Article

Understanding Regional Risk Factors for Cancer: A Cluster Analysis of Lifestyle, Environment and Socio-Economic Status in Poland

Dawid Majcherek [1,*] , **Marzenna Anna Weresa [2] and Christina Ciecierski [3]**

[1]   Collegium of World Economy, SGH Warsaw School of Economics, al. Niepodległości 162,
     02-554 Warsaw, Poland
[2]   World Economy Research Institute, Collegium of World Economy, SGH Warsaw School of Economics,
     al. Niepodległości 162, 02-554 Warsaw, Poland; mweres@sgh.waw.pl
[3]   Department of Economics, Northeastern Illinois University, 5500 North St. Louis Avenue,
     Chicago, IL 60625, USA; c-ciecierski@neiu.edu
*   Correspondence: dawid.majcherekk@gmail.com

check for updates

Abstract: To date, no results have been published regarding cluster analysis of risk factors for cancer in Poland. Many cancer deaths are preventable through the modification of cancer risk behaviours. This study explores the multidisciplinary connection between lifestyle, environment and socio-economic status (SES). Cluster analyses indicate that major metropolitan areas and large industrial regions differ significantly in terms of SES, lifestyle and environment when compared with other parts of Poland. Our findings show that in order for interventions to be effective, cancer-prevention policy should be addressed on both local and national scales. While anti-cancer policies in Poland's industrial regions should focus on air pollution, the country's northern regions should aim to curb smoking, increase sports activity and improve SES. Policy interventions must target the root causes of cancer in each region of Poland and must account for SES.

1. Introduction

Cancer is the second leading cause of death responsible for taking the lives of one in six or a total of 9.6 million people worldwide in 2018 [1]. The World Health Organization (WHO, Geneva, Switzerland) estimates that nearly a quarter of the global cancer burden occurs in Europe, where the probability of dying from cancer or any one of three leading non-communicable diseases (NCDs; include cardiovascular disease, diabetes and chronic respiratory disease) among adults between the ages of 30 and 70 ranges from 9.1% in Sweden to 23.6% in Bulgaria [2,3]. By comparison, the probability of dying from NCDs in the United States (US) is around 14.3% with 1.7 million newly diagnosed cases in 2018 [4]. In an effort to fast-track the overall aims of cancer research, including improved prevention, detection and new therapy development, US Congress allocated 500 million USD to the Cancer Moonshot Research Initiatives 2017–2025 [5]. Globally, the provision of policy options stems from the WHO Global Action Plan for the Prevention and Control of NCDs 2013–2020 which seeks to reduce cancer mortality rates by 20% by 2030 [6]. Efforts to fight cancer have also been undertaken at the European level. In 2009, the European Commission (EC) launched a joint action on "European Partnership for Action Against Cancer" (EPAAC), focused on facilitating cooperation and sharing experiences among the EU member states in the area of cancer control policies [7]. Intended to help policy-makers in the EU member states design national cancer control programmes [8], the key

output of this partnership was the release of the European Guide for Quality National Cancer Control Programmes. The EU has also introduced mission-oriented research [9,10] through its Horizon Europe Framework Programme which features cancer as one of five key missions for European research policy in 2021-2027. To facilitate effective implementation of these strategies, empirical evidence is needed to better understand the risk factors for cancer.

Poland has relatively high and growing cancer mortality rates. According to Poland's National Cancer Registry (KRN), the annual cancer morbidity growth rate averaged 2.4% between 2010 and 2016 [11], and the number of new cancer cases in 2017 reached 164,875. The number of new cancer cases is projected to increase among both women and men to an annual total incidence of approximately 176,000 cases by 2025 [12]. In 2017, cancer caused 99,644 or 260 cancer deaths per 100,000 people in Poland [11].

The current and forecasted rise in cancer incidence along with its negative social and economic consequences was addressed by the Parliament of the Republic of Poland on April 26, 2019 when the Project of the National Oncological Strategy Act was adopted. By November of that same year, its goal to improve epidemiological indicators over the years 2020–2030 was presented [13] and included:

1. Invest in human resources by increasing medical staff and improving the quality of oncological training and education;
2. Reduce cancer incidence through investments in education and primary prevention related to cancer disease (i.e., promotion of healthy lifestyles to curb cancer-risk behaviours);
3. Improve the effectiveness of secondary prevention through larger investments in screening;
4. Invest in bench sciences and healthcare innovation in order to introduce effective diagnostic and therapeutic solutions;
5. Fund improvements within the oncological care system in order to improve organization and processes whilst also providing patients with access to the highest quality of comprehensive, diagnostic and therapeutic care along the "patient path".

The National Oncological Strategy 2020–2030 was adopted by the Parliament in February 2020. As such, evidence-based strategies for cancer-risk factors in Poland are desperately needed to improve policy interventions and curb rising rates of cancer mortality. To date, only a limited number of studies have been published to help fill this void. Polak et al. [14] showed that socio-economic status (SES) is significantly correlated with lung cancer mortality in Poland. These findings, however, are constrained by the study's sole focus on lung cancer outcomes. Another study in Poland examined breast cancer and found a negative correlation between income, poverty and breast cancer incidence [15]. What is lacking in the literature are studies that take into account a variety of cancer types, coupled with geospatial considerations, cancer-risk behaviours and social and economic demographic conditions.

The WHO's International Agency for Research on Cancer (IARC) classifies cancer-risk factors into two main categories: internal and external. The former depends on genetic profiling and unhealthy behaviours (e.g., tobacco and alcohol use, physical inactivity and unhealthy diet). The latter is shaped by levels of physical, chemical and other biologically carcinogenic agents borne in air, water, soil, etc. [1]. Studies show that cancer-risk factors are unevenly distributed across regions and vary across gender, age and a variety of socio-economic factors [16–18].

Some cancer-risk behaviours, such as tobacco and/or alcohol use, have already been studied in Poland and across Central and Eastern Europe [14,19–21]. Thus far, no empirical evidence combining both internal and external cancer-risk factors across regions in Poland has been published. This study aims to fill this gap by investigating regional differences in SES (income and education levels) in Poland combined with lifestyle (tobacco and alcohol use, diet, sports activity) and environment (air pollution). This paper investigates the clustering of health-risk factors from data collected through the 2015 wave of the Social Diagnosis study (objective and subjective quality of life survey) and local pollution data obtained from Poland's Chief Inspectorate for Environmental Protection.

