

# COMPRIME - Conhecer Mais PaRa Intervir Melhor: Preliminary Mapping of Municipal Level Determinants of COVID-19 Transmission in Portugal at Different Moments of the 1st Epidemic Wave

Paulo Sousa<sup>a, b</sup> Nuno Marques da Costa<sup>c</sup> Eduarda Marques da Costa<sup>c</sup>  
Jorge Rocha<sup>c</sup> Vasco Ricoca Peixoto<sup>a, b</sup> Adalberto Campos Fernandes<sup>a, b</sup>  
Rogério Gaspar<sup>d</sup> Filipa Duarte-Ramos<sup>e</sup> Patrícia Abrantes<sup>c</sup> Andreia Leite<sup>a, b</sup>

<sup>a</sup>NOVA National School of Public Health, Public Health Research Center, Universidade NOVA de Lisboa, Lisbon, Portugal; <sup>b</sup>Comprehensive Health Research Center, Universidade NOVA de Lisboa, Lisbon, Portugal; <sup>c</sup>Centro de Estudos Geográficos, Universidade de Lisboa, Lisbon, Portugal; <sup>d</sup>Faculdade de Farmácia, Institute for Bioengineering and Biosciences, Universidade de Lisboa, Lisbon, Portugal; <sup>e</sup>Faculdade de Farmácia da Universidade de Lisboa e EPIUnit, Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal

## Keywords

Municipal level · COVID-19 · Pandemics · Linear model · Non-linear model

## Abstract

**Background:** The role of demographic and socio-economic determinants of COVID-19 transmission is still unclear and is expected to vary in different contexts and epidemic periods. Exploring such determinants may generate a hypothesis about transmission and aid the definition of prevention strategies. **Objectives:** To identify municipality-level demographic and socio-economic determinants of COVID-19 in Portugal. **Methods:** We assessed determinants of COVID-19 daily cases at 4 moments of the first COVID-19 epidemic wave in Portugal, related with lockdown and post-lockdown measures. We selected 60 potential determinants from 5 dimensions: population and settlement, disease, economy, social context, and mobility. We conducted a multiple linear regression (MLR) stepwise analysis ( $p < 0.05$ ) and an artificial neural network (ANN) analysis with the variables to identify

predictors of the number of daily cases. **Results:** For MLR, some of the identified variables were: resident population and population density, exports, overnight stays in touristic facilities, the location quotient of employment in accommodation, catering and similar activities, education, restaurants and lodging, some industries and building construction, the share of the population working outside the municipality, the net migration rate, income, and renting. In ANN, some of the identified variables were: population density and resident population, urbanization, students in higher education, income, exports, social housing buildings, production services employment, and the share of the population working outside the municipality of residence. **Conclusions:** Several factors were identified as possible determinants of COVID-19 transmission at the municipality level. Despite limitations to the study, we believe that this information should be considered to promote communication and prevention approaches. Further research should be conducted.

© 2021 The Author(s) Published by S. Karger AG, Basel on behalf of NOVA National School of Public Health

**COMPRIME - Conhecer Mais PaRa Intervir MELHOR:  
Análise preliminar de fatores determinantes da  
transmissão da COVID-19 em Portugal, a nível  
municipal, em diferentes momentos da 1ª onda  
epidémica**

**Palavra Chave**

Nível municipal · COVID-19 · Pandemia · Modelo linear ·  
Modelo não linear

**Resumo**

Contexto: O papel dos determinantes demográficos e socioeconómicos na transmissão do vírus SARS Cov2 ainda não é claro e acredita-se que varie em diferentes contextos e períodos da pandemia. A análise desses determinantes pode ajudar a gerar hipóteses sobre a transmissão e apoiar na definição de estratégias de prevenção. Objetivos: Identificar os determinantes demográficos e socioeconómicos que podem estar associados a maior transmissibilidade da COVID-19 ao nível do município em Portugal. Métodos: Pretende-se avaliar quais os determinantes que mais influenciam o número de casos diários de COVID-19 em 4 momentos entre março e junho (corresponde à primeira vaga da pandemia) em Portugal. Foram selecionados 60 indicadores de 5 dimensões: populacional, prevalência de doenças, economia, contexto social e mobilidade. Realizamos análises de regressão linear múltipla (RLM) ( $p < 0,05$ ) e análise de rede neural artificial (RNA) para identificar preditores do número de casos diários. Resultados: Para RML, algumas das variáveis identificadas foram: população residente e densidade populacional, exportações, dormidas em instalações turísticas, educação, restauração e alojamento, algumas indústrias e construção civil, proporção da população que trabalha fora do município, taxa de migração, entre outros. Na RNA, algumas das variáveis identificadas foram: densidade populacional e população residente, urbanização, alunos do ensino superior, exportações, edifícios de habitação social, emprego nos serviços de produção e parcela da população que trabalha fora do município de residência. Conclusões: Vários fatores foram identificados como possíveis determinantes da transmissibilidade da COVID-19 ao nível municipal. Apesar das limitações do estudo, acreditamos que estes resultados podem contribuir para apoiar tomadas de decisão e abordagens de comunicação e prevenção.

