Cavillot *et al. Archives of Public Health* (2024) 82:209
https://doi.org/10.1186/s13690-024-01437-8

Archives of Public Health

RESEARCH Open Access

# Unravelling demographic and socioeconomic patterns of COVID-19 death and other causes of death: results of an individual-level analysis of exhaustive cause of death data in Belgium, 2020

Lisa Cavillot[1,2*], Laura Van den Borre[1], Katrien Vanthomme[3,4], Aline Scohy[1], Patrick Deboosere[5], Brecht Devleesschauwer[1,6], Niko Speybroeck[2] and Sylvie Gadeyne[5]

## Abstract

**Background** The COVID-19 pandemic led to significant excess mortality in 2020 in Belgium. By using microlevel cause-specific mortality data for the total adult population in Belgium in 2020, three outcomes were considered in this study aiming at predicting sociodemographic (SD) and socioeconomic (SE) patterns of (1) COVID-19 specific death compared to survival; (2) all other causes of death (OCOD) compared to survival; and (3) COVID-19 specific death compared to all OCOD.

**Methods** Two complementary statistical methods were used. First, multivariable logistic regression models providing odds ratios and 95% confidence intervals were fitted for the three study outcomes. In addition, we computed conditional inference tree (CIT) algorithms, a non-parametric class of classification trees, to identify and rank by significance level the strongest predictors of the three study outcomes.

**Results** Older individuals, males, individuals living in collectivities, first-generation migrants, and deprived SE groups experienced higher odds of dying from COVID-19 compared to survival; living in collectivities was identified by the CIT as the strongest predictor followed by age and sex. Education emerged as one of the strongest predictors for individuals not living in collectivities. Overall, similar patterns were observed for all OCOD except for first- and second-generation migrants having lower odds of all OCOD compared to survival; age group was identified by the CIT as the strongest predictor. Older individuals, males, individuals living in collectivities, first- and second-generation migrants, and individuals with lower levels of education had higher odds of COVID-19 death compared to all OCOD; living in collectivities was identified by the CIT as the strongest predictor followed by age, sex, and migration background. Education and income emerged as among the strongest predictors among individuals not living in collectivities.

*Correspondence:
Lisa Cavillot
Lisa.Cavillot@sciensano.be

Full list of author information is available at the end of the article

**Conclusions** This study identified important SD and SE disparities in COVID-19 mortality, with living in collectivities highlighted as the strongest predictor. This underlines the importance of implementing preventive measures, particularly within the most vulnerable populations, in infectious disease pandemic preparedness to reduce virus circulation and the resulting lethality.

**Text box 1. Contributions to the literature**

- Thanks to the use of exhaustive individual level data on causes of death in Belgium in 2020, significant sociodemographic and socioeconomic disparities have been identified in COVID-19 mortality, aligning with the literature on COVID-19 from many other countries.
- Thanks to the use of classification trees, this study enables a hierarchical ranking of risk factors for COVID-19 mortality, making it possible to target at-risk groups in which the implementation of prevention measures is essential to reduce disparities in the context of the COVID-19 pandemic or other potential future pandemic caused by viral diseases.
- Future research should consider the impact of vaccination on mitigating sociodemographic and socioeconomic inequalities in COVID-19 mortality.

## Background

The Coronavirus Disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread rapidly around the world, resulting in a global pandemic as announced by the WHO in March 2020 [1]. The COVID-19 pandemic had major health consequences. A systematic review including 191 countries identified that the pandemic had led to a global excess mortality rate of 120.3 deaths per 100,000 persons over the period from January 1st 2020 to December 31st 2021 [2]. In Belgium, 126,850 deaths were recorded in 2020 of which 19,801 were due to COVID-19 [3, 4]. In the same year, 18,765 excess deaths were recorded leading to an excess mortality of 17.5% [5].

The COVID-19 pandemic was swiftly characterized as a syndemic pandemic, highlighting the interplay between the virus biological aspects and the social determinants of health, further intensifying existing social inequalities deeply rooted in our society [6, 7]. Indeed, beyond age and sex, which are well-established risk factors for COVID-19 mortality [8, 9], some international studies, based on either excess mortality or COVID-19 specific mortality, showed that ethnic minorities and deprived socioeconomic (SE) groups (e.g. with lower income and education levels, living in overcrowded households) experienced higher COVID-19 mortality or excess mortality during the pandemic [10–15]. In Belgium, research based on excess mortality identified similar social patterns during the first COVID-19 wave [16–19].

Due to delays in obtaining cause-specific mortality data so far, analyses on the SD and SE patterns of COVID-19 specific mortality have not yet been published in Belgium. Our paper aims to bridge this void by offering a nuanced and deeper understanding of the SD and SE inequalities in COVID-19 mortality using nationwide microlevel cause-specific mortality data for the entire Belgian adult population. This population-based study will use individual cause-specific mortality data for 2020 in Belgium to predict the SD and SE patterns of (1) COVID-19 specific deaths relative to survival; (2) all other causes of death (OCOD) relative to survival; and (3) COVID-19 specific deaths relative to all OCOD.

## Materials and methods

### Data sources and study population

Pseudonymized individual level data for the entire Belgian population (over 11 million individuals) were provided by Statistics Belgium (Statbel) who performed a record linkage between four exhaustive data sources linked at the individual level using the Belgian social security number: (1) the Belgian national register providing yearly stock files including demographic indicators and mortality data for all individuals officially residing in Belgium as of 1st January of each year; (2) the administrative census 2011 providing data on education and housing; (3) the tax register providing data on yearly net taxable personal income; and (4) death certificates providing data on causes of death.

As cause-specific mortality data on COVID-19 were available from January 1st 2020 to December 31st 2020 only, we limited our analysis to 2020, using the yearly stock file of 2020 and the individual cause-specific mortality data for that year. In order to determine the SE characteristics properly, the study population consisted of 8,254,632 adults aged 25 years and older officially residing in Belgium. A flowchart demonstrating the study population selection process is available in Supplementary Fig. 1.

### Variables

#### *Mortality outcomes*

Two mortality outcomes, occurring over the period from January 1st 2020 to December 31st 2020, were considered in our study: COVID-19 specific death and all other causes of death (OCOD). The individual cause-specific mortality data, classified according to the International Standard Classification of Disease and Related Health Problems 10th Revisions (ICD-10), enabled the identification of COVID-19 specific death using the ICD-10 codes U07.1 (*COVID-19, virus identified*) and U07.2 (*COVID-19, virus not identified*). We considered as COVID-19 deaths, deaths with one of these 2 codes as underlying cause of death. OCOD were identified by

any ICD-10 codes not falling within the aforementioned COVID-19 specific ICD-10 categories.

Figure 1 shows the daily number of COVID-19 deaths and all other causes of death during the year 2020 by age groups in Belgium. The first COVID-19 death recorded in Belgium occurred on 11th March 2020.

### *Sociodemographic and socioeconomic characteristics*

To identify correlates of COVID-19 specific death and all OCOD, we selected a set of relevant SD and SE variables regarding their association, previously demonstrated in several national and international studies, with COVID-19 mortality, excess mortality during the pandemic or all-cause mortality [16–18, 20–23]. Age group, sex, migration background, and living situation were included as SD characteristics. Migration background was based on the individual's country of birth, nationality of origin and nationality of citizenship, as well as the country of birth of the parents. Individuals born outside Belgium have been classified as 'First-generation migrants'. Individuals born in Belgium with a nationality of origin and a nationality of citizenship equivalent to Belgian, and whose parents were both born in Belgium were classified as 'Belgian natives'. Individuals born in Belgium with at least one nationality other than Belgian, or at least one parent born outside Belgium have been classified as 'Second-generation migrants'. The living situation, partially reflecting the social environment, was categorized into five groups: 'With a partner', 'Without a partner', 'Care homes', 'Other collectivities (no care home)' (e.g., individuals living in prison), and 'Other' (e.g. adult children still living with their parents, other co-residents, individuals living in atypical households). As SE predictors, we included education and income levels. Education level was classified into five categories based on the International Standard Classification of Education (ISCED): 'Primary or less' (ISCED 0 to ISCED 1), 'Lower secondary' (ISCED 2), 'Upper secondary' (ISCED 3 to ISCED 4), 'Higher education' (ISCED 5 to ISCED 8), and 'Missing'. Income data was available as the deciles of the yearly net taxable income per person and distinguished: 'Low income' (deciles 1 to 4), 'Middle income' (deciles 5 to 7), 'High income' (deciles 8 to 10), and 'Missing'.

### Statistical analyses

First, the descriptive analysis shows the distribution (frequencies and percentages) of all SD and SE characteristics in the total study population among individuals who died from COVID-19, from all OCOD, or survived as of 31st December 2020. Supplementary Table 1 shows the distribution (frequencies and percentages) of all SD characteristics in the study population with missing SE variables.

