A topological hypothesis behind the geographic differences in the Italian 'Covid19' spreading (25 May 2020)

Alessandro Barazzetti[1]

Federico Cecconi[2]

Marco Somaschini[3]

Introduction

The dynamics of infectious diseases represent one of the most ancient and rich areas of study of biomathematics; pathogens adapt and evolve over time, thus determining the appearance of new diseases or the regeneration of ancient forms of contagion which they thought were now extinct. It is clear that the human invasion of new ecosystems, factors such as global warming, environmental degradation, enhancement shifts and communications on an international scale, and changes in the economic field, have begun to provide more and more fertile ground for the development of new diseases.

The ongoing COVID-19 pandemic was first confirmed to have spread to Italy on 31 January 2020, when two Chinese tourists in Rome tested positive for the virus. One week later an Italian man repatriated back to Italy from the city of Wuhan, China, was hospitalised and confirmed as the third case in Italy. (Remuzzi & Remuzzi, 2020) (Calafiore, Novara, & Possieri, 2020) (Falco, Cioppa, Scafuri, & Tarantino, 2020) (Ren, et al., 2020) (Team, 2020)

A cluster of cases was later detected, starting with 16 confirmed cases in Lombardy on 21 February, and 60 additional cases and the first deaths on 22 February. By the beginning of March, the virus had spread to all regions of Italy. On 31 January, the Italian government suspended all flights to and from China and declared a state of emergency. In February, eleven municipalities in northern Italy were identified as the centres of the two main Italian clusters and placed under quarantine. The majority of positive cases in other regions traced back to these two clusters.

On 8 March 2020, Italian Prime Minister expanded the quarantine to all of Lombardy and 14 other northern provinces, and on the following day to all of Italy, placing more than 60 million people in quarantine. On 11 March 2020, Conte prohibited nearly all commercial activity except for supermarkets and pharmacies. On 21 March, the Italian government closed all non-essential businesses and industries, and restricted movement of people. (Galeazzi, et al., 2020) (Gollust, Nagler, & Fowler, 2020) (Gupta, Raghuwanshi, & Chanda, 2020) (Guzzi, Tradigo, & Veltri, 2020)

It has been hypothesized that in order to fully understand this complex phenomenon, the contribution of different and multiple disciplines was necessary: here mathematical models were extremely suitable for this purpose and began to play an incredibly significant role. The main goal of these studies resides in the attempt

---

[1] Head of R&D, QBT sagl – Professor, UCM Higher Education Institution – Dir. ScienceAdventure Sagl

[2] CSO, QBT Sagl - Professor, UCM Higher Education Institution - Head of R&D ScienceAdventure Sagl

[3] R&D, QBT Sagl

to develop models that can help evaluate the possible effects on populations, following the spread of a particular microorganism, and can consequently provide the necessary tools to control and predict the spread of infectious diseases. These analyzes they also allow us to understand which are the best treatment strategies and make important choices regarding immunization programs (vaccinations, therapies, prevention, control programs) and the distribution of monetary and technical resources.

A mathematical treatment is therefore appropriate and necessary, if the dynamics of these events require a quantitative analysis of the phenomenon, especially when, as in these cases, the agents causing the epidemics can be of the most varied types.

A crucial problem in defining such dynamics models of infectious diseases is the mathematical representation of the transmission mechanism of the infection. There is talk of an epidemic when the recurrence of the disease reaches times and proportions that are beyond normal expectations, up to exceeding a threshold point, or a critical value indicating the passage from a contagion situation restricted, to a large-scale diffusion. (Beretta & Takeuchi, 1995) (Gray, Greenhalgh, Hu, Mao, & Pan, 2011) (Xiao & Ruan, 2007)

Many network phenomena are well modeled as spreads of epidemics through a network. Prominent examples include the spread of worms and email viruses, and, more generally, faults. The spread of information can also often be modeled as the spread of anepidemic. An epidemic spreads along the underlying network topology from an initial set of infected nodes tosusceptible nodes.

In many cases infected nodes can be cured, reverting back to being susceptible and possibly being reinfected. It is imperative, then to understand how the topology affects the spread of an epidemic, in other words how the topology either impedes or facilitates its spread and maintenance.