© 2021 The Author(s) Published by S. Karger AG, Basel  
on behalf of NOVA National School of Public Health

**Introduction**

The COVID-19 pandemic represents a global threat and poses challenges for health, economy, and well-being. This highlights the importance of analysing robust and timely data to support decisions regarding the implementation of public health measures at the national, regional, and municipal level. It has thus been recommended that epidemiological studies should consider multi-level investigations of reliable and representative environmental, societal, and population determinants [1]. Behavioural, socioeconomic, and community factors, control measures, the effects of population mixing, and the use of appropriate spatial and temporal resolution and time frames all need to be carefully investigated [1].

The evidence from previous pandemics indicates that disadvantaged groups have been disproportionately affected [2, 3]. The determinants of COVID-19 transmission are still uncertain, but previous studies suggest that population density, overcrowding, mobility, and socio-economic status are potentially relevant [2, 4, 5]. These seem to vary according to contextual specific factors and moments in time, however. A small number of studies have attempted to identify municipality-level determinants of transmission using methods such as multiple linear regression (MLR) and neural network analysis [6, 7]. One study mapped county (municipality-level) determinants of COVID-19 transmission in nursing homes in the USA, and found that factors like per-capita income, average household size, population density, and minority composition were significant predictors of COVID-19 cases in nursing homes [6].

The social determinants of health are interrelated and likely to play a major role in the COVID-19 pandemic. Education level influences occupation, which determines economic stability and income level, which can, in turn, impact the type of health care and health-seeking behaviour. Simultaneously, education might influence in which neighbourhood an individual lives, i.e., determining their social and community context [7]. This intricate network makes the study of causality difficult and must be considered with caution. Nevertheless, it is relevant to assess the abovementioned factors and generate a hypothesis on how they influence the spread of COVID-19.

We thus aimed to identify the municipality-level determinants of COVID-19 cases in Portugal at 4 moments of the first epidemic wave.

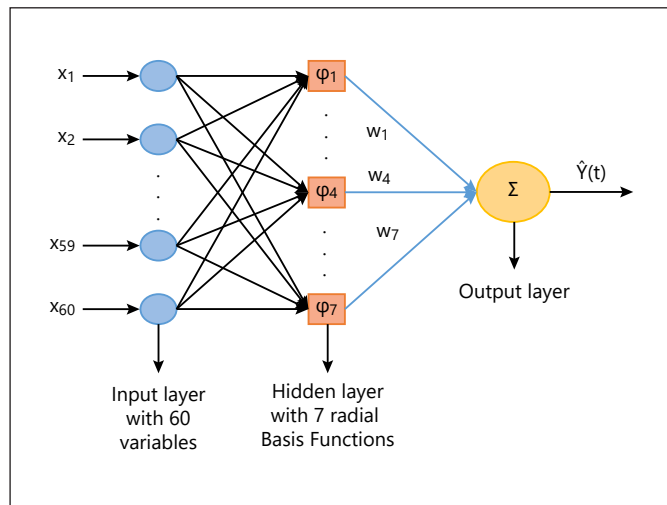
**Table 1.** Summary results of linear regression models

