

## ANALYSIS OF THE SIMILARITIES AND DIFFERENCES BETWEEN POVIATS IN THE MAŁOPOLSKIE VOIVODESHIP IN TERMS OF SARS-COV-2 INFECTIONS

Agata Sielska  <https://orcid.org/0000-0001-6880-8899>

Ewelina Nojszewska  <https://orcid.org/0000-0003-3176-0240>

Department of Applied Economics, Collegium of Management and Finances  
SGH Warsaw School of Economics

e-mail: asiels@sgh.waw.pl; ewelina.nojszewska@sgh.waw.pl

**Abstract:** The purpose of the paper is to study the genesis of SARS-CoV-2 infections in poviats in Małopolskie voivodeship from June 29 2020 till September 7 2020. At the beginning of September Małopolskie voivodeship was characterized by one of the highest numbers of confirmed cases of Covid-19 in Poland. It is a region of diversified structure. That fact can be reflected by the results of the study. The paper aims to verify whether the structure of infections is diversified in the region and to analyze the changes of these patterns in the studied period.

**Keywords:** SARS-CoV-2, Covid-19, hierarchical clustering, Ward's method

**JEL classification:** C38, I10

### INTRODUCTION

Coronavirus SARS-CoV-2 pandemic began at the end of 2019 in China. First cases were reported in the city Wuhan, capital of the Hubei province. In 2020 the virus spread throughout the globe, hitting Europe at first, then Americas. In Poland the first case of Covid-19 (disease caused by SARS-CoV-2 coronavirus) was reported on 4 of March in Lubuskie voivodship. In the paper we focus on Małopolskie voivodship where according to official statistics the disease reached on 9 of March. There are several reasons behind our decision. Firstly, at the time of preparing this study Małopolskie voivodeship was characterized by one of the highest numbers of confirmed cases of Covid-19 in Poland. Secondly, it is a region of diversified structure. It consists 22 poviats of different characteristics, including urban, urban-rural and rural communes and there is a question whether this structure

<https://doi.org/10.22630/MIBE.2020.21.2.8>

would be reflected by the results of the study. For example we may assume that in the poviats characterized by higher population density we may come across more cases. We aim to verify whether the structure of infections is diversified in the region and to analyze the changes of these patterns in the studied period.

The paper is organized as follows: in the first part we describe the evolution of SARS-CoV-2 pandemic in Małopolskie voivodship. Second part describes the methodology, while the third is dedicated to the presentation and discussion of the results. The paper ends with conclusions.

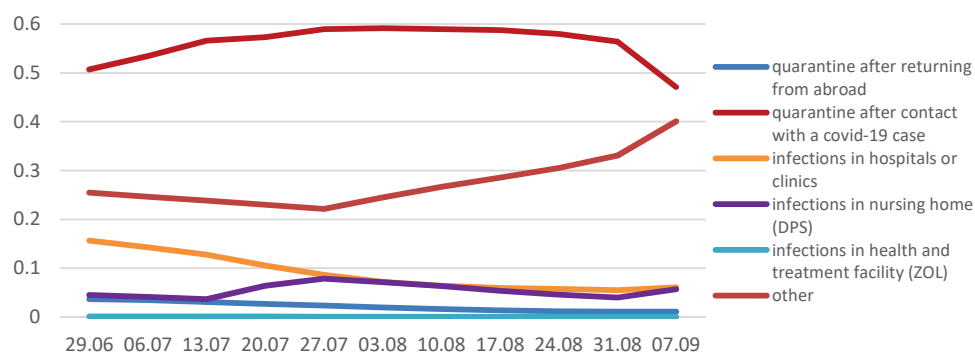
### SARS-COV-2 IN MAŁOPOLSKIE VOIVODSHIP

Poviats of Małopolska are diversified not only in terms of socio-economic conditions but also in terms of the number of cases and the structure of their sources.

First case of covid-19 in Małopolskie voivodship was reported on 09.03.2020, shortly after the virus reached Poland. In the study we focus on the cumulated (from March 9th) shares of infections from given sources.

