4.2%)</b>	<b>11 (4.7%)</b>	<b>52.4%</b>	
<b>3 poronienia</b>	<b>3 (1.3%)</b>	<b>1 (0.4%)</b>	<b>25.0%</b>	

\*\*\*\* zmienna wykorzystana w czterech modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (prostym, minimalnie dostosowanym, dużym i pełnym)

\*\*\* zmienna wykorzystana w trzech modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (minimalnie dostosowanym, dużym i pełnym)

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej zbudowanych dla liczby poronień samoistnych (dużym i pełnym)

\* zmienna wykorzystana w jednym modelu regresji logistycznej zbudowanym dla liczby poronień samoistnych (pełnym)

§ zmienna tworząca dopasowanie grupy badanej i kontrolnej

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

# nie uwzględnione w modelach

**Tabela A 8 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną” matek rodzących dziecko z ciążą drugą lub dalszą - zmienne ilościowe ciągle**

potencjalne czynniki ryzyka	kontrola populacyjna, N=236		CL±P, N=236		test Fishera-Snedecora	korekta Cochrana-Coxa testu t_Studenta
	średnia (odchlenie st.)	test Lillieforsa wartość p	średnia (odchlenie st.)	test Lillieforsa wartość p	wartość p	wartość p
wiek płodowy w tygodniach	39.0 (1.6)	<0.01	38.6 (2.5)	<0.01	1,774x10-11	<b>0.04959</b>
masa urodzeniowa w gramach	3456.0 (490.4)	0.1827 (NS)	3228.9 (649.9)	<0.01	0.00002	<b>0.00003</b>
wiek matki w latach	30.1 (4.9)	<0.01	30.1 (4.9)	<0.01	x	x
wiek ojca w latach	32.4 (5.4)	<0.01	32.6 (6.0)	<0.01	0.15649 (NS)	0.69127 * (NS)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

\*wartość p testu test t-Studenta

Tabela A 9 Czynniki ryzyka izolowanego CL±P w oparciu o grupę „kontrolną dopasowaną” matek rodzących dziecko z ciąży drugiej lub dalszej - zmienne dichotomiczne

potencjalne czynniki ryzyka	kontrola dopasowana, N=236		CL±P, N=236	test mid-p
		liczność (% z N)	liczność (% z N)	wartość p
pleć *				0.255244
	męska	127 (53,8%)	137 (58,1%)	
	żeńska	109 (46,2%)	95 (40,3%)	
	bd	0 (0,0%)	4 (1,7%)	
miejsce zamieszkania matki *				0.822041
	wieś	98 (41,5%)	100 (42,4%)	
	miasto	138 (58,5%)	135 (57,2%)	
	bd	0 (0,0%)	1 (0,4%)	
palenie tytoniu przez matkę **				0.000364
	nie	189 (80,1%)	174 (73,7%)	
	tak	8 (3,4%)	29 (12,3%)	
	bd	39 (16,5%)	33 (14,0%)	
infekcje ukł. oddech. matki podczas 1 <sup>szego</sup> trymestru *				0.014376
	nie	208 (88,1%)	174 (73,7%)	
	tak	6 (2,5%)	16 (6,8%)	
	bd	22 (9,3%)	46 (19,5%)	
inne infekcje matki podczas ciąży *				0.102193
	nie	226 (95,8%)	218 (92,4%)	
	tak	2 (0,8%)	7 (3,0%)	
	bd	8 (3,4%)	11 (4,7%)	
infekcje ukł. mocz. matki podczas 1 <sup>szego</sup> trymestru *				0.357305
	nie	200 (84,7%)	198 (83,9%)	
	tak	6 (2,5%)	3 (1,3%)	
	bd	30 (12,7%)	35 (14,8%)	
narażenie matki na promieniowanie podczas ciąży *				0.411835
	nie	188 (79,7%)	201 (85,2%)	
	tak	4 (1,7%)	2 (0,8%)	
	bd	44 (18,6%)	33 (14,0%)	
wcześniejsze poronienia sztuczne *				0.617013
	nie	231 (97,9%)	235 (99,6%)	
	tak	2 (0,8%)	1 (0,4%)	
	bd	3 (1,3%)	0 (0,0%)	
wcześniejsze porody martwe *				0.477775
	nie	219 (92,8%)	228 (96,6%)	
	tak	5 (2,1%)	3 (1,3%)	
	bd	12 (5,1%)	5 (2,1%)	

\*\* zmienna wykorzystana w dwóch modelach regresji logistycznej (dużym i pełnym)

