Socioeconomic factors and the probability of death by Covid-19 in Brazil

F. De Negri[1], R. Galiezz[2], P. Miranda[1], P. Koeller[1], G. Zucoloto[1], J. Costa[1], C. M. Farias[3], G. H. Travassos[4], R. A. Medronho[2]

[1]Institute for Applied Economic Research (IPEA), Rio de Janeiro, Brazil
[2]Faculty of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[3]Institute Tercio Pacitti of Computational Applications and Research, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[4]Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering (COPPE), Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
Address correspondence to Fernanda De Negri, E-mail: fernanda.denegri@ipea.gov.br

ABSTRACT

Background To design better measures to contain the Covid-19 epidemics, it is relevant to know whether socioeconomic factors are associated with a higher risk of death by Covid-19. This work estimates the effects of individual socioeconomic characteristics on the risk of death by Covid-19.

Methods Logistic models were estimated to assess the effect of socioeconomic characteristics (income, race/ethnicity, schooling, occupation and economic activity) on the risk of death from Covid-19. For this purpose, Covid-19 individual death records in Rio de Janeiro state, Brazil were combined with the Annual Register of Social Information, which contains socioeconomic information about formal workers.

Findings Workers employed in establishments in the health and public safety sectors present a risk of dying 2.46 and 2.25 times higher than those employed in other activities. Non-white people, men, and those who work in the Metropolitan Region are also more likely to die from Covid-19. People with higher education are 44% less likely to die from the disease.

Conclusions Some population groups are more vulnerable to the Covid-19 pandemic and individual socioeconomic conditions play a relevant role in the probability of death by the disease. That should be considered in the design of prevention policies to be adopted.

Keywords Covid-19, mortality rates, health and inequality, labor market, health policies

Introduction

Since the WHO declared the Covid-19 pandemic a public health emergency of international concern, researchers worldwide have sought to understand the dynamics of the virus' transmission and the epidemiological characteristics of the disease. In this context, a relevant question is whether certain socioeconomic characteristics are associated with a higher Covid-19 mortality rate.

This article aims to perform a detailed, although not exhaustive, analysis of how socioeconomic factors affect the probability of death by Covid-19 in the state of Rio de Janeiro, Brazil. To this purpose, the paper uses individual level socioeconomic data available for all workers in the formal sector of Rio de Janeiro state.

Methods

Data sources

The health database of the Rio de Janeiro State Department of Health (SES-RJ) used for this study contains individualized information on ∼130 000 confirmed cases of Covid-19 in the state of Rio de Janeiro from 8 March to 4 July 2020. The database contains information on the individual's address, age, date of birth, gender, race/ethnicity, besides some clinical data, such as the date of notification and outcome of the case.

These records were combined with the Annual Report of Social Information (RAIS) database, an administrative record of employer–employee data, to obtain other socioeconomic characteristics of the individuals, such as wage, schooling and occupation. It is noteworthy that RAIS information refers only to formal workers, which have more privileged

De Negri F., Researcher

Galiezz R., Professor

Miranda P., Researcher

Koeller P., Researcher

Zucoloto G., Researcher

Costa J., Researcher

Farias C. M., Professor

Travassos G. H., Professor

Medronho R. A., Professor

socioeconomic conditions than the average of the population or informal workers. Finally, it is important to note that the latest available dataset from RAIS is for 2018.

To merge the registers of Covid-19 cases and deaths and RAIS microdata, at first, the taxpayer identification number, which exists in both datasets has been used. Since the number of uninformed identification numbers is relatively high in Covid-19 records (∼26% of the database), individuals not linked in this way were merged using name and birth date. For this, we have calculated the Levenshtein distance between two text strings—which is given by the number of modifications required to transform one text *string* into another. When this distance is zero, the names on the two databases are the same. A small limit distance was used to restrict the merge to only very similar text strings. However, this can still join different people with similar names or even homonyms. Only records with the same date of birth were merged to avoid this type of error.

At the end of these procedures, 57 889 records were found in RAIS (44.7% of the cases of Covid-19 in the state), most of them merged through the identification number. The cases not found in RAIS result of at least three main factors: (1) individuals who were not in the formal labor market in 2018 (unemployed, self-employed, informal, individual entrepreneurs or business owners, for example); (2) those who are not in the labor market in 2018, such as elderly and retired, or children and adolescents; (3) to a lesser extent, records with incomplete or inaccurate information (identification numbers and names) that made the merge with RAIS impossible.