As shown in Figure 1, in the analysed period we may notice that the infections after the contact and subsequent quarantine dominated. It is important to mention that the shares of people ill after the quarantine after returning from abroad were falling down despite the holiday season. This trend is very likely a result of changes in tourist traffic caused by restrictions. The initial high share of infections from this group results from numerous quarantines of people returning from abroad at beginning of the pandemic. The share of people infected in hospitals and clinics was decreasing as well. That can be related to developing effective procedures by both medical personnel and patients. The share of other sources of infection was increasing which may result from the summer season and family meetings, etc. Despite the described tendencies, the median ranks calculated for the sources of infections show that during the whole analysed period 3 sources dominated: *Quarantine after contact*, *Infection in hospitals or clinics* and *Other*.

Figure 1. Shares of sources of Covid-19 cases in Małopolskie voivodeship



Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

## METHODOLOGY

### Time period

The study covers period 29.06.2020-07.09.2020. This period was chosen as a consequence of data availability. Data on the number of cases by different source of contagion are publicly available for Małopolskie voivodeship since 26 of June 2020. We don't use data from September due to the fact that the situation might have changed with the beginning of the school year. Because of the long incubation period, the effect of opening of schools might not yet be seen on 07.09.2020.

### Variables

Following variables were selected as the socio-economic characteristics of the poviats: number of hospital beds, doctors and nurses (midwives included; working personnel) per 10,000 population, gross remuneration, population density, unemployment rate, healthcare expenditures from regional budget (1,000 PLN per 10,000 population – sphere 851), share of population over 60 years old, number of stationary social care facilities per 10,000 population, number of nursing homes per 10,000 population, inhabitants of stationary social care facilities per 10,000 population, large enterprises (over 1000 employees) per 10,000 population.

We include healthcare related variables in order to represent possible places of infection (as one of the sources is *infection in hospitals or clinics*) and, at the same time, treatment opportunities for other disease which affects the populations' health condition. We assume that high population density facilitates the transmission of the virus and in consequence the number of cases. Situation on the labor market may be also of importance as high labor market participation goes in relation with more contacts and moving within or outside cities. Large share of older population may also contribute towards more cases. Older people have usually other health problems which may burden their bodies and immune systems making it easier not only to catch the virus, but also to develop the symptoms which increases the probability of being tested and confirming the disease (compare [Ioannidis et. al 2020]). It is related to the number of nursing homes and other care and treatment facilities. The residents of such places are often older, ill and grouped together which makes it easier to spread the virus. What is important the staff is often employed in several facilities at the same time (including clinics and hospitals), which also facilitates the spread in case of even a single outbreak.

### Data sources

We use two main data sources. Data on the number of cases and sources of infections come from provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków, available on [<https://wsse.krakow.pl/page/>]. Data we use has weekly intervals, i.e. the

observations from following days were taken into account: 29.06, 06.07, 13.07, 20.07, 27.07, 03.08, 10.08, 17.08, 24.08, 31.08 and 07.09.

Until the end of August the following groups of infection sources were reported: *Quarantine after returning from abroad*, *Quarantine after contact*, *Infection in hospitals or clinics*, *Infection in a nursing home* (DPS – Dom Pomocy Społecznej), *Infection in a care and treatment facility* (ZOL – Zakład Opiekuńczo-Lecznicy), *Other*. Since September the group *Other* has been presented in more detailed form and the sources of contagion previously grouped under this label have been divided into: *Workplace*, *School*, *University (and dormitory)*, *Wedding*, *Pubs/cafes/restaurants/clubs*, *Other*. In the study for 07.09.2020 we still use the 7 abovementioned sources aggregated into a group named *Other* in order to be consistent with the previous months.

Data on socio-economic variables come from Local Data Bank of Statistics Poland (Bank Danych Lokalnych Głównego Urzędu Statystycznego - BDL GUS) [bdl.stat.gov.pl]. We used the most current data, i.e. those for 2019. Only in case of the hospital beds data for 2018 were used.

