A STUDY OF THE PHENOMENON OF COVID-19 EPIDEMIOLOGICAL RISK

Tommaso Brancati, Francesco Dell'Orco, Stefania Fivizzani,
Veronica Iacono, Luana Sbardella

1. Introduction

The world population, between the end of the year 2019 and the beginning of 2020, has been affected by an epidemic, still in progress, known as "new coronavirus disease". Coronavirus are a large virus family giving having effects going to a common cold to harder diseases like Middle-East Respiratory Syndrome (MERS) or Severe Acute Respiratory Syndrome (SARS). The virus that causes the current coronavirus epidemic has been called "Severe Acute Respiratory Syndrome-CoronaVirus-2" (SARS-CoV-2), as told by the I*nternational Committee on Taxonomy of Viruses* (ICTV) that deals with the classification and the naming convention of the viruses in terms of species, genus, family etc. thanking to a group of experts in charge of studying the new coronavirus family. The disease caused by the new Coronavirus has been announced on 11th February of 2020 by the WHO General Director Tedros Adhanom Ghebreyesus, during a press briefing within a pause of the extraordinary forum dedicated to the virus[1].

The first cases have been found in Wuhan, in China, but the disease has been so quickly spread world-wide that WHO (World Health Organization) has declared that the coronavirus is a public health emergency of international concern and on 2020 11th of March has classified the emergency as pandemic.

COVID-19 symptoms are very similar to the flu: dry cough, breathing difficulty, weakness, fever. But in people with previous pathologies there have been found more severe symptoms like pneumonia and kidney failures often bringing to death.

Italy, hardly affected by the coronavirus, has fought the quick spreading of the epidemic by measures of containment, distancing and social isolation.

The purpose of this work is to build a composite index to analyze the COIVD-19 epidemiological risk. In particular it aims to describe the Sars-Cov-2 virus spreading and the related risk factors in Italy referred to a group of indicators measured prior to the spread of the virus and comparing the final result with an

[1] http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?id=228.

index representing the disease spread in the moment the infection has started to have a severe impact (on 03/26/2020, when the number of deaths from COVID-19 on Italian territory compared to the previous day, reached the absolute maximum).

The risk factors considered in this study to analyze the disease spread are air pollution, mobility, population density and people seniority, found in the provincial capital municipalities.

2. Synthesis of methodology

The selection of the elementary indicators, namely risk factors, has been dictated both by subjective considerations and by studies now in progress about the phenomenon being analyzed.

The data matrix contains the following elementary indicators.

- Rate of PM_{10} (particulate matter with a size less than or equal to 10 micrometers), which is the yearly average concentration of PM_{10} in $\mu g/m^3$ (it refers to year 2018) [2].
- Rate of NO_2 (nitrogen bioxide) , which is the yearly average concentration of NO_2 in $\mu g/m^3$ (it refers to year 2018)[3].
- Rate of passengers transported, which is the ratio between the number of passengers transported by local public transport in the provincial capital municipalities and the average resident population in the year (it refers to year 2017) [4].
- Vehicular density, which is the amount of vehicles per km^2 of urbanized area (it refers to year 2018). [5]
- Seniority rate, which is the ratio between the number of inhabitants over the age of 65 and the number of inhabitants over the age of 14, expressed as a percentage.
- Population density, which is the number of inhabitants per km^2.[6]

The selection of the elementary indicators comes from the following considerations.

- The areas the virus has shown the greatest impact have a high population density and a high productivity. In these lands there is the greatest number of high-growth and present in the international market companies.

[2] Data Source ISTAT.
[3] Data Source ISTAT.
[4] Data Source ISTAT.
[5] Data Source ISTAT.
[6] Data Source ISTAT.

