SLLD Exam Module 1 & Module 2

Ph.D. Data Science

Geographical and individual deprivation index to assess the risk of Sars-CoV-2 infection and disease severity in people at socioeconomic disadvantage

Professor: Francesca Chiaromonte

Student: Sara Mazzilli

Abstract

Background Despite the initial widespread opinion that the virus does not discriminate, it has been shown that COVID-19 affects people at socioeconomic disadvantage more strongly. This is perfectly explained by the covid syndemic concept, i.e., the biological, economic, and social interactions between non-communicable diseases and COVID-19 increase a person's susceptibility to infection and worse health outcomes. In this study, we explored the relationship among the socioeconomic deprivation and the risk of Sars-CoV-2 infection and disease severity in Apulia region, Italy.

Methods We linked individuals tested for Sars-CoV-2 in the surveillance data system on December 31, 2020, with the geographical deprivation index (DI) (1-5 scale) of their area of living. We calculated, using 5 census variable individual DI using 2 different techniques, a sum of z-score and the PCA. Through GLM logistic models we calculated the relative risk of acquiring Sars-CoV-2 infection, COVID-19-related hospitalization, and deaths of deprived individuals compared with people in the lower level of socioeconomic deprivation adjusting for gender, age and comorbidities.

Results In the study period, 195,016 individuals (49.3% female and 50.7% male) were tested for COVID-19: of those, 87,398 (44.8%) tested positive. Among those tested positive 9,027 (10.3%) have been hospitalised and 2,391 (2.7%) died. During the first epidemic wave less socioeconomic deprived individuals were at higher risk of testing positive, while during the second wave individuals with a higher level of socioeconomic deprivation had a higher likelihood of testing positive. For what concern the risk of being hospitalised if positive and dying if positive, it did not significantly change among groups with different level of socioeconomic deprivation when we consider the geographical DI. The GLMs using the individual DI showed that the risk of being hospitalised if positive and dying if positive increased with the increase of the level of individual deprivation during the second wave.

Conclusion According with our results, with the progression of the socioeconomic disadvantage the risk of acquiring Sars-CoV-2 infection increase, and the people with the highest DI are at higher risk of severe outcome when infected. Our study highlights the importance of adopting targeted health policies and action to protect those with the greatest socioeconomic vulnerability and enhance equity.

Table of Contents

Introduction

The newly emerged virus SARS-CoV-2 was initially reported in China in December, 2019 (1). On February 20, 2020, the first major COVID-19 outbreak in Europe was detected in the Lombardy region, Italy (2). On March 11, 2020, WHO declared the SARS-CoV-2 outbreak a pandemic (3). COVID-19, the disease resulting from SARS-CoV-2 infection, led to 6.3 million deaths worldwide and 166.000 in Italy by the end of May, 2022 (4,5). During the first pandemic year, the temporal course of the epidemic was characterized by 3 distinct phases: the first epidemic wave was from March to June 2020, followed by a summer period with a relatively low incidence, and a second wave that started in September and peaked in November 2020 (4).

Evidence shows that male, aged over 65 and smoking patients might face a greater risk of developing more critical or lethal conditions and comorbidities, such as hypertension, diabetes, cardiovascular disease or respiratory diseases, could also greatly affect the prognosis of the COVID-19 patients (6). In addition, a study carried out in a large community cohort has shown associations between adverse lifestyle and higher risk of COVID-19 (7). However, it is now well known that lifestyle plays a mediating role in the relationship between the socioeconomic position (SEP) and health (8).

Health inequalities appear to reflect social inequalities existing in the population. Investigation into health inequalities was inaugurated in 1980 when the 'Black Report' was published in the UK, notwithstanding this, health inequalities continue to be a current problem. In a pandemic context, health inequalities tend to become more marked. Especially during pandemics, social and economic factors play an important role in the context of the spread of diseases. Health and socioeconomic inequality mutually influence each other by triggering and feeding vicious circles (9). For instance, with regard to the COVID-19, the approach to social measures to contain virus transmission, i.e. hand-washing, physical distance, closure of businesses, proper use of masks, differed in the subpopulation groups with different SEP. A higher proportion of people who are from a lower SEP lives in overcrowded housing compared to wealthy individuals. Additionally, most deprived persons usually carry out manual labour professions or work in the informal sector and therefore, working from home was difficult or impossible for them. Unfortunately, the risk of infection increases by overcrowded living conditions and the inability to work from home, both barriers to social distancing.

Nevertheless, socio-economic inequalities also come into play in indirectly promoting the risk of complications in cases of COVID-19. Indeed, as mentioned above, socioeconomically disadvantaged people are known to be the most impacted by chronic non-communicable diseases (10,11). United States' studies indicated that social inequalities —including poverty, physical environment (e.g. homelessness), and race or ethnicity— in health are profoundly, and unevenly, impacting COVID-19 morbidity, mortality and transmission risk (12,13). In a large study conducted in UK, a striking gradient in risk of hospitalization for COVID-19 was noted to be related to race and a metric of socioeconomics deprivation (14).

The pandemic itself has accentuated these already existing social and health inequalities, widening the gap among individual with different SEP. Millions of people have become impoverished as a result of the pandemic. This is a crucial point for the present but also for the future. Social determinants of health cannot be left out of a pandemic response and prevention programme.

The available Italian deprivation index (DI) is a multidimensional measure of the disadvantage in the ownership of both social and material resources of geographical area (the census sections, which are comparable to neighbourhoods) residents. Methods for calculating this index has been extensively described elsewhere (15,16). The DI is used as a proxy for the level of individual social disadvantage,

albeit with the inevitable potential ecological bias that can emerge by attributing a collective measure to an individual. In this study, using census information at individual level and two statistical methods we calculated the individual DI.

The general aim of this research was to assess the impact of socioeconomic position on the spread, morbidity, and mortality of COVID-19 in Apulia region from the beginning of the pandemic until December 2020. Moreover, it explored the type of association between SEP and COVID-19 outcomes assessed using the geographic DI (standard Italian DI) and the individual DI.

Methods

Study setting

The Italian 'Servizio Sanitario Nazionale' (SSN) was introduced in 1978 to ensure that healthcare is accessible to all Italian citizens without socio-economic barriers, according to a principle of horizontal equity (17). The Italian health-care system is a regionally based national health service (NHS). The system is organized into three levels: national, regional, and local. The national level is responsible for establishing the general objectives and fundamental principles of the NHS. The nineteen regions and two autonomous provinces (R&AP) are then responsible for organizing and delivering health care (18). In this scenario, through ministerial decrees, the Ministry of Health has taken the lead in the fight against COVID-19 epidemic. Then, the R&AP were in charge of organizing and implementing the prevention strategy at the local level based on national guidelines.

Data Sources

Data were provided by the Apulia regional health agencies (AreSS - Agenzia Regionale Strategica per la Salute ed il Sociale). Data on individuals tested for COVID-19 collected through the regional information system was available. The above mentioned anonymized dataset contained data on patients' demographic characteristics, partial census individual information, Italian standard DI, outcome of the diagnostic test, and for people resulted positive: place and time of infections, outcome of the infection (hospitalisation and death). An assessment of comorbidity was linked to study participants through Charlson comorbidities index that has been extensively described previously (19). Health care workers were removed from the analysis since their risk of infection acquisition is strongly related to their professional exposure to the virus.

Definition of the individual deprivation index

For the definition of the individual deprivation index we took the 5 census variable shown in table 1.

Z scores individual deprivation index

In order to calculate the z score deprivation index: (1) the median of the census variable of interest has been calculated; (2) for each observation, the z scores for each census variable of interest has been calculated (3) for each observation, the sum of the z scores for each census variable of interest has been calculated; (4) based on the score resulted the sample population has been divided in quintiles.

$$IDI = \sum_{i=1}^{5} z_i \qquad z_i = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}}$$

PCA individual deprivation index

Using the 5 census variable shown in table 1, after their scaling, a principal component analysis (PCA) has been carried out. We took the coordinate of the observation over the PC1 and based on this value we divided the observation in quintile.

Table 1. Census variable used to define the individual deprivation index.

Number of family members	Citizenship	Family Type	Employment Status	Education
Numerical value	1 Italian	1 Couple without children	1 Employed	1 Academic Diploma 2nd Level/Academy of Fine Arts Diploma/Master degree
	2 Foreigner	2 Couple with children	2 Recipient(s) of one or more pensions due to previous employment or investment income	2 Academic 1st Level Diploma/Bachelor degree
		3 Single parent: father	3 Students/Housewives/In other status	3 High school graduation
		4 Single parent: mother	4 First-time job seekers	4 Middle school graduation
			5 Unemployed	5 Elementary school diploma
				6 Illiterates/Literate without educational qualification

Statistical Analysis

Generalized linear model logistic regression models were used to test associations between COVID-19 outcome (tested positive, being hospitalised, death) and "geographical" or individual deprivation index. Age, gender, Charlson comorbidity index and provinces of residence were included in the model as covariates. Different models were used for the first and the second epidemic wave.