2. Cancer Risk Factors in Poland

2.1. Lifestyle

Cancer incidence is highly correlated with a number of lifestyle factors [22], which include tobacco use, diet, physical inactivity, obesity and alcohol consumption [23–25]. Tobacco use increases the risk of fourteen types of cancer [26–28] including lung cancer, which is the leading cause of cancer deaths in Poland [29]. One-third of annual cancer mortalities in the US are caused by tobacco exposure and use, while another one-third of cancer deaths are associated with diet, physical inactivity and obesity [30]. Obesity increases the risk of thirteen types of cancer [30–37], while physical activity has been proven to reduce the risk of four of these cancer types [30,31,34,38], and regular physical activity combined with a healthy caloric intake helps to maintain a healthy body weight, thus reducing the overall risk of cancer [30]. Excessive regular alcohol consumption (more than two drinks for men and one drink for women) is an established cause of many cancers, including mouth, pharynx, larynx, esophagus and liver [39–41]. In the US, multiple studies have shown that cancer-risk behaviours often co-occur [23,42–44], and only a very small percentage of adults (<5%) achieve a healthy lifestyle that includes no tobacco use, low alcohol consumption, regular physical activity as well as a healthy diet and weight [43,45].

Some studies have found evidence of health-risk lifestyles clustering within geographic areas and across gender, age and socio-economic groups. For example, high rates of the co-occurrence of smoking and alcohol have been identified among both females [24] and males [46]. A study from Spain examined the relationship between smoking, alcohol consumption, physical activity, and the diet of college students and found that clusters of smoking, diet, and physical inactivity occur together [47]. Other studies show a high prevalence of unhealthy diet and inactivity among young adults in college [40,41].

2.2. Air Quality

The impacts of environmental pollution on health and cancer incidence are studied continuously, and results vary depending on geographical region, pollutant type and environmental medium (e.g., air, soil, water). Because research conducted by the European Environment Agency (EEA) finds that "air pollution is a major cause of premature death and disease and is the single largest environmental health risk" [48], this paper focuses on air pollution in Poland. Key air pollutants with standards set by WHO air quality guidelines and EU Ambient Air Quality Directives include particulate matter (PM), carbon monoxide (CO), hydrocarbons, such as benzo[a]pyrene (BaP), sulphur oxides (SO_2), nitrogen oxides (NO_2), ozone (O_3), benzene (C_6H_6), certain metals (Pb, As, Cd, Ni) and polycyclic aromatic hydrocarbons [48].

Numerous epidemiological studies conducted over the last few decades have provided evidence that risk of cancer is related to air pollution [14,49–51]. Particulate matter (PM_{10}, $PM_{2.5}$) and benzo[a]pyrene (BaP) are regarded as pollutants of particularly high risk [48,52]. In general, PM consists of complex mixtures of solid and liquid particles in the air. Particles with a diameter of 10 microns or less (PM_{10}) can penetrate and stay inside the lungs. Even more damaging to health are particles with a diameter of 2.5 microns or less ($PM_{2.5}$), which can penetrate lung barriers and enter the circulatory system. Chronic exposure to these types of particle matter increases the risk of developing cardiovascular and respiratory diseases, as well as lung cancer. Limits aim to achieve the lowest concentrations of PM possible [53]. Therefore, these air pollutants were included in our analysis of cancer risk factors.

2.3. Socio-Economic Factors

Many studies provide evidence of socio-economic differences (i.e., income, household wealth gains and losses, education) on individual health [54–58] as well as on the cost of illness [59]. Studies also show disparities in cancer risk and mortality related to socio-economic factors, such as income or educational level. Income plays a significant role in shaping social living conditions since higher income allows for better housing, increased schooling and more sophisticated forms of recreation [60].

Education may influence the risk of cancer in a variety of ways, including but not limited to the choice of profession, sports activities, smoking and alcohol consumption, and participation in health screening programmes [61].

There is a common agreement among scholars that the role of SES in shaping health is largely indirect. SES shapes individual consumer behaviour (i.e., smoking, diet, physical activity, obesity) and thus impacts health. SES also influences health through better or worse access to the health care system.

Significant disparities in cancer death rates due to differences in income were identified through a cross-sectional study of 3135 counties in the US [62] and were found to be influenced by mediators such as health-risk behaviours, quality and costs of medical care and food insecurity. Similar findings regarding reasons for the growing rates of cancer mortality disparities in US counties are provided by Withrow et al. [63], who conclude that smoking, obesity and high alcohol consumption are highly correlated with income and may contribute to differences in cancer mortality.

Another study uses survey data collected in France (N = 3359) to examine the relationship between perceptions of cancer risk factors and SES. Results from this analysis show that the perception of cancer risk factors depends on the SES of respondents. Here, higher perceptions of both behavioural and psychosocial factors emerge for respondents with higher SES, with respondents with intermediate levels of SES being more likely to stress environmental risk factors [17].

Polak et al. [14] confirm the correlation between socio-economic factors and death from lung cancer and other respiratory diseases in 66 regions in Poland for the period 2010-2014. The authors measured SES using an index derived from five variables: average monthly salary, percentage of the population with higher education, percentage of people employed in finance and real estate, unemployment rate, and percentage of the population receiving social support due to poverty. Using multiple weighted linear regression models, the authors estimated a relationship between SES and mortality resulting from respiratory diseases and lung cancer. The results confirm a correlation between current SES measured by the SES index and mortality from both respiratory diseases and lung cancer at the population level. However, changes in SES were found to be important only for mortality caused by respiratory diseases. An increase in SES was associated with a decrease in mortality from respiratory disease but did not matter for lung cancer mortality.

Tanaka et al. examine links between the risk of cancer-related mortality and economic disparities among 21,000 patients diagnosed with various types of cancer between 2010 and 2012 in the Aomori Prefect of Japan. To measure cancer incidence, the authors employ two indicators: (1) an age-standardized incidence rate and (2) the incidence rate ratio. Using a multivariable Cox regression approach, the authors investigate the relationship between patient survival time and income disparities. They also examine how specific factors, such as stage of cancer diagnosis and type of treatment, influence the rate of cancer death at a particular point in time (hazard rate). The authors find that patient income levels differentiate age-standardized cancer incidence rates for breast, cervical and prostate cancer but not for stomach, lung and liver cancer [64]. These results are in-line with findings from Korea, where a study used Korean National Health Insurance Cancer Registration data to investigate the relationship between cancer incidence and family income. The age-standardized incidence rates for major cancer types were calculated and adjusted for age, number of family members and residential area. Using a logistic model, the authors prove that there is a negative and significant relationship between income and cancer incidence risk. In general, cancer incidence risk is higher for low-income families when compared to higher-income counterparts. However, this study also shows that within the same income group, cancer-risk differs across types of cancer [65].

The results of previous studies of the relationship between socio-economic factors and cancer incidence are mixed. The impact of socio-economic factors varies by country or district and may be different for various types of cancer. However, the majority of studies confirm that socio-economic factors have an indirect impact on the incidence of certain cancers. Therefore, studies on cancer risk factors must be situated in a socio-economic context.

3. Research Questions, Data and Methods

3.1. Research Questions

Based on the literature review, the following research questions were developed:

- Are clusters distinguishable when internal factors (lifestyle and socio-economic status) are combined with external factors (environment) into a single spatial analysis?
- What are the implications of spatial cluster analysis of cancer-risk factors for health policy interventions in Poland?