(Arino & Van den Driessche, 2003) (Colizza & Vespignani, 2008) (Iwami, Takeuchi, & Liu, 2007) (Zhou, Fu, & Wang, 2005)

‘Incidence’ data

It is quite difficult to understand the impact of an epidemic: it is possible to measure the number of infected cases obtained through a diagnostic test. Or observe how many of the people who died in a certain period are positive for the same type of test. Both methods are sensitive to measurement errors, in particular thinking of a virus that has a high lethality mainly among the elderly. It becomes complex to understand what to attribute the reason for the death, besides the two dynamics as well as being very noisy in the measure, flow over different time horizons.

One solution, which is the one used in this work, is to compare the number of deaths in two different periods, the first when you imagine that the epidemic did not run (for example the first three months of 2019) and those during a period in which we assume that the epidemic is taking place (the first three months of 2020). It is a question of comparing the number of deaths without specifying what they depend on. If we observe a substantial difference in the two measures, then we are probably observing an epidemic in action

We have observed a distinctive geographical distribution by measuring the difference between the number of deaths in the first three months of this year compared to the same period in 2019. We call ‘***incidence’*** the difference in the number of deaths divided by the number of inhabitants for each municipality involved in the survey.

The graph shows a distribution with some outliers (see Figure 1)

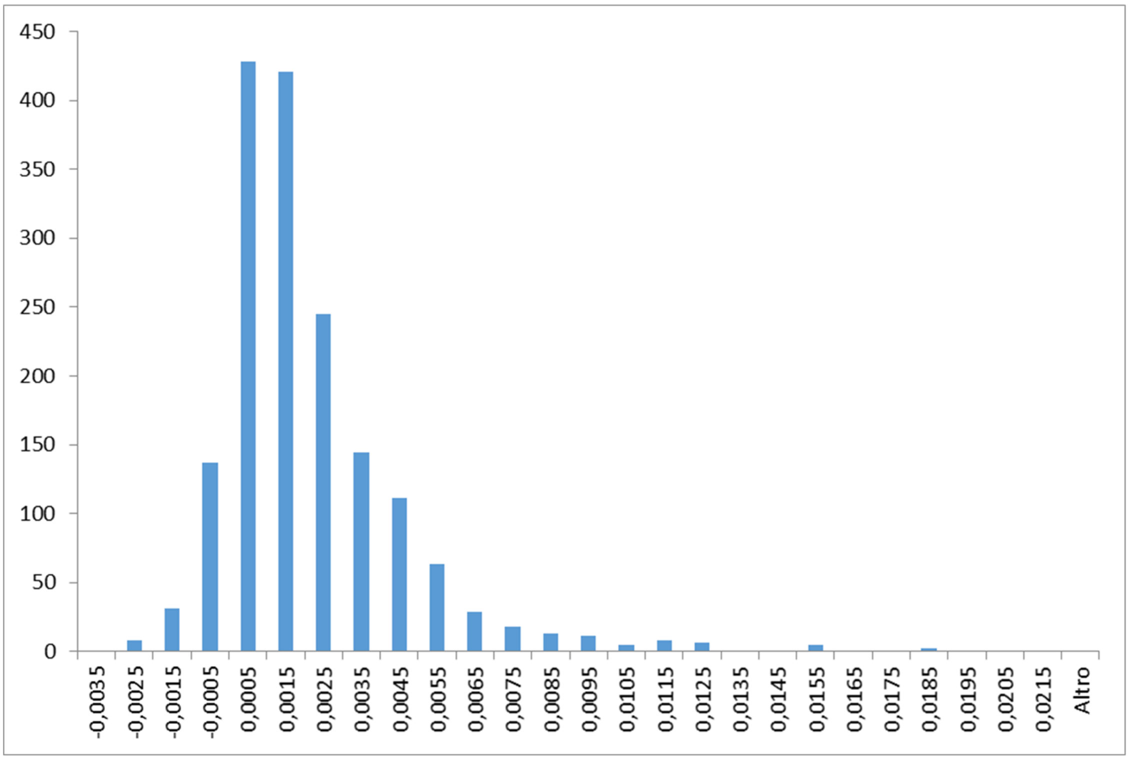

Figure 1. The distribution of incidence for municipality (tot. 1688)

We have associated municipality to the region to which it belongs. The result is really difficult to understand, albeit confirmed by the measures regarding the spread of the pandemic: the ***incidence*** is much higher in some regions, in particular in the northern regions (see Figure 2).