Results

In the study period 195,016 individuals (49.3% female and 50.7% male) were tested for COVID-19: 9,230 (4.7%) during the first wave, 27,941 (14.3%) during the summer and 157,845 (80.9%) during the second wave. In our study population 22.0% were aged below 30, 54.2% were aged from 30 to 69 years old and 23.8% were aged above 69. The level of deprivation was 1 for 31,248 (18.2%) individuals, 2 for 30,970 (18.0%), 3 for 33830 (19.7%), 4 for (21.8%), 5 for 38591 (22.4%). In our sample, 87,398 (44.8%) tested positive: 3,459 (4.0%) during the first wave, 2,128 (2.4%) during the summer and 81,811 (93.6%) during the second wave. Among those tested positive 9,027 (10.3%) have been hospitalised and 2,391 (2.7%) died.

Individual deprivation index

The census variables selected to build the individual DI were available for a subpopulation of 169,841 individuals and before starting calculating the individual DI, the distribution of the geographical DI in this sub-sample was checked. Table 2 shows that the sample selected for the availability of the census variables had the same distribution of the geographical DI as the initial population.

Table 2. Distribution of the geographical DI in the initial population and in the sub-sample of observations for which

	Geo DI 1	Geo DI 2	Geo DI 3	Geo DI 4	Geo DI 5
Total sample	0.19	0.19	0.2	0.21	0.21
Sample with the selection of census variables	0.19	0.19	0.20	0.21	0.21

Z scores individual deprivation index

Table 3. shows the cross table of the geographical DI and the z-score individual DI. In census areas with a lower level of deprivation a higher proportion of people with a lower z-score individual DI lives. Vice versa, in census areas with higher socio-economic deprivation a higher proportion of people with a high z-score individual DI lives.

Table 3. cross table of the geographical DI and the z-score individual DI.

Z-score individual DI | Geographical DI

	1	2	3	4	5	Row Total
1	7493	6187	6076	6018	4525	30299
	0.247	0.204	0.201	0.199	0.149	0.232
2	3908	4092	4419	4903	4542	21864
	0.179	0.187	0.202	0.224	0.208	0.168
3	4280	4381	4798	5036	4806	23301
	0.184	0.188	0.206	0.216	0.206	0.179
4	4371	4940	5607	6519	6871	28308
	0.154	0.175	0.198	0.230	0.243	0.217
5	3215	3930	4791	6132	8510	26578
	0.121	0.148	0.180	0.231	0.320	0.204
Column Total	23267	23530	25691	28608	29254	130350

PCA individual deprivation index

The variance explained by the PC1 is 26.9% and the variance explained by the PC2 is 22.3% (Figure 1). Figure 2 shows how the census variables are oriented in the first two PC of PCA.

Table 4. PCA variable loadings.

	PC1	PC2	PC3	PC4	PC5
Citizenship	0.02312791	-0.1383537	0.9846104966	-0.09751823	-0.03682013
N family members	0.44756510	-0.5055956	-0.0037278953	0.66663229	0.31566774
Education	-0.65676728	-0.2397384	0.0006490758	-0.07744466	0.71076307
Type of family	0.47067763	-0.4176653	-0.1345254304	-0.73490554	0.21409111
Employment	-0.38245247	-0.7023621	-0.1114933187	0.00278589	-0.58989746

Figure 1. Scree plot showing the PCA results.

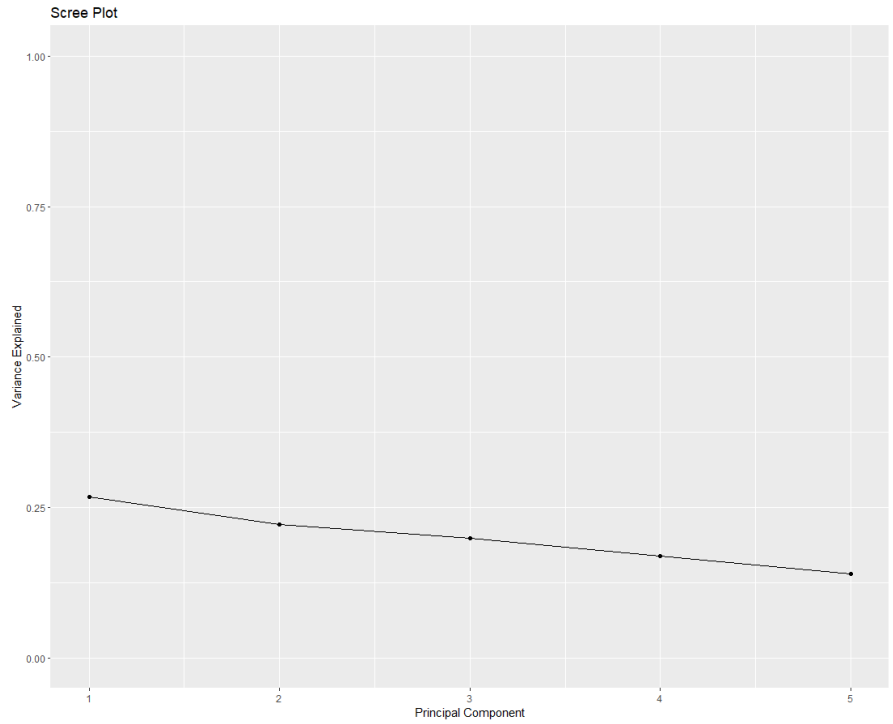
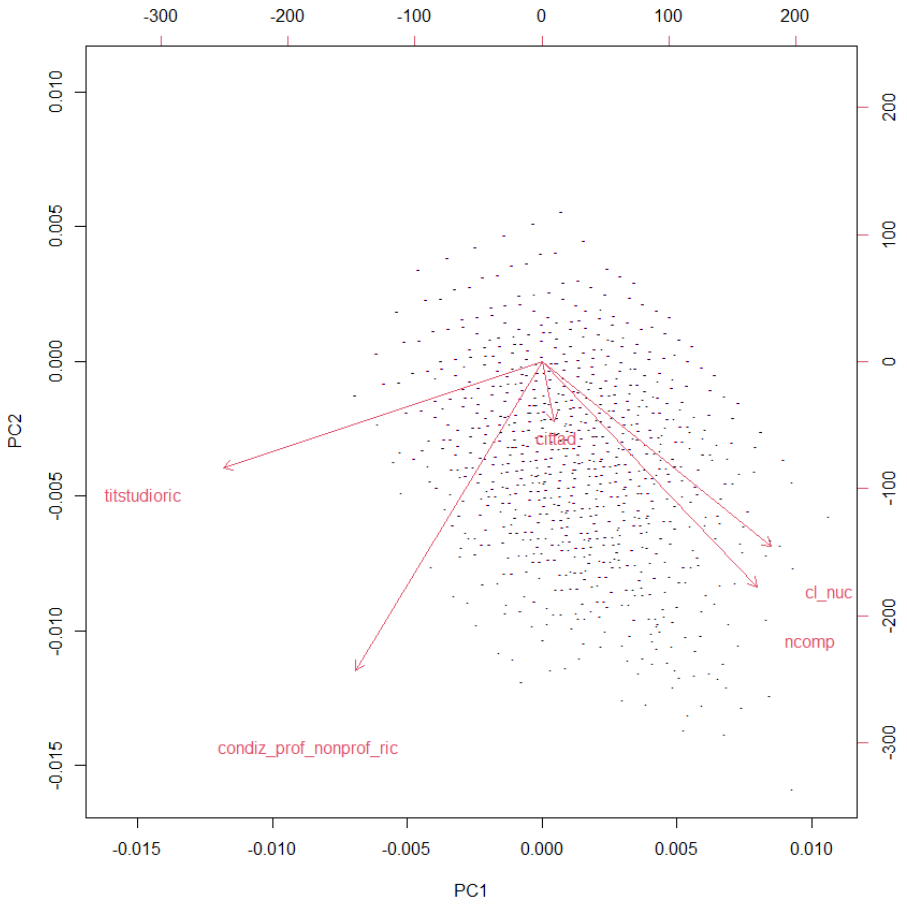