3.2. Data

The Social Diagnosis (SD) study is an appreciably long-standing survey developed by an interdisciplinary team of professionals seeking to measure significant aspects of household life in Poland, including attitudes, mindsets and behaviour. The survey focuses on collecting economic information (e.g., income, material wealth, savings and finance) as well as behavioural, attitudinal, demographic and socio-economic details (e.g., including but not limited to educational attainment, medical care, approach to problem-solving, stress, psychological well-being, lifestyle, pathologies, engagement in cultural events and the arts, use of communication technologies) from participants aged 16 or older and living in households across all regions in Poland. The dataset was derived from a panel study where the sample of households was first drawn and surveyed in 2000. This same sample of households was re-surveyed in 2003, and then every two years consecutively.

In this study, we merged the 2015 SD sample (19,769 respondents) with 2014 local air pollution data received from the Office of the Chief Inspectorate for Environmental Protection, which provides hourly measurements of various air pollutants taken at approximately 2000 air-quality measuring stations from across Poland in the years 2000–2018 [66,67]. The data was aggregated and merged using the 66 Nomenclature of Territorial Units for Statistics (NUTS-3) region classification in Poland. [68]. The sample characteristics of regions are provided in Table 1.

Table 1. Descriptive statistics (*n* = 66 regions).

Variable	Mean	Std Dev	Minimum	Maximum
cigarettes_perday	3.49	0.83	1.79	5.83
Sport (% of responders)	0.35	0.07	0.19	0.53
Alcohol (% of responders)	0.05	0.02	0.01	0.11
BMI	26.30	0.41	25.11	27.09
household_income [PLN]	3953.48	557.71	2971.14	6223.23
years_of_edu	11.98	0.87	10.41	15.08
PM_{10}_avg [$\mu g/m^3$]	32.73	6.19	20.92	52.05
PM_{10}_no_of_days	53.69	22.21	17.41	128.00
PM_{25}_avg [$\mu g/m^3$]	23.59	5.142	13.33	35.76
BaP_avg [$\mu g/m^3$]	4.42	2.09	0.94	10.41

Note: This table presents descriptive statistics of 66 districts based on the Social Diagnosis dataset and local dataset on pollution from the Chief Inspectorate for Environmental Protection.

Four cancer-risk behaviours were identified using data from the SD on the regional level: average daily cigarette use, alcohol consumption, sports activity, and being overweight. Cigarettes per day is a continuous variable constructed from the self-reported average number of cigarettes smoked per day. Alcohol consumption is a dichotomous measure that was defined as true if the respondent reported "drinking too much" in the last year and false otherwise. The sports activity condition was set to be true if the respondent reported engaging in any sports activity including aerobics, running/jogging/Nordic walking, gym, cycling, skiing or other winter sport, swimming, football or other team sport, yoga, martial arts and another sport or type of physical activity. The measure of overweight/obesity is represented by the calculation of the body-mass index (BMI). It is a continuous measure that reflects

the combination of the respondent's self-reported height and weight measured in kilograms and centimetres, respectively.

Four continuous measures of air quality were merged with the SD survey data and include: level of benzoapyrene (BaP), airborne particle matter concentrations with a diameter of either 10 or less and 2.5 or less microns (PM_{10} and $PM_{2.5}$, respectively) and duration (measured in days) of elevated PM_{10} concentrations (PM_{10} days) above 50 $\mu g/m^3$. In general, the lower the ambient air measure, the better the air quality.

In this sample, average tobacco use is an average of 3.5 cigarettes per day. Nearly 5% of respondents declared that they consume too much alcohol. The average national BMI measures 26.3, and 36% of respondents confirmed completing at least one sports activity per week. The average number of years of completed education for the sample is around 12, indicating that most of the sample earned a high school diploma or its equivalent. Monthly net household income was around 4000 PLN (approx. 1000 USD, with an exchange rate of 1 USD = 4 PLN). In our sample, the annual particulate matter for larger particles, PM_{10}, averaged around 32 $\mu g/m^3$, which exceeds WHO guidelines (20 $\mu g/m^3$). In addition, although concentrated PM_{10} levels should not exceed 35 days per year, our air-quality measurement shows that the PM_{10} concentration exceeded WHO limit guidelines for over 53 days. Annual measures of fine particulate matter concentrations ($PM_{2.5}$) reached 23.60 $\mu g/m^3$, which also exceeds the recommended WHO annual limit of 10 $\mu g/m^3$ [69,70]. Moreover, the average annual levels of benzoapyrene measured 4.42 $\mu g/m^3$, which is very close to the recommended limit (5 $\mu g/m^3$) [69].

3.3. Methods

Analysis of air quality conditions, socio-economic factors and cancer-risk behaviours was completed using Ward's method [71] through which cluster centroids are performed based on minimum variance criterion, and the distance between two objects is calculated using Euclidean distance. In the beginning, each cluster is a single observation. During each step of Ward's method, the algorithm seeks pairs that lead the minimum variance to increase upon merging. Through this method, the sum of squares begins at zero but grows as the algorithm merges observations. Ward's method does not provide guidelines on the number of clusters to select, thereby allowing for some theoretical argumentation as to the appropriate number of clusters. According to earlier studies, there is no one decision tree for selecting the proper number of clusters [72–75]. The selection of the number of clusters is a multi-decision problem, and algorithms need to be developed in order to automatically choose adequate numbers. In this study, the best fit was based on the R-square parameter (>0.7), and the cubic clustering criterion, which leads to the selection of 3 to 5 clusters. It was empirically determined that the 5-cluster solution yielded the best match and was most easily interpretable. This solution was selected for further exploration. The PROC CLUSTER and PROC FASTCLUS procedures in SAS (9.4) were used to perform cluster analyses. By implementing a cluster analysis involving all variables, a standardization was performed with Z-scores in order to limit the probability of variables with large variance (e.g., income) from yielding larger effects on results relative to variables with smaller variance (e.g., number of cigarettes per day) [76,77]. Standardization (sometimes called normalization) is important as it helps ensure that groups are defined based on the distance between observations. Final results are interpreted as deviation from the mean value for all regions versus the national average. It is necessary to perform standardization tests in order to eliminate outliers and measure all variables on a similar scale. This helps avoid situations in which one variable has a larger impact on cluster analysis relative to another one. The graphical representation of clusters was performed using R software's *sp* package [78] and the NUTS-3 classification. [68]. In order to measure potential impact on raw cancer morbidity data, we conducted multiple regression modelling [79]. The dependent variable was the number of newly diagnosed cancer cases without age adjustment, and the explanatory variables included SES, lifestyle and environment.