- The air pollution is a risk factor for the lower respiratory tract infections, in particular in vulnerable subjects, the elderly and people with previous pathologies and these conditions are also those that have characterized the Covid-19 epidemic. Some preliminary studies made by some Harvard University researchers show that an increase of the Pm levels affects the complications of coronavirus disease. Since the study is at a preliminary phase, it must be taken with due caution and it will need to be deepened.
- Mobility is a sure risk factor of disease spreading. La mobilità è sicuramente un fattore di rischio di diffusione della malattia. Think of how much a large concentration of people in a means of public transport, for example, can affect the contagion and spread of the disease.
- Data from the Italian Istituto Superiore di Sanità regarding the Covid-19 epidemic in Italy show how the elderly are the most affected people by the disease. In particular we have observed an increased mortality in the ages between 70 and 90 years old, confirmed all over the world by the WHO.

Table 1 - *Table of elementary indicators.*

Provincial capitalities	Rate of passengers transported	Density of population	Vehicular density	Rate of NO_2	Rate of PM_{10}	Seniority index	Incidence rate
Alessandria	14,6	203,6	2786,7	31,0	37,0	208,3	0,23
Ancona	107,6	124,8	3631,4	17,0	25,0	216,0	0,183
Andria	11,4	402,9	6846,6	23,0	23,0	120,0	
Aosta	12,6	21,4	3821,0	28,0	18,0	221,4	0,322
Arezzo	40,6	384,7	2355,6	36,0	23,0	205,2	0,072
Ascoli Piceno	13,1	158,0	2865,3	15,0	20,0	262,7	0,053
Asti	36,5	151,3	2774,2	31,0	36,0	205,6	0,13
Avellino	52,5	30,6	3980,4	24,0	35,0	199,5	0,043
Bari	75,6	117,4	3980,3	37,0	27,0	198,9	0,031
Barletta	6,0	149,4	5968,5	19,0	22,0	132,9	0,022
Belluno	81,0	147,2	2310,2	27,0	22,0	234,2	0,152
Benevento	21,7	130,8	4074,8	16,0	10,0	188,8	0,005
Bergamo	153,2	40,2	4068,5	41,0	30,0	197,9	0,672
Biella	12,5	46,7	2429,9	22,0	25,0	264,2	0,188
Bologna	289,5	140,9	3785,9	49,0	26,0	211,0	0,124
Bolzano - Bozen	112,3	52,3	6894,1	40,0	20,0	167,4	0,173
Brescia	276,0	90,3	2947,6	57,0	33,0	189,8	0,549
Brindisi	18,8	333,0	4018,7	23,0	25,0	174,5	0,031
Cagliari	158,4	85,0	4876,9	30,0	36,0	283,3	0,015
Caltanissetta	5,2	421,3	3781,0	30,0	26,0	162,5	0,019
Campobasso	50,7	56,1	4718,6	34,0	17,0	215,4	0,038
Catania	54,0	182,9	5186,5	50,0	27,0	152,0	0,031