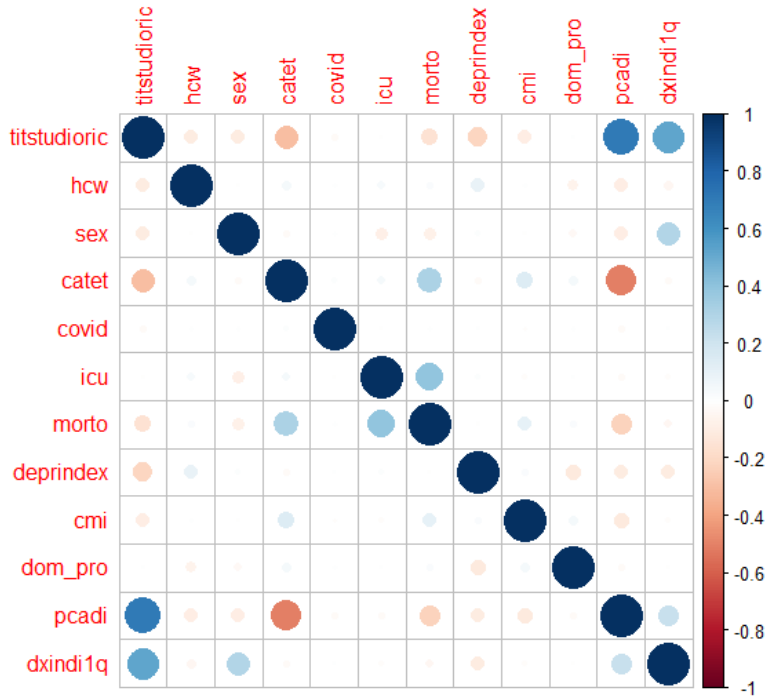
Figure 2. PCA biplot of the variable over the PC1 and PC2.



GLM

Figure 3 shows that there are not strong correlations among the variables included in the GLM, with the only exception of a moderate correlation among the age groups and de PCA individual DI.

Figure 3. Correlation matrix for variables included in the GLM



1st epidemic wave

According with the results of the logistic GLMs, during the 1st epidemic wave the risk of testing positive for Sars-CoV-2 infection was lower for people with high level of socioeconomic deprivation when considering the geographical DI and the z-score individual DI. While being hospitalised if positive and dying if positive did not significantly change in population groups with different level of socioeconomic deprivation. Results were homogeneous despite the type of deprivation index used in the GLM, with the only exception of the risk of dying of the population with the highest socio-economic deprivation, that is significantly higher when the index used is the PCA individual DI (Figure 4).

2nd epidemic wave

According with the results of the logistic GLMs, during the 2nd epidemic wave the risk of testing positive for Sars-CoV-2 infection was significantly higher in people with a level of deprivation higher than 1 when compared with the less deprived population group, with homogeneous results disregarding the deprivation index used in the model (Figure 5.1). For what concern the risk of being hospitalised if positive and dying if positive, it did not significantly change among groups with different level of socioeconomic deprivation when we consider the geographical DI. The GLM using z-score or PCA individual DI results showed that the risk of being hospitalised if positive and dying if positive increased with the increase of the level of individual deprivation (Figure 5.2B, 5.2C, 5.3B, 5.3C).

Figure 4. Results on the association among COVID-19 outcome and the level of socioeconomic deprivation during the 1st wave. 1. Relative risk of testing positive (0: negative, 1: positive) for population group with different level of geographical (1A) or individual (1B z-score; 1C PCA) socioeconomic deprivation index. 2. Relative risk of being hospitalised if positive (0: no hospitalization, 1: hospitalization) for population group with different level of geographical (2A) or individual (2B z-score; 2C PCA) socioeconomic deprivation index. 3. Relative risk of dying if positive (0: alive, 1: death) for population group with different level of geographical (3A) or individual (3B z-score; 3C PCA) socioeconomic deprivation index.

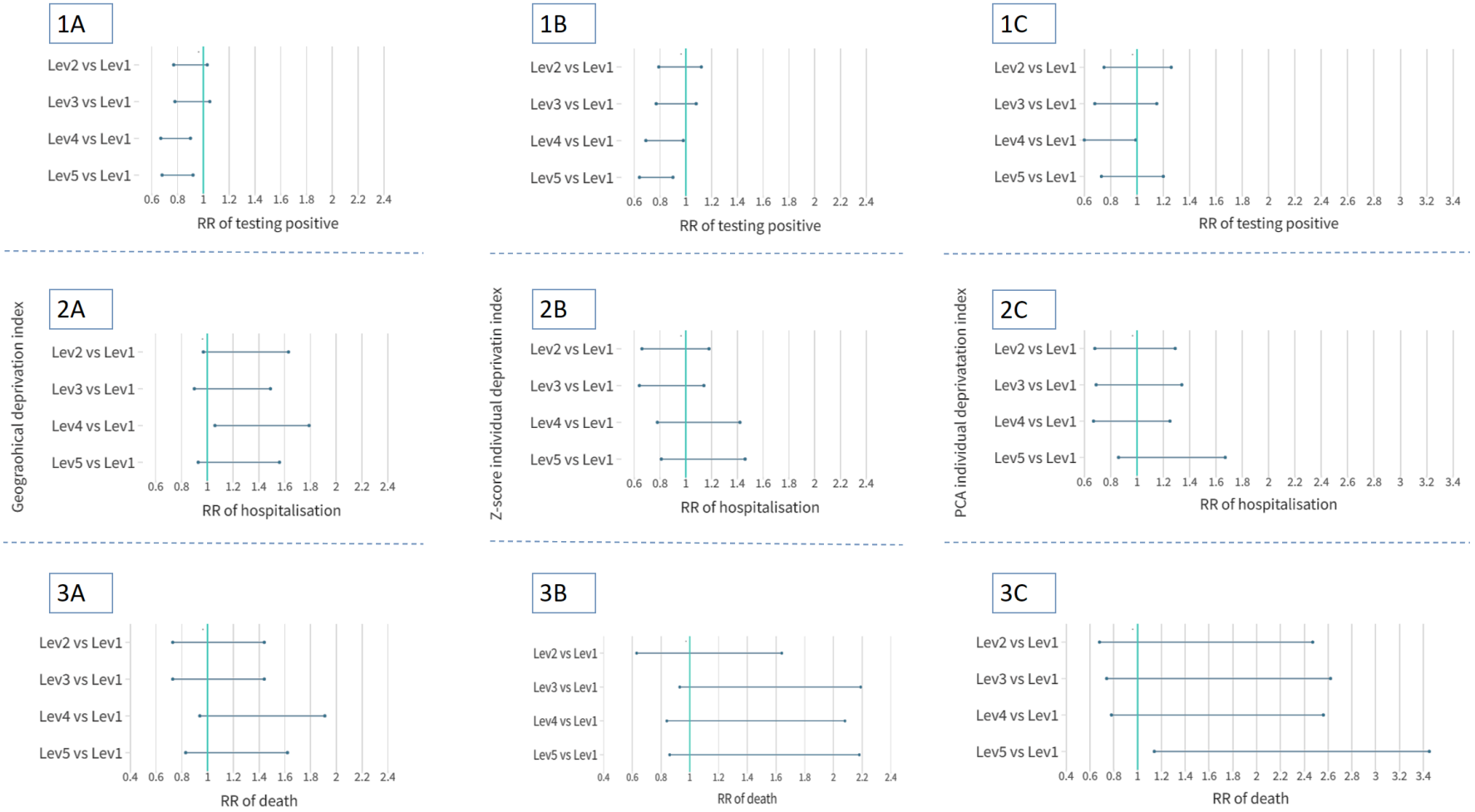
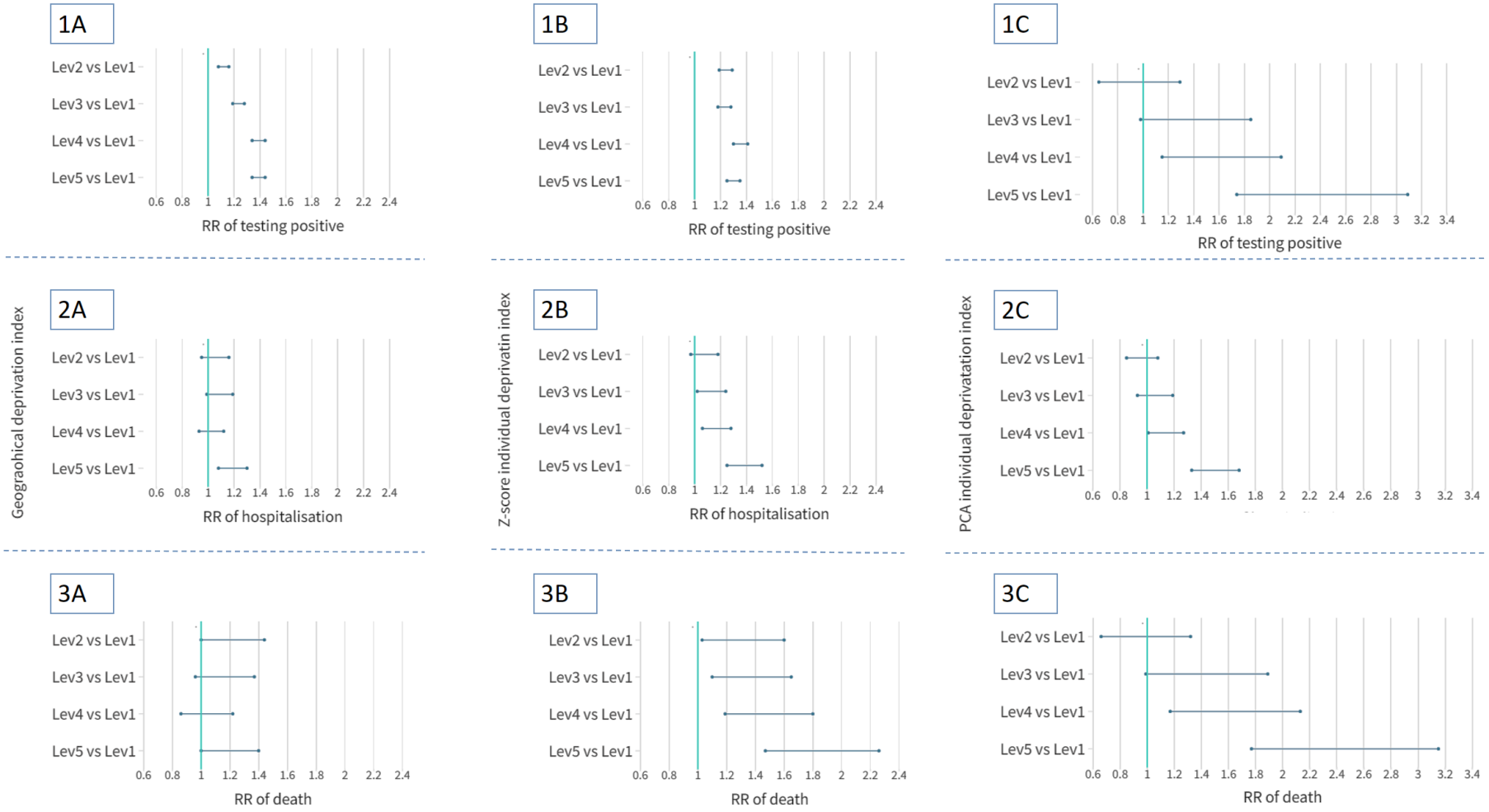