4. Results

4.1. Cluster Analysis

Our cluster analysis captured three groups of factors: environment, lifestyle and SES. Table 2 presents five clusters based on a combined analysis of environmental, lifestyle and SES. The first cluster is the city of Kraków (green district). This cluster is characterized by the second-highest level of household income, years of education and sports activity. In this cluster, measures of air pollution are very high and exceed WHO limits. The second cluster (yellow) covers the largest part of Poland. This cluster neighbours industrial and metropolitan areas. When it comes to lifestyle factors, there is a low sports activity level and low BMI, which falls below the national average level. In this cluster, SES is low and constitutes the lowest household income across all clusters. In Figure 1, the third cluster (purple) represents a district that includes the Silesia region and the city of Bełchatów. This cluster is characterized by limit-exceeding levels of benzopyrene and fairly high measures of $PM_{2.5}$ and PM_{10}. Additionally, these regions are characterized by high BMI and appreciable levels of educational attainment and income. Figure 1 reveals that district 4 (red), which represents the metropolitan areas of Warsaw, Wrocław, Szczecin, Poznań and Gdańsk, differ significantly from other areas of Poland. Cluster 4 is characterized by the highest levels of household net monthly income, most years of education, and the highest declared sports activity. In addition, this metropolitan cluster includes a fairly high percentage of respondents who declare consuming too much alcohol. It appears that these respondents are dually aware of both their healthy and non-healthy behaviours. Cluster 5 (blue), which encompasses the northern part of Poland, is the cleanest area in terms of air pollution. Nevertheless, it presents with the nation's highest average prevalence of cigarette use, over-consumption of alcohol, low sports activity and high BMI. The difference between clusters 2 (yellow) and 5 (blue) stems from their respective measures of air pollution. Areas near industrial areas (i.e., old central industrial regions of Silesia) as well as those in and near Kraków appear more heavily polluted than the northern region of Poland.
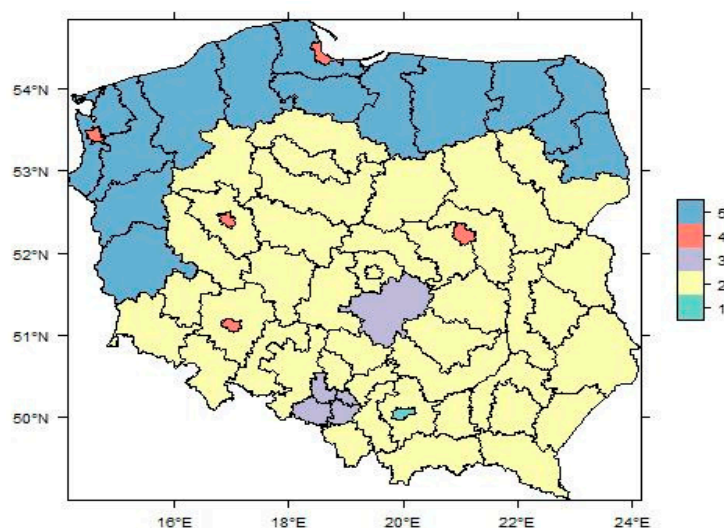
Figure 1. Spatial illustration of clusters in three categories of variables: environmental, lifestyle and socio-economic status (SES).

Table 3 reveals the number of newly diagnosed cancer cases per 100,000 citizens in each cluster in 2014. The highest number of new incidence occurs in clusters 1 and 4, which include metropolitan areas and the city of Kraków. These are characterized by high levels of SES. The lowest number of newly diagnosed cases occurs in cluster 5, which is located in the northern Poland and is characterized by the lowest levels of air pollution.

Table 2. Results of the cluster analysis of all cancer risk factors. BMI: body-mass index.

Cluster	Cigarettes _Perday	Sport	Alcohol	BMI	Household _Income	Years _of_Edu	PM_{10} _Avg	PM_{10}_No _of_Days	PM_{25} _Avg	BaP _ Avg
1	−1.85	1.76	−0.17	−2.83	1.07	2.86	3.11	3.34	2.36	1.54
2	−0.13	−0.18	−0.32	−0.07	−0.21	−0.24	0.17	0.16	0.16	0.14
3	0.01	0.59	1.40	0.87	0.04	0.46	1.77	1.69	1.40	1.64
4	−0.28	1.90	0.89	−1.53	2.00	2.36	−0.70	−0.67	−0.69	−1.01
5	0.68	−0.51	0.17	0.70	−0.18	−0.50	−1.20	−1.18	−1.22	−0.88

Note: This table presents cluster means after the standardization of variables in relation to national average results based on the Social Diagnosis (SD) dataset and local dataset on pollution from the Chief Inspectorate for Environmental Protection.

Table 3. Cancer morbidity by clusters.

Cluster	New Cases Per 100,000
1 (green)—Kraków	487
2 (yellow)—neighbouring industrial areas	407
3 (purple)—Silesia + Bełchatów (industrial areas)	398
4 (red)—metropolitan area	487
5 (blue)—northern part of Poland	390
National	414

4.2. Multiple Linear Regression Model of Cancer Morbidity

Standard multiple linear regression was performed in order to measure the causality of cancer morbidity. Cancer morbidity was measured per 100,000 citizens based on the National Cancer Registry (KRN) data [80] at the NUT-3 regional level.

In order to discuss the potential effects of three groups of factors (environment, lifestyle and SES), it is necessary to understand the impact of these cancer risk factors on cancer morbidity. Firstly, there is a strong and significantly positive impact of the benzopyrene (BaP) level on cancer morbidity. If the BaP level rises by 1 μg/ m^3, cancer morbidity increases by 24 incidences per 100,000 citizens. When it comes to lifestyle measure, there is no significant impact of this group of variables. As presented in Table 4, cigarette consumption has a positive effect on cancer morbidity while sports activity decreases the number of newly diagnosed cancer cases. With respect to the two variables capturing SES, higher household income has a negative effect on cancer morbidity, while additional years of education result in higher morbidity. It is necessary to analyse these measures according to cluster segments and higher SES is often associated with industry areas where environmental aspects play a significant role. The R-squared statistic was 0.5673 which means that almost 57% of the variation in cancer morbidity is explained by SES, lifestyle and environment variables in this regression model. A significant limitation of the regression model is the sample size with observation included from only 66 regions in one period of time. As such, regression models may lend insightful findings in addition to cluster analysis.

Table 4. Multiple linear regression results of cancer morbidity.

Variable	Estimate	Std. Error	t Value	Pr(>|t|)
(Intercept)	371.1916	656.5338	0.5654	0.5751
cigarettes_perday	11.562	10.9768	1.0533	0.2989
BMI	−22.0325	21.342	−1.0324	0.3084
BaP_avg	24.1219	6.3011	3.8282	0.0005 ***
PM_{25}_avg	−4.1596	3.5321	−1.1777	0.2463
PM_{10}_avg	−8.2597	5.6905	−1.4515	0.1549
PM_{10}_no_of_days	0.879	1.2331	0.7129	0.4803
household_income	−0.0474	0.0184	−2.5724	0.0141 *
years_of_edu	89.2808	18.6942	4.7759	0.001 ***
Sport	−162.3632	155.2515	−1.0458	0.3023
Alcohol	−594.1154	431.0968	−1.3781	0.1762

Significance: *** <0.001, ** <0.01, * <0.05.