Table 1 shows the distribution, by age group, of the total number of Covid-19 cases and deaths in the state of Rio de Janeiro and the sample associated with RAIS. The representativeness of the sample is bigger in the age ranges going from 20- to 49-years old. For those individuals above 50 years, there are fewer cases and deaths found in the RAIS dataset.

Besides that, in all age groups, the percentage of deaths obtained in the sample of formal workers is lower than the percentage of cases, which makes the lethality in the sample of formal workers lower than in the population as a whole, in all age groups. While workers who were in the formal market in 2018 accounted for 44.7% of registered cases, they represent only 13.5% of total deaths in the state.

Econometric methods

This study will use logistic models to estimate the Covid-19 mortality risk in all workers in the formal labor market. The hypothesis to be tested is that there is a set of individual socioeconomic conditions that affected the probability of death due to Covid-19 during the SARS-COV2 epidemic in Brazil.

These models use the following logistic equation:

$g(\pi) = \alpha + \beta x_i$, where:

(i) $g(\pi) = \Pr(Y = 1 \mid x)$, Y being the dummy variable indicating death due to Covid-19;

(ii) α is the intercept;

(iii) x_i is the vector of explanatory variables and;

(iv) β are the estimated coefficients.

The explanatory variables are:

(a) age and age squared;

(b) man;

(c) race/ethnicity: dummy variable identifying non-white people and another identifying those whose race/ethnicity has not been declared in the RAIS;

(d) Education: dummy variables for individuals with complete primary, secondary and higher education;

(e) Metropolitan region: dummy for establishments located in the Metropolitan Region of Rio de Janeiro;

(f) the logarithm of the average annual labor earnings of the individual in 2018;

(g) dummies for different types of occupation;

(h) dummy indicating vulnerable occupations: health, protective and transport workers;

(i) dummies for economic activity of the company/organization where the individual works;

(j) dummy for essential activities: human health, public order and safety, freight transport, courier and support activities for transportation, passenger transport, essential wholesale and retail trade, and other essential services (Food and beverage, banks, cleaning, funeral and others).

The economic activity and occupations variables were used interchangeably in different models' specifications, especially due to the high correlation between some occupations and certain economic activities.

The individual's occupation was obtained from the Brazilian Classification of Occupations (CBO), present in the RAIS database, based on the *International Standard Classification of Occupations (ISCO-88),* and aggregated into 19 groups.

The essential economic activities were defined according to the Government of Rio de Janeiro state's rules. The economic activities were divided into 13 groups of essential activities and three groups of non-essential activities, organized based on the Brazilian National Classification of Economic Activities revision 2.0 (CNAE 2.0), which is based on the International Standard Industrial Classification (ISIC, rev. 4).

Covid-19 deaths can be considered a rare event since the 1686 deaths identified in the RAIS correspond to less than

Table 1 The number of Covid-19 cases and deaths by age groups in the state of Rio de Janeiro and the RAIS: March–July 2020

Age group	Total number of cases	Total number of deaths	Number of cases in the RAIS	% of total cases in the RAIS	Number of deaths in the RAIS	% of total deaths in the RAIS
0–19	4.563	57	60	1.3%		0.0%
20–29	14.375	149	7.184	50.0%	33	22.1%
30–39	29.317	418	19.074	65.1%	160	38.3%
40–49	28.546	1.026	17.050	59.7%	353	34.4%
50–59	21.873	1.792	10.697	48.9%	528	29.5%
60–69	14.417	2.907	3.264	22.6%	452	15.5%
70–79	8.613	2.998	505	5.9%	134	4.5%
80 +	6.675	3.145	55	0.8%	26	0.8%
Uninformed	1.065	41		0.0%		0.0%
Total	129.444	12.533	57.889	44.7%	1.686	13.5%

Source: Annual Report of Social Information (RAIS) and Rio de Janeiro State Department of Health (SES-RJ).

Source: Own estimation based on RAIS and SES-RJ data.

0.1% of the total workers. King and Zeng[1] studied problems related to the statistical estimation of rare events, binary-dependent variables whose values equal to one, such as wars or epidemics, are thousands of times less numerous than the values equal to zero (the 'non-events'). They argue that traditional statistical procedures underestimate the probabilities of occurrence of this type of event. To overcome this problem, the authors suggest a Monte-Carlo type of random sampling procedure, in which samples are extracted from the original population so that the number of events is equal to that of non-events. Therefore, one thousand random samples were performed, in which the number of deaths is equal to the number of non-deaths. The sample always contained all the cases in which death from Covid-19 was equal to one, performing random sampling, without replacement, of the cases where death did not occur. The coefficients presented in these models are the average of the estimates made in each of these one thousand balanced samples.