Figure 5. Results on the association among COVID-19 outcome and the level of socioeconomic deprivation during the 2nd wave. 1. Relative risk of testing positive (0: negative, 1: positive) for population group with different level of geographical (1A) or individual (1B z-score; 1C PCA) socioeconomic deprivation index. 2. Relative risk of being hospitalised if positive (0: no hospitalization, 1: hospitalization) for population group with different level of geographical (2A) or individual (2B z-score; 2C PCA) socioeconomic deprivation index. 3. Relative risk of dying if positive (0: alive, 1: death) for population group with different level of geographical (3A) or individual (3B z-score; 3C PCA) socioeconomic deprivation index.



Discussion

In this study, we find that the association between the risk of testing positive for Sars-CoV-2 and the level of socioeconomic deprivation changed between the first and second wave in Apulia region. While during the first epidemic wave less socioeconomic deprived individuals were at higher risk of testing positive, during the second wave individuals with a higher level of socioeconomic deprivation had a higher likelihood of testing positive.

This may be explained by the fact that during the early stages of the epidemic outbreak, with affected geographical areas still circumscribed, mostly of the cases in Apulia region were due to the returning residents (20). Indeed, the gradual implementation of control measures in Italy, sparked substantial movement of people travelling from northern regions at the epicentre of the epidemic toward other regions, such as Apulia. Then, the swift extension of lock-down to the entire country (21) mitigated the impact of these COVID-19 seeding events and epidemic in Apulia was successfully contained. For these reasons, during the first wave in Apulia the epidemic spread mainly among individuals who had the financial means to travel or who were economic migrants in Lombardy or in northern regions, areas of intense economic activities.

To our knowledge this is the first study in Italy developing an individual DI to assess the relationship between COVID-19 outcomes and the socioeconomic deprivation of individuals. A previous study investigated the association between deprivation at municipality level and COVID-19 outcomes in Italy (22). In accordance with our results, Mateo-Urdiales and colleagues found a higher incidence of cases in the most deprived municipalities compared with the least deprived ones and no differences in case-hospitalisation and case-fatality according to deprivation were observed in any period under study. This is consistent with our results when we consider the geographical DI. However, we observed that for the 2nd pandemic waves, the risk of hospitalization and death increased at the increase of the level of socioeconomic deprivation when it was measured through the individual DI. These results are in line with other similar studies carried out in EU/EEA (14,23–25).

PCA is a widely used technique for computing indices of deprivation (26–28), it is mostly used for the calculation of geographical deprivation index or is based on dichotomous asset measures, but in this work we have shown how it can also be used to calculate the individual DI by using categorical variables.

Governments and health-care systems should address this pandemic of inequality by taking measures to reduce health inequalities in their response to the SARS-CoV-2 pandemic and more in general during pandemic events.

References

1. Cash R, Patel V. Has COVID-19 subverted global health? Lancet. 2020;395(10238):1687–8.

2. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, et al. The emergence of SARS-CoV-2 in Europe and North America. Science (80-). 2020;370(6516):564–70.

3. WHO. World Health Organization. Coronavirus disease (COVID-19) Situation Report – 148 Data as received by WHO from national authorities by 10:00 CEST, 16 June 2020. Fam Pract Guidel [Internet]. 2020;(June):1–18. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

4. Italian National Institute of Health (Istituto Superiore di Sanità). COVID-19 integrated surveillance: key national data [Internet]. 2021 [cited 2022 May 25]. Available from: https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-integrated-surveillance-data

5. The Johns Hopkins University of Medicine. COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University [Internet]. [cited 2022 May 27]. Available from: https://coronavirus.jhu.edu/map.html

6. Zheng Z, Peng F, Xu B, Zhao J., Liu H., Peng H. , Li Q., Jiang C., Zhou Y., Liu S. YC. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. 2020;(January).

7. Hamer M, Kivimäki M, Gale CR, Batty GD. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. Brain, Behav Immun J. 2020;87:184–187.

8. Wang J, Geng L. Effects of socioeconomic status on physical and psychological health: Lifestyle as a mediator. Int J Environ Res Public Health. 2019;16(2).

9. Horton R. Offline: COVID-19 is not a pandemic. Lancet [Internet]. 2020;396(10255):874. Available from: http://dx.doi.org/10.1016/S0140-6736(20)32000-6

10. Boes S, Sabariego C, Bickenbach J, Stucki G. How to capture the individual and societal impacts of syndemics: The lived experience of COVID-19. BMJ Glob Heal. 2021;6(10):1–6.

11. World Health Organization. COVID-19 and the social determinants of health and health equity: evidence brief. World Health Organization. 2021. 1–32 p.

12. Ramírez IJ, Lee J. COVID-19 Emergence and Social and Health Determinants in Colorado: A Rapid Spatial Analysis. Int J Environ Res Public Health. 2020;17(11).

13. Turner-Musa J, Ajayi O, Kemp L. Examining Social Determinants of Health, Stigma, and COVID-19 Disparities Jocelyn. Healthcare. 2020;8(168):1–7.

14. Patel P, Paranjpe M., Kathiresan NP, M.A. Rivas, Khera AV. Race, socioeconomic deprivation, and hospitalization for COVID-19 in English participants of a national biobank. Int J Equity Health. 2020;19(114).

15. Caranci N, Biggeri A, Grisotto L, Pacelli B, Spadea T, Costa G. L ' indice di deprivazione italiano a livello di sezione di censimento : definizione , descrizione e associazione con la mortalità The Italian deprivation index at census block level : definition , with general mortality. 2010;34(4):167–76.

16. Rosano A, Pacelli B, Zengarini N, Costa G, Cislaghi C, Caranci N. Aggiornamento e revisione dell ' indice di deprivazione italiano 2011 a livello di sezione di censimento Update and review of the 2011 Italian deprivation index calculated at the census section level. 2020;44:162–70.

17. France G, Taroni F, Donatini A. The Italian health-care system. Health Econ. 2005;14(SUPPL. 1):187–202.

18. Ferré F, de Belvis AG, Valerio L, Longhi S, Lazzari A, Fattore G, et al. Italy: Health system review. Health Syst Transit. 2014;16(4):1–166.

19. E. Charlson M, Pompei P, L. Ales K, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis. 1987;40(5):373–83.

20. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci U S A. 2020;117(19):10484–91.