Dimensions	Indicators	March 23	May 28	June 8	June 27
Disease (2019, by group)	Infectious				
	Cancer				
	Blood and other diseases involving the immune system				
	Endocrine, nutritional, and metabolic				
	Cardiovascular				
	Mental illness and substance abuse	+	+		
	Respiratory			+	
	Digestive				
Urinary and genital system					
Economy	Exports (% of the of the country) 2019	+		+	
	Overnight stays 2019	+	+		
	Food industry location quotient 2018				+
	Beverage industry location quotient 2018		+	+	
	Wholesale and retail location quotient 2018				
	Trade location quotient, maintenance and repair of motor vehicles and motorcycles 2018				
	Wholesale location quotient 2018				
	Transport and storage location quotient 2018				
	Location quotient of storage activities and transport ancillary activities				
	Location quotient of accommodation, catering and similar activities 2018	+	+		
	Restoration location quotient and similar 2018			+	
	Construction location quotient 2018		+	+	+
	Agriculture location quotient, animal production, hunting, forestry, and fisheries 2018				
	Textile manufacturing location quotient 2018				
	Location quotient for the manufacture of chemicals and synthetic or artificial fibres				
	Location quotient of the manufacture of electrical equipment 2019		+	+	
	Education location quotient 2018			+	+
	Location quotient of social support activities with housing 2018				
	Employment in agriculture, animal production, hunting, forestry, and fisheries (% at national level) 2018				
	Employment in manufacturing industries (% at national level) 2018				
Employment in traditional industries (% at national level) 2018					
Employment in housing, catering, etc. (% at national level) 2018					
Employment in production services (% at national level) 2018					
Mobility	% of the population working in another parish of the municipality 2011				
	% of the population working outside the municipality 2011	+			+
	% of movements with car use 2011				
	Infectious diseases 2019				
Population and settlement	% of the population aged 25–64 years 2019		+		
	% of the population aged 65–74 years 2019				
	% of the population aged ≥75 years 2019				
	Ageing index 2019				
	Net migration rate 2019		+	+	+
	% of legal immigrants 2019				
	Resident population 2019	+	+	+	+
	Population density 2019	+	+	+	+
Urbanization rate 2013					

**Table 1** (continued)

Dimensions	Indicators	March 23	May 28	June 8	June 27
Social context	Number of pensioners/1,000 inhabitants 2018				
	Declared gross income deducted from the IRS settled by taxable person (in EUR) 2018			-	-
	Pre-school/1,000 inhabitants 2018/2019				
	Students in 1 <sup>st</sup> cycle of basic education/1,000 inhabitants 2018/2019				
	Students in 2nd cycle of basic education/1,000 inhabitants 2018/2019				
	Students in 3rd cycle of basic education/1,000 inhabitants 2018/2019				
	Students in secondary education/1,000 inhabitants 2018/2019	-	-		
	Students in higher education/1,000 inhabitants 2018/2019				
	Rate of basic education 2011				
	Rate of superior education 2011	+			
	Rental prices 2011		-	-	-
	% of individuals with socially valued professions 2011				
	Average family size 2011				
	Buildings in need of repair/100 inhabitants 2011				
Social housing buildings/100 inhabitants 2011					
Number of independent variables in the final model		9	12	12	8
R		0.86	0.94	0.95	0.96
R <sup>2</sup>		0.73	0.88	0.91	0.93
R <sup>2</sup> adjusted		0.72	0.87	0.91	0.92

+/-, Positively/negatively associated statistically significant variables in final models selected through backward elimination. To know more about these variables, see the Project Online Dashboard [7] (<https://www.comprime-compri-mov.com/index.html>).



**Fig. 1.** Neural clustering method.

## Materials and Methods

We conducted an ecological study to analyse the association of 65 municipal-level variables from official statistics drawn from 5 dimensions, i.e., population and settlement, disease, economy, social context, and mobility, and the number of daily cases per municipality at 4 pre-defined moments (taken from the official surveillance system). The dates for this analysis were selected accord-

ing to the public health measures in place, i.e., the date of publication of guidelines/legal documents and the maximum number of cases (determined by the 3-day moving average) occurring in the following 2 weeks. Four periods were selected, starting on March 23 (the 1st day with information available per county [lockdown phase]), May 28 (the 1st phase of the gradual resumption of activities), June 8 (to evaluate the effects of the 2nd phase of the gradual resumption of activities), and June 27 (the gradual resumption of activities after the 3rd phase).

For each moment of analysis, we used a multivariate linear model (MLR) and a nonlinear model, i.e., artificial neural networks (ANN). MLR identified the strength of association between each independent variable and the outcome (number of cases). The variables presented in the final MLR were selected by backward elimination until all remaining variables had a  $p$  value  $< 0.05$ . Results were summarized for each moment showing variables included in the final models.