## Method

Before the clustering the data on sources of Covid-19 cases in poviats of Małopolskie voivodship were standardized according to the formula:

$$z_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, \quad (1)$$

where  $x_i$  denotes the value of variable  $x$  for  $i$ -th poviat.

In the next step poviats were clustered based on the similarities in the sources of infections using the Ward's method (Ward 1963). Assuming maximum number of clusters equal to 10, optimal number of clusters was identified based on Dunn index [Dunn 1974]. The calculations were done using R Software [R Core Team 2020], packages NbClust [Charrad et al. 2014] and cluster [Maechler et al. 2019].

## RESULTS

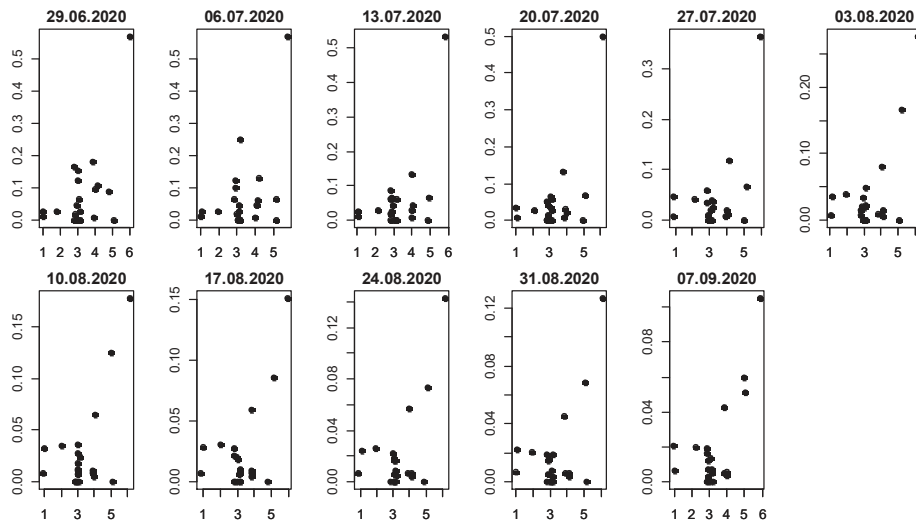
### Identified clusters – Covid-19 characteristics

Ward's method with Dunn index resulted in identification of 6 clusters. First cluster consisted of poviats: bocheński and tarnowski. Second cluster included only one poviat: brzeski, while third one was the largest and included poviats: chrzanowski, miechowski, suski, myślenicki, oświęcimski, nowotarski, krakowski, Kraków, gorlicki, tatrzański, limanowski and wadowicki. Dąbrowski, nowosądecki, wielicki and Nowy Sącz made up the fourth group. Fifth cluster was made of poviats: oluski and proszowicki, while Tarnów was identified as a separate cluster.

The share of infections from the group: *quarantine after returning from abroad*, was high in Tarnów (6th cluster), in 5th group (olkuski and proszowicki) it

was relatively high since the beginning of August (at first in one poviát, on 7<sup>th</sup> of September in both – Figure 2).

Figure 2. Comparison of groups of poviát with respect to shares of Covid-19 cases, source: quarantine after returning from abroad



Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

The share of infections from the group: quarantine *after contact with a Covid-19 case* (Figure 3) was lowest in groups 2 (brzeski) and fifth (olkuski and proszowicki). It was relatively highest in 4<sup>th</sup> cluster, made up of dąbrowski, nowosądecki, wielicki and Nowy Sącz. In clusters 1 and 3 these shares were on a medium level.