Provincial capitalities	Rate of passengers transported	Density of population	Vehicular density	Rate of NO_2	Rate of PM_{10}	Seniority index	Incidence rate
Catanzaro	38,9	112,7	2759,8	21,0	26,0	174,6	0,015
Chieti	50,1	59,6	2924,6	16,0	24,0	239,9	0,045
Como	88,8	37,1	2323,4	44,0	29,0	208,4	0,127
Cosenza	68,6	37,9	4569,2	26,0	23,0	196,2	0,015
Cremona	18,9	70,5	3467,4	33,0	34,0	227,7	0,938
Crotone	15,7	182,0	2403,7	25,0	30,0	123,8	0,046
Cuneo	40,0	119,7	2557,1	23,0	21,0	204,1	0,083
Enna	40,1	358,8	4034,4	3,0	15,0	217,4	0,079
Ferrara	63,9	405,2	2022,0	38,0	31,0	266,9	0,061
Firenze	246,3	102,3	4897,3	60,0	30,0	217,7	0,075
Foggia	23,6	509,3	6282,5	18,0	16,0	163,3	0,049
Forlì	53,2	228,2	2502,7	29,0	26,0	194,2	0,13
Frosinone	17,6	46,9	2065,8	41,0	41,0	189,9	0,032
Genova	233,9	240,3	5706,7	60,0	25,0	255,5	0,09
Gorizia	21,4	41,3	2168,2	23,0	18,0	249,4	0,054
Grosseto	12,6	473,6	3472,9	37,0	27,0	199,5	0,078
La Spezia	134,0	51,4	4010,5	42,0	22,0	226,4	0,098
L'Aquila	41,5	473,9	1934,9	17,0	17,0	190,4	0,068
Latina	7,8	277,6	2388,3	27,0	23,0	153,5	0,021
Lecce	14,0	241,0	2357,1	29,0	23,0	201,6	0,036
Lecco	51,9	45,1	3776,5	37,0	23,0	206,2	0,022
Livorno	56,4	104,5	3838,9	39,0	23,0	216,9	0,06
Lodi	14,7	41,4	3988,2	34,0	38,0	188,7	0,858
Lucca	20,9	185,8	1550,7	25,0	25,0	205,8	0,123
Macerata	35,2	92,5	3606,4	13,0	17,0	214,6	0,127
Mantova	98,6	63,8	3525,1	28,0	30,0	225,7	0,303
Massa	16,2	93,8	1864,1	15,0	14,0	217,6	0,17
Messina	46,5	213,8	4528,8	30,0	23,0	179,5	0,024
Milano	480,1	181,7	6521,3	59,0	35,0	178,0	0,035
Modena	83,7	183,2	3258,6	40,0	33,0	177,7	0,215
Monza	29,0	33,1	5158,5	37,0	33,0	196,2	0,239
Napoli	112,6	119,0	7120,7	57,0	35,0	139,1	0,201
Novara	79,0	103,1	2978,7	45,0	27,0	184,8	0,021
Nuoro	27,8	192,1	4233,7	24,0	19,0	207,2	0,156
Oristano	4,7	84,6	3401,1	24,0	22,0	264,6	0,033
Padova	126,9	93,0	2265,9	38,0	37,0	216,6	0,004
Palermo	42,3	160,6	6373,2	52,0	36,0	145,3	0,19
Parma	124,8	260,6	3355,0	36,0	32,0	171,9	0,016
Pavia	63,7	63,2	3425,7	35,0	35,0	235,3	0,359
Perugia	120,0	449,5	2350,7	25,0	25,0	183,0	0,308
Pesaro	24,1	126,8	4174,6	19,0	26,0	214,8	0,091

Provincial capitalities	Rate of passengers transported	Density of population	Vehicular density	Rate of NO_2	Rate of PM_{10}	Seniority index	Incidence rate
Pescara	76,8	34,4	3801,2	34,0	28,0	208,7	0,409
Piacenza	89,7	118,2	2801,9	37,0	33,0	187,3	0,138
Pisa	74,6	185,2	2786,2	32,0	26,0	229,5	0,772
Pistoia	44,6	236,2	2528,0	22,0	19,0	218,1	0,083
Pordenone	50,2	38,2	2363,4	27,0	23,0	200,1	0,09
Potenza	26,7	175,4	2507,9	6,0	19,0	198,1	0,095
Prato	38,7	97,4	3563,7	30,0	25,0	153,5	0,023
Ragusa	6,8	444,7	2957,9	12,0	30,0	171,9	0,074
Ravenna	46,1	653,8	3198,7	30,0	27,0	202,1	0,009
Reggio Calabria	32,4	239,0	2518,0	19,0	21,0	165,4	0,115
Reggio Emilia	85,0	230,7	2789,2	35,0	35,0	143,8	0,022
Rieti	40,0	206,5	2486,5	21,0	19,0	217,3	0,319
Rimini	98,8	135,7	3996,4	39,0	31,0	181,9	0,026
Roma	331,6	1287,4	4658,9	58,0	30,0	170,0	0,353
Rovigo	12,2	108,8	2216,0	30,0	32,0	228,0	0,036
Salerno	42,6	59,9	5109,3	18,0	31,0	210,1	0,046
Sassari	95,6	547,0	2664,7	30,0	25,0	202,4	0,022
Savona	69,1	65,3	4989,6	32,0	21,0	256,2	0,099
Siena	213,9	118,5	2683,3	36,0	18,0	243,0	0,075
Siracusa	4,0	207,8	3772,6	23,0	35,0	162,9	0,08
Sondrio	3,7	20,9	3756,3	24,0	23,0	224,8	0,019
Taranto	55,2	249,9	3194,3	29,0	28,0	187,4	0,179
Teramo	34,8	152,8	3773,6	21,0	24,0	195,6	0,085
Terni	41,1	212,4	2921,3	29,0	35,0	229,0	0,078
Torino	320,1	130,0	6469,2	56,0	39,0	213,5	0,136
Trapani	25,7	273,1	2754,9	26,0	19,0	176,8	0,011
Trento	180,7	157,9	3998,2	44,0	24,0	167,5	0,241
Treviso	111,3	55,6	2470,8	30,0	34,0	212,9	0,142
Trieste	327,7	85,1	4116,1	28,0	20,0	255,9	0,164
Udine	110,6	57,2	1766,7	24,0	21,0	223,8	0,087
Varese	71,7	54,8	2401,0	36,0	24,0	222,5	0,056
Venezia	822,6	415,9	1913,1	51,0	37,0	244,2	0,108
Verbania	45,7	37,5	2700,2	28,0	15,0	244,5	0,16
Vercelli	11,7	79,8	3557,9	26,0	30,0	218,3	0,193
Verona	188,7	198,9	3822,9	26,0	31,0	203,5	0,152
Vibo Valentia	8,0	46,6	2326,3	21,0	23,0	157,5	0,017
Vicenza	79,7	80,6	3090,1	34,0	34,0	195,0	0,104
Viterbo	21,8	406,2	4608,0	23,0	18,0	183,1	0,038