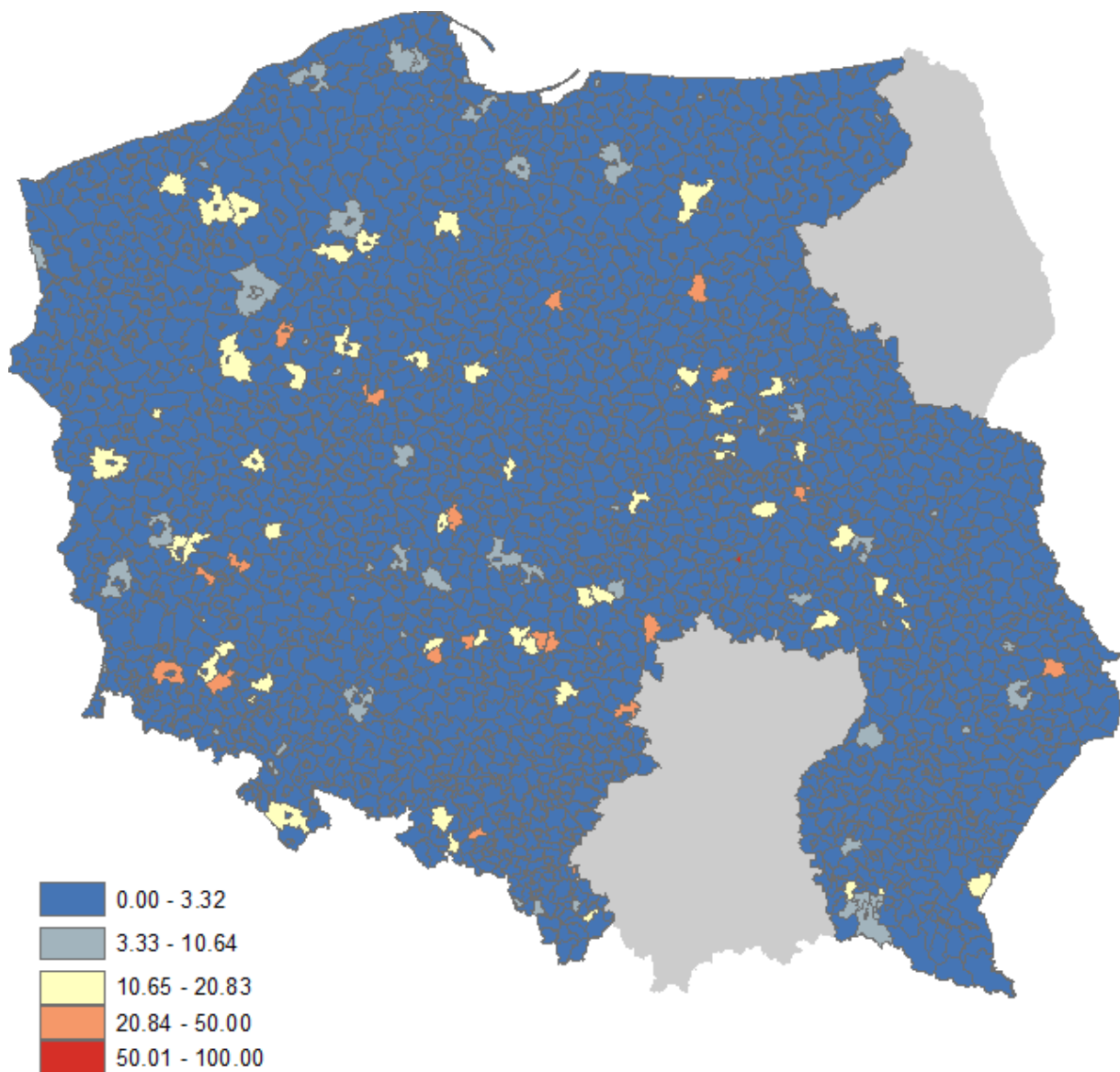
\* zmienna wykorzystana w jednym modelu regresji logistycznej (pełnym)