5. Discussion

This is the first attempt to examine a comprehensive set of cancer risk factors at the regional level in Poland. In 2017, the number of newly diagnosed cancer cases exceeded 164,000. According to the National Cancer Registry, the average growth rate of newly diagnosed cases in 2011–2016 measured around 2.4% yearly [29]. The negative impact of tobacco use, poor diet, physical inactivity and obesity has already been documented by several earlier studies [23,25,45,81]. Tobacco use accounts for nearly 22% of cancer deaths [1] and increases the risk of fourteen types of cancer [82], particularly lung cancer, which is the leading cause of cancer death in Poland [29]. Moreover, obesity linked with physical inactivity, alcohol consumption and an unhealthy diet increases the risk of cancer [24,25]. Another key finding is that air pollution increases the risk of cancer [49–51]. In addition, low SES is associated with an increased probability of tobacco use and related exposure to carcinogenic substances [17,83]. One study from Poland examined the relationship between SES and mortality resulting from lung cancer and respiratory diseases and confirmed a correlation between current SES and mortality from both respiratory diseases and lung cancer [14]. Considering WHO-defined cancer risk factors [1], this paper focuses not only on smoking and declared alcohol consumption but also takes into account internal factors such as SES, diet and sports activity combined with external determinants such as the environment. With this extensive approach in mind, a regional analysis of cancer risk factors was also undertaken. In order to develop cluster analysis on the aggregated regional NUT-3 level [68] data, the SD dataset with 19,769 respondents was merged with a local data bank on pollution [66,67].

The main research question examines the impact of combined environment, lifestyle and SES on the risk of cancer while taking the spatial split into account. The results of the final cluster analysis are presented in Figure 1 and Table 2. The first cluster (green) occurs in the city of Kraków, which is an interesting cluster, where citizens present similarly in terms of SES and sports activity to those living in other large metropolitan areas across Poland. However, Kraków is very heavily polluted due to coal, rubbish and wood heating methods as well as heavy and congested traffic. Moreover, Kraków is located in the Vistula River valley, which significantly limits natural air circulation. The second cluster (yellow) is the area neighbouring industrial regions. Table 2 shows that cluster 2 has slightly higher air pollution compared to the national average. Moreover, people living in the second cluster declare low sports activity and have relatively low SES. The third cluster (purple) includes the Silesia region and metropolitan area of Bełchatów. Like Krakow, this area is heavily polluted due to coal-mining and heavy industry. Unlike Kraków, however, the people living in this area exhibit less sports activity, which implies a higher BMI level. The fourth cluster (red) is a metropolitan district characterized by high SES and high sports activity ratio. On the other hand, people in large cities are more aware of risky behaviours and declare a fairly high level of alcohol consumption. People here are highly educated, which implies a higher awareness of non-healthy behaviours. District 5 (blue) in Figure 1 covers northern Poland and is characterized by the lowest levels of pollution in the country. However, the people of this region are heavy consumers of tobacco and alcohol, with low engagement in sports activity and a high BMI level.

In addition to cluster analysis, the multiple regression model was performed in order to help draw conclusions about the impact of the proposed interventions on cancer morbidity. Taking into account the significant limitations of our regression model (a low number of districts used), the regression analysis shows the direction of impact of the environment, lifestyle and SES on cancer morbidity. While smoking and air pollution have positive impacts on cancer morbidity, sports activity and household income carry opposite results. The regression modelling is an ecological analysis, and causality interpretations should be made with caution.

This study relied on self-reported survey data from the SD study and was therefore subject to response bias. A second potential limitation is related to the SD alcohol consumption data. In conducting a detailed review of the available dataset, there were no precise data on alcohol consumption per capita at the regional NUT-3 level in Poland. A third possible limitation relates to cancer risk specifications. That is, this analysis did not take into account differences in cancer risk factors as they relate to different

types of cancer. For example, 90% of lung cancer deaths are related to cigarette smoking [84]. Finally, this analysis lacks in the differentiation of large metropolitan cities, where population density is high, but more detailed data are not available. For this reason, our study aggregated all data to the NUT-3 level. Nonetheless, the results of this study examine global cancer risk factors [1] and can be referenced when seeking to improve the effectiveness of cancer-prevention policy in Poland.

6. Conclusions

This study focused on identifying clusters where to better address appropriate local cancer prevention policy. Policy interventions should not only consider anti-smoking campaigns but may also include tailored fiscal and social policy tools that are specific to geographical clusters. These tools should be designed to focus on issues related to air pollution, consumer lifestyle and SES in target regions. Our main recommendations are presented in Table 5.

Table 5. Policy recommendations for selected clusters.

Cluster	Cancer-Risk Factors—Focus Areas
1 (green)—Kraków	air pollution—$PM_{2.5}$, PM_{10}, benzoapyrene
2 (yellow)—neighbouring industrial areas	sports activity; air pollution—$PM_{2.5}$, PM_{10}, benzoapyrene; SES
3 (purple)—Silesia + Bełchatów (industrial areas)	air pollution—benzoapyrene, $PM_{2.5}$, PM_{10}; alcohol; BMI
4 (red)—metropolitan area	alcohol; air pollution—$PM_{2.5}$, PM_{10},
5 (blue)—northern part of Poland	smoking; sports activity; alcohol; BMI; SES

In order to improve their effectiveness, national cancer prevention campaigns should focus on local cancer-risk factors. Regarding which national policy tackles broadly defined problems such as cigarette smoking or air pollution, both national and local policy-makers must work together to focus on matching proper policy tools to the cancer-risk factors that dominate a particular jurisdiction. Moreover, better methods of communication must be developed as a means to adequately address the different needs of each of the various regions of Poland.

Abbreviations

BMI	body mass index
household_income	net household income, monthly
PM_{10}	particulate matter with a diameter between 2.5 μm and 10 μm
$PM_{2.5}$	fine particulate matter with a diameter of 2.5 μm or less
WHO	World Health Organization
NCDs	non-communicable diseases
SES	socio-economic status
BaP	Benzoapyrene
SD	Social Diagnosis
PLN	Polish zloty

References

1. World Health Organization Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 5 October 2019).

2. Mendis, S.; World Health Organization. *Global Status Report on Noncommunicable Diseases 2014*; WHO: Geneva, Switzerland, 2014; ISBN 978-92-4-156485-4.

3. World Health Organization. *Global Status Report on Noncommunicable Diseases: 2010*; World Health Organization: Geneva, Switzerland, 2011; ISBN 978-92-4-156422-9.

4. National Cancer Institute Cancer Statistics. Available online: https://www.cancer.gov/about-cancer/understanding/statistics (accessed on 22 December 2019).

5. Blue Ribbon Panel Report. Cancer Moonshoot. 2016; p. 74. Available online: https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/blue-ribbon-panel/blue-ribbon-panel-report-2016.pdf (accessed on 31 October 2020).

6. World Health Organization Global Action Plan for the Prevention and Control of NCDs. 2013–2020. Available online: http://www.who.int/nmh/events/ncd_action_plan/en/ (accessed on 6 September 2019).