There have not been considered the cities of Trani, Isernia, Cesena, Carbonia, Agrigento and Fermo because they are not included in each data source, Imperia because the rate of NO_2 is missing, Matera because the rate of PM_{10} is missing.Only for the city of Benevento the rate of NO_2 has been reported as the maximum among the yearly averages of the control units at the URL http://www.arpacampania.it/web/guest/dati-validati-annuali.

To deal with the data matrix it has been used COMIC (COMposite Indices Creator), a program made at Italian Statistical Institute (Istat), within the BES (Benessere Equo e Sostenibile) group.

This software calculates composite indicators choosing among various several synthesis methods, build rankings and output reports with the main achieved results.

After the definition of the starting theoretical framework and the selection of the elementary indicators, the procedure to create a composite index involves the normalization of the elementary indicators to make them comparable (they are often expressed in different units of measure and may have different polarities, so in the case of this phenomenon, each indicator has a negative polarity and the phenomenon is negative itself, so the polarity of the phenomenon matches the polarity of the indicators), the aggregation of the normalized indices with the assignment of a weight (in the case under study, weight 1 was given to all the indicators), the choice of the synthesis technique and the validation of the composite index, i.e. the verification of consistency with the theoretical framework of reference.

In this program there have been implemented functions that allow to perform the operations described above.

In particular to create, analyze and validate the composite index we are dealing with there have been done

- an exploratory data analysis,
- the rankings construction of statistical units,
- the influence analysis.

The synthesis method used is the Mazziotta-Pareto Index (MPI).

MPI purpose is to give a synthetic measure of a multidimensional phenomenon, in the hypothesis each component is not replaceable with the other ones. This takes to the insertion of a 'penalty' for the units not having balanced values. It is a composite index that guarantees high robustness against each other existing indexes because it is invariant when input data vary.

In order to implement MPI the following steps are performed.

1) Standardization:

starting from X matrix having its elements x_{ij} we calculate the normalized values, having i as its row index, j as its column index and n its number of statistical units

$$z_{ij} = \frac{x_{ij} - M_{x_j}}{S_{x_j}} S + M \quad (1)$$

where

$$S=10,\ M=100,\ M_{x_j} = \frac{\sum_{i=1}^{n} x_{ij}}{n},\ S_{x_j} = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - M_{x_j})^2}{n}}$$

2) Aggregation.