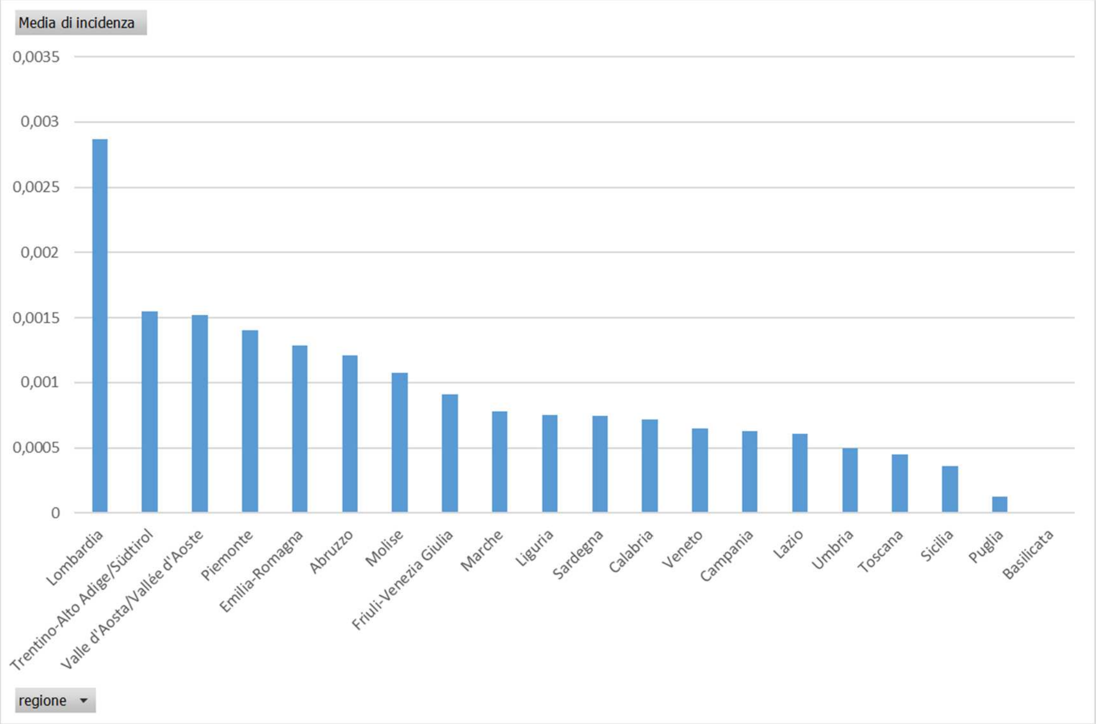

Figure 2. The distribution of the average of incidence for regions

We ask ourselves, why have many more people died in proportion to the population density in Lombardy than in Lazio, or in Campania? We have already shown with simulation models based on agents how some socio-cultural characteristics of northern Italy have been able to trigger dynamics favorable to contagion. (Cecconi & Barazzetti, 2020) In this work we show how topological factors exist, which together with the dynamics related to commuting; it can explain the distribution of the incidence.

Municipality network

We calculated the distance between all the municipalities involved in the survey, obtaining a distance matrix. Two municipalities are to be considered connected to each other if the distance is less than a threshold, θ=0.428 (about 3 hours by car). The resulting network has the following topological properties

	Value	***Notes***
Number of nodes	1,688	*The municipality*
Number of links	48,370	*The links are undirected*
Average degree	0.27	
Clustering coeff.	0.96	*The probability that two of my neighbors are close to each other*
Diameter	21	