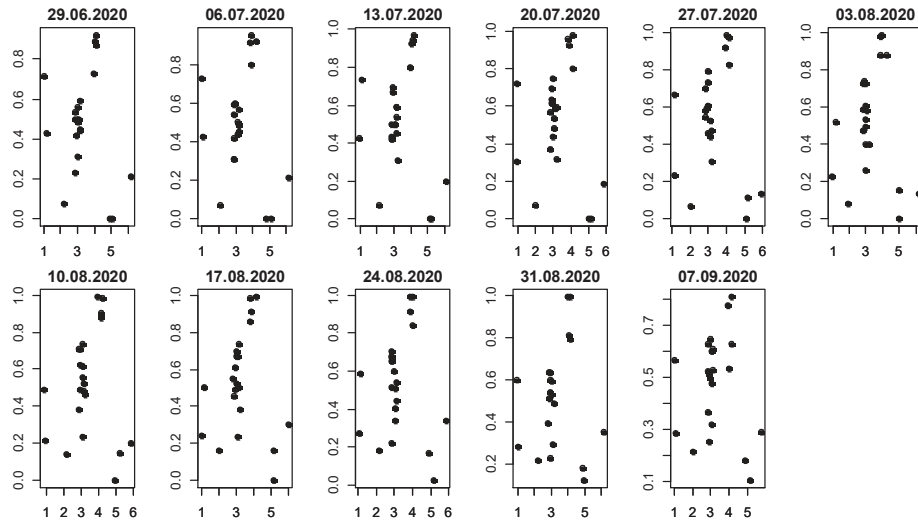
As shown in Figure 4, the share of infections in hospitals or clinics was highest in 2<sup>nd</sup> cluster (brzeski) and relatively high in the fifth one (olkuski and proszowicki).

Infections in nursing homes were reported rarely, at first only in poviát from the first group (bocheński and tarnowski – Figure 5), since the end of July also in the fifth group (proszowicki poviát). In September also three poviát from the fourth cluster (wadowicki, Nowy Sącz and nowosądecki) reported such cases.

As shown in Figure 6, *infections in a care and treatment facility (ZOL – Zakład Opiekuńczo-Leczniczy)* were reported only in second cluster which consists only of poviát brzeski.

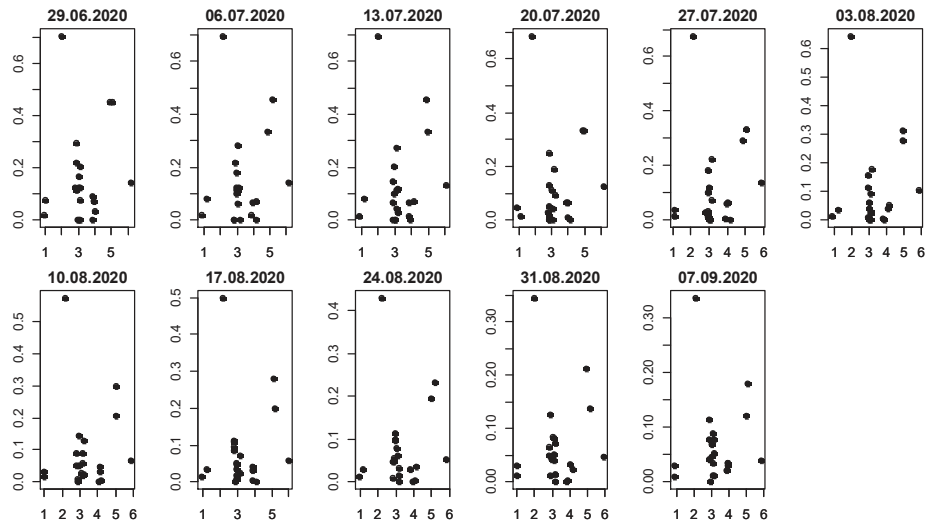
Throughout the period relatively highest share of infections coming from other sources (Figure 7) was reported in fifth group (olkuski and proszowicki) and the lowest – in fourth group (till September). This source was of growing relative importance in all the groups, while in groups 2 (brzeski) and sixth (Tarnów) this tendency was most visible. In the third group, the largest one, the share was also relatively high as compared with other clusters.