ANN constitute a non-linear parametric model, with the advantage of implicitly detecting non-linear relationships between the outcome and explanatory variables. ANN have been used to identify risk factors for different health outcomes, including reported incidences of COVID-19 at the county/municipality level [8–10]. As there is no need for independence and normality of the variables, applying ANN in the analysis of epidemiological data is attractive. In addition, neural processing is able to extract relationships from input variables directly over high-dimensional spaces, making such processing a valuable tool in complex pattern recognition problems. The selected non-linear approximation implemented is depicted in Figure 1. For details, please refer to the project website [11].

**Table 2.** Results of the artificial neural network analysis

March 23 <sup>rd</sup>		May 28 <sup>th</sup>		June 8 <sup>th</sup>		June 27 <sup>th</sup>		
variables	importance standardized importance, %	variables	importance standardized importance, %	variables	importance standardized importance, %	variables	importance standardized importance, %	
Urbanization rate 2013	0.036	100.00	0.029	100.00	0.041	100.00	0.036	100.00
Buildings not needing repair 2011	0.036	100.00	0.028	98.60	0.037	90.60	0.036	99.90
Buildings needing minor repairs 2011	0.032	89.40	0.028	97.50	0.035	85.80	0.032	89.90
Buildings needing repair 2011	0.031	85.70	0.025	86.00	0.03	74.90	0.031	87.50
Buildings in need of medium repairs 2011	0.029	79.90	0.023	79.10	0.03	73.60	0.027	76.00
Buildings needing major repairs 2011	0.028	78.00	0.022	76.90	0.029	71.00	0.027	76.00
Resident population 2019	0.028	77.40	0.021	74.70	0.028	69.30	0.025	70.30
Population density 2019	0.026	72.10	0.021	73.30	0.028	68.50	0.023	63.60
Buildings in need of extensive repairs 2011	0.026	71.00	0.021	72.20	0.027	67.40	0.023	62.90
Infectious diseases 2019	0.025	70.20	0.021	72.20	0.026	64.40	0.022	60.70
Students in higher education (n) 2018/2019	0.025	69.40	0.021	71.60	0.025	60.70	0.021	60.00
Rate of higher education 2011	0.024	67.00	0.02	70.70	0.024	59.50	0.02	56.80
Students in 1st cycle of basic education/1,000 inhabitants, 2018/19	0.023	64.00	0.02	70.50	0.023	57.80	0.02	54.90
Income/inhabitant	0.023	63.60	0.02	69.90	0.022	54.20	0.018	51.50
% (of the whole country) of people employed in production services 2019	0.023	63.20	0.019	66.90	0.02	48.00	0.018	50.90

**Table 2** (continued)

March 23rd	May 28th		June 8th		June 27th			
variables	importance	standardized importance, %	variables	importance	standardized importance, %	variables	importance	standardized importance, %
% of the population with a socially valued profession	0.023	62.70	Buildings in need of medium repairs 2011	0.019	65.40	Textile manufacturing employment - QL	0.019	46.20
% (of the whole country) of exports by the municipality 2019	0.02	56.20	% (of the whole country) of exports by the municipality	0.019	65.40	Exports	0.019	45.80
% of people working in the parish of their residence	0.018	50.30	% of legal immigrants in the total population, 2018	0.019	65.00	Income	0.017	42.50
						Exports	0.018	50.90
						Urbanization rate	0.017	47.10
						Blood disorders and other diseases of the immune system	0.017	47.00

## Results

For MLR, some of the identified variables (Table 1) were: resident population and population density, exports, overnight stays in touristic facilities, the location quotient of employment in accommodation, catering and similar activities, education, restaurants and lodging, some industries and building construction, the share of the population working outside the municipality, the net migration rate, income, and renting. For ANN, some of the identified variables (Table 2) were: population density and resident population, urbanization, students in higher education, income, exports, social housing buildings, production services employment, and the share of the population working outside their municipality of residence. There is a communality of factors identified at different epidemic moments by both methods and specific ones emerged for each epidemic moment.

## Discussion

Our results attempted to identify municipality-level determinants of COVID-19 transmission using complementary approaches. Variables identified as being associated with the number of cases reported changed over time, emphasizing the dynamic nature of this communicable disease.

