



Unexpected positive correlation between human development index and risk of infections and deaths of COVID-19 in Italy

Kai Liu^a, Mu He^b, Zian Zhuang^c, Daihai He^{c,*}, Huaichen Li^d

^a Department of Statistics and Mathematics, Institute of Data Science and Interdisciplinary Studies, Shanghai Lixin University of Accounting and Finance, Shanghai, China

^b Department of Mathematics and Statistics, McMaster University, Canada

^c Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR, China

^d Department of Respiratory and Critical Care Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong 250021, China

ABSTRACT

In this analysis, we observed that human development index (an integrated index of life expectation, education and living standard) correlates with infection rate (proportion of confirmed cases among the population) and the fatality rate of COVID-19 in Italy based on data as of May 15, 2020. Further analysis showed that HDI is negatively correlated with cigarette consumption, whereas it is positively correlated with chronic disease and average annual gross salary. These factors may partially explain why unexpected positive correlation is observed between human development index and risk of infections and deaths of COVID-19 in Italy.

To the editor,

The coronavirus disease 2019 (COVID-19) broke out quickly in Italy since March 2020 when the epidemic got controlled in China. Reasons of rapid breakout and overall case-fatality rate in Italy have been studied and reported in literature [1–3]. Obvious differences in epidemic spread and fatality rates among regions exist, but factors related to these spatial differences are unclear. It is of interest to study this regional heterogeneity and the related factors.

Global data of COVID-19 have been integrated by researchers and available publicly from R package nCov2019 [4]. We downloaded and extracted the data of Italy by regions for our study. As of May 15, 2020, Lombardy ranked top 1 with 83,820 cumulative confirmed cases among the 20 regions, while the number of cumulative confirmed cases in Basilicata was the smallest (389 cases). The number of death ranged from 22 to 15,296, corresponding to regions of Molise and Lombardy, respectively. Demographical data including population, area, population density and human development index (HDI) by regions of Italy 2019 were downloaded from https://en.wikipedia.org/wiki/Regions_of_Italy. The regional infection rates (the proportion of confirmed cases among regional population) ranged from 0.0006 to 0.009 with a median of 0.0025, while the regional death rates (proportion of deaths among regional population) ranged from 0.00005 to 0.00152 with a median of 0.00026. HDI [5] is an integrated index of healthy long life, education and living standard, measured by life expectancy, expected/mean years of schooling, gross national income per capita, respectively.

The median HDI is 0.891 with a range from 0.845 to 0.919. Fig. 1 presents the log odds of infection rates and death rates as of May 15, 2020 against HDI by regions. It showed a linear pattern between the log odds and HDI. To quantify the association between infection rates and death rates with HDI, we performed a univariate logistic regression.

It is reasonable to assume people in the same region are independent and identical with the same probability of being infected and diagnosed as a confirmed case. Under this assumption, we performed a univariate logistic regression between the cumulative confirmed cases and HDI. We found that HDI is statistically significant (log odds = 28.6476, p -value < 2×10^{-16}). If HDI increases by 0.1, the odd of a confirmed case (that is, the probability that a person is a confirmed case against the probability that a person is not a confirmed infected) increase exponentially by $\exp(2.8648) = 17.5448$.

Many literature have studied the case-fatality rate. Case-fatality rate is defined as the proportion of death among the confirmed cases. However, not all infected people are diagnosed and counted into the confirmed cases. We assume people in the same region has the same probability to get infected and die due to COVID-19 while the risk of death is different among different regions. An univariate logistic regression to study the relation between the cumulative death and HDI is also performed. HDI is again significant (log odds = 36.7946, p -value < 2×10^{-16}). An increase of 0.1 in HDI associates with an exponential increase of 39.6230 in odds of death.

It is interesting to note that high HDI is associated with high infection rate and high fatality rate. To further explore how each of the

* Corresponding author.

E-mail addresses: liuk0809@lixin.edu.cn (K. Liu), hem11@math.mcmaster.ca (M. He), larry.zhuang@polyu.edu.hk (Z. Zhuang), daihai.he@polyu.edu.hk (D. He), lihuaichen@163.com (H. Li).