7. European Commission Communication from the Commission on European Action against Cancer. Available online: http://ec.europa.eu/health/ph_information/dissemination/diseases/docs/com_2009_291.en.pdf (accessed on 29 December 2019).

8. Albreht, T.; Martin-Moreno, J.M.; Jelenc, M.; Gorgojo, L.; Harris, M. *European Guide for Quality National Cancer Control Programmes*; National Institute of Public Health: Ljubljana, Slovenia, 2015; ISBN 978-961-6911-60-3.

9. European Commission. *Towards a Mission-Oriented Research and Innovation Policy in the European Union—An ESIR Memorandum*; Publications Office of the European Union: Luxembourg, 2017. [CrossRef]

10. Mazzucato, M. *Mission-Oriented Research & Innovation in the European Union: A Problem-Solving Approach to Fuel Innovation-Led Growth*; Publications Office of the European Union: Luxembourg, 2018; ISBN 978-92-79-79918-1.

11. Didkowska, J.; Wojciechowska, U.; Czaderny, K.; Olasek, P.; Ciuba, A. *Nowotwory złośliwe w Polsce w 2017 roku | Cancer in Poland in 2017*; Warszawa, Poland. 2019. Available online: http://onkologia.org.pl/wp-content/uploads/Nowotwory_2017.pdf (accessed on 31 October 2020).

12. Didkowska, J. Prognozy rozwoju chorób nowotworowych w Polsce (Forecasts of the development of cancer diseases in Poland). In *Zachorowalność i umieralność na nowotwory a sytuacja demograficzna Polski*; Rządowa Rada Ludnościowa: Warsaw, Poland, 2014.

13. Ministerstwo Zdrowia Projekt uchwały Rady Ministrów w sprawie przyjęcia Narodowej Strategii Onkologicznej na lata 2020–2030 [Draft of resolution of the Cabinet regarding the adoption of the National Oncology Strategy for 2020–2030]. Available online: https://www.gov.pl/web/zdrowie/projekt-uchwaly-rady-ministrow-w-sprawie-przyjecia-narodowej-strategii-onkologicznej-na-lata-2020-2030 (accessed on 29 December 2019).

14. Polak, M.; Genowska, A.; Szafraniec, K.; Fryc, J.; Jamiołkowski, J.; Pająk, A. Area-Based Socio-Economic Inequalities in Mortality from Lung Cancer and Respiratory Diseases. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1791. [CrossRef] [PubMed]

15. Orlewska, K.; Sliwczynski, A.; Orlewska, E. An ecological study of the link between the risk of most frequent types of cancer in Poland and socioeconomic variables. *Int. J. Public Health* **2018**, *63*, 777–786. [CrossRef] [PubMed]

16. Comba, P.; Bianchi, F.; Fazzo, L.; Martina, L.; Menegozzo, M.; Minichilli, F.; Mitis, F.; Musmeci, L.; Pizzuti, R.; Santoro, M.; et al. Cancer mortality in an area of Campania (Italy) characterized by multiple toxic dumping sites. *Ann. N. Y. Acad. Sci.* **2006**, *1076*, 449–461. [CrossRef] [PubMed]

17. Peretti-Watel, P.; Fressard, L.; Bocquier, A.; Verger, P. Perceptions of cancer risk factors and socioeconomic status. A French study. *Prev. Med. Rep.* **2016**, *3*, 171–176. [CrossRef] [PubMed]

18. World Health Organization. World Health Statistics 2019: Monitoring Health for the SDGs. Available online: http://www.who.int/gho/publications/world_health_statistics/2019/en/ (accessed on 6 September 2019).

19. Kubik, A.K.; Parkin, D.M.; Plesko, I.; Zatonski, W.; Kramarova, E.; Möhner, M.; Friedl, H.P.; Juhasz, L.; Tzvetansky, C.G.; Reissigova, J. Patterns of cigarette sales and lung cancer mortality in some central and eastern European countries, 1960–1989. *Cancer* **1995**, *75*, 2452–2460. [CrossRef]

20. Tyczyński, J. Lung cancer mortality and tobacco smoking in Poland—Will we observe further changes in mortality in light of recent smoking patterns? *Nowotw. J. Oncol.* **2004**, *54*, 333–341.

21. Zatoński, W.; Jha, P. *The Health Transformation in Eastern Europe after 1990: A Second Look*; M. Skłodowska-Curie Memorial Cancer Center and Institute of Oncology: Warsaw, Poland, 2000.

22. National Cancer Institute Cancer Disparities. Available online: https://www.cancer.gov/about-cancer/understanding/disparities (accessed on 22 December 2019).

23. Arminger, G.; Clogg, C.C.; Sobel, M.E. (Eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*; Springer: Berlin/Heidelberg, Germany, 1995; ISBN 978-0-306-44805-8.

24. Collins, L.M.; Lanza, S.T. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences | Wiley. Available online: https://www.wiley.com/en-us/Latent+Class+and+Latent+Transition+Analysis%3A+With+Applications+in+the+Social%2C+Behavioral%2C+and+Health+Sciences-p-9780470228395 (accessed on 22 December 2019).

25. Courtney, K.E.; Polich, J. Binge drinking in young adults: Data, definitions, and determinants. *Psychol. Bull.* **2009**, *135*, 142–156. [CrossRef]

26. Kuper, H.; Boffetta, P.; Adami, H.-O. Tobacco use and cancer causation: Association by tumour type. *J. Intern. Med.* **2002**, *252*, 206–224. [CrossRef]

27. Gallaway, M.S.; Henley, S.J.; Steele, C.B.; Momin, B.; Thomas, C.C.; Jamal, A.; Trivers, K.F.; Singh, S.D.; Stewart, S.L. Surveillance for Cancers Associated with Tobacco Use - United States, 2010–2014. *MMWR Surveill Summ* **2018**, *67*, 1–42. [CrossRef]

28. Office of the Surgeon General (US); Office on Smoking and Health (US). *The Health Consequences of Smoking: A Report of the Surgeon General*; Reports of the Surgeon General; Centers for Disease Control and Prevention (US): Atlanta, GA, USA, 2004.

29. Krajowy Rejestr Nowotworów [National Cancer Registry]. Raporty. 2018. Available online: http://onkologia.org.pl/raporty/ (accessed on 31 October 2020).

30. Kushi, L.H.; Doyle, C.; McCullough, M.; Rock, C.L.; Demark-Wahnefried, W.; Bandera, E.V.; Gapstur, S.; Patel, A.V.; Andrews, K.; Gansler, T.; et al. American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J. Clin.* **2012**, *62*, 30–67. [CrossRef]

31. Vainio, H.; Kaaks, R.; Bianchini, F. Weight control and physical activity in cancer prevention: International evaluation of the evidence. *Eur. J. Cancer Prev.* **2002**, *11* (Suppl. 2), S94–S100.