Initially, more affected areas presented international relations associated with tourism or exports (in MLR and ANN) and the socio-economic conditions of the population (more evident in ANN). Later, during the lifting of the lockdown, the epidemic surged in suburban areas with lower incomes and a higher number of immigrants, thus emphasizing the role of the socio-economic and cultural determinants of transmission (e.g., crowded housing conditions and the concentration of specific economic sectors with a high concentration of employment – building construction, beverage, and storage). Finally, at moment 4, higher-education students, 1st-cycle (of basic education) students, and urbanization became relevant. Population density and the share of people working outside their municipality of residence were identified as factors at all 4 moments and in both methods.

It has been stated that responding to COVID-19 requires continuous monitoring of environmental and societal determinants to implement adequate prevention strategies [1]. Only a few studies have attempted to relate transmission levels to community-level determinants [6, 7, 12]. One study found that per-capita income, average

household size, population density, and minority population composition were significant predictors of COVID-19 cases in nursing homes [6]. Another identified age, disability, language, race, occupation, and urban status as predictors [12]. Areas with more deprived populations and social vulnerability have been reported to have worse outcomes in terms of COVID-19 transmission [13], also at the county level [14]. Reports of deaths disproportionately affecting specific groups, e.g., those with a non-white ethnic background, have also been published [15]. Some reports are calling COVID-19 a “sindemic” due to the concurrence of social, economic, and health vulnerabilities and the exponential increase of the pandemic [3, 16]. The European Centre for Disease Prevention and Control (ECDC) also identified clusters of occupational economic activities and outbreaks in health care, food packaging and processing, factories/manufacturing, building and construction, and educational facilities [17]. Our findings are in line with these other studies.

This is a preliminary approach to the study of municipality-level determinants in Portugal and some study limitations need to be acknowledged. First, the ecological design limited the ability to determine causal relationships [18]. Second, the definition of the outcome as the daily number of cases might not have fully captured the spread of the disease; alternative definitions, e.g., changes in cases over time, could be considered in future analyses. Third, the number of COVID-19 cases identified is also influenced by surveillance system sensitivity and testing strategies [19, 20]. Accounting for these was not feasible but could be investigated in future studies, to ensure comparability over time. Finally, the definition of initially selected variables might have been too broad, as these were official statistics readily available for analysis.

There is still a lot of uncertainty regarding the actual significance of these findings and the exact role of each variable in the causal network [21]. Nonetheless, the fact that our results were consistent with those of previous studies was reassuring. Further studies should consider a more extensive analysis of several waves of the pandemic and both space and time patterns. Individual-based studies would be important to shed further light on both the determinants of transmission and the underlying mechanisms at work.

In conclusion, several factors were identified as possible determinants of COVID-19 transmission at the municipality level. Aspects regarding the socio-economic characteristics of the population showed varying relationships with COVID-19 cases, while population density and mobility-related aspects were consistently associ-

ated at all 4 moments analysed. Despite some study limitations, we believe that these preliminary results should be considered to support decisions regarding COVID-19 prevention and control measures. More studies are required to enhance the robustness of this methodological approach and its results.

### Statement of Ethics

This was an observational, ecological study using data from a secondary data source. Ethics approval and consent to participate are not applicable to this study.

### Conflict of Interest Statement

There are no conflicts of interest.

### Funding Sources

This paper presented the main results of the COMPRIME (Conhecer Mais PaRa Intervir MELhor) study, funded by the FCT (Fundação para a Ciência e Tecnologia) “RESEARCH 4 COVID-19” special support, 1st ed. ID. 596685735 (coordinated by Paulo Sousa).

### Author Contributions

P.S., E.M.C., A.C.F., and R.G.: conception and design of the work. P.S., V.R.P., and A.L.: introduction. N.M.C., J.R., P.A., and P.S.: collecting data, selection of variables, methodological implementation, and results. P.S., E.M.C., A.C.F., R.G., F.D.R., N.M.C., J.R., V.R.P., and A.L.: discussion of results. P.S., V.R.P., A.L., and A.C.F.: conclusions. All authors read and approved the final manuscript.