<https://doi.org/10.1016/j.onehlt.2020.100174>

Received 1 July 2020; Received in revised form 27 August 2020; Accepted 4 September 2020

Available online 29 September 2020

2352-7714/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

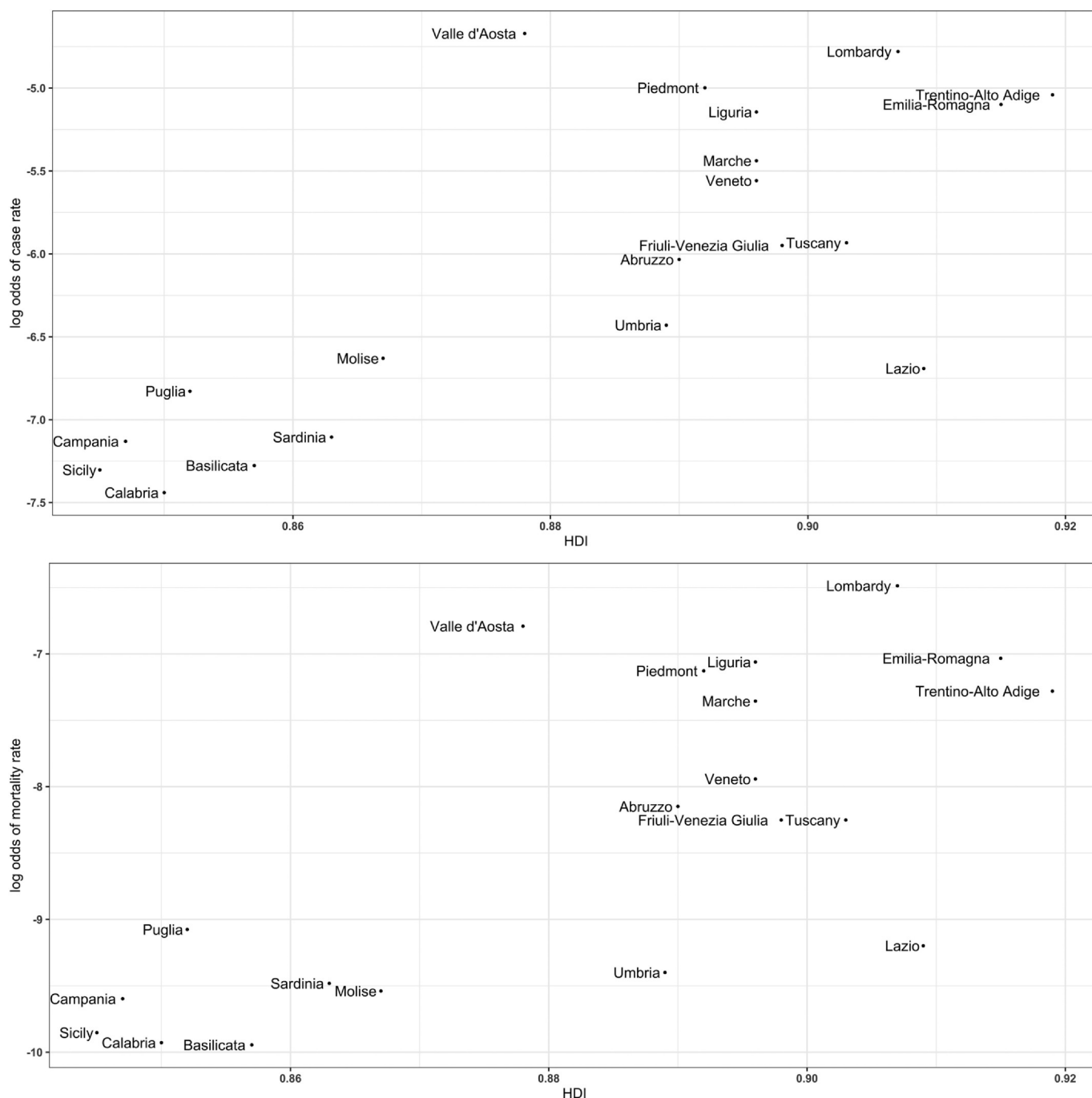


Fig. 1. Log odds of infection rates and log odds of mortality rates against HDI by May 15, 2020.

components of HDI associates with infection rates and fatality rates, we downloaded health data, including smoking data in 2019 and chronic disease data in 2018 from <http://dati.istat.it/?lang=en#>. Average annual gross salaries by regions in 2019 is also downloaded from <https://www.statista.com/statistics/708972/average-annual-nominal-wages-of-employees-italy-by-region/>. Number of cigarettes per day per 100 persons with the same characteristics (cigarette smokers aged 14 years and over) and number of persons with at least one chronic disease per 100 people are used as surrogate index for healthy life for two reasons, one is that they are publicly available and the other reason is that they are associated with life expectancy. The median of number of cigarettes per day per 100 persons is 11.10 (range from 9.2 to 12.5). The number of persons with at least one chronic disease per 100 people ranges from 32.7 to 47.8 with a median of 41. Average annual gross salary by

regions is used to present living standard. Its median is 27,962 with a range from 24,308 to 31,446.