Table 1. The topological properties of municipality network

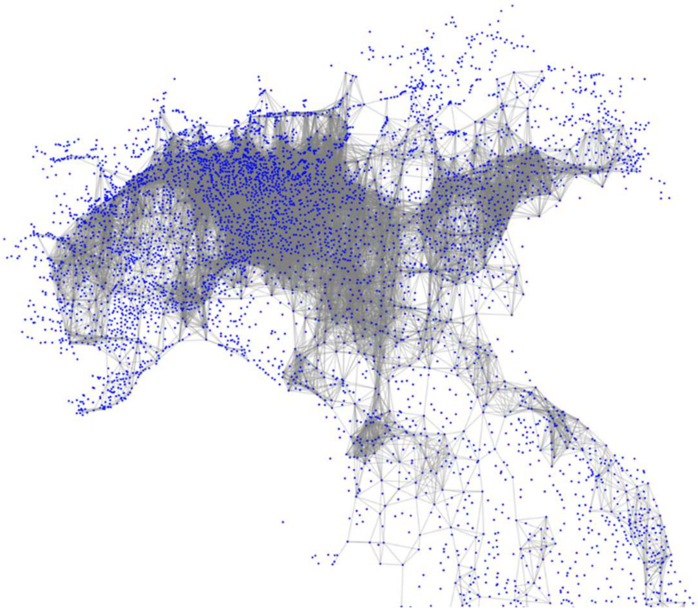

Figure 3. The municipality network

Mobility index

On a network made up of municipalities such as the one just described, it is possible to calculate a mobility index using two factors: (1) topological properties of the network and (2) data from Istat concerning the amount of work-related commuting, internal and outside the region.

In particular, the ***M*** mobility index for each municipality *i* is obtained **summing two separate indices:**

(a) M_i^1, or the average of the number of inhabitants of the municipalities connected to and (b) M_i^2 , for each of them the average the number of inhabitants of their neighbors.

This is obviously a recursive process, which requires using a fixed number of recursion levels (in our case 2) and weighing the coefficients differently (in our case the weight for the second level is 0.1). In details, for each municipality i we have latitude and longitude, and we compute the distance between municipality *i* and *j* by

$$D_{i,j} = \left|Lat_i - Lat_j\right| + \left|Long_i - Long_j\right|$$

and we build a network, *G*. We build *G* putting a link between i and j if $D_{i,j} < \theta$.

M_i^1 is computed as $M_i^1 = \langle w_{i,J} \rangle$ where *J* is the set of the network neighbors of *i* on G and *w* the number of inhabitants for each one. The same computation the same calculation occurs on the neighbors of the first neighbors of *i* (i.e. on the set *J*).

This result is (after multiply by scaling factor 0.1), M_i^2.

Results and conclusions

We have calculated the average of the mobility index ***M***, for each of the regions involved in the survey. We then used the data on mobility linked to commuting for each region, thus obtaining a second index, a mobility index enriched with data on commuting movement. In Figure 4 and Figure 5 we show the correlation between the two indices and the incidence for each region.

The data clearly show how a correlation exists between these two indices and the incidence, and that therefore it can be said that one of the factors that caused such a marked difference between the north and south in the impact of the pandemic is to be referred to socio-economic structures.

In other words, what we find is that there is a correlation between a purely 'structural' property of the environment in which the virus moves, and the presence of a greater number of deaths.

So, since the greatest number of deaths is logically associated with the spread of the epidemic, we can associate a structural property such as the number of municipalities and how they are topologically linked to each other to the spread. This does not deny that there may be many other factors that justify the greater number of deaths in the northern regions.

But surely the topology of the territory, understood as the number of municipalities and their connectivity, plays a role.