21. Decree of the Prime Minister. Ulteriori disposizioni attuative del decreto-legge 23 febbraio 2020, n. 6, recante misure urgenti in materia di contenimento e gestione dell'emergenza epidemiologica da COVID-19, applicabili sull'intero territorio nazionale. (20A01605). GU Ser Gen n 64 del 11 marzo 2020 [Internet]. Available from: http://www.trovanorme.salute.gov.it/norme/dettaglioAtto?id=73643

22. Mateo-Urdiales A, Fabiani M, Rosano A, Vescio MF, Del Manso M, Bella A, et al. Socioeconomic patterns and COVID-19 outcomes before, during and after the lockdown in Italy (2020). Heal Place [Internet]. 2021;71(March):102642. Available from: https://doi.org/10.1016/j.healthplace.2021.102642

23. Riou J, Panczak R, Althaus CL, Junker C, Perisa D, Schneider K, et al. Socioeconomic position and the COVID-19 care cascade from testing to mortality in Switzerland: a population-based analysis. Lancet Public Heal [Internet]. 2021;6(9):e683–91. Available from: http://dx.doi.org/10.1016/S2468-2667(21)00160-2

24. Politi J, Martín-Sánchez M, Mercuriali L, Borras-Bermejo B, Lopez-Contreras J, Vilella A, et al. Epidemiological characteristics and outcomes of COVID-19 cases: Mortality inequalities by socioeconomic status, Barcelona, Spain, 24 February to 4 May 2020. Eurosurveillance [Internet]. 2021;26(20):1–12. Available from: http://dx.doi.org/10.2807/1560-7917.ES.2021.26.20.2001138

25. Drefahl S, Wallace M, Mussino E, Aradhya S, Kolk M, Brandén M, et al. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. Nat Commun [Internet]. 2020;11(1):1–7. Available from: http://dx.doi.org/10.1038/s41467-020-18926-3

26. Lamnisos D, Lambrianidou G, Middleton N. Small-area socioeconomic deprivation indices in Cyprus: development and association with premature mortality. BMC Public Health. 2019;19(627):395–410.

27. Meijer M, Engholm G, Grittner U, Bloomfield K. A socioeconomic deprivation index for small areas in Denmark. Scand J Public Health. 2013;41(6):560–9.

28. Panczak R, Galobardes B, Voorpostel M, Spoerri A. A Swiss neighbourhood index of socioeconomic position : development and association with mortality. J Epidemiol Community Heal. 2012;66(12):1129–36.

Supplemental material

eTable 1. cross table of the geographical DI and the PCA individual DI.

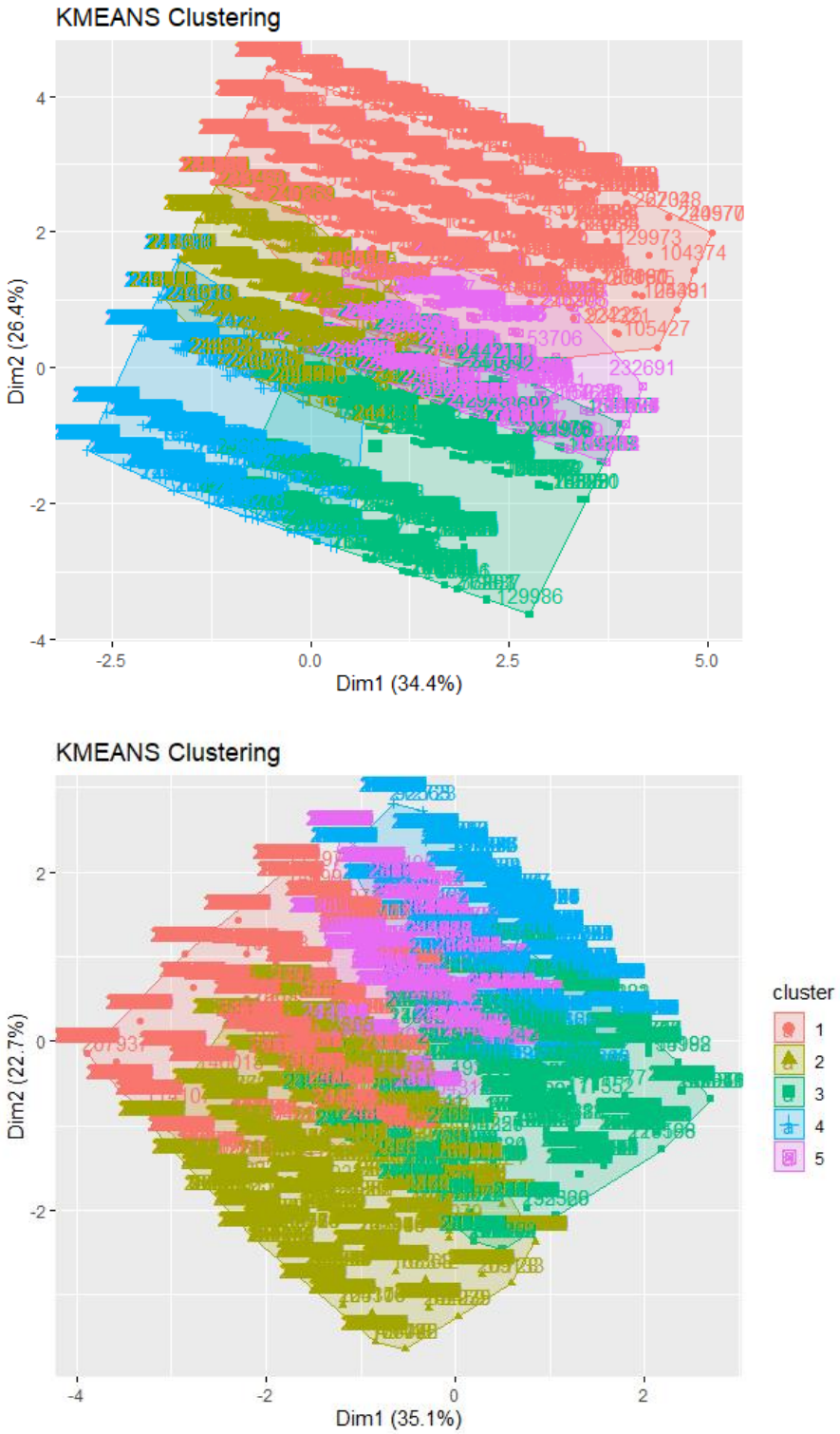
PCA individual DI	Geographical DI 1	2	3	4	5	Row Total
1	4722	3648	3814	4034	3398	19616
	0.241	0.186	0.194	0.206	0.173	0.150
2	4733	4782	5072	5407	4720	24714
	0.192	0.193	0.205	0.219	0.191	0.190
3	3713	4060	4532	5269	5275	22849
	0.163	0.178	0.198	0.231	0.231	0.175
4	4822	5435	6262	7210	7933	31662
	0.152	0.172	0.198	0.228	0.251	0.243
5	5277	5605	6011	6688	7928	31509
	0.167	0.178	0.191	0.212	0.252	0.242
Column Total	23267	23530	25691	28608	29254	130350

eTable 2. cross table of the geographical DI and the PCA individual DI.

PCA individual DI	Z-score individual DI 1	2	3	4	5	Row Total
1	8751	4134	1618	2565	2548	19616
	0.446	0.211	0.082	0.131	0.130	0.150
2	15499	1231	2816	2387	2782	24715
	0.627	0.050	0.114	0.097	0.113	0.190
3	603	8729	5936	929	6654	22851
	0.026	0.382	0.260	0.041	0.291	0.175
4	2715	4023	4661	14694	5572	31665
	0.086	0.127	0.147	0.464	0.176	0.243
5	2732	3748	8272	7738	9029	31519
	0.087	0.119	0.262	0.246	0.286	0.242

We attempt to use another unsupervised technique to calculate the individual DI, the K means clustering, but in the end we prefer to continue the analysis just with the PCA individual DI, since the use of K means clustering for calculating the DI is not reported in literature.

eFigure 1. Graphical representation of K-means clustering results using two different set of census variables.



eTable 3: GLM output

Time period: 1st wave
Outcome: 0: negative, 1: positive
DI: geographical

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3835  -0.9491  -0.8463   1.2936   2.1400