Second, the methodology includes two complementary statistical methods allowing a more in-depth exploration of the data: multivariable logistic regression models on the one hand and conditional inference tree (CIT) algorithms on the other hand.

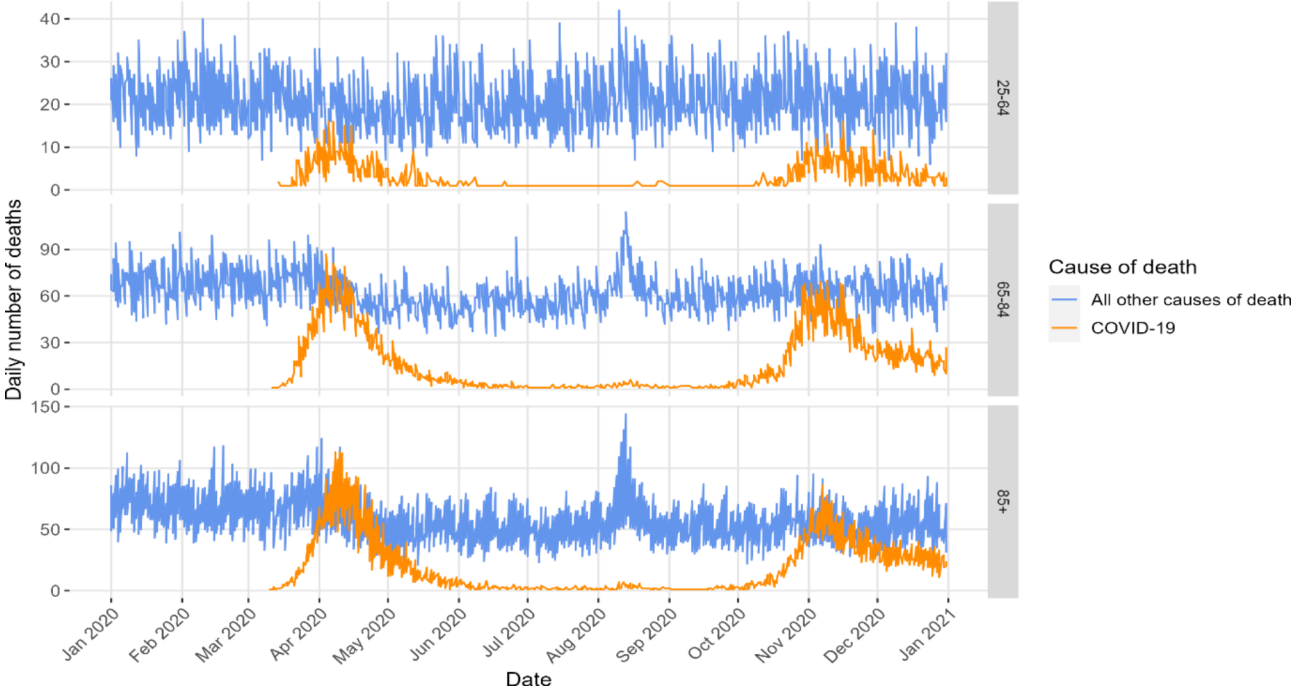


**Fig. 1** : Daily number of COVID-19 deaths and all other causes of death by age groups in 2020 in Belgium

#### *Multivariable logistic regression models*

In a first step, adjusted odds ratios (ORs) and Wald 95% confidence intervals (CIs) were estimated by fitting multivariable logistic regression models. A first logistic regression model was fitted considering COVID-19 specific deaths as cases and survivors as controls, while excluding individuals who died from all OCOD from the model. A second regression model was fitted considering all OCOD as cases and survivors as controls, while excluding individuals who died from COVID-19 from the model. Finally, a third regression model, applied on all deaths occurring in 2020 only, was fitted considering COVID-19 specific deaths as cases and all OCOD as controls. All three multivariable logistic regression models were adjusted for all SD and SE covariates included in our study.

#### *Conditional inference tree algorithm*

In a second step, CIT algorithms were computed, complementing logistic regression models. CIT algorithm, a non-parametric class of classification tree, is a valuable machine learning method to assess the predictive association between a dependent variable and a set of predictors. This method relies on the concept of statistical significance, employing p-values. Specifically, we used conditional recursive partitioning tree algorithms, a subtype of CIT. This method uses binary recursive partitioning within a conditional inference framework to model a regression relationship. The p-values used in the algorithm construction are obtained from conditional permutation tests [24–26]. The algorithm works as follows [24, 25, 27]:

1) The algorithm initiates with the entire sample and seeks to identify the optimal binary initial split to form the root node. For this purpose, it assesses the correlation ($\rho_{X_jY}$) between the response variable ($Y$) and any random variable from the set of predictors ($X_j$). The null hypothesis ($H_0$: $\rho_{X_jY} = 0$) between $Y$ and $X_j$ is assessed using a statistical test (i.e. conditional permutation test). If $H_0$ is rejected ($H_1$ : $\rho_{X_jY} \neq 0$), the predictors $X_j$ with the strongest association (i.e. the lowest p-value) to $Y$ is selected and constitutes the root node.
2) For each subset created, the algorithm recursively applies the conditional permutation tests for all predictors $X_j$ and selects the one that maximizes the pre-specified significance level α for splitting that subset. The algorithm stops when the stopping criteria are met.
3) The algorithm checks whether the specific stopping criteria are met at each recursive step. When the stopping criteria are met, the algorithm stops splitting the subset and the current subsets become the terminal nodes of the tree. The stopping criterium can be, for example, that $H_0$ can no longer be rejected at a pre-specified level α, or that a specific tree depth or a minimum number of observations in a terminal node has been achieved.

Logistic regression and CIT offer a comprehensive examination of the data, each method possessing distinct advantages. Logistic regression is useful to quantify the association between a response variable and covariates, providing probability levels and confidence intervals for regression coefficients [28]. However, multicollinearity is not handled and the inclusion of interactions can become complicated to interpret. Classification tree, serving as a valuable complement to standard regression methods, addresses these limitations by handling complex, nonlinear interactions between variables, in which subgroups are formed through optimal splitting variables, leading to the natural emergence of interactions as new starting population [26]. The multicollinearity issues are managed by the algorithm by selecting the most important of the two colinear variables and dealing with the non-selected variables by calculating importance scores, indicating their role as a substitute for primary divisions [29]. While some classification trees are prone to overfitting (e.g. Classification and Regression Tree) [24, 29], the CIT algorithm mitigates the overfitting issue by employing a statistical testing approach at each node. Instead of growing the tree without constraints, the CIT algorithm used the significance of a test (i.e. permutation tests) to determine whether a split is statistically significant, preventing the algorithm from creating overly complex structures that may not generalize well to new data. This approach helps maintain a more accurate and less overfitted model, enhancing the reliability of predictions [24, 25]. Finally, the classification tree provides a clear and easy hierarchical understanding of the relationship between the predictors and the outcome that can help policymakers target their interventions [26].

In our analyses, three CIT algorithms were computed to identify and rank in order of importance the SD and SE predictors of cause-specific mortality. Following exactly the same approach applied to the multivariable logistic regression models, the first CIT considered COVID-19 deaths as cases and survivors as controls while the second CIT considered all OCOD as cases and survivors as controls. The third CIT was applied only on all death occurring in 2020 and considered COVID-19 deaths as cases and all OCOD as controls.

By applying these two complementary statistical methods to each outcome, this methodological approach allows first to explore separately the SD and SE patterns associated with COVID-19 death and all OCOD in comparison with survivors, respectively. Additionally,

comparing COVID-19 death with all OCOD further helps characterize and understand how COVID-19 death differs from all OCOD, contributing to the identification of its specific patterns that distinguish it from 'regular' mortality in 2020. This approach deepens the analysis and contributes to a more robust understanding of the multi-faceted dynamics underlying mortality during the early stages of the COVID-19 pandemic.

#### *Sensitivity analyses*

To test the robustness of our results, we estimated sub-distribution hazard ratios (SHR) and 95% confidence intervals using Fine and Gray competing risk models to identify the SD and SE predictors of COVID-19 death and all OCOD. Three competing risk models were performed. The first model (Supplementary Table 2) considered COVID-19 death as cases and survival as the competing event, while excluding individuals who died from all OCOD. The second model (Supplementary Table 2) considered all OCOD as cases and survival as the competing event, while excluding individuals who died from COVID-19. The third model (Supplementary Table 3) considered COVID-19 deaths as cases and all OCOD as the competing event, while excluding survivors. Due to the use of large datasets leading to computation constraints, the first and second competing risk models could only be performed on a 50% random subset representative of the total study population.