Results

Table 2 presents the result of five different models for the probability of death due to Covid-19 among formal workers in Rio de Janeiro state. The difference between them is in the occupation and economic activity variables.

The main explanatory variables are quite consistent among the various specifications. In all models, age is statistically significant with the expected sign. Furthermore, the growth rate of this probability decreases as age increases, which is represented by the negative sign in age squared. As anticipated by the previous empirical research,[2,3] men have a significantly higher likelihood to die than women.

A fact that has already been mentioned in analyses with aggregated data is the greater vulnerability of the non-white populations to the epidemic. [4,5,6,7] This variable was positive and significant in all the specifications of the models.

People with higher education have lower risks of dying than people with no education or without primary level education, although there is no difference between those who have finished primary or secondary levels and those with lower schooling. Working in the Metropolitan Region of the city of Rio de Janeiro also results in higher risks of death from Covid-19.

The logarithm of the average annual labor income, despite presenting a positive sign, was not statistically significant in all the specifications.

The occupations or economic activities in which individuals work are the variables that differ among the models. In the first model, in addition to the basic variables, the individual occupation was included as an explanatory variable. In the second model, there is a dummy variable for essential economic activities in addition to occupation. In the third and fourth models, the occupation variables were replaced by economic activity. The choice to use either occupation or activity is because many occupations are concentrated in a few activities, generating a very high correlation between these two sets of variables. In the fourth model, a dummy was added for vulnerable occupations. Finally, the fifth model creates a new aggregation of economic activities, based on the previous one, but where health professionals in the public administration were incorporated into the Human health activities, and

Table 2 Logistic model with correction for rare events: the probability of dying of COVID-19 in the State of Rio de Janeiro-03/08/2020-07/04/2020

Variables	Model 1		Model 2		Model 3		Model 4		Model 5	
Age	0.165	***	0.165	***	0.165	***	0.159	***	0.163	***
Age 2	−0.001	***	−0.001	***	−0.001	***	−0.001	***	−0.001	***
Men	0.878	***	0.873	***	0.855	***	0.801	***	0.856	***
Non-white people	0.336	***	0.330	***	0.329	***	0.321	***	0.331	***
Uninformed race/ethnicity	0.118	*	0.148	*	−0.093	ns	−0.089	ns	−0.014	ns
Complete primary education	−0.117	ns	−0.120	ns	−0.098	ns	−0.102	ns	−0.112	ns
Complete secondary education	−0.028	ns	−0.021	ns	0.044	ns	0.044	ns	0.020	ns
Complete higher education	−0.446	***	−0.430	***	−0.439	***	−0.388	***	−0.418	***
Metropolitan region	0.853	***	0.847	***	0.904	***	0.895	***	0.880	***
Log (average annual labor earnings)	0.012	ns	0.013	ns	0.034	ns	0.014	ns	0.018	ns
Essential activities			0.121	*						
Vulnerable occupations (security, health e-transport)							0.436	***		
Occupations										
Police officers, firefighters and protective workers	0.542	**	0.490	**						
Health professionals	1.021	***	0.977	***						
Customer services clerks	0.643	**	0.599	**						
Other occupations	ns		ns							
Economic activities										
Essential wholesale and retail trade					0.266	**	0.219	*	0.263	**
Information and communications services					0.401	**	0.393	**	0.396	**
Manufacture of essential products					−0.382	*	−0.451	*	−0.382	*
Human health activities					0.782	***	0.534	***	0.899	***
Public order and safety activities					0.596		0.374		0.810	***
Other essential services					0.328	**	0.209	*	0.324	**
Public administration and defense					0.595	***	0.465	***	0.316	**
Freight transport, courier and support activities for transportation					0.275	*	0.133		0.273	*
Other activities					ns		ns		ns	

Source: Own estimation based on RAIS and SES-RJ data. Statistical significance level: *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$; ns = not significant.
Note: Intercepts not reported.

protective workers in the public administration were shifted to the public order safety activities.

The two dummies that aggregate vulnerable occupations or essential activities are always positive and significant, revealing that the workers of these occupations or activities had a greater risk of dying of Covid-19 compared to other workers. This effect is greater for the set of vulnerable occupations than for essential activities, showing that professionals in these occupations (health, protective and transport) were more affected by the epidemic.