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.36897    0.09980  -3.697 0.000218 ***
sex1        -0.08209    0.04722  -1.739 0.082093 .
catet2      -0.14867    0.08455  -1.758 0.078675 .
catet3      -0.18378    0.08682  -2.117 0.034286 *
cmi1        -0.68157    0.08274  -8.238  < 2e-16 ***
cmi2        -0.99089    0.16493  -6.008 1.88e-09 ***
cmi3        -1.81275    0.61142  -2.965 0.003029 **
dom_pro2     0.34647    0.07426   4.666 3.08e-06 ***
dom_pro3     0.55239    0.08866   6.230 4.66e-10 ***
dom_pro4     0.84146    0.06491  12.962  < 2e-16 ***
dom_pro5    -0.01135    0.07910  -0.144 0.885867
dom_pro6     0.19217    0.10002   1.921 0.054703 .
deprindex2  -0.11695    0.07622  -1.534 0.124960
deprindex3  -0.09793    0.07498  -1.306 0.191534
deprindex4  -0.25398    0.07687  -3.304 0.000953 ***
deprindex5  -0.23145    0.07535  -3.071 0.002130 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10704  on 8117  degrees of freedom
Residual deviance: 10380  on 8102  degrees of freedom
  (1112 observations deleted due to missingness)
AIC: 10412

Number of Fisher Scoring iterations: 4
```

Time period: 1st wave
Outcome: 0: negative, 1: positive
DI: z-score individual

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + dxindi1q,
    family = "binomial", data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3241  -0.9469  -0.8168   1.2802   2.1937

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.55430    0.15261  -3.632 0.000281 ***
sex1        -0.07862    0.05716  -1.375 0.169033
catet2      -0.01532    0.13951  -0.110 0.912537
catet3      -0.09795    0.14131  -0.693 0.488194
cmi1        -0.69567    0.09712  -7.163 7.89e-13 ***
cmi2        -1.29929    0.20776  -6.254 4.00e-10 ***
cmi3        -1.81791    0.74500  -2.440 0.014681 *
dom_pro2     0.47454    0.08649   5.487 4.09e-08 ***
dom_pro3     0.38863    0.10564   3.679 0.000234 ***
dom_pro4     0.90816    0.07557  12.017  < 2e-16 ***
dom_pro5     0.01624    0.09051   0.179 0.857608
dom_pro6     0.41372    0.10956   3.776 0.000159 ***
dxindi1q2   -0.06233    0.08883  -0.702 0.482840
dxindi1q3   -0.09616    0.08751  -1.099 0.271865
dxindi1q4   -0.19549    0.08728  -2.240 0.025109 *
dxindi1q5   -0.28143    0.08739  -3.220 0.001280 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7944.4  on 6066  degrees of freedom
Residual deviance: 7659.5  on 6051  degrees of freedom
  (3163 observations deleted due to missingness)
AIC: 7691.5

Number of Fisher Scoring iterations: 4
```

Time period: 1st wave
Outcome: 0: negative, 1: positive
DI: PCA individual

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3333  -0.9323  -0.8261   1.2855   2.1597

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.83141    0.15745  -5.281 1.29e-07 ***
sex1        -0.11347    0.05551  -2.044 0.040941 *
catet2       0.06254    0.13747   0.455 0.649162
catet3       0.02359    0.14848   0.159 0.873757
cmi1        -0.70506    0.09694  -7.273 3.52e-13 ***
cmi2        -1.30875    0.20764  -6.303 2.92e-10 ***
cmi3        -1.81072    0.74511  -2.430 0.015093 *
dom_pro2     0.46719    0.08650   5.401 6.61e-08 ***
dom_pro3     0.38656    0.10562   3.660 0.000252 ***
dom_pro4     0.90467    0.07546  11.988  < 2e-16 ***
dom_pro5     0.02162    0.09046   0.239 0.811124
dom_pro6     0.42859    0.10957   3.912 9.17e-05 ***
pcadi2       0.01624    0.08593   0.189 0.850133
pcadi3       0.10423    0.09537   1.093 0.274430
pcadi4       0.22352    0.09732   2.297 0.021636 *
pcadi5       0.13177    0.10006   1.317 0.187877
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7944.4  on 6066  degrees of freedom
Residual deviance: 7664.7  on 6051  degrees of freedom
  (3163 observations deleted due to missingness)
AIC: 7696.7

Number of Fisher Scoring iterations: 4
```

Time period: 1st wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: geographical

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0246  -0.9969  -0.3361   1.0122   2.7285

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.81968    0.26664 -10.575  < 2e-16 ***
sex1         -0.65193    0.08382  -7.777  7.4e-15 ***
catet2        2.19226    0.25085   8.739  < 2e-16 ***
catet3        3.48968    0.25465  13.704  < 2e-16 ***
cmi1          0.42765    0.15504   2.758 0.005811 **
cmi2          0.91625    0.37308   2.456 0.014054 *
cmi3         13.82599  263.69790   0.052 0.958185
dom_pro2     -0.41302    0.13336  -3.097 0.001955 **
dom_pro3      0.18124    0.14908   1.216 0.224108
dom_pro4      0.30553    0.10761   2.839 0.004520 **
dom_pro5      0.54035    0.14327   3.772 0.000162 ***
dom_pro6      0.66910    0.17775   3.764 0.000167 ***
deprindex2    0.23145    0.13165   1.758 0.078741 .
deprindex3    0.14497    0.12885   1.125 0.260543
deprindex4    0.32079    0.13395   2.395 0.016628 *
deprindex5    0.18685    0.13047   1.432 0.152106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4149.4  on 3008  degrees of freedom
Residual deviance: 3521.6  on 2993  degrees of freedom
  (450 observations deleted due to missingness)
AIC: 3553.6

Number of Fisher Scoring iterations: 12
```

Time period: 1st wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: z-score individual

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + dxindi1q, family = "binomial",
    data = df_gml1)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.240  -1.019  -0.422   1.020   2.542

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.44091    0.44497  -5.486 4.12e-08 ***
sex1         -0.80041    0.10174  -7.867 3.63e-15 ***
catet2        2.20881    0.43456   5.083 3.72e-07 ***
catet3        3.59554    0.43692   8.229  < 2e-16 ***
cmi1          0.44267    0.18415   2.404  0.01622 *
cmi2          1.42092    0.57420   2.475  0.01334 *
cmi3         14.24899  287.82261   0.050  0.96052
dom_pro2     -0.70215    0.15616  -4.496 6.92e-06 ***
dom_pro3     -0.03322    0.18315  -0.181  0.85608
dom_pro4      0.17882    0.12318   1.452  0.14660
dom_pro5      0.16099    0.16403   0.981  0.32635
dom_pro6      0.49249    0.19007   2.591  0.00957 **
dxindi1q2    -0.12466    0.14681  -0.849  0.39582
dxindi1q3    -0.15239    0.14645  -1.041  0.29806
dxindi1q4     0.05069    0.15222   0.333  0.73913
dxindi1q5     0.08369    0.15051   0.556  0.57820
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3038.4  on 2197  degrees of freedom
Residual deviance: 2617.3  on 2182  degrees of freedom
  (1261 observations deleted due to missingness)
AIC: 2649.3

Number of Fisher Scoring iterations: 12
```

Time period: 1st wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: PCA individual

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3021  -1.0361  -0.4195   1.0104   2.5480

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.15772    0.44918  -4.804 1.56e-06 ***
sex1         -0.77175    0.09754  -7.913 2.52e-15 ***
catet2        2.09262    0.43133   4.851 1.23e-06 ***
catet3        3.34610    0.44220   7.567 3.82e-14 ***
cmi1          0.42419    0.18490   2.294   0.0218 *
cmi2          1.38825    0.57122   2.430   0.0151 *
cmi3         14.14522  291.27836   0.049   0.9613
dom_pro2     -0.68824    0.15578  -4.418 9.96e-06 ***
dom_pro3     -0.01014    0.18327  -0.055   0.9559
dom_pro4      0.18077    0.12315   1.468   0.1421
dom_pro5      0.17725    0.16398   1.081   0.2797
dom_pro6      0.48745    0.18983   2.568   0.0102 *
pcadi2       -0.27684    0.15023  -1.843   0.0654 .
pcadi3       -0.22849    0.16517  -1.383   0.1665
pcadi4       -0.24894    0.16546  -1.504   0.1325
pcadi5       -0.18392    0.16914  -1.087   0.2769
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3038.4  on 2197  degrees of freedom
Residual deviance: 2616.9  on 2182  degrees of freedom
  (1261 observations deleted due to missingness)
AIC: 2648.9

Number of Fisher Scoring iterations: 12
```

Time period: 1st wave
Outcome: 0: alive, 1: death
DI: geographical

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.57198  -0.71507  -0.31026  -0.00015   2.72559