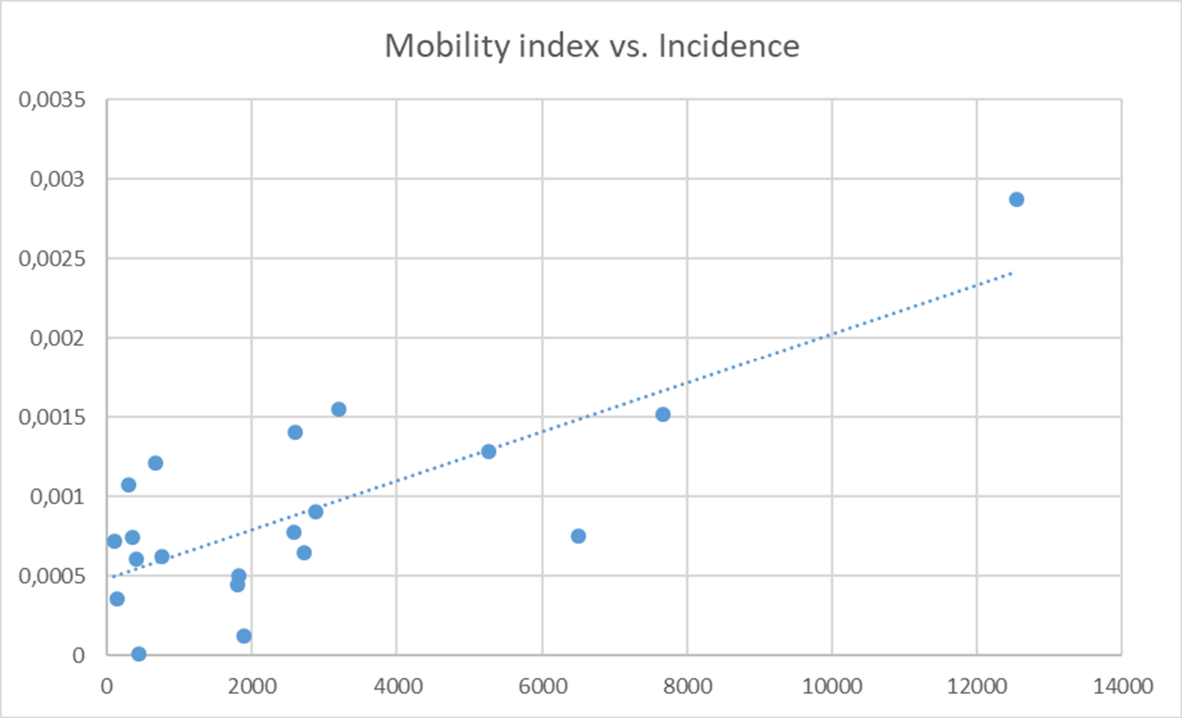

Figure 4. The correlation between mobility index and incidence

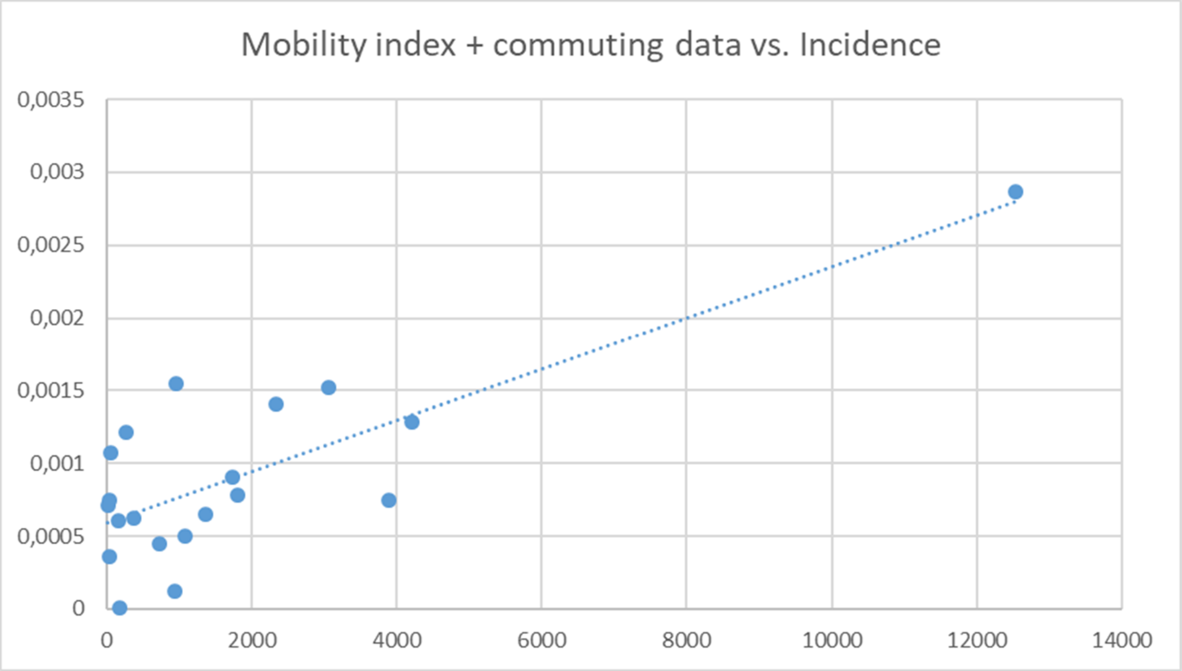

Figure 5. The correlation beetween mobility index weigthed by commuting data and incidence

References

Arino, J., & Van den Driessche, P. (2003). A multi-city epidemic model. *Mathematical Population Studies, 10*, 175-193.