In our analyses, we set a significance level α of 0.05 and imposed a maximum tree depth of four levels. A four-level tree allowed to maintain accuracy in the information provided by the CIT while avoiding an excessive number of terminal nodes making visual interpretation and the message to be conveyed to policymakers more complex. For each CIT, the prevalence of cases in the study population of interest was used as the predictive threshold. CIT were trained on 70% of the data, 30% of the data was used to validate the performance of the algorithm. CIT performances were measured by the means of sensitivity [95%CI] and specificity [95%CI]. All analyses were performed in R version 4.2.0 [30]. The package "Partykit" was used for computing the CIT algorithms [31].

**Table 1** Characterization of the study population, January 1st 2020 – December 31st 2020, Belgium

| | **All** ***n*** **= 8 254 632 (100%)** | **COVID-19 specific death** ***n*** **= 21 941 (0.27%)** | **All other causes of death** ***n*** **= 103 591 (1.25%)** | **Survivors** ***n*** **= 8 129 100 (98.48%)** |
|---|---|---|---|---|
| Age groups, n(%) | | | | |
| 25–64 | 6 050 325 (73.30) | 1 421 (6.48) | 15 204 (14.68) | 6 033 699 (74.22) |
| 65–84 | 1 869 239 (22.64) | 9 483 (43.22) | 45 823 (44.23) | 1 813 933 (22.31) |
| 85+ | 335 069 (4.06) | 11 037 (50.30) | 42 564 (41.09) | 281 468 (3.46) |
| Sex, n (%) | | | | |
| Females | 4 247 959 (51.46) | 11 253 (51.29) | 52 889 (51.06) | 4 183 817 (51.47) |
| Males | 4 006 673 (48.54) | 10 688 (48.71) | 50 702 (48.94) | 3 945 283 (48.53) |
| Living situation, n (%) | | | | |
| With partner | 5 226 987 (63.32) | 6 713 (30.6) | 39 082 (37.73) | 5 181 192 (63.74) |
| Without partner | 2 166 454 (26.25) | 6 307 (28.75) | 37 106 (35.82) | 2 123 041 (26.12) |
| Care homes | 96 464 (1.17) | 7 283 (33.19) | 19 825 (19.14) | 69 356 (0.85) |
| Other collectivities (no care home) | 35 857 (0.43) | 655 (2.99) | 2 021 (1.95) | 33 181 (0.41) |
| Other | 728 870 (8.83) | 983 (4.48) | 5 557 (5.36) | 722 330 (8.89) |
| Migration background, n (%) | | | | |
| Belgian natives | 5 668 832 (68.67) | 17 175 (78.28) | 83 628 (80.73) | 5 568 029 (68.50) |
| Second-generation migrants | 923 599 (11.19) | 1 630 (7.43) | 8 034 (7.76) | 913 935 (11.24) |
| First-generation migrants | 1 662 201 (20.14) | 3 136 (14.29) | 11 929 (11.52) | 1 647 136 (20.26) |
| Income, n (%) | | | | |
| Low | 2 733 998 (33.12) | 9 992 (45.54) | 46 624 (45.01) | 2 677 382 (32.94) |
| Middle | 2 319 805 (28.10) | 8 814 (40.17) | 40 005 (38.62) | 2 270 986 (27.94) |
| High | 2 632 537 (31.89) | 2 534 (11.55) | 13 864 (13.38) | 2 616 139 (32.18) |
| Missing | 568 292 (6.88) | 601 (2.74) | 3 098 (2.99) | 564 593 (6.95) |
| Education, n (%) | | | | |
| Primary or less | 934 378 (11.32) | 8 581 (39.11) | 34 985 (33.77) | 890 812 (10.96) |
| Lower secondary | 1 731 746 (20.98) | 5 279 (24.06) | 26 069 (25.17) | 1 700 398 (20.92) |
| Upper secondary | 2 470 213 (29.93) | 3 188 (14.53) | 18 309 (17.67) | 2 448 716 (30.12) |
| Higher education | 1 981 898 (24.01) | 2117 (9.65) | 12 437 (12.01) | 1 967 344 (24.2) |
| Missing | 1 136 397 (13.77) | 2776 (12.65) | 11 791 (11.38) | 1 121 830 (13.8) |

## Results

### Description of the study population

Our study population was composed of 8,254,632 individuals aged 25 years and over officially residing in Belgium as of 1st January 2020 (see Supplementary Fig. 1). Table 1 characterizes the study population in detail. As of January 1st 2020, the study population included 8,254,632 individuals officially residing in Belgium aged 25 years and over. Individuals who died from COVID-19 accounted for 0.27% ($N$=21,941) of the total study population, while individuals who died from all OCOD accounted for 1.25% ($N$=103,591) of the total study population. Individuals who survived as of December 31st 2020 accounted for 98.48% ($N$=8,129,100) of the total study population.

Supplementary Table 1 characterizes the study population including individuals with missing SE information. The majority of individuals with missing SE information was aged 25 to 64 years (93.58%), was females (49.32%), lived with a partner (45.05%), and a first-generation migration background (89.61%).

### Predictors of COVID-19 specific death and other causes of death vs. survival, respectively

#### *Multivariable logistic regression model results*

Table 2 shows the fully adjusted ORs and 95% CIs estimated resulting from the multivariable logistic regression models to identify the association of SD and SE variables with COVID-19 death and all OCOD compared to survival, respectively.

A higher odds of dying from COVID-19 was found among the middle-aged (OR 19.53 [18.22–20.96]) and the elderly (OR 77.45 [71.92–83.40]), compared to the youngest age group. Men had a two times higher mortality from COVID-19 (OR 2.19 [2.11–2.26]) than women. Individuals living with a partner have lower odds of dying

**Table 2** Fully adjusted odds ratios (ORs) and their 95% confidence intervals (CIs) for the sociodemographic and socioeconomic characteristics associated with COVID-19 death and all other causes of death compared to survival, respectively

| | COVID-19 specific death* | | All other causes of death** | |
|---|---|---|---|---|
| | OR [95%CI] | P-value | OR [95%CI] | P-value |
| Age group (in years) | | | | |
| 25–64 | 1 | 1 | 1 | 1 |
| 65–84 | 19.53 [18.22–20.96] | < 0.001 | 8.48 [8.29–8.69] | < 0.001 |
| 85+ | 77.45 [71.92–83.40] | < 0.001 | 32.64 [31.78–33.51] | < 0.001 |
| Sex | | | | |
| Females | 1 | 1 | 1 | 1 |
| Males | 2.19 [2.11–2.26] | < 0.001 | 1.83 [1.80–1.86] | < 0.001 |
| Living situation | | | | |
| With partner | 1 | 1 | 1 | 1 |
| Without partner | 1.58 [1.51–1.65] | < 0.001 | 1.62 [1.59–1.65] | < 0.001 |
| Care homes | 12.63 [12.04–13.27] | < 0.001 | 6.41 [6.25–6.58] | < 0.001 |
| Other collectivities (no care home) | 7.29 [6.59–8.06] | < 0.001 | 3.93 [3.71–4.17] | < 0.001 |
| Other | 1.72 [1.59–1.87] | < 0.001 | 1.48 [1.42–1.53] | < 0.001 |
| Migration background | | | | |
| Belgian natives | 1 | 1 | 1 | 1 |
| Second-generation migrants | 1.02 [0.96–1.08] | 0.468 | 0.79 [0.77–0.81] | < 0.001 |
| First-generation migrants | 1.24 [1.18–1.30] | < 0.001 | 0.92 [0.87–0.94] | < 0.001 |
| Income | | | | |
| High | 1 | 1 | 1 | 1 |
| Low | 1.32 [1.24–1.40] | < 0.001 | 1.52 [1.48–1.56] | < 0.001 |
| Middle | 1.18 [1.12–1.25] | < 0.001 | 1.29 [1.25–1.32] | < 0.001 |
| Missing | 1.28 [1.14–1.44] | < 0.001 | 1.40 [1.32–1.46] | < 0.001 |
| Education | | | | |
| Higher education | 1 | 1 | 1 | 1 |
| Primary or less | 1.75 [1.65–1.87] | < 0.001 | 1.46 [1.42–1.57] | < 0.001 |
| Lower secondary | 1.43 [1.35–1.53] | < 0.001 | 1.24 [1.20–1.27] | < 0.001 |
| Upper secondary | 1.20 [1.12–1.28] | < 0.001 | 1.08 [1.05–1.12] | < 0.001 |
| Missing | 1.51 [1.40–1.63] | < 0.001 | 1.26 [1.22–1.31] | < 0.001 |

* Individuals who died from COVID-19 were considered as cases ($N$=21,941) and survivors as controls ($N$=8,129,100). All ORs were fully adjusted for all SD and SE variables

** Individuals who died from all other causes of death were considered as cases ($N$=103,591) and survivors as controls ($N$=8,129,100). All ORs were fully adjusted for all SD and SE variables

from COVID-19 compared to individuals in any other type of living situation; individuals living in care homes presenting the highest odds of dying from COVID-19 (OR 12.63 [12.04–13.27]). Compared to Belgian natives, first-generation migrants had higher odds of dying from COVID-19 (OR 1.24 [1.18–1.30]). A higher odds of dying from COVID-19 was identified among individuals with low (OR 1.32 [1.24–1.40]) and middle (OR 1.18 [1.12–1.25]) income, compared to individuals with high income. Compared to individuals with higher education, individuals with primary or less education (OR 1.75 [1.65–1.87]), lower secondary education (OR 1.43 [1.35–1.53]), and upper secondary education (OR 1.20 [1.12–1.28]) had higher odds of dying from COVID-19, highlighting a clear gradient in COVID-19 mortality by education level.