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -18.564975 380.471135  -0.049 0.961083
sex1         -0.522946   0.111609  -4.686 2.79e-06 ***
catet2       15.528438 380.471129   0.041 0.967444
catet3       17.957559 380.471116   0.047 0.962355
cmi1          0.576162   0.163960   3.514 0.000441 ***
cmi2          1.160969   0.319004   3.639 0.000273 ***
cmi3          2.518918   1.599514   1.575 0.115302
dom_pro2     -0.130256   0.184720  -0.705 0.480714
dom_pro3     -0.008505   0.195748  -0.043 0.965346
dom_pro4      0.303372   0.146249   2.074 0.038046 *
dom_pro5      0.311293   0.192179   1.620 0.105273
dom_pro6      0.035769   0.240295   0.149 0.881669
deprindex2    0.026893   0.172494   0.156 0.876107
deprindex3    0.027269   0.173420   0.157 0.875054
deprindex4    0.293541   0.179575   1.635 0.102124
deprindex5    0.147628   0.171979   0.858 0.390667
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2714.8  on 2961  degrees of freedom
Residual deviance: 2090.2  on 2946  degrees of freedom
  (497 observations deleted due to missingness)
AIC: 2122.2

Number of Fisher Scoring iterations: 17
```

Time period: 1st wave
Outcome: 0: alive, 1: death
DI: z-score individual

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + dxindi1q,
    family = "binomial", data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7120  -0.3984  -0.3030  -0.2283   2.7386

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -17.55536  403.31175  -0.044 0.965281
sex1         -0.65788    0.14935  -4.405 1.06e-05 ***
catet2       14.57908  403.31173   0.036 0.971164
catet3       16.87493  403.31172   0.042 0.966626
cmi1          0.86529    0.19824   4.365 1.27e-05 ***
cmi2          1.47531    0.42966   3.434 0.000596 ***
cmi3          1.96762    1.95124   1.008 0.313265
dom_pro2     -0.37271    0.22682  -1.643 0.100340
dom_pro3     -0.18019    0.25737  -0.700 0.483848
dom_pro4      0.11575    0.17586   0.658 0.510420
dom_pro5      0.12705    0.23849   0.533 0.594205
dom_pro6     -0.09855    0.26145  -0.377 0.706210
dxindi1q2     0.01644    0.24260   0.068 0.945970
dxindi1q3     0.35548    0.21824   1.629 0.103351
dxindi1q4     0.28083    0.23080   1.217 0.223691
dxindi1q5     0.31457    0.23816   1.321 0.186555
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1879.7  on 2166  degrees of freedom
Residual deviance: 1440.4  on 2151  degrees of freedom
  (1292 observations deleted due to missingness)
AIC: 1472.4

Number of Fisher Scoring iterations: 16
```

Time period: 1st wave
Outcome: 0: alive, 1: death
DI: PCA individual

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7453  -0.4366  -0.2960  -0.2114   2.7634

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -16.90380  403.18353  -0.042 0.966558
sex1         -0.66020    0.14132  -4.672 2.99e-06 ***
catet2       14.45809  403.18351   0.036 0.971394
catet3       16.57434  403.18351   0.041 0.967209
cmi1          0.84880    0.19828   4.281 1.86e-05 ***
cmi2          1.45571    0.42706   3.409 0.000653 ***
cmi3          1.99778    2.04105   0.979 0.327679
dom_pro2     -0.35037    0.22616  -1.549 0.121327
dom_pro3     -0.13767    0.25713  -0.535 0.592363
dom_pro4      0.14310    0.17563   0.815 0.415195
dom_pro5      0.15081    0.23841   0.633 0.527021
dom_pro6     -0.09179    0.26005  -0.353 0.724103
pcadi2       -0.33973    0.19857  -1.711 0.087110 .
pcadi3       -0.35291    0.23153  -1.524 0.127444
pcadi4       -0.42294    0.26187  -1.615 0.106293
pcadi5       -0.68438    0.28285  -2.420 0.015538 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1879.7  on 2166  degrees of freedom
Residual deviance: 1436.4  on 2151  degrees of freedom
  (1292 observations deleted due to missingness)
AIC: 1468.4

Number of Fisher Scoring iterations: 16
```

Time period: 2nd wave
Outcome: 0: negative, 1: positive
DI: geographical

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6457  -1.2016   0.8084   1.0487   2.1421

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.10677    0.02015   5.300 1.16e-07 ***
sex1        -0.02059    0.01120  -1.838   0.0661 .
catet2      -0.10354    0.01457  -7.105 1.20e-12 ***
catet3      -0.37892    0.01780 -21.292  < 2e-16 ***
cmi1        -0.57809    0.02908 -19.880  < 2e-16 ***
cmi2        -0.85234    0.06786 -12.560  < 2e-16 ***
cmi3        -1.78704    0.29980  -5.961 2.51e-09 ***
dom_pro2    -0.49383    0.02087 -23.660  < 2e-16 ***
dom_pro3     0.38723    0.02005  19.309  < 2e-16 ***
dom_pro4     0.61975    0.01681  36.858  < 2e-16 ***
dom_pro5    -0.97013    0.01914 -50.694  < 2e-16 ***
dom_pro6     0.09614    0.01794   5.358 8.41e-08 ***
deprindex2   0.11018    0.01894   5.818 5.97e-09 ***
deprindex3   0.20926    0.01859  11.255  < 2e-16 ***
deprindex4   0.32894    0.01843  17.845  < 2e-16 ***
deprindex5   0.32782    0.01858  17.641  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 191197  on 138224  degrees of freedom
Residual deviance: 181068  on 138209  degrees of freedom
  (19620 observations deleted due to missingness)
AIC: 181100

Number of Fisher Scoring iterations: 4
```

Time period: 2nd wave
Outcome: 0: negative, 1: positive
DI: z-score individual

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + dxindi1q,
    family = "binomial", data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6284  -1.1947   0.8144   1.0520   2.1578

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.004621   0.028776   0.161  0.87241
sex1        -0.071111   0.013382  -5.314 1.07e-07 ***
catet2       0.107068   0.024892   4.301 1.70e-05 ***
catet3      -0.253779   0.027192  -9.333  < 2e-16 ***
cmi1        -0.612193   0.032600 -18.779  < 2e-16 ***
cmi2        -0.942392   0.076753 -12.278  < 2e-16 ***
cmi3        -1.850297   0.322541  -5.737 9.66e-09 ***
dom_pro2    -0.539141   0.024103 -22.369  < 2e-16 ***
dom_pro3     0.460919   0.022579  20.414  < 2e-16 ***
dom_pro4     0.603803   0.019113  31.592  < 2e-16 ***
dom_pro5    -1.033980   0.021484 -48.127  < 2e-16 ***
dom_pro6     0.065823   0.020477   3.214  0.00131 **
dxindi1q2    0.217884   0.020391  10.685  < 2e-16 ***
dxindi1q3    0.202803   0.020788   9.756  < 2e-16 ***
dxindi1q4    0.301656   0.020013  15.073  < 2e-16 ***
dxindi1q5    0.261839   0.020314  12.890  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144445  on 104468  degrees of freedom
Residual deviance: 137086  on 104453  degrees of freedom
  (53376 observations deleted due to missingness)
AIC: 137118

Number of Fisher Scoring iterations: 4
```

Time period: 2nd wave
Outcome: 0: negative, 1: positive
DI: PCA individual

```
Call:
glm(formula = covid ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6342  -1.1829   0.8161   1.0442   2.1343

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.34968    0.03184  10.981  < 2e-16 ***
sex1        -0.04548    0.01301  -3.497 0.000471 ***
catet2       0.07396    0.02429   3.045 0.002327 **
catet3      -0.35221    0.02970 -11.860  < 2e-16 ***
cmi1        -0.61907    0.03262 -18.977  < 2e-16 ***
cmi2        -0.94559    0.07679 -12.314  < 2e-16 ***
cmi3        -1.84948    0.32279  -5.730 1.01e-08 ***
dom_pro2    -0.54074    0.02411 -22.426  < 2e-16 ***
dom_pro3     0.45803    0.02259  20.275  < 2e-16 ***
dom_pro4     0.60643    0.01912  31.719  < 2e-16 ***
dom_pro5    -1.03026    0.02149 -47.947  < 2e-16 ***
dom_pro6     0.06100    0.02049   2.977 0.002912 **
pcadi2      -0.05640    0.02178  -2.590 0.009607 **
pcadi3      -0.05600    0.02356  -2.377 0.017448 *
pcadi4      -0.19110    0.02352  -8.125 4.48e-16 ***
pcadi5      -0.39364    0.02450 -16.069  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144445  on 104468  degrees of freedom
Residual deviance: 136993  on 104453  degrees of freedom
  (53376 observations deleted due to missingness)
AIC: 137025