NS: brak istotności statystycznej - dwustronna wartość-p >=0.05

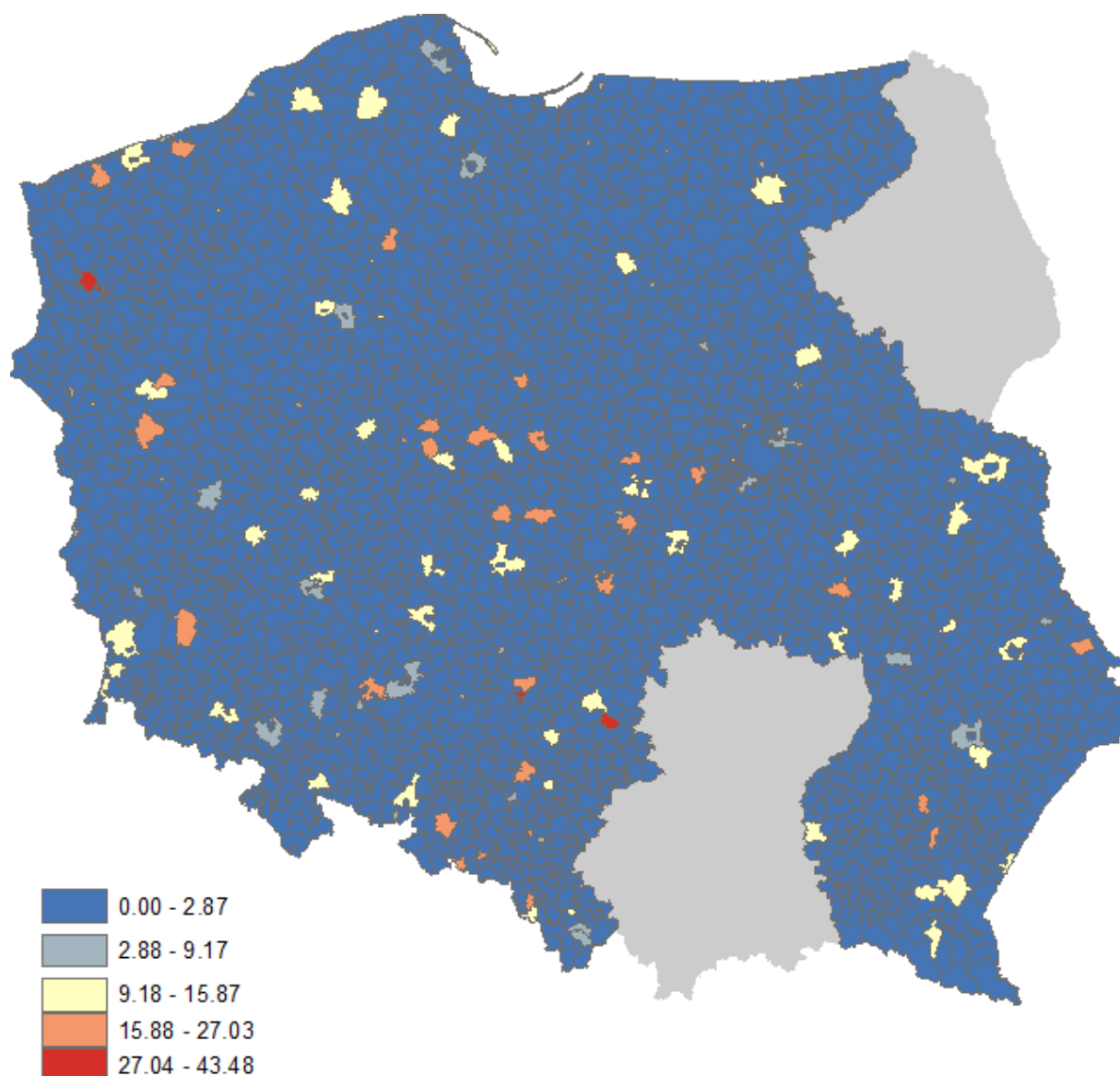
---

## 15 ZAŁĄCZNIK 5

---

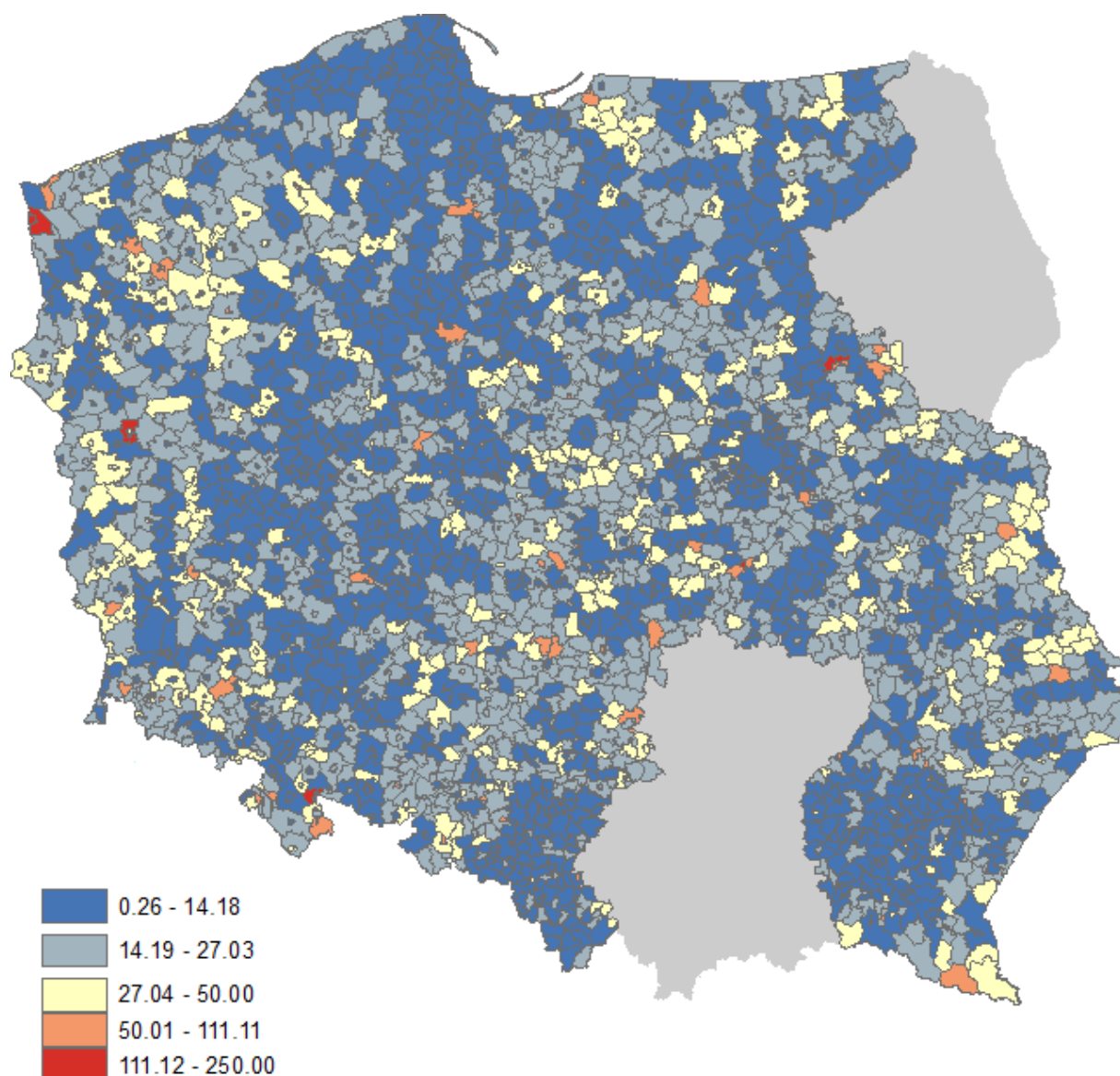


Rycina A 1 Surowe współczynniki występowania izolowanego CL±P w roku 2005 w badanych województwach, przedstawione według naturalnego podziału Jenksa