Overall, the SD and SE patterns identified in all OCOD were similar to those observed in COVID-19 mortality. An important difference was reflected by the much lower ORs for individuals aged 65 to 84 and aged 85 and over (OR 8.48 [8.29–8.69] and OR 32.64 [31.78–33.51] compared to individuals aged 25 to 64, respectively), as well as for individuals living in care homes and those living in other collectivities (no care homes) (OR 3.93 [3.71–4.17] and OR 6.41 [6.25–6.58] compared to individuals living with a partner, respectively) (Table 2). Another difference between all OCOD and COVID-19 death was observed for the migration background variable. Compared to Belgian natives, second-generation migrants (OR 0.79 [0.77–0.81]) and first-generation migrants (OR 0.92 [0.87–0.94]) had lower odds of dying from OCOD.

### *Conditional recursive partitioning tree algorithm results*

Figure 2 presents the CIT algorithm considering COVID-19 specific deaths as cases and survivors as controls to identify and rank in order of importance the SD and SE predictors of COVID-19 mortality. 'Living situation', 'Age group', 'Sex', and 'Education' emerged as the strongest predictors of COVID-19 death, ranked according to their statistical significance. The variable 'Living situation' was assigned to the root node and constituted the first split by dividing the total study population in two subgroups (i.e. individuals living in collectivities vs. those not living in collectivities). The tree continued to grow while 15 terminal nodes were determined. The proportion of COVID-19 deaths relative to the sample size of each terminal node is indicated in each of them. When the proportion of COVID-19 deaths in the terminal nodes exceeded its prevalence in the study population (i.e. 0.27%), they were categorized as 'COVID-19 death' and colored in red. Conversely, all terminal nodes with a proportion of COVID-19 deaths below 0.27% were categorized as 'Survival' and colored in light yellow.

The left branch of the tree, composed of individuals not living in collectivities, was secondly partitioned based on 'Age group' between individuals aged 25 to 84 and those aged 85 and over. The subgroup composed of individuals aged 25 to 84 was thirdly partitioned between those aged

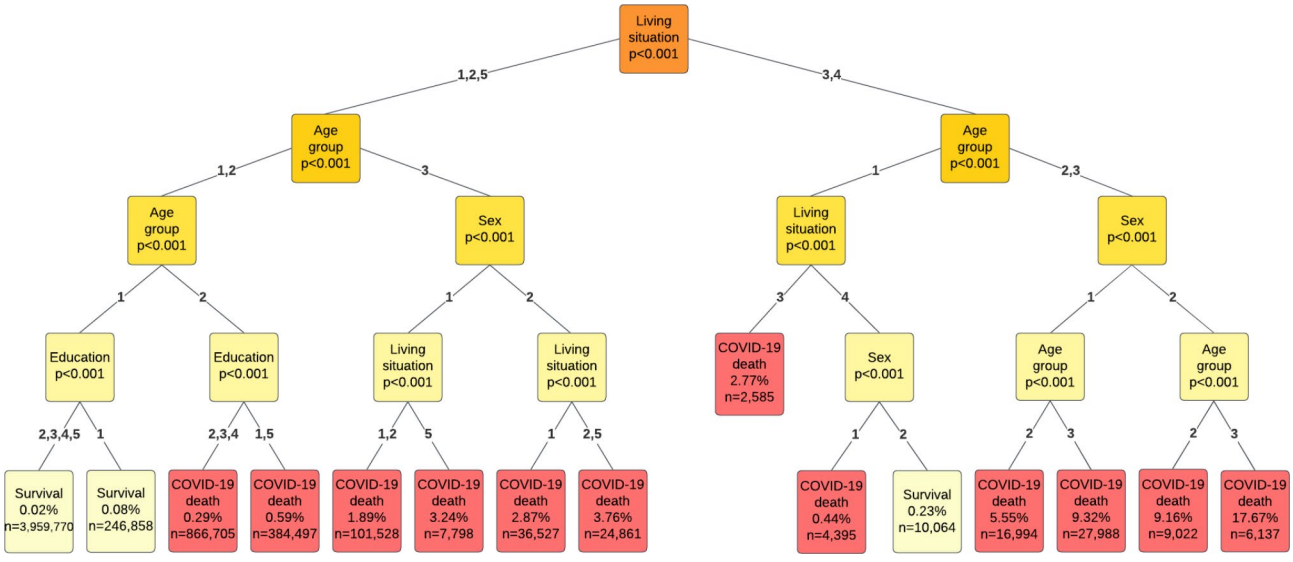


**Fig. 2** Conditional recursive partitioning tree algorithm considering COVID-19 specific death as cases and survivors as controls. The variable 'Living situation' was categorized as follows: 1='With partner', 2='Without partner', 3='Care homes', 4='Other collectivities (no care home)', 5='Other'. The variable 'Age group' (in years) was categorized as follow: 1='25–64', 2='65–84', 3='85+'. The variable 'Sex' was categorized as follow: 1='Females', 2='Males'. The variable 'Education' was categorized as follows: 1='Primary or less', 2='Lower secondary', 3= 'Upper secondary', 4='Higher education', 5='Missing'. The percentages in the terminal nodes indicate the percentage of COVID-19 deaths relative to the sample size of each terminal node, denoted 'n'. When the proportion of COVID-19 deaths in the terminal nodes exceeded its prevalence in the study population (i.e. 0.27%), they were categorized as 'COVID-19 death' and colored in red

25 to 64 and those aged 65 to 84 while the subgroup composed of individuals aged 85 and over was thirdly partitioned between males and females. 'Education' finally partitioned the subgroups of individuals aged 25 to 64 and 65 to 84 years in four terminal nodes, with the lowest levels of education predicting highest proportions of COVID-19 deaths in both subgroups. The two terminal nodes composed of individuals aged 65 to 84 years predicted a twice higher proportion of COVID-19 deaths among those with the lowest or an unknown level of education (0.59% of COVID-19 deaths) compared to the terminal node composed of those with the highest levels of education (0.29% of COVID-19 deaths). 'Living situation' finally partitioned the subgroups of males and females aged 85 and over in four terminal nodes. Those four terminal nodes predicted COVID-19 death with the highest proportion of COVID-19 deaths among males living without a partner or in other type of households (3.76% of COVID-19 deaths).

The right branch of the tree, composed of individuals living in collectivities, was secondly partitioned between individuals aged 25 to 64 and individuals aged 65 and over. The subgroup composed of individuals aged 25 to 64 was thirdly partitioned based on 'Living situation' between those living in care homes, leading directly to a terminal node predicting COVID-19 death (2.77% of COVID-19 deaths), and those living in other type of collectivities. The latter subgroup was finally partitioned based on 'Sex' with the terminal node composed of females predicting COVID-19 death (0.44% of COVID-19 deaths). The subgroup composed of individuals aged 65 and over was further splitted into four terminal nodes all predicting COVID-19 death based on 'Sex' and 'Age group' with the highest proportion of COVID-19 deaths among males aged 85 years and over (17.67% of COVID-19 deaths). The sensitivity [95%] and specificity [95%] of the CIT were 0.956 [0.943–0.961] and 0.744 [0.740–0.749], respectively.

Figure 3 presents the CIT algorithm considering all OCOD as cases and survivors as controls to identify and rank in order of importance the SD and SE predictors of all OCOD. 'Age group', 'Living situation', and 'Sex' emerged as the three strongest predictors of all OCOD, ranked according to their statistical significance. The variable 'Age group' was assigned to the root node and constituted the first split by dividing the study population into two subgroups (i.e. individuals aged 25 to 84 years vs. individuals aged 85 years and over). The tree continued to grow resulting in 16 terminal nodes. The proportion of all OCOD relative to the sample size of each terminal node is indicated in each of them. When the proportion of all OCOD in the terminal nodes exceeded its prevalence in the study population (i.e. 1.25%), they were categorized as 'All OCOD' and colored in red. Conversely, all terminal nodes with a proportion of all OCOD below 1.25% were categorized as 'Survival' and colored in light yellow.