When occupations are disaggregated, health and protective workers present a higher risk of death than other workers, besides customer services clerks. When activities are considered, the Information and communications services and the freight transport, courier activities, and support activities for transportation present a higher probability of death than the other activities. The public order and safety activities are only significant in the last specification when protective workers previously allocated in the public administration (especially police officers and firefighters) are allocated for the public order and safety activities. The manufacture of essential products seems to have a lower risk of death than the other activities, at a 10% level of significance and the public administration present a higher risk, even when health and protective workers are removed from this activity (model 5).
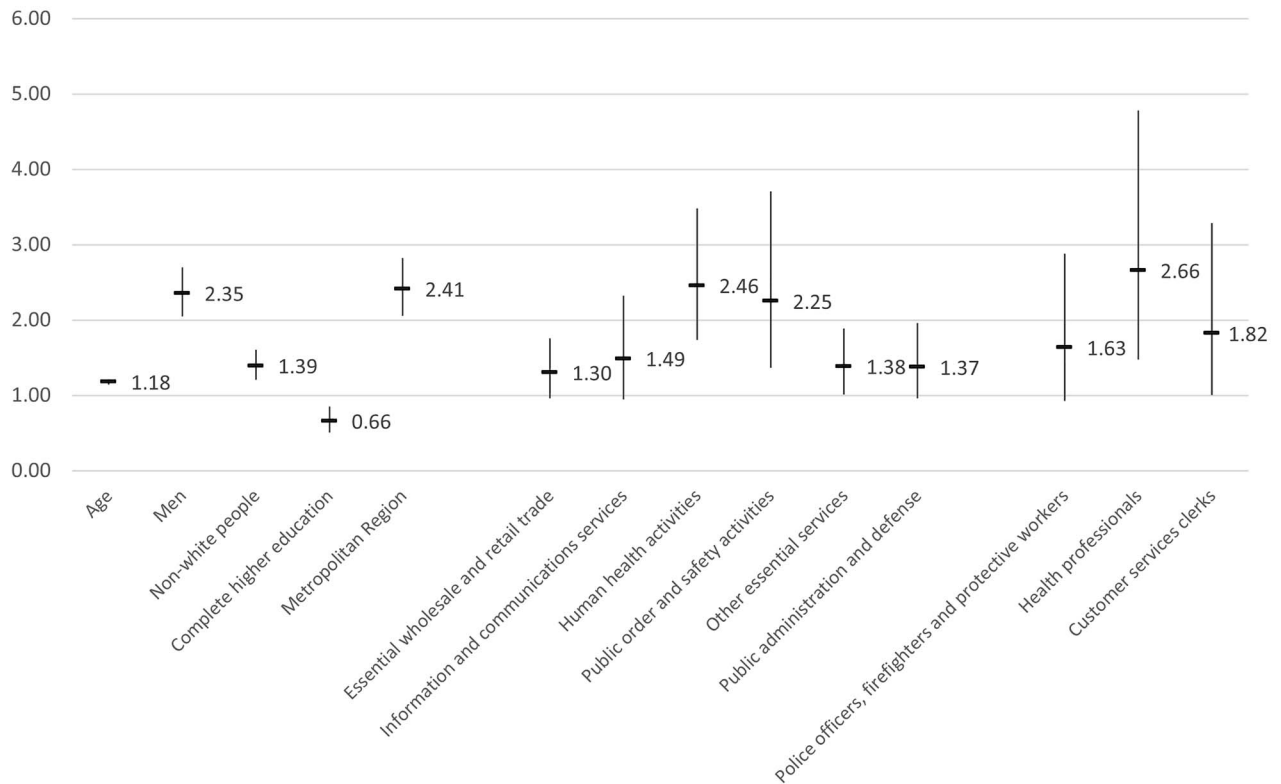
Fig. 1 Odds ratio of mortality from COVID-19 among formal workers in the State of Rio de Janeiro. Source: Own estimation based on RAIS and SES-RJ data.

Figure 1 summarizes these results by presenting the odds ratios of death from Covid-19 for the different individual and occupational characteristics statistically significant by at least 5%. The results presented are those obtained by model 5 and, in the case of occupations, by model 2.

The results show that with each additional year of age, the risk of death from Covid-19 increases by 18%. Men are 135% more likely to die than women (or 2.35 times more). For non-white people, this risk is 39% higher than that of white, while those with a university degree have a 44% lower risk of dying of the disease. Those who work in the Metropolitan Region are 141% more likely to die.

Considering activities, workers employed in establishments in the human health and public safety sectors present, respectively, a risk of dying 2.46 and 2.25 times higher than those employed in other activities. Workers in the essential trade, information and communication, and other essential services have, respectively, 30%, 49% and 38% more risk of death than others. Finally, workers in the public administration have 37% more risk of dying of the disease in the state of Rio de Janeiro.

Regarding occupations, the results of Model 2 show that health professionals from any economic activity have a 166% higher risk of dying of Covid-19 than others. For protective workers and customer services clerks, the probability of

death from the disease is 63 and 82% higher than for other workers.