Number of Fisher Scoring iterations: 4
```

Time period: 2nd wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: geographical

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3243  -0.4154  -0.3212  -0.1667   3.1965

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.32122    0.08720 -49.557  < 2e-16 ***
sex1        -0.57183    0.02769 -20.651  < 2e-16 ***
catet2       1.83535    0.07957  23.066  < 2e-16 ***
catet3       3.39790    0.08008  42.433  < 2e-16 ***
cmi1         0.65675    0.05274  12.453  < 2e-16 ***
cmi2         0.97408    0.11973   8.136 4.09e-16 ***
cmi3         1.10212    0.63236   1.743 0.081357 .
dom_pro2    -0.20983    0.06354  -3.302 0.000959 ***
dom_pro3    -0.13520    0.04654  -2.905 0.003672 **
dom_pro4     0.11876    0.03516   3.378 0.000731 ***
dom_pro5     0.01267    0.05820   0.218 0.827642
dom_pro6    -0.03034    0.04435  -0.684 0.493943
deprindex2   0.05176    0.04993   1.037 0.299864
deprindex3   0.07909    0.04798   1.648 0.099299 .
deprindex4   0.02252    0.04699   0.479 0.631714
deprindex5   0.16947    0.04596   3.687 0.000227 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 44205  on 72937  degrees of freedom
Residual deviance: 37996  on 72922  degrees of freedom
  (8873 observations deleted due to missingness)
AIC: 38028

Number of Fisher Scoring iterations: 7
```

Time period: 2nd wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: z-score individual

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + dxindi1q, family = "binomial",
    data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4355  -0.4170  -0.3368  -0.2880   3.1040

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.96950    0.12119 -32.753  < 2e-16 ***
sex1        -0.64359    0.03260 -19.742  < 2e-16 ***
catet2       1.45130    0.11608  12.502  < 2e-16 ***
catet3       3.00705    0.11613  25.893  < 2e-16 ***
cmi1         0.67491    0.05939  11.365  < 2e-16 ***
cmi2         1.11099    0.13673   8.126 4.45e-16 ***
cmi3         0.80396    0.71579   1.123  0.26136
dom_pro2    -0.19624    0.06893  -2.847  0.00441 **
dom_pro3    -0.11705    0.05001  -2.341  0.01924 *
dom_pro4     0.11966    0.03880   3.084  0.00205 **
dom_pro5    -0.04839    0.06259  -0.773  0.43948
dom_pro6    -0.07925    0.04894  -1.620  0.10533
dxindi1q2    0.06460    0.04958   1.303  0.19260
dxindi1q3    0.11759    0.04922   2.389  0.01689 *
dxindi1q4    0.14982    0.04905   3.055  0.00225 **
dxindi1q5    0.32068    0.04958   6.468 9.93e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35445  on 55380  degrees of freedom
Residual deviance: 31298  on 55365  degrees of freedom
  (26430 observations deleted due to missingness)
AIC: 31330

Number of Fisher Scoring iterations: 6
```

Time period: 2nd wave
Outcome: 0: no hospitalization, 1: hospitalization
DI: PCA individual

```
Call:
glm(formula = osp ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3980  -0.4148  -0.3438  -0.2880   3.0719

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.44057    0.12181 -28.246  < 2e-16 ***
sex1        -0.62390    0.03117 -20.016  < 2e-16 ***
catet2       1.35817    0.11539  11.771  < 2e-16 ***
catet3       2.72529    0.11855  22.989  < 2e-16 ***
cmi1         0.66834    0.05942  11.248  < 2e-16 ***
cmi2         1.10216    0.13683   8.055 7.96e-16 ***
cmi3         0.72211    0.71610   1.008  0.31327
dom_pro2    -0.19927    0.06898  -2.889  0.00387 **
dom_pro3    -0.12034    0.05004  -2.405  0.01619 *
dom_pro4     0.11823    0.03884   3.044  0.00233 **
dom_pro5    -0.05282    0.06265  -0.843  0.39919
dom_pro6    -0.09281    0.04901  -1.894  0.05827 .
pcadi2      -0.27909    0.04533  -6.157 7.41e-10 ***
pcadi3      -0.35223    0.05190  -6.787 1.14e-11 ***
pcadi4      -0.44556    0.05399  -8.253  < 2e-16 ***
pcadi5      -0.40308    0.05831  -6.913 4.75e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35445  on 55380  degrees of freedom
Residual deviance: 31254  on 55365  degrees of freedom
  (26430 observations deleted due to missingness)
AIC: 31286

Number of Fisher Scoring iterations: 6
```

Time period: 2nd wave
Outcome: 0: alive, 1: death
DI: geographical

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + deprindex,
    family = "binomial", data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9090  -0.1439  -0.1177  -0.0881   4.2462

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.42174    0.58155 -14.481  < 2e-16 ***
sex1        -0.69573    0.05298 -13.133  < 2e-16 ***
catet2       3.56841    0.57902   6.163 7.14e-10 ***
catet3       6.27777    0.57708  10.878  < 2e-16 ***
cmi1         0.57457    0.07959   7.219 5.22e-13 ***
cmi2         0.84815    0.16513   5.136 2.80e-07 ***
cmi3         0.77785    1.09688   0.709   0.4782
dom_pro2     0.09006    0.11724   0.768   0.4424
dom_pro3     0.12084    0.08844   1.366   0.1718
dom_pro4     0.44170    0.06443   6.855 7.12e-12 ***
dom_pro5     0.04957    0.11774   0.421   0.6737
dom_pro6    -0.08106    0.08953  -0.905   0.3652
deprindex2   0.18371    0.09328   1.969   0.0489 *
deprindex3   0.13771    0.09082   1.516   0.1294
deprindex4   0.02474    0.08984   0.275   0.7831
deprindex5   0.16753    0.08665   1.933   0.0532 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 15714  on 68861  degrees of freedom
Residual deviance: 12047  on 68846  degrees of freedom
  (12949 observations deleted due to missingness)
AIC: 12079

Number of Fisher Scoring iterations: 10
```

Time period: 2nd wave
Outcome: 0: alive, 1: death
DI: z-score individual

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + dxindi1q,
    family = "binomial", data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0157  -0.1515  -0.1167  -0.0846   3.9662

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.83803    0.71445 -10.971  < 2e-16 ***
sex1        -0.96634    0.06829 -14.150  < 2e-16 ***
catet2       2.87813    0.71088   4.049 5.15e-05 ***
catet3       5.51530    0.70875   7.782 7.16e-15 ***
cmi1         0.60874    0.09187   6.626 3.44e-11 ***
cmi2         0.99040    0.18820   5.263 1.42e-07 ***
cmi3         0.79492    1.11169   0.715 0.474574
dom_pro2     0.08891    0.13081   0.680 0.496724
dom_pro3     0.08488    0.09813   0.865 0.387080
dom_pro4     0.34002    0.07440   4.570 4.88e-06 ***
dom_pro5    -0.09181    0.13339  -0.688 0.491239
dom_pro6    -0.10087    0.10058  -1.003 0.315892
dxindi1q2    0.25151    0.11088   2.268 0.023314 *
dxindi1q3    0.29477    0.10407   2.832 0.004621 **
dxindi1q4    0.38169    0.10593   3.603 0.000314 ***
dxindi1q5    0.59939    0.10918   5.490 4.02e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11961.2  on 52325  degrees of freedom
Residual deviance:  9323.6  on 52310  degrees of freedom
  (29485 observations deleted due to missingness)
AIC: 9355.6

Number of Fisher Scoring iterations: 10
```

Time period: 2nd wave
Outcome: 0: alive, 1: death
DI: PCA individual

```
Call:
glm(formula = morto ~ sex + catet + cmi + dom_pro + pcadi, family = "binomial",
    data = df_gml3)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9766  -0.1476  -0.1133  -0.0863   3.8650

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.87617    0.71253  -9.650  < 2e-16 ***
sex1        -0.93084    0.06466 -14.395  < 2e-16 ***
catet2       2.76211    0.71058   3.887 0.000101 ***
catet3       5.04828    0.71027   7.108 1.18e-12 ***
cmi1         0.59642    0.09201   6.482 9.06e-11 ***
cmi2         0.99699    0.18867   5.284 1.26e-07 ***
cmi3         0.67871    1.11106   0.611 0.541291
dom_pro2     0.08272    0.13101   0.631 0.527762
dom_pro3     0.07496    0.09829   0.763 0.445682
dom_pro4     0.33828    0.07455   4.537 5.70e-06 ***
dom_pro5    -0.09836    0.13369  -0.736 0.461896
dom_pro6    -0.12900    0.10071  -1.281 0.200254
pcadi2      -0.40056    0.08609  -4.653 3.27e-06 ***
pcadi3      -0.54596    0.10878  -5.019 5.19e-07 ***
pcadi4      -0.93162    0.13553  -6.874 6.25e-12 ***
pcadi5      -0.85907    0.14706  -5.842 5.16e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11961.2  on 52325  degrees of freedom
Residual deviance:  9271.9  on 52310  degrees of freedom
  (29485 observations deleted due to missingness)
AIC: 9303.9

Number of Fisher Scoring iterations: 10
```