Figure 3. Comparison of groups of poviats with respect to shares of Covid-19 cases, source: quarantine after contact with a Covid-19 case



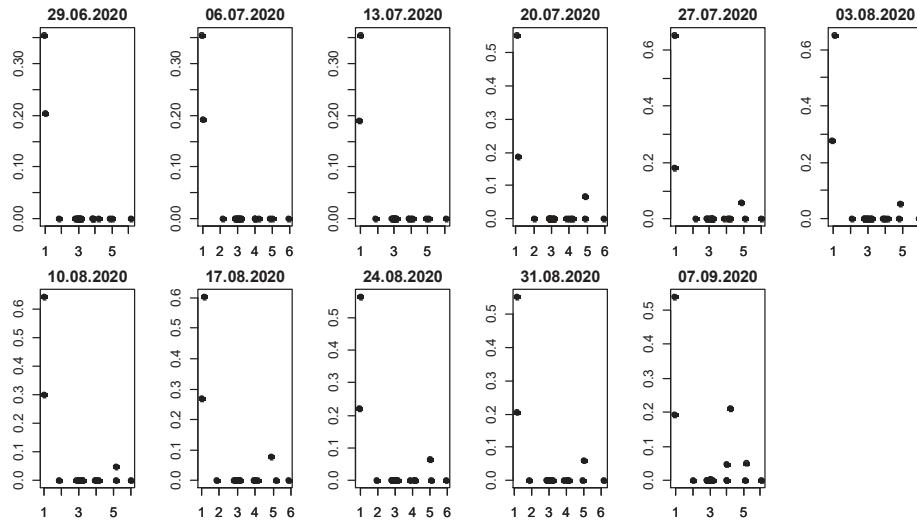
Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

Figure 4. Comparison of groups of poviats with respect to shares of Covid-19 cases, source: hospitals or clinics



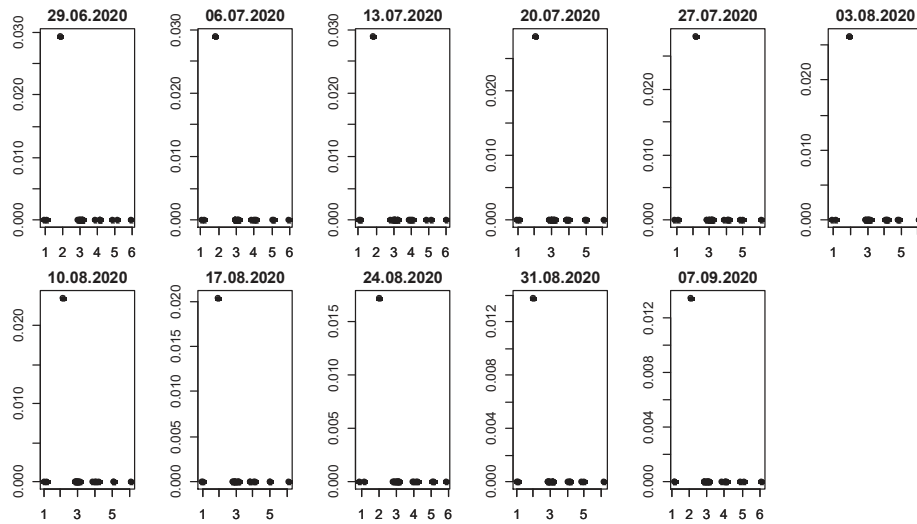
Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

Figure 5. Comparison of groups of poviats with respect to shares of Covid-19 cases, source: nursing homes (DPS)



