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

Kai Liu1, Mu He2, Zian Zhuang3, Daihai He3 and Huaichen Li4,5

1 Department of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance.

2 Department of Mathematics and Statistics, McMaster University, Canada.

3 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR China

4 Department of Respiratory and Critical Care Medicine, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, 250021, China. lihuaichen@163.com.

5 School of Medicine and Life Sciences, University of Jinan-Shandong Academy of Medical Sciences, Jinan, 250062, Shandong, China. lihuaichen@163.com.

Email of authors

Kai Liu [liuk0809@lixin.edu.cn](mailto:liuk0809@lixin.edu.cn)

Mu He [hem11@math.mcmaster.ca](mailto:hem11@math.mcmaster.ca)

Zian Zhuang [larry.zhuang@polyu.edu.hk](mailto:larry.zhuang@polyu.edu.hk)

Daihai He [daihai.he@polyu.edu.hk](mailto:daihai.he@polyu.edu.hk)

huaichen Li lihuaichen@163.com

Correspondence to

Daihai He, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR China, [daihai.he@polyu.edu.hk](mailto:daihai.he@polyu.edu.hk)

Highlights:

* Human development index is positively correlated infection and death rates of COVID-19;
* 0.1 increase in human development index results in 17.5448 and 39.6230 exponential increase in odds of infections and deaths, respectively
* Regions with smaller human development index consumes more cigarettes, but have lower average annual gross salaries and smaller chronic diseases population
* The positive correlation between human development index and infections and deaths rates is partially explained by the smoking habits, health status and income level.

Abstract (95 words)

In this analysis, we observed that human development index (an integrated index of life expectation, education and living standard) correlates with case rate (proportion of confirmed cases among the population) and the fatality rates 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.

Main text: 1077 words

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, 2, 3]. Obvious differences in epidemic spread and fatality rates among regions exist, but factors related 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 ranks top 1 with 83820 cumulative confirmed cases among the 20 regions, while the number of cumulative confirmed cases in Basilicata is the smallest (389 cases). The number of death ranges from 22 to 15296, 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 case rates (the proportion of confirmed cases among regional population) range from 0.0006 to 0.009 with a median of 0.0025, while the death rate (proportion of deaths among regional population) ranges 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. Figure 1 presents the log odds of case rates and death rates as of May 15, 2020 against HDI by regions. It presents a linear pattern between the log odds and HDI. To quantify the association between cases rates and death rates with HDI, we perform 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 confirmed. 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 literatures 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. A 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 increase of 39.6230 in odds of death.

it is interesting to note that high HDI is associated with high case rate and high fatality rate. HDI is an integrated index of healthy life, education and living standard. To further explore how each of the components of HDI associates with case 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#](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 they are public and the other reason is 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 oone 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 27962 with a range from 24308 to 31446.

Initial graphical and quantitative analysis of relations between HDI and smoking data, chronic disease data and average annual salary are performed. It shows 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 are performed to study the direct effect of the three factors on case 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 case rates and death rates.

Multiple logistic regression is 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 decreased but remain positive. All the factors are statistically significant. The results are consistent with our knowledge that regions with more cigarettes comsuption and more persons with chronic disease is positively associated with case rates and mortality rates. However, regions with higher average annual gross salary are also positively correlated with case 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 confirmation odds (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 results in 1.34, 1.35, 1.72 exponential increase in odds of confirmed case, respectively. The interpretation is similar for odds of confirmed case.

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 case rates and mortality rate. By further study, we observe 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 case rates and death rates. This may partially explain why unexpected positive effect of HDI on case rate and mortality rates.

**Declarations**

**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 an Alibaba-Hong Kong Polytechnic University Collaborative Research project. 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.

**Acknowledgements**

None.

**Conflict of interest statement**

The authors declare no competing interests.

Reference

[1] Andrea Remuzzi and Giuseppe Remuzzi. COVID-19 and Italy: what next? *The Lancet.* 2020, 395:1225-28.

# [2] Edward Livingston and Karen Bucher. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA.* 2020, 323(14):1335. https://jamanetwork.com/journals/jama/fullarticle/2763401

[3] Graziano Onder, Giovanni Rezza and Silvio Brusaferro. Case-fatality rate and characteristics of patients Dying in relation to COVID-19 in Italy. 2020, 323(18):1775-1776.https://jamanetwork.com/journals/jama/fullarticle/2763667

# [4] Guangchuang YU, <https://github.com/GuangchuangYu/nCov2019>. Accessed on May 18, 2020.

# [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/>

# Figure 1 Log odds of case rate and log odds of mortality rate against HDI by May 15, 2020

# Table 1 log odds effects of different factors from logistic regression

|  |  |  |  |
| --- | --- | --- | --- |
| 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) |