Initial graphical and quantitative analysis of relations between HDI and smoking data, chronic disease data and average annual salary were performed. It showed that HDI is negatively correlated with smoking data (correlation = -0.6428 , p -value = 0.0022), is positively correlated with chronic disease but not statistically significant (correlation = 0.3275 , p -value = 0.1587), is positively correlated with average annual gross salary (correlation = 0.6521 , p -value = 0.0018).

Simple logistic regressions were performed to study the direct effect of the three factors on infection rate and death rate. The results are summarized in Table 1 in terms of log odds estimates and standard errors. It turns out all the factors are significantly associated with infection rate and death rate.

Table 1

Log odds effects of different factors from logistic regression (values in parenthesis are standard errors).

Univariate/multivariate	Factors	Log odds of case rates	Log odds of death rate
Univariate	HDI	28.6476 (0.1214)	36.7946 (0.3735)
Univariate	Average annual gross salary	0.0003 (0.000001)	0.0005 (0.000004)
Univariate	Number of cigarettes per day per 100 persons with the same characteristics 2019	−0.3057 (0.0027)	−0.2411 (0.0072)
Univariate	Number of persons with at least a chronic disease per 100 people with the same characteristics 2018	0.1604 (0.0009)	0.1831 (0.0024)
Multivariate	HDI	17.97 (0.15)	22.81 (0.5067)
	Average annual gross salary	0.0002 (0.000002)	0.0003 (0.000005)
	Number of cigarettes per day per 100 persons with the same characteristics 2019	0.0695 (0.0004)	0.0307 (0.0139)
	Number of persons with at least a chronic disease per 100 people with the same characteristics 2018	0.0554 (0.0014)	0.0545 (0.0044)

Multiple logistic regression was further performed to investigate the effect of HDI after adjustment of other factors. The estimates of log odds are presented in Table 1. The effect of HDI decreases but remain positive. All the factors are statistically significant. The results are consistent with our knowledge that regions with more cigarettes consumption and more persons with chronic disease have higher infection rates and mortality rates. However, regions with higher average annual gross salary also have higher infection rates and mortality rates, but with small magnitude. More specifically, assume other factors remain same, 0.1 increase in HDI results in 6.03 exponential increase in odds of confirmed case ($p < 0.001$), and 9.78 exponential increase in death odds. Similarly, 1000 increase in average annual gross salary, 10 increase in number of cigarettes per day per 100 persons and 10% increase in the number of persons with at least one chronic disease result in 1.34, 1.35, 1.72 exponential increase in odds of confirmed case, respectively. The interpretation is similar for odds of death.

In summary, though high HDI means longer life expectation, better education and better living standard, it is surprising to note that it associates with higher infection rate and mortality rate. By further study, we observed that regions with high HDI normally has higher number of persons with more than one chronic disease, less cigarette consumption and higher average annual gross salary. Multiple logistic regression analysis shows that these three factors take some effects of HDI on the infection rate and death rate. This may partially explain the unexpected positive effect of HDI on infection rate and mortality rate.

Ethics approval and consent to participate

The ethical approval or individual consent was not applicable.

Availability of data and materials

All data and materials used in this work were publicly available.

Consent for publication

Not applicable.

Funding

DH was supported by General Research Fund (15205119) of Research Grants Council of Hong Kong and Alibaba (China) Co. Ltd. Collaborative Research Project (P0031768). The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Author contribution

All authors conceived the study, carried out the analysis, discussed the results, drafted the first manuscript, critically read and revised the manuscript, and all authors gave final approval for publication.

Declaration of Competing Interest

DH received grants from Alibaba (China) Co. Ltd. Collaborative Research project. Other authors declare no competing interests.

Acknowledgements

None.

References

- [1] Andrea Remuzzi, Giuseppe Remuzzi, COVID-19 and Italy: what next? *Lancet* 395 (2020) 1225–1228.
- [2] Edward Livingston, Karen Bucher, Coronavirus disease 2019 (COVID-19) in Italy, *JAMA* 323 (14) (2020) 1335 <https://jamanetwork.com/journals/jama/fullarticle/2763401>.
- [3] Graziano Onder, Giovanni Rezza, Silvio Brusaferro, Case-fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy, 323(18) (2020), pp. 1775–1776 <https://jamanetwork.com/journals/jama/fullarticle/2763667>.
- [4] Y.U. Guangchuang, <https://github.com/GuangchuangYu/nCov2019>.
- [5] Average Annual Gross Salary in Italy in 2019, by Region, <https://www.statista.com/statistics/708972/average-annual-nominal-wages-of-employees-italy-by-region/>.