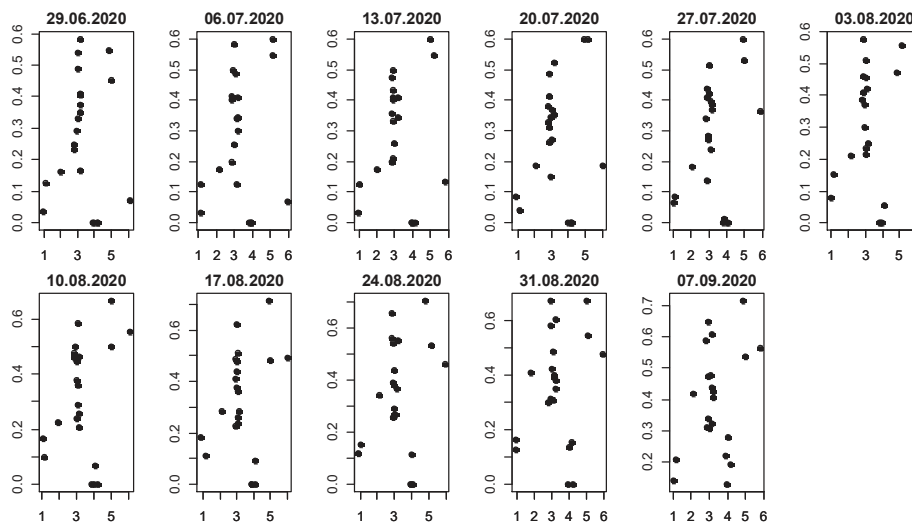
Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

Figure 6. Comparison of groups of poviats with respect to shares of Covid-19 cases, source: care and treatment facilities (ZOL)



Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

Figure 7. Comparison of groups of poviats with respect to shares of Covid-19 cases, source: other



Source: authors' own work based on provincial sanitary and epidemiological station (Wojewódzka Stacja Sanitarno-Epidemiologiczna – WSSE) in Kraków data

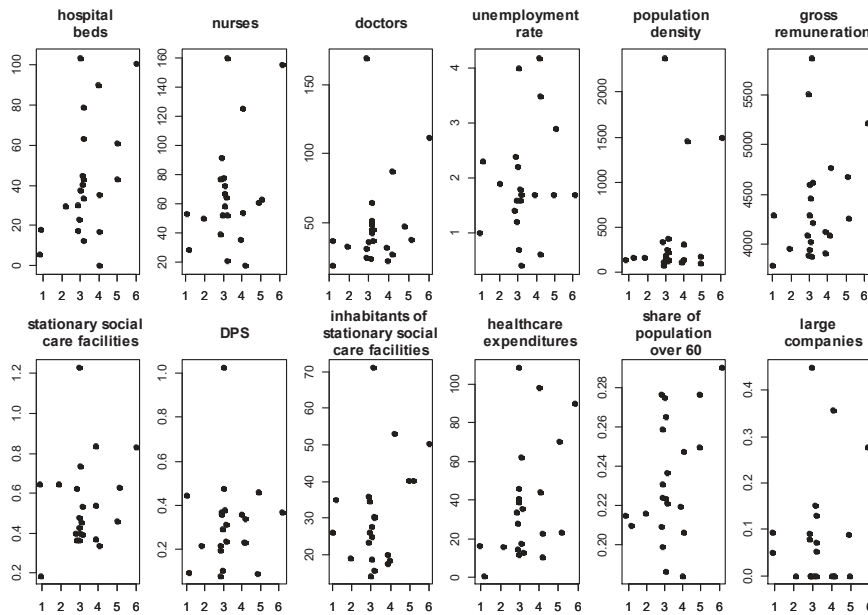
### Identified clusters – socio-economic characteristics

Socio-economic differences between poviats of Małopolskie voivodship are a background of Covid-19 pandemic and our analysis. In this part of the paper we present the differences and similarities between cluster identified in the previous step in terms of their socio-economic characteristics (Figure 8). As for the number of beds per 10,000 population clear differences may be seen. Values of this variable are relatively low in case of groups 1, 2 and most of the poviats classified into 4<sup>th</sup> cluster (all except for Nowy Sącz). Relatively (medium-) high values may be noticed in poviats olkuski and proszowicki (fifth group) and high in Tarnów (sixth cluster). In case of the unemployment rate, the differences are not so clear. With few exceptions most of the poviats are characterized by medium levels that are not differentiated clearly between clusters. Clusters are more differentiated in terms of gross remuneration. Relatively low values may be noticed in first two clusters and in 3 out of 4 poviats classified into the fourth one. In case of proszowicki and olkuski poviats (i.e. fifth group), values were on a relatively (medium-)high level. Similar statement can be made in case of Tarnów (sixth group). The spread in the third group is the highest which results from high values in krakowski and Kraków. Taking into account the number of stationary social care facilities it can be noticed that the only group which may be clearly different from the others is the sixth one, but it is made up of only one poviat. Low values of this ratio may be found in bocheński, but this poviat was classified together with tarnowski, characterized by medium values. Similarly with DPSs. Most poviats was characterized by similar values. As for the