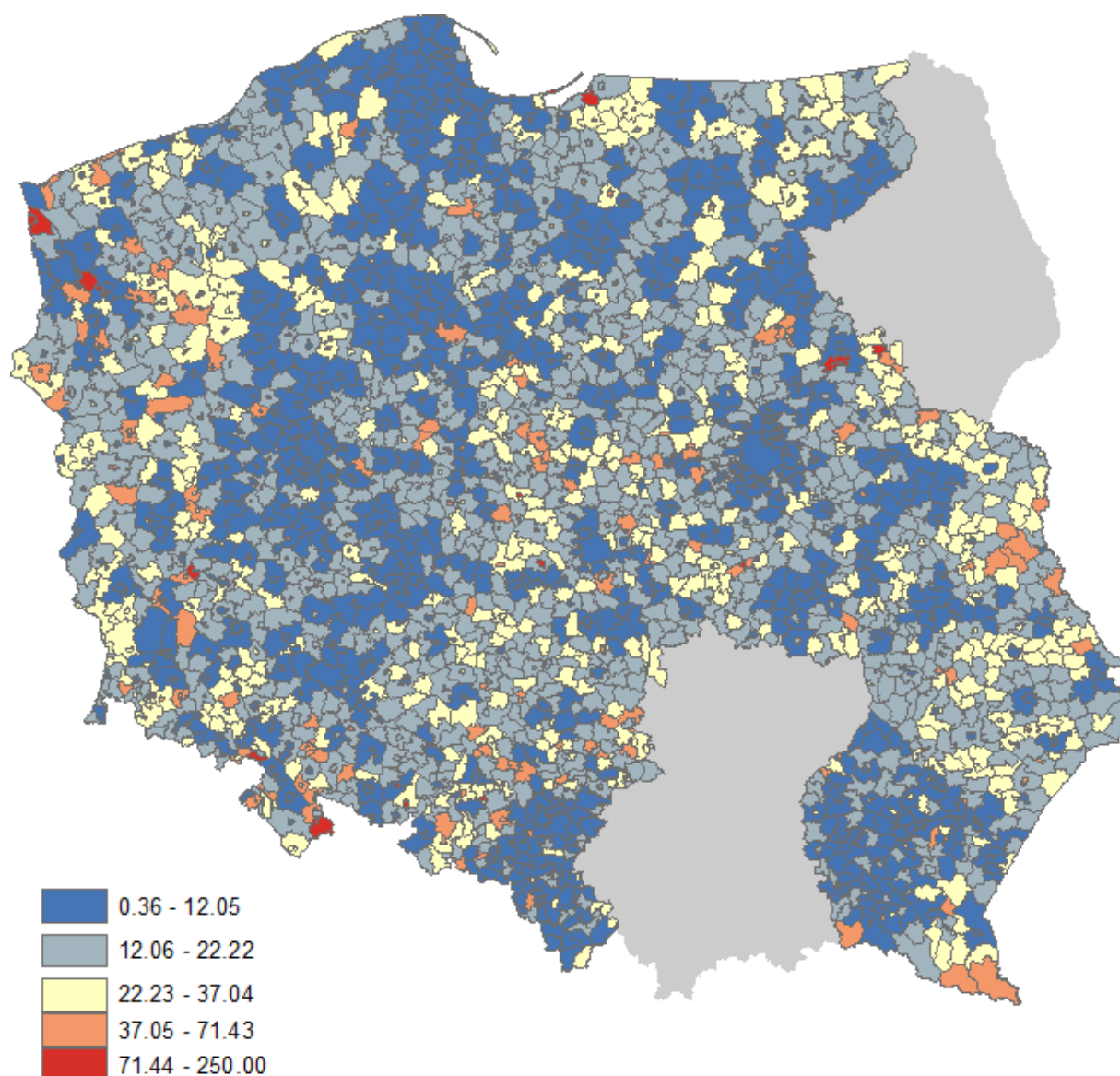


Rycina A 2 Surowe współczynniki występowania izolowanego CL±P w roku 2006 w badanych województwach, przedstawione według naturalnego podziału Jenksa





Rycina A 3 Surowe współczynniki występowania izolowanego CL±P w roku 2005 w badanych województwach – zmodyfikowane zgodnie z metodą Cressie, przedstawione według naturalnego podziału Jenksa

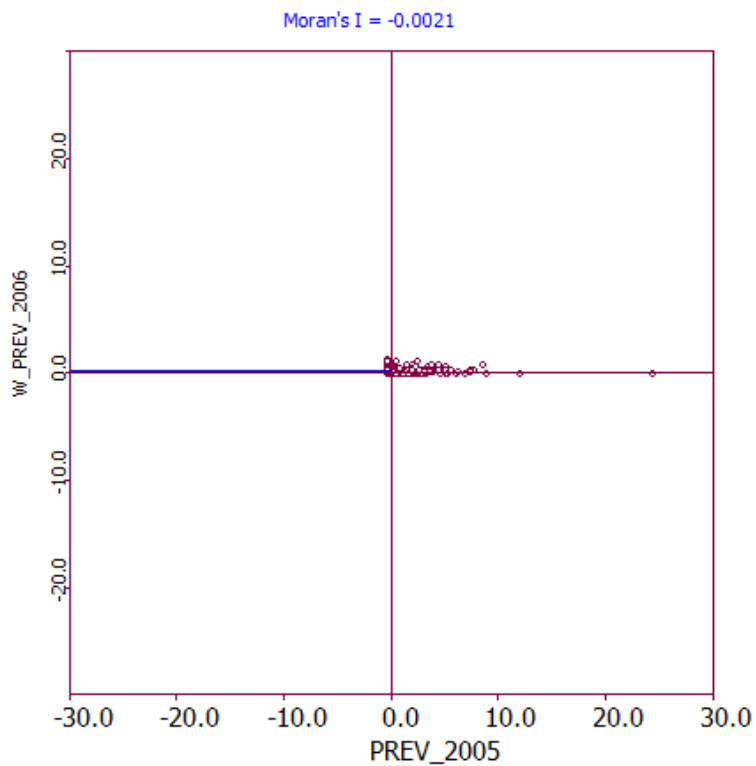


Rycina A 4 Surowe współczynniki występowania izolowanego CL±P w roku 2006 w badanych województwach – zmodyfikowane zgodnie z metodą Cressie, przedstawione według naturalnego podziału Jenksa