Starting from the normalized values matrix Z, called m the number of elementary indicators, we can define the vector

$$cv_i = \frac{S_{z_i}}{M_{z_i}} \quad (2)$$

where $M_{z_i} = \frac{\sum_{j=1}^{m} z_{ij}}{m}$, $S_{z_i} = \sqrt{\frac{\sum_{j=1}^{m} \left(z_{ij} - M_{z_i}\right)^2}{m}}$

3) Calculation of the composite index.

$$MPI_i^{+/-} = M_{z_i}(1 \pm cv_i^2) = M_{z_i} \pm S_{z_i} cv_i \quad (3)$$

where + or – sign depends on the polarity of the indicator. If it is positive, the indicator growth is linked to a positive perception and the other way round.

3. Results

The correlation matrix does not show significant correlations between the indicators analyzed apart from the NO_2 rate which has a slight correlation with the PM_{10} rate (r=0.500) and the rate of transported passengers (r=0.586). Please see how the seniority index is negatively correlated with each other analyzed indicators apart from the rate of transported passengers.

The influence analysis describes the indicators that have the greatest influence on the composition of the rankings of the provincial capitals. From Table 2, the macroareas that “weigh” most in the epidemiological risk phenomenon are the

vehichular density (Root Mean Squared Error 10.551), PM_{10}rate (RMSE 10.181) and the density of population (RMSE 8.920).

Table 2 - *Matrix of correlation among the indicators.*

Base indicator removed	Mean	Standard deviation	C.V.
Rate of transported passengers	4.515	4.930	1.092
Density of population	7.980	8.920	1.118
Vehicular density	9.248	10.551	1.141
NO2 rate	7.673	7.631	0.995
PM10 rate	10.277	10.181	0.991
Seniority index	8.832	7.893	0.894
Mean	8.087	8.351	
Standard deviation	1.809	1.868	

In Table 3 the value of the composite indicator in the Italian provincial capitals is shown, followed by a map representation of the index. In particular the best 10, the worst 10 in this special ranking and some relevant provincial capitals are shown.

Table 3 - *Result IC.*

Province Capitals	MPI	Province Capitals	MPI
Vibo Valentia	108.492	Catania	96.808
Massa	108.112	Bologna	93.39
Potenza	107.389	Brescia	92.735
Benevento	106.83	Palermo	92.39
Reggio di Calabria	106.249	Cagliari	90.398
Crotone	105.622	Napoli	89.91
Macerata	105.261	Firenze	89.778
Latina	105.246	Genova	86.869
Catanzaro	105.165	Torino	84.492
Pordenone	104.84	Milano	82.628
Bari	99.176	Roma	74.574
		Venezia	74.391

4. Conclusion

The spread of Covid-19 does not seem to be linked to the risk factors taken in count in the analysis because the epidemiological risk does not allow for predictions about the spread of the disease. From the analysis through MPIs it has been found that the epidemiological risk is not an exhaustive dimension to explain the dynamics of the contagion spread across Italy. From the graph in figure 1 it can be seen that there is an uneven concentration of points that does not follow a linear trend, further supporting the demonstration that there is no correlation between the composite index and the incidence rate.