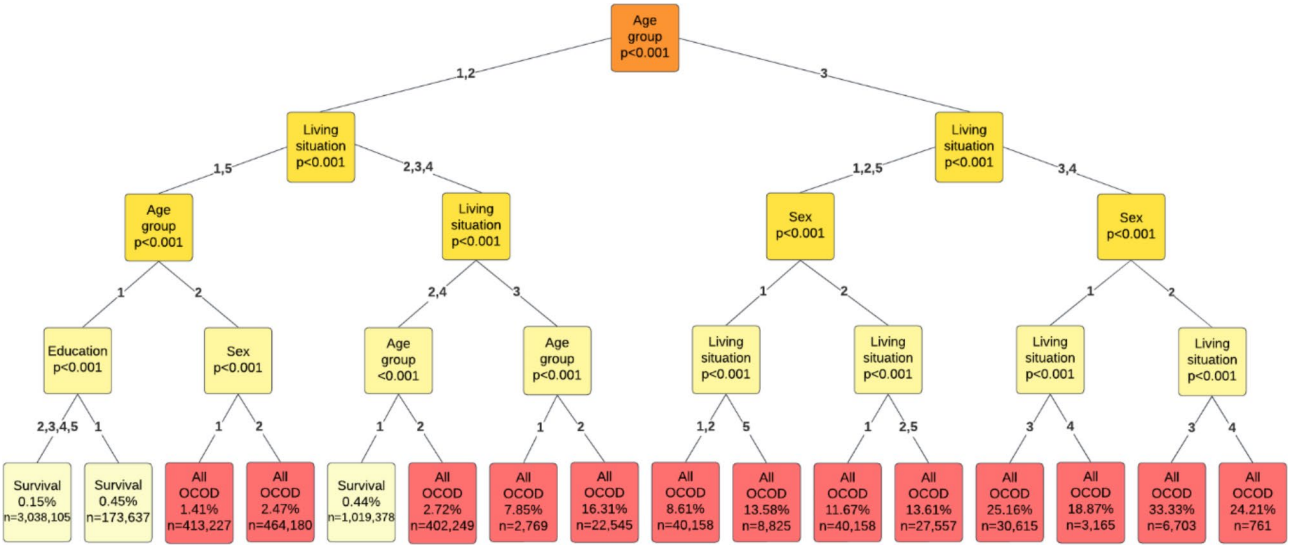


**Fig. 3** Conditional recursive partitioning tree algorithm considering all other causes of death (OCOD) as cases and survivors as controls. The variable 'Age group' (in years) was categorized as follows: 1='25–64', 2='65–84', 3='85+'. The variable 'Living situation' was categorized as follow: 1='With partner', 2='Without partner', 3='Care homes', 4='Other collectivities (no care home)', 5='Other'. The variable 'Sex' was categorized as follow: 1='Females', 2='Males'. The variable 'Education' was categorized as follows: 1='Primary or less', 2='Lower secondary', 3= 'Upper secondary', 4='Higher education', 5='Missing'. The percentages in the terminal nodes indicate the percentage of all OCOD relative to the sample size of each terminal node, denoted 'n'. When the proportion of OCOD deaths in the terminal nodes exceeded its prevalence in the study population (i.e. 1.25%), they were categorized as 'COVID-19 death' and colored in red

The left branch of the tree, composed of individuals aged 25 to 84, was secondly partitioned between those living with a partner or in the category 'Other' and those living without a partner or in collectivities. The subgroup of individuals living with a partner or in the category 'Other' was thirdly partitioned in four terminal nodes based on 'Age group' and divided between those aged 25 to 64 and those aged 65 to 84. Among individuals aged 25 to 64, the variable 'Education' performed the final split, with a three-fold higher proportion of all OCOD among those with the lowest level of education (0.45% of all OCOD), compared to those with highest levels of education (0.15% of all OCOD). Among individuals aged 65 to 84, the final split was performed by the variable 'Sex' with the terminal nodes composed of males and females predicting all OCOD (2.47% and 1.41% of all OCOD, respectively). The subgroup of individuals living without a partner or in collectivities was further partitioned based on 'Living situation' and 'Age group' in four terminal nodes. Out of these four nodes, three predicted all OCOD, with the highest proportion of all OCOD identified among individuals living in care homes aged 65 to 84 (16.31% of all OCOD).

The right branch of the tree, composed of individuals aged 85 and over, was secondly partitioned between individuals living in collectivities and those not living in collectivities. The subgroup composed of individuals not living in collectivities was further partitioned into four terminal nodes based on 'Sex' and 'Living situation'. Those four terminal nodes predicted all OCOD with the highest proportions of all OCOD identified among males living without a partner or in other types of household (13.61% of all OCOD), followed by females living in other types of household as well (13.58% of all OCOD). The subgroup composed of individuals living in collectivities was further partitioned based on 'Sex' and 'Living situation' into four terminal nodes, all predicting all OCOD, with the highest proportion of all OCOD among males living in care homes (33.33% of all OCOD). The sensitivity [95%CI] and specificity [95%CI] were 0.708 [0.697–0.712] and 0.815 [0.810–0.817], respectively.

**Table 3** Fully adjusted odds ratios (ORs) and their 95% confidence intervals (CIs) for the sociodemographic and socioeconomic characteristics associated with COVID-19 specific death compared to all other causes of death

| | OR [95%CI] | *P*-value |
|---|---|---|
| Age groups (in years) | | |
| 25–64 | 1 | 1 |
| 65–84 | 2.01 [1.87 – 2.16] | < 0.001 |
| 85+ | 2.19 [2.03 – 2.35] | < 0.001 |
| Sex | | |
| Females | 1 | 1 |
| Males | 1.22 [1.17 – 1.27] | < 0.001 |
| Living situation | | |
| With partner | 1 | 1 |
| Without partner | 1.01 [0.96 – 1.06] | 0.474 |
| Collectivities (care home) | 2.04 [1.94 – 2.15] | < 0.001 |
| Collectivities (no care home) | 1.88 [1.68 – 2.10] | < 0.001 |
| Other | 0.98 [0.89 – 1.07] | 0.62 |
| Migration background | | |
| Belgian natives | 1 | 1 |
| Second-generation migrants | 1.42 [1.35 – 1.50] | 0.016 |
| First-generation migrants | 1.50 [1.41 – 1.59] | < 0.001 |
| Income | | |
| High | 1 | 1 |
| Low | 0.97 [0.91 – 1.04] | 0.371 |
| Middle | 0.99 [0.91 – 1.03] | 0.320 |
| Missing | 1.10 [0.97 – 1.25] | 0.142 |
| Education | | |
| Higher education | 1 | 1 |
| Primary or less | 1.25 [1.16 – 1.34] | < 0.001 |
| Lower secondary | 1.16 [1.08 – 1.24] | < 0.001 |
| Upper secondary | 1.08 [1.01 – 1.16] | 0.181 |
| Missing | 1.18 [1.09 – 1.28] | < 0.001 |

Individuals who died from COVID-19 ($N$=21,941) were considered as cases. Individuals who died from all other causes of death ($N$=103,591) were considered as controls. All ORs were fully adjusted for all SD and SE variables

## Predictors of COVID-19 death vs. all other causes of death

### *Multivariable logistic regression model results*

Table 3 shows the fully adjusted ORs and 95% CIs estimated with a logistic regression model applied to all deaths occurring in 2020 to quantify the association between SD and SE characteristics and COVID-19 specific death compared to all OCOD. Compared to individuals aged 25 to 64, those aged 65 to 84 (OR 2.01 [1.87–2.16]) and 85 and over (OR 2.19 [2.03–2.35]) presented higher odds of dying from COVID-19 than from OCOD. Compared to females, higher odds of COVID-19 death than from OCOD was found among males (OR 1.22 [1.17–1.27]). Individuals living in collectivities, either in care home (OR 2.04 [1.94–2.15]) or not (OR 1.88 [1.68–2.10]), had higher odds of dying from COVID-19 than from OCOD, compared to those living with a partner. Second-generation migrants (OR 1.42 [1.35–1.50]) and first-generation migrants (OR 1.50 [1.41–1.59]) presented higher odds of dying from COVID-19 than from OCOD, compared to their Belgian native counterparts. Moreover, individuals with lower education, specifically those with primary or less education (OR 1.25 [1.16–1.34]) and lower secondary education (OR 1.16 [1.08–1.24]) demonstrated higher odds of COVID-19 death than from OCOD in contrast to those with higher educational attainment.