Discussion

Main findings of this study

This study demonstrated the extent to which individuals' socioeconomic characteristics, such as schooling, race/ethnicity and working in the metropolitan region, affect the probability of death by Covid-19. Furthermore, the analysis by occupation and economic activity made it clear that workers in essential activities such as health and security and those who could not work remotely, became more vulnerable to death by Covid-19.

That is further evidence suggesting that social distancing measures managed to protect the portion of workers for which it was possible to either interrupt work activities or carry them out remotely. These differences must be taken into account to design prevention policies to be adopted by governments.

What is already known on this topic

There is some evidence that socioeconomic conditions, such as income, could be related to a higher level of exposure and

risk of death by Covid-19,[3,8] although most of it is based on aggregated data. There is also a group of studies[9,10,11] that claims the importance of taking occupation into account in public policy to deal with the pandemic.

Our findings are aligned with Baker *et al.*[12], who found that U.S. workers most affected are those in health, security sectors, occupations in administrative support and offices, education, social and community services, and construction and repair.

What this study adds

Although there are other similar findings in the literature, to the best of the authors' knowledge, this is the first study in which these relationships are measured by controlling for a larger set of socioeconomic characteristics of the individual, such as wage, schooling and type of job.

Limitations of this study

The analysis considered the 2018 set of formal workers in Rio de Janeiro state, representing approximately half of the Covid-19 cases in the state. This group has characteristics that make it less vulnerable than the rest of the population, be it the informal, unemployed or elderly. It is also a more homogeneous group regarding several socioeconomic variables.

One question to be investigated is whether the results obtained would disappear after the reopening measures underway in the state.

Acknowledgments

Funding sources

References

1 King G, Zeng L. Logistic regression in rare events data. *Political Anal* 2001;**9**(**2**):137–63. doi: 10.1093/oxfordjournals.pan.a004868.

2 Batista A, Antunes B, Faveret G *et al.* Análise socioeconômica da taxa de letalidade da COVID-19 no Brasil. *Núcleo de Operações e Inteligência em Saúde (NOIS) Nota Técnica* 2020;11–27.

3 Bermudi, M. M., P. Lorenz, C., Souza de Aguiar, B. *et al*. (2020). *Spatiotemporal Dynamic of COVID-19 Mortality in the City of Sao Paulo, Brazil: Shifting the High Risk from the Best to the Worst Socio-economic Conditions*. (2020): arXiv-2008. Available at: https://arxiv.org/abs/2008.02322.

4 McLaren, J. (2020). *Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data (No. w27407)*. National Bureau of Economic Research. Available at: https://ssrn.com/abstract=3632634.

5 Millett GA, Jones AT, Benkeser D *et al.* Assessing differential impacts of COVID-19 on black communities. *Ann Epidemiol* 2020;**47**:37–44ISSN 1047-2797. doi: 10.1016/j.annepidem. 2020.05.003

6 Raifman MA, Raifman JR. Disparities in the population at risk of severe illness from covid-19 by race/ethnicity and income. *Am J Prevent Med* 2020;**59**(**1**):137–9. doi: 10.1016/j.amepre.2020.04.003.

7 Benitez J, Courtemanche C, Yelowitz A. Racial and ethnic disparities in Covid-19: evidence from six large cities. *IZA Discussion Paper No. 13521* 2020. https://ssrn.com/abstract=3660268.

8 Pires, L. N., Carvalho, L., & Xavier, L. D. L. (2020). COVID-19 e desigualdade: a distribuição dos fatores de Risco no Brasil. *Experiment Findings*. Available at: https://www.researchgate. net/publication/340452851.

9 Adams-Prassl A, Boneva T, Golin M *et al.* Work that can be done from home: evidence on variation within and across occupations and industries. IZA Discussion Papers No. 2020;13374.

10 Mongey, S., Pilossoph, L., & Weinberg, A. (2020). *Which workers bear the burden of social distancing policies?* (No. w27085). National Bureau of Economic Research. Available at: https://www.nber.org/papers/w27085.

11 Akbarpour, M., Cook, C., Marzuoli, A. *et al.* (2020). *Socioeconomic Network Heterogeneity and Pandemic Policy Response*. University of Chicago, Becker Friedman Institute for Economics, Chicago. Working Paper, (2020-75).

12 Baker MG, Peckham TK, Seixas NS. Estimating the burden of United States workers exposed to infection or disease: a key factor in containing risk of COVID-19 infection. *PLoS One* 2020;**15**(**4**):e0232452. doi: 10.1371/journal.pone.0232452.