Figure 1 – *Scatter Plot of MPI vs. Covid Incidence Rate.*

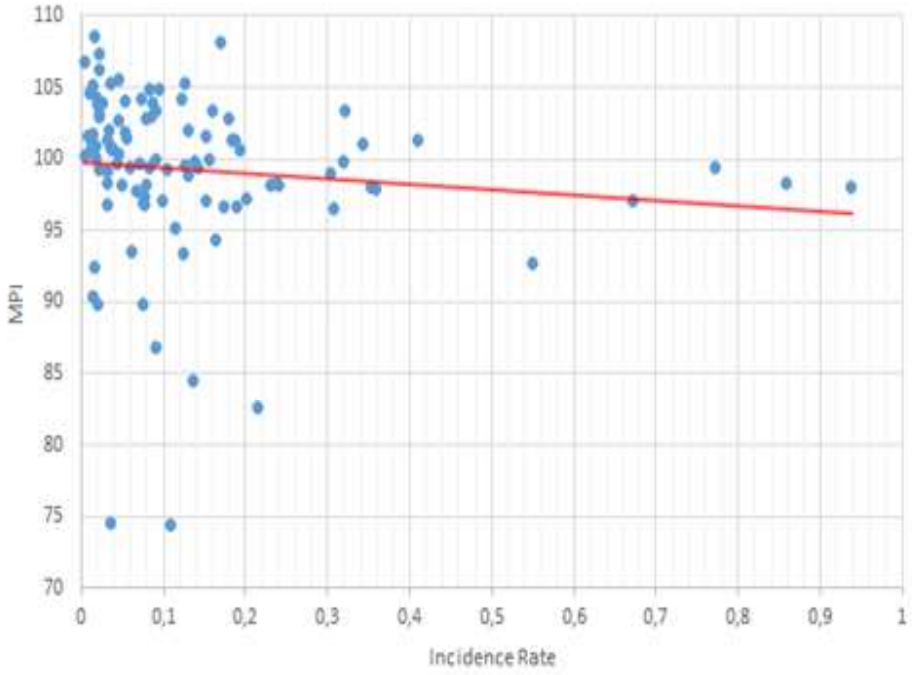
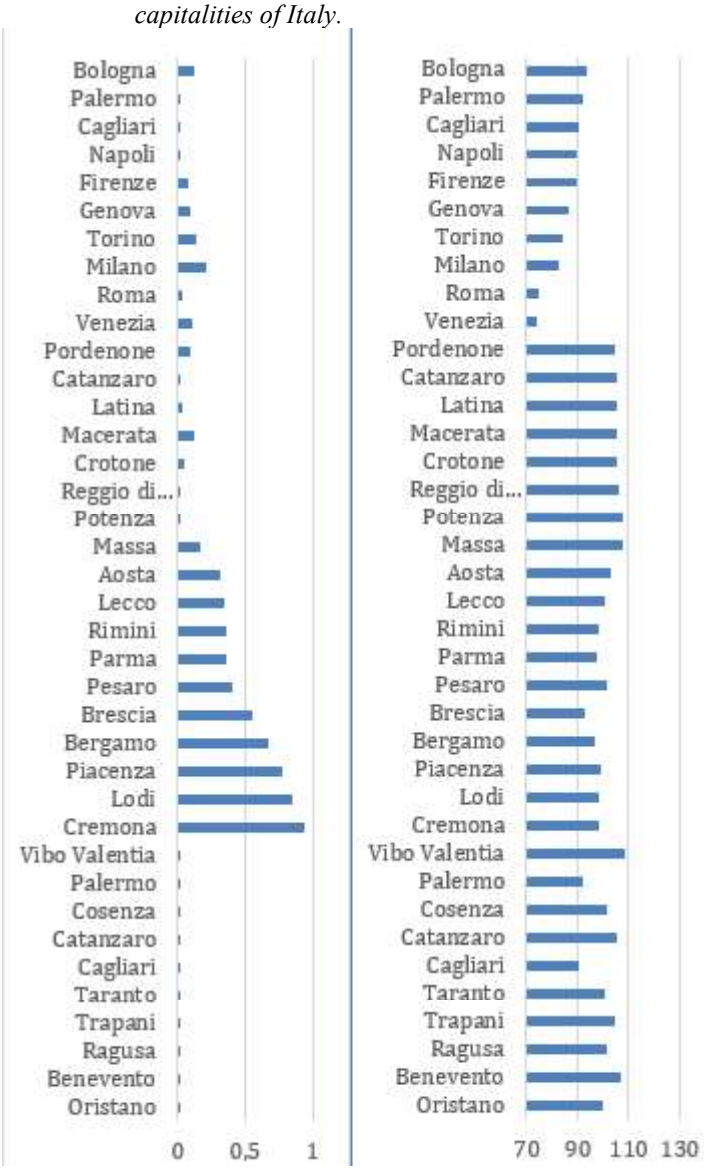


Figure 2 – *Incidence rate (on the left) and MPI calculated (on the right) in province capitalities of Italy.*



According to the graphics in figure 2 some of the provincial capitals with a higher epidemiological risk have a relatively low incidence rate from Covid (Benevento,

Oristano, Potenza and Vibo Valentia). There are also cities such as Bergamo, Brescia and Cremona with a high incidence rate and a low epidemiological risk.
A definitive proof of the lack of correlation between the composite index created and the incidence rate, is provided by the calculation of the Pearson index whose value, equal to -0.11, demonstrates the total lack of correlation.