32. Calle, E.E.; Rodriguez, C.; Walker-Thurmond, K.; Thun, M.J. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N. Engl. J. Med.* **2003**, *348*, 1625–1638. [CrossRef] [PubMed]

33. Berrington de Gonzalez, A.; Sweetland, S.; Spencer, E. A meta-analysis of obesity and the risk of pancreatic cancer. *Br. J. Cancer* **2003**, *89*, 519–523. [CrossRef] [PubMed]

34. Patel, A.V.; Rodriguez, C.; Bernstein, L.; Chao, A.; Thun, M.J.; Calle, E.E. Obesity, recreational physical activity, and risk of pancreatic cancer in a large U.S. Cohort. *Cancer Epidemiol. Biomarkers Prev.* **2005**, *14*, 459–466. [CrossRef] [PubMed]

35. Lindblad, M.; Rodríguez, L.A.G.; Lagergren, J. Body mass, tobacco and alcohol and risk of esophageal, gastric cardia, and gastric non-cardia adenocarcinoma among men and women in a nested case-control study. *Cancer Causes Control* **2005**, *16*, 285–294. [CrossRef]

36. Amling, C.L.; Riffenburgh, R.H.; Sun, L.; Moul, J.W.; Lance, R.S.; Kusuda, L.; Sexton, W.J.; Soderdahl, D.W.; Donahue, T.F.; Foley, J.P.; et al. Pathologic variables and recurrence rates as related to obesity and race in men with prostate cancer undergoing radical prostatectomy. *J. Clin. Oncol.* **2004**, *22*, 439–445. [CrossRef] [PubMed]

37. Freedland, S.J.; Terris, M.K.; Platz, E.A.; Presti, J.C. Body mass index as a predictor of prostate cancer: Development versus detection on biopsy. *Urology* **2005**, *66*, 108–113. [CrossRef] [PubMed]

38. Patel, A.V.; Callel, E.E.; Bernstein, L.; Wu, A.H.; Thun, M.J. Recreational physical activity and risk of postmenopausal breast cancer in a large cohort of US women. *Cancer Causes Control* **2003**, *14*, 519–529. [CrossRef]

39. Dodd, L.J.; Al-Nakeeb, Y.; Nevill, A.; Forshaw, M.J. Lifestyle risk factors of students: A cluster analytical approach. *Prev. Med.* **2010**, *51*, 73–77. [CrossRef]

40. El Ansari, W.; Stock, C.; John, J.; Deeny, P.; Phillips, C.; Snelgrove, S.; Adetunji, H.; Hu, X.; Parke, S.; Stoate, M.; et al. Health promoting behaviours and lifestyle characteristics of students at seven universities in the UK. *Cent. Eur. J. Public Health* **2011**, *19*, 197–204. [CrossRef]

41. Laska, M.N.; Pasch, K.E.; Lust, K.; Story, M.; Ehlinger, E. Latent class analysis of lifestyle characteristics and health risk behaviors among college youth. *Prev. Sci.* **2009**, *10*, 376–386. [CrossRef] [PubMed]

42. American College of Health Association. The American College Health Association National College Health Assessment (ACHA-NCHA), Spring 2003 Reference Group report. *J. Am. Coll. Health* **2005**, *53*, 199–210.

43. Mays, D.; Peshkin, B.N.; Sharff, M.E.; Walker, L.R.; Abraham, A.A.; Hawkins, K.; Tercyak, K.P. Correlates of Adherence to a Telephone-Based Multiple Health Behavior Change Cancer Preventive Intervention for Teens: The Healthy for Life Program (HELP). *Health Educ. Behav.* **2012**, *39*, 18–26. [CrossRef] [PubMed]

44. Rebholz, C.E.; Rueegg, C.S.; Michel, G.; Ammann, R.A.; von der Weid, N.X.; Kuehni, C.E.; Spycher, B.D. Clustering of health behaviours in adult survivors of childhood cancer and the general population. *Br. J. Cancer* **2012**, *107*, 234–242. [CrossRef] [PubMed]

45. Hill, K.G.; White, H.R.; Chung, I.J.; Hawkins, J.D.; Catalano, R.F. Early adult outcomes of adolescent binge drinking: Person- and variable-centered analyses of binge drinking trajectories. *Alcohol. Clin. Exp. Res.* **2000**, *24*, 892–901. [CrossRef] [PubMed]

46. Elliott, M.; Arbogast, K.; Durbin, D. A Latent Class Analysis of Injury Patterns Among Rear-Seated, Seat-Belted Children. *J. Trauma Inj. Infect. Crit. Care* **2006**, *61*, 1244–1248. [CrossRef] [PubMed]

47. Moreno-Gómez, C.; Romaguera-Bosch, D.; Tauler-Riera, P.; Bennasar-Veny, M.; Pericas-Beltran, J.; Martinez-Andreu, S.; Aguilo-Pons, A. Clustering of lifestyle factors in Spanish university students: The relationship between smoking, alcohol consumption, physical activity and diet quality. *Public Health Nutr.* **2012**, *15*, 2131–2139. [CrossRef]

48. European Environment Agency Air Quality in Europe. 2018. Available online: https://www.eea.europa.eu/publications/air-quality-in-europe-2018 (accessed on 6 September 2019).

49. Hemminki, K.; Pershagen, G. Cancer risk of air pollution: Epidemiological evidence. *Environ. Health Perspect.* **1994**, *102*, 187–192.

50. Briggs, D. Environmental pollution and the global burden of disease. *Br. Med. Bull.* **2003**, *68*, 1–24. [CrossRef]

51. Homaei Shandiz, F.; Hadizadeh Talasaz, Z. The relationship between breast cancer and air pollution: Review article. *Rev. Clin. Med.* **2017**, *4*, 136–139. [CrossRef]

52. Jędrak, J.; Konduracka, E.; Badyda, A.J.; Dąbrowiecki, P. *Wpływ zanieczyszczeń powietrza na zdrowie [Impact of air pollution on health]*; Krakowski Alarm Smogowy: Kraków, Poland, 2017; Available online: https://depot.ceon.pl/bitstream/handle/123456789/15434/ksi%c4%85%c5%bcka%20o%20zdrowiu.pdf?sequence=1&isAllowed=y (accessed on 31 October 2020).

53. World Health Organization Ambient (Outdoor) Air Pollution. Available online: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (accessed on 3 March 2020).