### *Conditional recursive partitioning tree algorithm results*

Figure 4 presents the results of the CIT algorithm applied to all deaths occurring in 2020 and considering COVID-19 specific death as cases and all OCOD as controls to identify and rank in order of importance the SD and SE predictors specifically characterizing COVID-19 mortality patterns compared to 'regular' mortality patterns in 2020. 'Living situation', 'Age group', 'Sex', 'Migration background', 'Income' and 'Education' emerged as the strongest predictors of COVID-19 death compared to all OCOD, ranked according to their statistical significance. The variable 'Living situation' was assigned to the root node by dividing the study population between individuals living in collectivities and those not living in collectivities. The tree continued to grow and was finally partitioned into 14 terminal nodes. The proportion of COVID-19 deaths relative to the sample size of each terminal node is indicated in each of them. When the proportion of COVID-19 deaths in the terminal nodes exceeded its prevalence in the study population of interest (i.e. 17.5% : number of COVID-19 deaths out of all deaths occurring in 2020), they were categorized as COVID-19 death and colored in red. Conversely, all terminal nodes with a proportion of COVID-19 deaths out of all deaths occurring in 2020 below 17.5% were categorized as all OCOD and colored in light yellow.

The left branch of the three, composed of individuals not living in collectivities, was further partitioned based

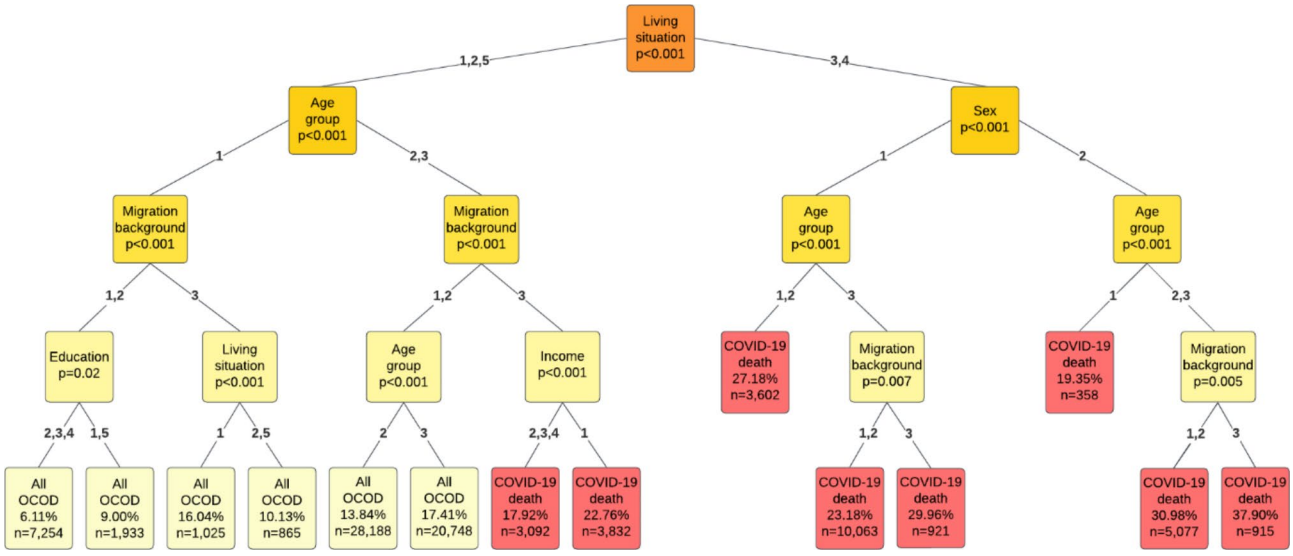


**Fig. 4** Conditional recursive partitioning tree algorithm considering COVID-19 specific death as cases and all other causes of death (OCOD) as controls. The variable 'Living situation' was categorized as follows: 1='With partner', 2='Without partner', 3='Care homes', 4='Other collectivities (no care home)', 5 ='Other'. The variable 'Age group' (in years) was categorized as follow: 1='25–64', 2='65–84', 3='85+'. The variable 'Sex' was categorized as follow: 1='Females', 2='Males'. The variable 'Education' was categorized as follows: 1='Primary or less', 2='Lower secondary', 3= 'Upper secondary', 4='Higher education', 5='Missing'. The variable migration background was categorized as follows: 1='Belgian natives', 2='Second-generation migrants', 3='First-generation migrants'. The variable 'Income' was classified as follows : 1='Low', 2='Middle', 3='High', 4='Missing'. The percentages in the terminal nodes indicate the percentage of COVID-19 deaths relative to all deaths occurring in 2020 in each terminal subgroups whose sample size is denoted by 'n'. When the proportion of COVID-19 deaths in the terminal nodes exceeded its prevalence in the study population of interest (i.e. 17.5%), they were categorized as 'COVID-19 death' and colored in red

on 'Age group', 'Migration background', 'Living situation', 'Education' and 'Income' into eight terminal nodes. Out of those eight terminal nodes, two predicted COVID-19 death. Those two terminal nodes were composed of first-generation migrants aged 85 and over with the lowest income level (22.76% of COVID-19 death) and those with higher or missing income levels (17.92% of COVID-19 death).

The right branch of the three, composed of individuals living in collectivities, was further partitioned based on 'Sex', 'Age group', and 'Migration background' into six terminal nodes, all predicting COVID-19 death. Among males, the highest proportion of COVID-19 deaths was found among first-generation migrant aged 65 and over (37.90% of COVID-19 death). Among females, the highest proportion of COVID-19 deaths was found among first-generation migrant aged 85 and over (29.96% of COVID-19 deaths). The sensitivity [95%] and specificity [95%] of the CIT were 0.647 [0.601–0.662] and 0.764 [0.764–0.765], respectively (Fig. 4).

#### Sensitivity analyses

Supplementary Table 2 shows the SHRs and 95% CI using Fine and Gray competing risk models identifying SD and SE predictors of COVID-19 death and all OCOD, respectively, taking into account survival as competing event, performed on a 50% random subset representative of the total study population. The results show similar trends to those obtained using the logistic regression models (Table 2). Supplementary Table 3 shows the SHRs and 95% CI using Fine and Gray competing risk models identifying SD and SE predictors of COVID-19 death while taking into account all OCOD as competing event. The results show similar trends to those obtained using the logistic regression model (Table 3). This reinforces the robustness of our main results.

## Discussion

Thanks to the implementation of two complementary statistical methods, i.e. multivariable logistic regression models and CIT algorithms, the identification and ranking of SD and SE predictors of (1) COVID-19 death compared to survival; (2) all OCOD compared to survival; and (3) COVID-19 death compared to all OCOD were performed providing more reliable estimates and an easy hierarchical visual way of understanding the patterns characterizing mortality during the early phases of the COVID-19 pandemic in Belgium.

Compared to individuals not living in collectivities, our results showed higher odds of dying from COVID-19 among individuals living in collectivities (approximatively 12-fold higher and 7-fold higher for individuals living in care homes and those living in other type of collectivities, respectively). These results were confirmed by the CIT algorithm, highlighting that living in collectivities was the strongest predictor of COVID-19 death across all age groups, whether compared to survival or all OCOD. This aligns with several studies reporting important excess mortality during the COVID-19 pandemic in collectivities, whether in care homes or other settings (e.g. prisons) [18, 21, 32, 33]. This higher rate of COVID-19 deaths can be attributed on one hand to a higher comorbidity burden among residents of certain collectivities, such as prisons [34] and especially care homes, given the frailty profile of the residents [33, 35, 36]. On the other hand, higher virus transmission in collective settings in addition to pandemic-induced shortage of medical staff and resources may have led to higher COVID-19 related lethality [33, 37, 38] especially during the first year of the pandemic characterized by shortage masks, hydroalcoholic gels, and the unavailability of the vaccine at that time [39, 40].

Once individuals reside in collective settings, the CIT algorithm reveals that SD factors, i.e. age and sex, are the most important predictors of COVID-19 mortality, whether compared to survival or all OCOD. Age and gender are well known risk factors for COVID-19 death [8, 9, 41] and for mortality in general [42]. Ageing increases the risk of mortality following COVID-19 via many biological and behavioral factors such as: the overexpression of Angiotensin converting enzyme-2 (ACE-2) (receptor enzyme allowing the SARS-CoV-2 to enter the cell and replicate, leading to an acceleration of the replication of the virus in the lungs and thus a more severe infection) [22]; the presence of aging-associated comorbidities (e.g., diabetes, cardiovascular diseases) which are determinants in the progression towards a severe form of COVID-19 [43]; a change in lifestyle habits (malnutrition, decreased physical activity) also contributing to a more severe progression of the disease [22, 43]. The variation in susceptibility to infectious diseases between men and women can be explained by hormonal factors (e.g. higher plasma levels of ACE-2 in males) [20], genetic factors (e.g. high density of immune-related genes are located on the X chromosome giving women stronger immune response) [20], and behavioral factors (e.g. high-risk behaviors among men such as decreased perception of risks, higher rate of tobacco and alcohol use) [44].