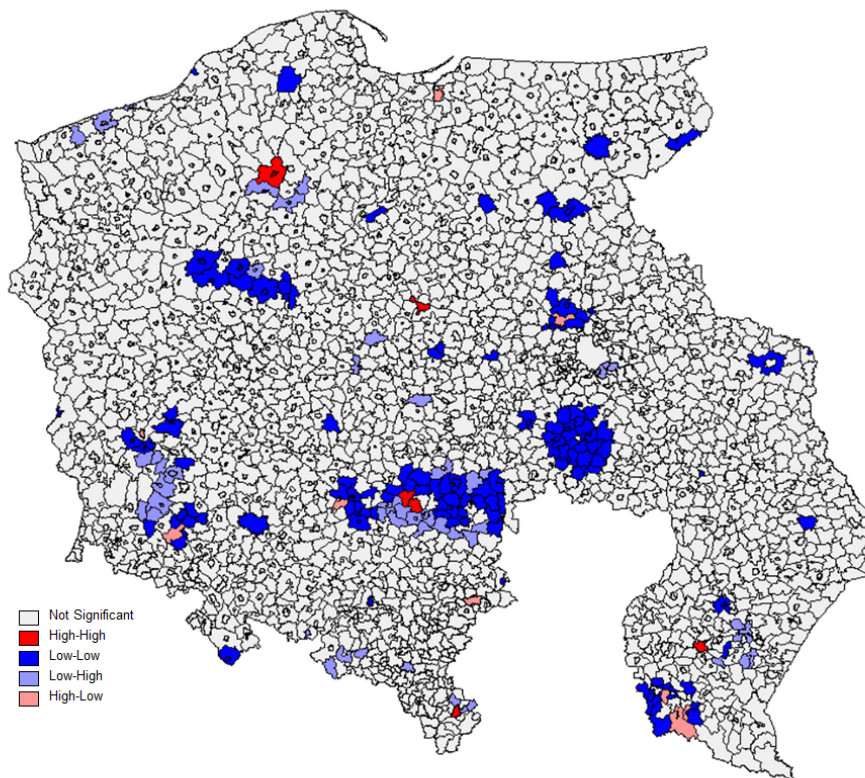
---

# 16 ZAŁĄCZNIK 6

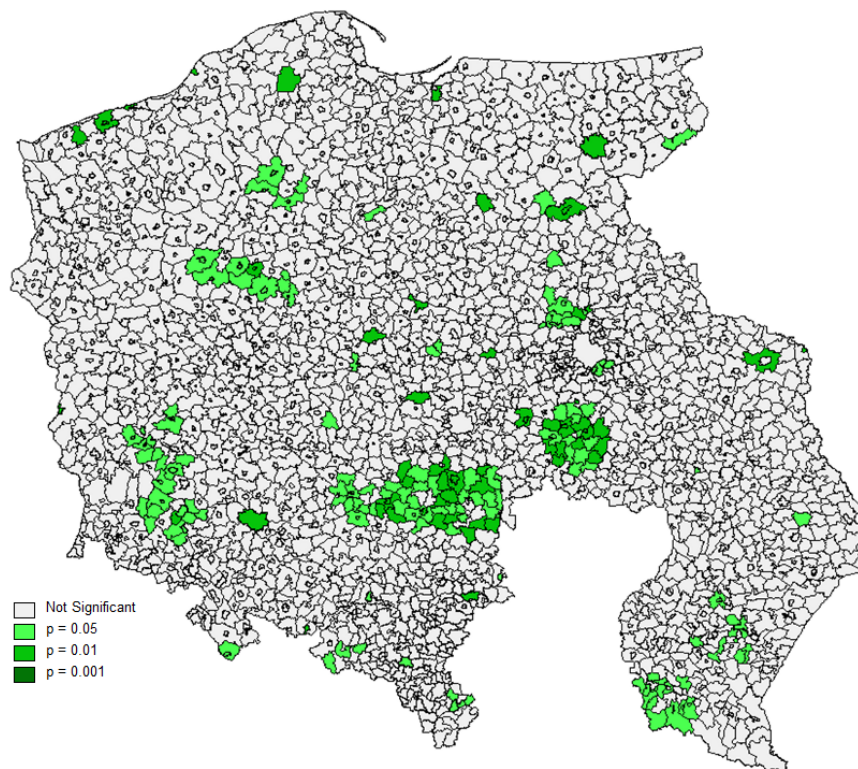
---



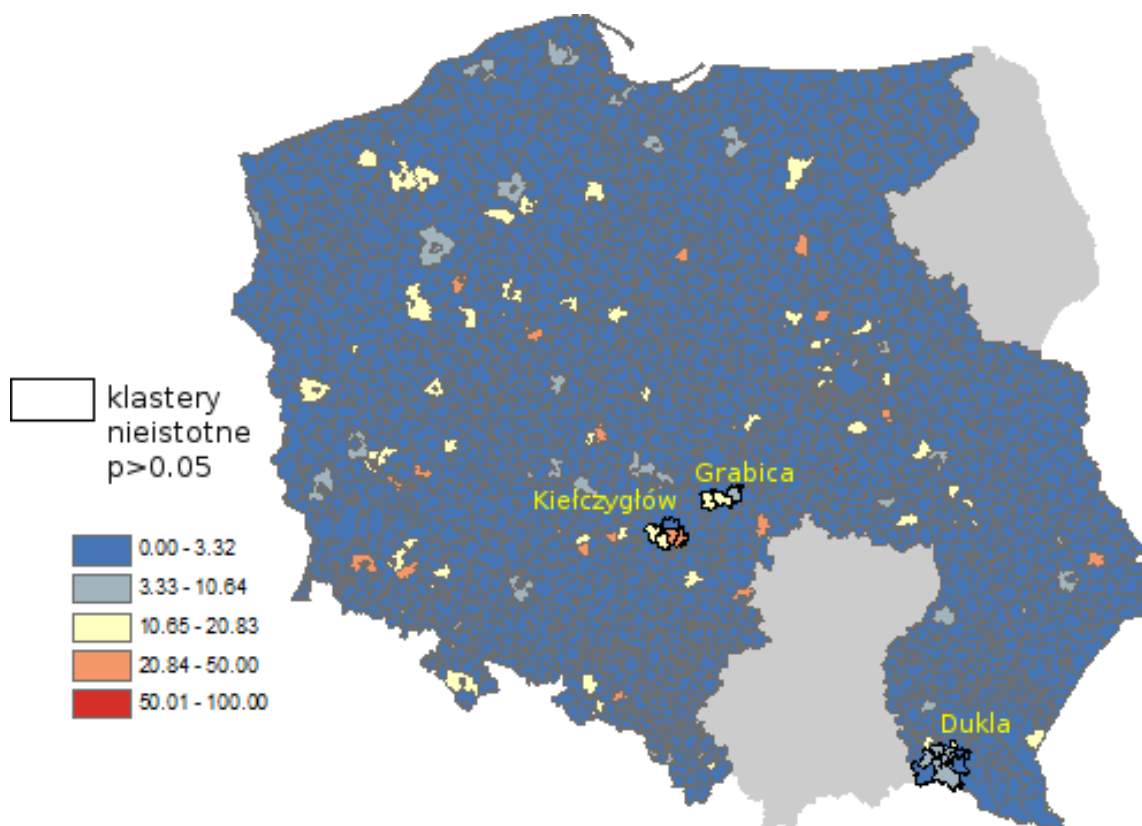
**Rycina A 5 Wykres punktowy statystyki I Morana surowych współczynników występowania izolowanego CL±P w badanych województwach, rok 2005 kontra 2006. P=0.999**