54. Deuchert, E.; Cabus, S.; Tafreschi, D. A Short Note on Economic Development and Socioeconomic Inequality in Female Body Weight. *Health Econ.* **2014**, *23*, 861–869. [CrossRef] [PubMed]

55. Fichera, E.; Gathergood, J. Do Wealth Shocks Affect Health? New Evidence from the Housing Boom. *Health Econ.* **2016**, *25*, 57–69. [CrossRef] [PubMed]

56. Goding Sauer, A.; Siegel, R.L.; Jemal, A.; Fedewa, S.A. Current Prevalence of Major Cancer Risk Factors and Screening Test Use in the United States: Disparities by Education and Race/Ethnicity. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 629–642. [CrossRef] [PubMed]

57. Hagen, T.P.; Häkkinen, U.; Iversen, T.; Klitkou, S.T.; Moger, T.A. Socio-economic Inequality in the Use of Procedures and Mortality among AMI Patients: Quantifying the Effects along Different Paths. *Health Econ.* **2015**, *24*, 102–115. [CrossRef] [PubMed]

58. Link, B.G.; Phelan, J. Social Conditions as Fundamental Causes of Disease. *J. Health Soc. Behav.* **1995**, 80–94. [CrossRef]

59. Brodszky, V.; Beretzky, Z.; Baji, P.; Rencz, F.; Péntek, M.; Rotar, A.; Tachkov, K.; Mayer, S.; Simon, J.; Niewada, M.; et al. Cost-of-illness studies in nine Central and Eastern European countries. *Eur. J. Health Econ.* **2019**, *20*, 155–172. [CrossRef]

60. Adler, N.E.; Newman, K. Socioeconomic Disparities in Health: Pathways and Policies. *Health Aff.* **2002**, *21*, 60–76. [CrossRef] [PubMed]

61. Hemminki, K.; Li, X. Level of Education and the Risk of Cancer in Sweden. *Cancer Epidemiol. Biomark. Prev.* **2003**, *12*, 796–802.

62. O'Connor, J.M.; Sedghi, T.; Dhodapkar, M.; Kane, M.J.; Gross, C.P. Factors Associated With Cancer Disparities Among Low-, Medium-, and High-Income US Counties. *JAMA Netw. Open* **2018**, *1*, e183146. [CrossRef]

63. Withrow, D.R.; Berrington de González, A.; Spillane, S.; Freedman, N.D.; Best, A.F.; Chen, Y.; Shiels, M.S. Trends in Mortality Due to Cancer in the United States by Age and County-Level Income, 1999–2015. *J. Natl. Cancer Inst.* **2019**, *111*, 863–866. [CrossRef] [PubMed]

64. Tanaka, R.; Matsuzaka, M.; Sasaki, Y. Influence of Income on Cancer Incidence and Death among Patients in Aomori, Japan. *Asian Pac. J. Cancer Prev.* **2018**, *19*, 3193–3202. [CrossRef]

65. Kim, J.-M.; Kim, H.-M.; Jung, B.-Y.; Park, E.-C.; Cho, W.-H.; Lee, S.-G. The Association between Cancer Incidence and Family Income: Analysis of Korean National Health Insurance Cancer Registration Data. *Asian Pac. J. Cancer Prev.* **2012**, *13*, 1371–1376. [CrossRef]

66. Chief Inspectorate for Environmental Protection Local Dataset about Pollution. Available online: http://powietrze.gios.gov.pl/pjp/archives?lang=en (accessed on 5 October 2019).

67. Czapiński, J.; Panek, T. Social Diagnosis—Objective and Subjective Quality of Life in Poland. Available online: http://www.diagnoza.com/ (accessed on 5 October 2019).

68. European Commission. *Commission Regulation(EU) No 31/2011 of 17 January 2011 Amending Annexes to Regulation (EC) No 1059/2003 of the European Parliament and of the Council on the Establishment of a Common Classification of Territorial Units for Statistics (NUTS)*; Official Journal of the European Union: Brussels, Belgium, 2011; Volume 31, Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R0031&from=EN (accessed on 31 October 2020).

69. Główny Inspektor Ochrony Środowiska [Central Environmental Protection Inspectorate] Poziomy dopuszczalne zanieczyszczeń w powietrzu ze względu na ochronę zdrowia ludzi i ochronę roślin, terminy ich osiągnięcia oraz okresy, dla których uśrednia się wyniki pomiarów—GIOŚ [Admissible Levels of Pollutants in the Air Due to the Protection of Human Health and Plant Protection, Deadlines for Their Achievement and Periods for Which Measurement Results Are Averaged]. Available online: https://powietrze.gios.gov.pl/pjp/content/annual_assessment_air_acceptable_level (accessed on 5 October 2019).

70. World Health Organization Air Quality Guidelines Global Update. 2005. Available online: https://apps.who.int/iris/handle/10665/107823 (accessed on 12 October 2020).

71. Ward, J.H., Jr. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]

72. Henry, D.; Dymnicki, A.B.; Mohatt, N.; Allen, J.; Kelly, J.G. Clustering Methods with Qualitative Data: A Mixed Methods Approach for Prevention Research with Small Samples. *Prev. Sci.* **2015**, *16*, 1007–1016. [CrossRef] [PubMed]

73. Ketchen, D.J.; Shook, C.L. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [CrossRef]

74. Caruso, G.; Gattone, S.A.; Fortuna, F.; Di Battista, T. Cluster Analysis for mixed data: An application to credit risk evaluation. *Socio Econ. Plan. Sci.* **2020**, 100850. [CrossRef]

75. Caruso, G.; Gattone, S.A.; Balzanella, A.; Di Battista, T. Cluster Analysis: An Application to a Real Mixed-Type Data Set. In *Models and Theories in Social Systems*; Flaut, C., Hošková-Mayerová, Š., Flaut, D., Eds.; Studies in Systems, Decision and Control; Springer International Publishing: Cham, Switzerland, 2019; pp. 525–533. ISBN 978-3-030-00084-4.

76. Milligan, G.W.; Cooper, M.C. A study of standardization of variables in cluster analysis. *J. Classif.* **1988**, *5*, 181–204. [CrossRef]

77. Standardization in Cluster Analysis. Available online: https://community.alteryx.com/t5/Alteryx-Designer/Standardization-in-Cluster-Analysis/ta-p/302296 (accessed on 22 July 2020).

78. R: A Package Providing Classes and Methods for Spatial Data. Available online: http://127.0.0.1:16776/library/sp/html/00sp.html (accessed on 21 July 2020).

79. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed.; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2003; ISBN 978-0-8058-2223-6.

80. Krajowy Rejestr Nowotworów [National Cancer Registry] Zachorowania na nowotwory złośliwe w powiatach—Otwarte Dane [Cancer Incidence by Districts—open Data]. Available online: https://dane.gov.pl/pl/dataset/1793,zachorowania-na-nowotwory-zosliwe-w-powiatach (accessed on 12 October 2020).

81. Brener, N.D.; Collins, J.L. Co-occurrence of health-risk behaviors among adolescents in the United States. *J. Adolesc. Health* **1998**, *22*, 209–213. [CrossRef]

82. Elk, R.; Landrine, H. *Cancer Disparities: Causes and Evidence-Based Solutions*; Springer Publishing Company: Berlin/Heidelberg, Germany, 2012; ISBN 978-0-8261-0882-1.

83. Merletti, F.; Galassi, C.; Spadea, T. The socioeconomic determinants of cancer. *Environ. Health* **2011**, *10*, S7. [CrossRef] [PubMed]

84. Centers for Disease Control and Prevention What Are the Risk Factors for Lung Cancer? Available online: https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm (accessed on 22 December 2019).

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.