Beretta, E., & Takeuchi, Y. (1995). Global stability of an SIR epidemic model with time delays. *Journal of mathematical biology, 33*, 250-260.

Calafiore, G. C., Novara, C., & Possieri, C. (2020). A Modified SIR Model for the COVID-19 Contagion in Italy. *CoRR, abs/2003.14391*. Retrieved from https://arxiv.org/abs/2003.14391

Cecconi, F., & Barazzetti, A. (2020). Agent-based simulation model applied to social behaviors determining the dynamics ofpandemics. *academia.edu*.

Colizza, V., & Vespignani, A. (2008, 2 25). Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett. 99, 148701 (2007)*. doi:10.1103/PhysRevLett.99.148701

Falco, I. D., Cioppa, A. D., Scafuri, U., & Tarantino, E. (2020). Coronavirus Covid-19 spreading in Italy: optimizing an epidemiological model with dynamic social distancing through Differential Evolution. *CoRR, abs/2004.00553*. Retrieved from https://arxiv.org/abs/2004.00553

Galeazzi, A., Cinelli, M., Bonaccorsi, G., Pierri, F., Schmidt, A. L., Scala, A., . . . Quattrociocchi, W. (2020). Human Mobility in Response to COVID-19 in France, Italy and UK. *CoRR, abs/2005.06341*. Retrieved from https://arxiv.org/abs/2005.06341

Gollust, S. E., Nagler, R. H., & Fowler, E. F. (2020, 5). The Emergence of COVID-19 in the U.S.: A Public Health and Political Communication Crisis. *Journal of health politics, policy and law*. doi:10.1215/03616878-8641506

Gray, A., Greenhalgh, D., Hu, L., Mao, X., & Pan, J. (2011). A stochastic differential equation SIS epidemic model. *SIAM Journal on Applied Mathematics, 71*, 876-902.

Gupta, S., Raghuwanshi, G. S., & Chanda, A. (2020, 4). Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020. *The Science of the total environment, 728*, 138860. doi:10.1016/j.scitotenv.2020.138860

Guzzi, P. H., Tradigo, G., & Veltri, P. (2020, 5). Spatio-Temporal Resource Mapping for Intensive Care Units at Regional Level for COVID-19 Emergency in Italy. *International journal of environmental research and public health, 17*(10). doi:10.3390/ijerph17103344

Iwami, S., Takeuchi, Y., & Liu, X. (2007). Avian--human influenza epidemic model. *Mathematical biosciences, 207*, 1-25.

Remuzzi, A., & Remuzzi, G. (2020). covid-19 and italy: what next? *The Lancet*.

Ren, H., Zhao, L., Zhang, A., Song, L., Liao, Y., Lu, W., & Cui, C. (2020, 4). Early forecasting of the potential risk zones of COVID-19 in China's megacities. *The Science of the total environment, 729*, 138995. doi:10.1016/j.scitotenv.2020.138995

Team, C. D.-1. (2020, 4). Geographic Differences in COVID-19 Cases, Deaths, and Incidence - United States, February 12-April 7, 2020. *MMWR. Morbidity and mortality weekly report, 69*(15), 465-471. doi:10.15585/mmwr.mm6915e4

Xiao, D., & Ruan, S. (2007). Global analysis of an epidemic model with nonmonotone incidence rate. *Mathematical biosciences, 208*, 419-429.

Zhou, T., Fu, Z.-Q., & Wang, B.-H. (2005, 8 15). Epidemic dynamics on complex networks. *Progress in Natural Science, 16(5): 452-457 (2006)*.

Declarations

Ethics approval and consent to participate
“Not applicable”

Consent for publication
“Not applicable”

Availability of data and materials
The data that support the findings of this study are available from ISTAT www.istat.it and Protezione Civile Italiana www.salute.gov.it
The datasets generated during and/or analysed during the current study are available in the geodiffusion_paper_32 repository, can be requested sending an email to the authors.

Competing interests
"The authors declare that they have no competing interests"

Funding
No funding for the study

Authors' contributions
AB formulated the hypotheses of the study and verified the correctness of the procedure.
FC developed the analysis model and selected the data sources.
MS created the data dataset used for the study and applied the model to the data.

Acknowledgements
"Not applicable"