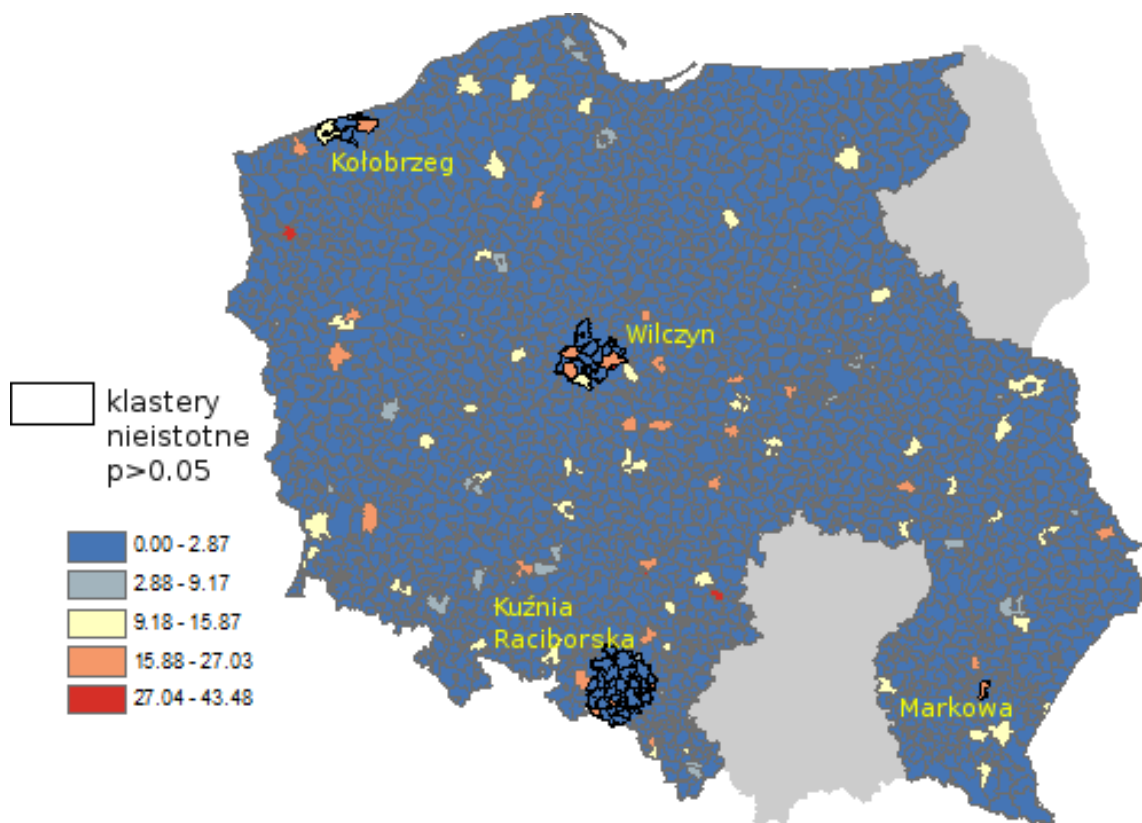
**Rycina A 6** Przynależność gmin do ćwiartek wykresu punktowego Morana na poziomie istotności 0.05. Rok 2005 kontra 2006



**Rycina A 7** Poziom istotności dla mapy przynależności gmin do ćwiartek wykresu punktowego Morana. Rok 2005 kontra 2006



Rycina A 8 Najbardziej prawdopodobne, ale nie istotne statystycznie klaster wyznaczone poprzez statystykę skanującą Kulldorff'a oraz surowe współczynniki występowania izolowanego CL±P w latach 2005 w badanych województwach, przedstawione według naturalnego podziału Jenksa



Rycina A 9 Najbardziej prawdopodobne, ale nie istotne statystycznie klaster wyznaczone poprzez statystykę skanującą Kulldorff'a oraz surowe współczynniki występowania izolowanego CL±P w latach 2006 w badanych województwach, przedstawione według naturalnego

Tabela A 10 Wyniki analizy skanującej Kulldorff'a. Rok 2005 kontra 2006

	Najbardziej prawdopodobny klaster	Kolejne prawdopodobne klaster					
		2005	2005	2006	2006	2005	2006
<b>Gmina środka okręgu</b>	Kiełczygłów	Dukla	Wilczyn	Kuźnia Raciborska	Grabica	Markowa	Kołobrzeg
<b>Województwo środka okręgu</b>	Łódzkie	Podkarpackie	Wielkopolskie	Śląskie	Łódzkie	Podkarpackie	Zachodniopomorskie
<b>Środek okręgu wsp. X</b>	775560	985805	715687	742154	816751	1023563	533038
<b>Środek okręgu wsp. Y</b>	5684263	5504122	5817148	5565720	5711889	5563358	5998821
<b>Długość promienia (w metrach)</b>	11420.41	21677.77	23546	26571.88	10856.54	0	14907.02
<b>Liczność - populacja</b>	256	945	1008	7686	251	72	654
<b>Obserwowana liczba przypadków</b>	4	6	6	16	3	2	4
<b>Oczekiwana liczba przypadków</b>	0.18	0.66	0.71	5.56	0.18	0.05	0.45
<b>Współczynnik częstości na 1000</b>	15.625	6.3492063	5.952381	2.081707	11.952191	27.777778	6.116208
<b>Relatywne ryzyko</b>	22.239	9.194	8.589	2.946	16.39	37.317	9.03
<b>Statystyka LR</b>	8.570725	7.931272	7.569387	6.59182	5.564048	5.28818	5.230889
<b>Wartość p</b>	0.230 (NS)	0.387 (NS)	0.479 (NS)	0.791 (NS)	0.986 (NS)	0.995 (NS)	0.996 (NS)

NS: brak istotności statystycznej -- wartość p >= 0.05