For individuals not living in collective settings during the pandemic, the CIT identified, in addition to age and sex, education as one of the strongest predictors of COVID-19 mortality among individuals aged 65 to 84. Indeed, compared to individuals with higher levels of education, individuals with the lowest or an unknown level of education had a two times higher percentage of COVID-19 death. This was confirmed by the multivariable logistic regression model, which indicated that, compared to individuals with the highest level of education,

those with the lowest level of education had a 75% higher COVID-19 mortality compared to survivors. Our multivariable logistic regression model results also showed a 32% higher COVID-19 mortality among individuals with the lowest level of income, compared to those with the highest. Numerous studies documented social disparities in COVID-19 mortality or excess mortality during the pandemic [16–18, 45–50]. In Italy, Bello et al. identified a protective role of education on excess mortality during the first major COVID-19 outbreak. Their results showed that when the proportion of residents with at least secondary education increased by 10%, deaths per month decreased by 0.426 in the Northern region and by 0.081 in the Southern region [46]. In Canada, van Ingen et al. found an adjusted relative COVID-19 mortality risk two times higher in the neighborhood with a high proportion of individuals with less than high school education [51]. In Belgium, similar SE patterns were observed in excess mortality during the first COVID-19 wave with a 23.4% and 21.3% excess mortality among females with the lowest level of education not living in collectivities aged 25 to 64 and 65 to 84 years, respectively [18]. Regarding income inequalities, a study conducted in the Eastern Mediterranean region found a 3.9% increase in COVID-19 death per million population with a one unit increase in the Gini coefficient (a measure of income inequalities) [52]. In Belgium, during the first COVID-19 wave, similar patterns were observed by Decoster et al. who found a significant negative income gradient in excess mortality among people aged over 65 years, for both men and women [16]. In the Netherlands, Wouterse et al. found that, across all-age groups, COVID-19 mortality was more concentrated among low-income groups, compared to all other causes of mortality [53]. In our study, we found no significant difference in the multivariable logistic regression model by income level when comparing COVID-19 mortality to all OCOD. However, we did find significant differences by education level, with lower levels of education significantly increasing the odds of dying from COVID-19, compared with all OCOD. In addition, the CIT algorithm reported that income was a predictor of COVID-19 mortality compared with all OCOD in first-generation migrants aged 85 years and over.

A study conducted by Albani et al. identified that inequalities in transmission and in vulnerability were the two main factors explaining the highest proportion of COVID-19 mortality by deprivation [54]. Indeed, SE disadvantaged groups are more at risk of SARS-CoV-2 virus transmission and related severe COVID-19 outcomes due to increased vulnerability, susceptibility, exposure, and transmission [6]. SE disadvantaged groups are more vulnerable in terms of health conditions, as they are experiencing a higher burden of comorbidities [55], which has been shown to increase COVID-19 mortality [56]. They are more susceptible to be infected and developed severe related COVID-19 outcomes due to poor living conditions, isolation and chronic stress weakening the immune system, even without preexisting underlying health conditions [6, 57]. Working in essential sectors (e.g. social sectors, education, defense, logistic and transportation, manufacturing, facilities) during the pandemic has been shown to led to higher rates of SARS-CoV-2 infection and COVID-19 death [58, 59]. Individuals with lower levels of education are more likely to hold essential jobs or work in lower-skilled occupations, limiting their ability to work remotely during the pandemic. This reduced access to remote work created obstacles to implementing preventive measures, potentially contributing to higher exposure to the virus among SE disadvantaged groups [6, 60, 61]. In addition, a lower adoption of preventive measures has been identified as highly correlated with the level of education due to decreased perception of the risk and trust in the effectiveness of preventive measures [62, 63]. Finally, SE disadvantaged groups experienced an increased transmission by living in overcrowded households and neighborhood with higher population density with restricted access to outside spaces [6, 60, 64].

We identified in the multivariable logistic regression model a lower risk of all OCOD compared to survival up to 21% among first- and second-generation migrants. The lower migrant mortality, or *migrant mortality advantage*, is often explained by a selection effect (i.e. the healthiest individuals are able to immigrate) and healthier lifestyle habits (e.g. the Mediterranean diet leads to lower rates of chronic diseases) [65–67], leading to lower mortality in the destination country, at least in the early stages of immigration [65]. Conversely, we identified in the multivariable logistic regression model results a 24% higher COVID-19 mortality, compared to survival, among first-generation migrants, even when controlling for SE characteristics. This is in line with numerous studies reporting a higher risk of COVID-19 death among ethnic minorities [10, 13, 17, 68–70]. In addition, when comparing COVID-19 death to all OCOD, this percentage rises to 50% in the multivariable logistic regression model. This may suggest that, in the case of highly transmissible infectious diseases such as the SARS-CoV-2, the migrant mortality advantage is not applicable. This can be explained by poorer migrant's living conditions leading to a higher SARS-CoV-2 virus transmission and resulting in higher mortality from COVID-19 [71, 72]. Indeed, two studies reported that living in area with high density population, favoring virus exposure and infection, was the main factor explaining up to 60% of the disparities in COVID-19 mortality among ethnic minorities [13, 68]. Migrant's working conditions may also contribute to these disparities, as they are disproportionately

represented in essential and public-facing jobs, which increases their exposure to the virus [73]. In addition, once infected, difficulties in accessing health care services among migrants may have arisen due to language and cultural barriers as well as a lack of support system or digital literacy [74–76]. The CIT algorithm also identified the variable 'Migration background' as one of the strongest predictors of COVID-19 death compared to all OCOD among first-generation migrants aged 85 and over not living in collectivities and first-generation migrants aged 65 and over living in collectivities. This is probably due to the fact that advanced age leads to frailty combined with an attenuation of the migrant mortality advantage once individuals have immigrated for some time.

Our study has several limitations. Firstly, we were unable to adjust for the health status prior to COVID-19 outbreak (i.e. comorbidities), an essential factor to control for when analyzing cause-specific mortality data. Adjusting for some mediating factors of the socioeconomic disparities in COVID-19 severe outcomes such as the type of occupation, the adoption of preventive measures (e.g. mask wearing), and access to healthcare resources and services would have better captured the underlying pathways of cause-specific mortality during the COVID-19 pandemic. Second, the dataset used in this study contains individuals officially residing in Belgium. We therefore missed unregistered individuals, a more vulnerable part of the population (e.g. undocumented migrants), who may have had higher exposure to the SARS-CoV-2 virus and higher related COVID-19 mortality leading to an underestimation of cause-specific mortality for certain SD and SE groups. Third, the exact cause of death of individuals who died abroad were not available in the death certificate, deaths recorded abroad were therefore automatically categorized as all OCOD. This has potentially led to an underestimation of COVID-19 deaths among our study sample given that 21,792 deaths were recorded abroad, or 17% of the total number of deaths recorded in 2020. Fourth, the lack of testing at the start of the pandemic, and the fact that healthcare professionals might be still unfamiliar with the symptomatology of the COVID-19, which could potentially be confused with another respiratory virus, may have also led to an underestimation of COVID-19 deaths. Conversely, an overestimation of COVID-19 deaths during the first and second COVID-19 waves can be expected in care homes. Finally, the use of the income variable, which is based on taxes records, may not accurately represent certain population groups, such as those working in the informal economy, individuals earning below certain eligibility thresholds, or wealthier people benefiting from special tax arrangements. However, the direction and magnitude of this bias on the results remain uncertain.

Our study has several strengths. First, using an exhaustive dataset on the Belgian adult population allows to generalize our results at the national level. Second, the use of microlevel cause-specific mortality data allowed to distinguish COVID-19 specific death from all OCOD in our analyses. Belgium recorded COVID-19 related deaths with fairly good accuracy compared with other European countries, despite limited testing capacity during the first wave of the COVID-19 pandemic and difficulties in determining cause of death due to overlap with comorbidities [77]. Third, the use of two complementary methods provided a more in-depth investigation of the data. Thanks to the CIT algorithm, we were able to account for multicollinearity and interactions between the different SD and SE predictors, leading to robust conclusions (also with regard to the good performances of the CIT algorithms measured in terms of sensitivity and specificity). As the vaccination campaign started on 28 December 2020 in Belgium, the effect of the vaccine on cause-specific mortality is beyond the scope of this article. Given the SD and SE disparities identified in the COVID-19 vaccine uptake in Belgium [78], future research should focus on the effect of the vaccination campaign on SD and SE disparities in COVID-19 specific mortality during the pandemic.

## Conclusion

In conclusion, we identified important sociodemographic and socioeconomic disparities in COVID-19 mortality. Living in collectivities mainly predicts the risk of COVID-19 death across all age groups, highlighting the importance of preventive measures in reducing the transmission of the virus. For young and middle-aged individuals not living in collective settings during the early phases of the pandemic, having a low level of education plays a key role in predicting COVID-19 death. Our results are useful for future pandemic preparedness and suggest that policymakers should implement political measures and communication strategies promoting preventive behaviors adapted and targeted to the most vulnerable populations more exposed to the virus.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13690-024-01437-8.