### References

- 1 Zeka A, Tobias A, Leonardi G, Bianchi F, Lauriola P, Crabbe H, et al.; Scientific Committee of the International Network of Public Health and Environmental Tracking. Responding to COVID-19 requires strong epidemiological evidence of environmental and societal determining factors. *Lancet Planet Health*. 2020 Sep;4(9):e375–6.
- 2 Bambra C, Riordan R, Ford J, Matthews F. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health*. 2020 Nov;74(11):964–8.
- 3 Kawachi I. COVID-19 and the ‘rediscovery’ of health inequities. *Int J Epidemiol*. 2020 Oct; 49(5):1415–8.
- 4 Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al.; Open COVID-19 Data Working Group. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. 2020 May; 368(6490):493–7.

- 5 Li R, Richmond P, Roehner BM. Effect of population density on epidemics. *Physica A*. 2018;510:713–24.
- 6 Sugg MM, Spaulding TJ, Lane SJ, Runkle JD, Harden SR, Hege A, et al. Mapping community-level determinants of COVID-19 transmission in nursing homes: A multi-scale approach. *Sci Total Environ*. 2021 Jan;752:141946.
- 7 Singu S, Acharya A, Challagundla K, Byraredy SN. Impact of social determinants of health on the emerging COVID-19 pandemic in the United States. *Front Public Health*. 2020 Jul;8:406.
- 8 Mollalo A, Rivera KM, Vahedi B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *Int J Environ Res Public Health*. 2020 Jun;17(12):1–13.
- 9 Sherriff A, Ott J; ALSPAC Study Team. Artificial neural networks as statistical tools in epidemiological studies: analysis of risk factors for early infant wheeze. *Paediatr Perinat Epidemiol*. 2004 Nov;18(6):456–63.
- 10 Chen J, Pan QS, Hong WD, Pan J, Zhang WH, Xu G, et al. Use of an artificial neural network to predict risk factors of nosocomial infection in lung cancer patients. *Asian Pac J Cancer Prev*. 2014;15(13):5349–53.
- 11 Consortium COMPRIME/COMPRI\_MOV. Projetos COMPRIME e COMPRI\_MOV. Lisboa: Consortium COMPRIME/COMPRI\_MOV; 2020. Available from <https://www.comprime-compri-mov.com/index.html> [cited November 13, 2020].
- 12 Andersen LM, Harden SR, Sugg MM, Runkle JD, Lundquist TE. Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Sci Total Environ*. 2021 Feb;754:142396.
- 13 Iacobucci G. Covid-19: deprived areas have the highest death rates in England and Wales. *BMJ*. 2020 May;369:m1810.
- 14 Dasgupta S, Bowen VB, Leidner A, Fletcher K, Musial T, Rose C, et al. Association between social vulnerability and a county's risk for becoming a COVID-19 hotspot, United States, June 1–July 25, 2020. *MMWR Morb Mortal Wkly Rep*. 2020 Oct;69(42):1535–41.
- 15 UK. Office for National Statistics (ONS). Updating ethnic contrasts in deaths involving the coronavirus (COVID-19), England and Wales: deaths occurring 2 March to 28 July 2020. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/births-deathsandmarriages/deaths/articles/updating-ethnic-contrasts-in-deaths-involving-the-coronavirus-covid-19-england-and-wales/deaths-occurring-2-march-to-28-july-2020> [cited December 22, 2020].
- 16 Horton R. Offline: COVID-19 is not a pandemic. *Lancet*. 2020 Sep;396(10255):874.
- 17 European Centre for Disease Prevention and Control. COVID-19 clusters and outbreaks in occupational settings in the EU/EEA and the UK. Stockholm: ECDC; 2020. Available from <https://www.ecdc.europa.eu/en/publications-data/covid-19-clusters-and-outbreaks-occupational-settings-eueea-and-uk> [cited November 9, 2020].
- 18 Sedgwick P. Ecological studies: advantages and disadvantages. *BMJ*. 2014 May;348:g2979.
- 19 Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CA, et al; CMMID COVID-19 working group. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med*. 2020 Oct;18(1):332.
- 20 Ricoca-Peixoto V, Nunes C, Abrantes A. Epidemic surveillance of Covid-19: considering uncertainty and under-ascertainment. *Port J Public Health*. 2020;38(1):23–9.
- 21 Ahrens W. Commentary: Socioeconomic status: more than a confounder? *Int J Epidemiol*. 2004 Aug;33(4):806–7.