inhabitants of these facilities, the situation is more clear. High values can be found (apart from sixth group – Tarnów) in the fifth cluster i.e. poviats olkuski and proszowicki. Healthcare expenditures are low in groups 1-2 and high in 6th, while share of population over 60 is relatively low in groups 1-2 and high in 5-6.

Figure 8. Comparison of groups of poviats with respect socio-economic characteristics



Source: authors' own work based BDL GUS data

## CONCLUSIONS

Social and economic environment is diverse in the Małopolskie Voivodeship. Despite the differences it is possible to find cluster characterized by similar structure of the sources of infection in summer 2020.

First cluster can be described by relatively low shares of infections after quarantine after returning from abroad. Apart from the fact that infections in nursing homes were reported rarely, they were reported by both poviats from this group. Second cluster consists only of one poviat, and therefore it's characteristics are more clear. Poviat Brzeski is characterized by the lowest share of cases from the group: quarantine *after contact with a Covid-19 case*. On the other hand the share of infections in hospitals or clinics was highest. It was also the only poviat which reported *infections in a care and treatment facility (ZOL – Zakład Opiekuńczo-Leczniczy)*. The share of infections coming from other sources was of clearly growing relative importance. Third cluster is the largest one and the poviats assigned to it are differentiated. However we may conclude, that in this group the share of

cases from the last group: *other* was relatively high as compared with other clusters. The fourth group was characterized by the relatively high share of infections from the group: quarantine *after contact with a Covid-19 case*. In September also three poviats from the fourth cluster (wadowicki, Nowy Sącz and nowosądecki) reported cases of infections in nursing homes. Till September relatively lowest share of infections coming from other sources was reported in fourth group. In the fifth cluster the share of infections from the quarantine *after returning from abroad*, was relatively high since the beginning of August. The share of infections in hospitals or clinics was also relatively high. This group is also characterized by the highest share of infections coming from other sources throughout the period. In case of the sixth cluster, the share of infections from the quarantine *after returning from abroad*, was high. The growing importance of other sources of infections was clearly visible.

Finally, it is worth emphasizing that the division into groups is diverse. On the one hand, there is a fourth group with 12 poviats, and on the other hand, there are two groups with one poviat. However, sometimes the similarities between the groups are greater than in one group of several poviats.

There is no linkage between Covid-19 and socio-economic characteristics for the created clusters.

Current study has some limitations. Firstly socioeconomic data come from 2019. Secondly, in August 2019 additional restrictions were introduced in some of the Małopolska poviats. Limited volume of the paper does not allow us to include those in the study, however, due to the weekly data intervals, we believe, that some of that impact is included, even if it is not directly addressed.

## REFERENCES

- Charrad M., Ghazzali N., Boiteau V., Niknafs A. (2014) NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6), 1-36. URL: <http://www.jstatsoft.org/v61/i06/>
- Dunn J. (1974) Well Separated Clusters and Optimal Fuzzy Partitions. *Journal Cybernetics*, 4(1), 95-104.
- Ioannidis J., Axfors C., Contopoulos-Ioannidis D. G. (2020) Population-level COVID-19 Mortality Risk for Non-elderly Individuals overall and for Non-elderly Individuals without Underlying Diseases in Pandemic Epicenters. *Environmental Research*, 188, 109890. <https://doi.org/10.1016/j.envres.2020.109890>
- Maechler M., Rousseeuw P., Struyf A., Hubert M., Hornik K. (2019) cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.
- R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Ward J. H. (1963) Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236-244.  
<https://wsse.krakow.pl/page/> (in Polish).  
<https://bdl.stat.gov.pl/BDL/start>