The lack of correlation makes the spread of the virus escape from the intuitive logic represented by the indicators we have selected, considering the Covid incidence rate related to the period it has been taken in. The global crisis coming from the pandemic also rises up from the lack of understanding the factors that most influence the contagion and, in general, the context in which the virus takes root.

Aknowledgments

We thank Professor Matteo Mazziotta for his availability for continuous comparison and for his contribution in the validation of some important aspects of the paper: without his indispensable contribution it would not have been possible to publish this article.

References

ALAIMO L., CELARDO L., MAGGINO F., MORRONE A., OLIVIERI F. .2019., Monitoring SDGs at territorial level: the case of Lombardy. *Book of short Papers SIS 2019*

CHE COS'È IL NUOVO CORONAVIRUS, http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?id=228.

CIENCEWICKI J., JASPERS I. .2008. Air Pollution and Respiratory Viral Infection. *Inhalation Toxicology*, available at https://bit.ly/3hIQ7DM

CORONAVIRUS DISEASE (COVID-19) DATA, https://www.who.int/data/gho

GOSCÉ L., ANDERS J. 2018. Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground, available at https://ehjournal.biomedcentral.com/articles/10.1186/s12940-018-0427-5

ISTAT, http://dati.istat.it. Risk levels & key metrics, https://covidactnow.org/covid-risk-levels-metrics.

MAZZIOTTA M., PARETO A. .2012. Indici sintetici per confronti Spazio-Temporali: Un'applicazione alla dotazione infrastrutturale. Report presented at the XXXIII Italian Conference of Regional Sciences.

OECD. 2008. Handbook on Constructing Composite Indicators. Methodology and user guide. OECD Publications, Paris.
SETTI L., PASSARINI F., *et al.*, Relazione circa l'effetto dell'inquinamento da particolato atmosferico e la diffusione di virus nella popolazione. Università di Bologna, Università di Bari, Società Italiana Medicina Ambientale, available at https://bit.ly/2N8Ndu9.

SUMMARY

A study of the phenomenon of covid-19 epidemiological risk

Epidemiology is a word coming from the greek ἐπί, «over», δῆμος, «people», and λόγος, «speech, study» therefore it literally means "study of everything happens over the people". According to the explanation of the World Health Organization, "epidemiology is the study of the distribution and determinants of health-related states and events, and the e the application of this study to the control of diseases and other health problems[7]". Primary goals of epidemiology are:

- Limiting the disease spread within a community;
- Studying the disease etiology and the contagion modes.

The purpose of this paper is to proof that some of the facts the common man thinks to be the most common epidemiological risk factors, like air pollution, mobility, population density and seniority, observed within the italian province capital cities, are not related to the covid 19 spread. To reach this goal a composite index has been realized and it has been compared with a covid incidence rate, proofing that a correlation between these two indexes does not exist. The composite index has been realized using the PM10 rate, the NO_2 rate, the carried passengers rate, the vehicular density, the seniority rate, the population density. The synthesis method used for the creation of the composite index has been the Mazziotta-Pareto Index (MPI).

Tommaso BRANCATI, Author, brancatit@gmail.com
Francesco DELL'ORCO, Author, francesco.dellorco@gmail.com
Stefania FIVIZZANI, Author, stefania.fivizzani@gmail.com
Veronica IACONO, Author, iacono.veronica@gmail.com
Luana SBARDELLA, Author, luana.sbardella@gmail.com

[7] http://www.who.int/topics/epidemiology/en/.