Supplementary Material 1

**Acknowledgements**

We would like to thank Patrick Lusyne and Elias Neirynck from Statistic Belgium for their valuable advices. In addition, we would like to thank Johan Surkyn, the BRISPO data manager, who requested the data.

**Author contributions**

LC reviewed the literature; LC, LVdB, KV, AS, BD, NK, and SG conceived the study; LC, LVdB, KV, AS, BD, NK, and SG selected the population; LC, LVdB,

KV, AS, and SG reviewed all available data; LC, LVdB, KV, AS, BD, NK, and SG designed the statistical methodology; LC conducted the statistical analyses; LC, LVdB, KV, AS, BD, NK, and SG interpret the findings; LC wrote the first draft of the paper. All authors revised the text. LC, LVdB, KV, AS, and SG have directly access to the data of the present study. All authors approved the final version of this manuscript and accepted responsibility for its submission for publication.

**Funding**
The Belgian Science Policy Office (BELSPO) funded this research within the BRAIN-be 2.0 framework supporting pillar 3 Federal societal challenges (grant number B2/202/P3/HELICON). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data availability**
The analyses are based on data from a census-linked mortality follow-up study and cannot be made available due to privacy issues. Researchers can gain full access to the data by submitting an application to the Privacy Commission Belgium. In order to get permission to use data from the Belgian population register linked to census data an authorization request (in Dutch or French) needs to be submitted to the Belgian Data Protection Authority. The authorization request includes an application form and additional forms regarding data security. The necessary forms for the authorization request can be downloaded from the Data Protection Authority website (https://www.dataprotectionauthority.be). Next to information on the applicant and a list of requested data, the authorization request should specify why the data from the population register are necessary, for which time span data will be stored, and who will have access to the data.

## Declarations

**Ethics approval and consent to participate**
This research adhere to the ethical code of scientific research in Belgium, see: http://www.belspo.be/belspo/organisation/publ/pub_ostc/Eth_code/ethcode_nl.pdf.

**Consent for publication**
Not applicable.

**Competing interests**
Brecht Devleesschauwer is the editor in chief of Archives of Public Health. The other authors have no competing interests to declare that are relevant to the content of this article.

**Author details**
[1]Department of Epidemiology and Public Health, Sciensano, Rue Ernest Blerot 1, Anderlecht, Brussels 1070, Belgique
[2]Research Institute of Health and Society, University of Louvain, Brussels, Belgium
[3]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium
[4]Interface Demography, Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium
[5]Brussels Institute for Social and Population Studies, Vrije Universiteit Brussel, Brussels, Belgium
[6]Department of Translational Physiology, Infectiology and Public health, Ghent University, Merelbeke, Belgium

Received: 16 July 2024 / Accepted: 28 October 2024
Published online: 13 November 2024

**References**

1. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed. 2020;91:157–60.
2. Wang H, Paulson KR, Pease SA, Watson S, Comfort H, Zheng P, et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. Lancet. 2022;399:1513–36.
3. Belgium COVID-. 19 death Dashboard - Sciensano [Internet]. Google Data Studio. [cited 2021 Aug 23]. http://datastudio.google.com/reporting/c14a5cfc-cab7-4812-848c-0369173148ab/page/QTSKB?feature=opengraph
4. Décès. | Statbel [Internet]. [cited 2023 Nov 28]. https://statbel.fgov.be/fr/themes/population/mouvement-de-la-population/deces
5. Bustos Sierra N, Bossuyt N, Braeye T, Haarhuis F, Peeters I, Proesmans K et al. Excess mortality during the first and second wave of the COVID-19 epidemic in Belgium (data from 10 March 2020 to 14 February 2021) [Internet]. Brussels, Belgium: Sciensano; 2021 p. 47. Report No.: D/2021/14.440/62. https://epistat.wiv-isp.be/momo/
6. Bambra C, Lynch J, Smith KE. The Unequal Pandemic: COVID- 19 and Health Inequalities [Internet]. Policy Press; 2021 [cited 2023 May 13]. https://library.oapen.org/handle/20.500.12657/51451
7. Horton R, Offline. COVID-19 is not a pandemic. Lancet. 2020;396:874.
8. Pijls BG, Jolani S, Atherley A, Derckx RT, Dijkstra JIR, Franssen GHL, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. BMJ Open. 2021;11:e044640.
9. Booth A, Reed AB, Ponzo S, Yassaee A, Aral M, Plans D, et al. Population risk factors for severe disease and mortality in COVID-19: a global systematic review and meta-analysis. PLoS ONE. 2021;16:e0247461.
10. 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. 2020;11:5097.
11. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584:430–6.
12. Ribeiro KB, Ribeiro AF, Veras MA, de SM MC. Social inequalities and COVID-19 mortality in the city of São Paulo, Brazil. Int J Epidemiol. 2021;50:732–42.
13. Nafilyan V, Islam N, Mathur R, Ayoubkhani D, Banerjee A, Glickman M, et al. Ethnic differences in COVID-19 mortality during the first two waves of the Coronavirus Pandemic: a nationwide cohort study of 29 million adults in England. Eur J Epidemiol. 2021;36:605–17.
14. Seligman B, Ferranna M, Bloom DE. Social determinants of mortality from COVID-19: a simulation study using NHANES. PLoS Med. 2021;18:e1003490.
15. Krieger N, Waterman PD, Chen JT. COVID-19 and overall mortality inequities in the Surge in Death Rates by Zip Code characteristics: Massachusetts, January 1 to May 19, 2020. Am J Public Health. 2020;110:1850–2.
16. Decoster A, Minten T, Spinnewijn J. The income gradient in mortality during the Covid-19 Crisis: evidence from Belgium. J Econ Inequal. 2021;19:551–70.
17. Vanthomme K, Gadeyne S, Lusyne P, Vandenheede H. A population-based study on mortality among Belgian immigrants during the first COVID-19 wave in Belgium. Can demographic and socioeconomic indicators explain differential mortality? SSM - Popul Health. 2021;14:100797.
18. Gadeyne S, Rodriguez-Loureiro L, Surkyn J, Hemelrijck WV, Nusselder W, Lusyne P, et al. Are we really all in this together? The social patterning of mortality during the first wave of the COVID-19 pandemic in Belgium. Int J Equity Health. 2021;20:258.
19. Vanthomme K, Gadeyne S, Devleesschauwer B, Van den Borre L. Excess mortality among native Belgians and migrant groups in Belgium during the first three COVID-19 waves: the evolving dynamics of social inequalities. J Public Health (Berl) [Internet]. 2023 [cited 2024 Feb 22]; https://doi.org/10.1007/s10389-023-02180-0
20. Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol. 2016;16:626–38.
21. Davies B, Parkes BL, Bennett J, Fecht D, Blangiardo M, Ezzati M, et al. Community factors and excess mortality in first wave of the COVID-19 pandemic in England. Nat Commun. 2021;12:3755.
22. Wong LSY, Loo EXL, Kang AYH, Lau HX, Tambyah PA, Tham EH. Age-related differences in immunological responses to SARS-CoV-2. J Allergy Clin Immunol Pract. 2020;8:3251–8.
23. Balaj M, Henson CA, Aronsson A, Aravkin A, Beck K, Degail C, et al. Effects of education on adult mortality: a global systematic review and meta-analysis. Lancet Public Health. 2024;9:e155–65.
24. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference Framework. J Comput Graphical Stat. 2006;15:651–74.
25. Hothorn T, Hornik K, Zeileis A. ctree: Conditional Inference Trees.
26. Kuncheva LI. Combining Pattern Classifiers.
27. ctree function -. RDocumentation [Internet]. [cited 2023 Oct 5]. https://www.rdocumentation.org/packages/partykit/versions/1.2-20/topics/ctree
28. Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med. 2011;18:1099–104.

29. Speybroeck N. Classification and regression trees. Int J Public Health. 2012;57:243–6.
30. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing. 2020. https://www.R-project.org/
31. Hothorn T, Seibold H, Zeileis A. partykit: A Toolkit for Recursive Partytioning [Internet]. 2023 [cited 2023 Nov 28]. https://cran.r-project.org/web/packages/partykit/index.html
32. Gulliford MC, Prevost AT, Clegg A, Rezel-Potts E. Mortality of Care Home residents and Community-Dwelling controls during the COVID-19 pandemic in 2020: Matched Cohort Study. J Am Med Dir Assoc. 2022;23:923–e9292.
33. Hardy OJ, Dubourg D, Bourguignon M, Dellicour S, Eggerickx T, Gilbert M, et al. A world apart: levels and determinants of excess mortality due to COVID-19 in care homes: the case of the Belgian region of Wallonia during the spring 2020 wave. Demographic Res. 2021;45:1011–40.
34. Ali K, Rao S, Berdine G, Test V, Nugent K. A retrospective analysis and comparison of prisoners and Community-based patients with COVID-19 requiring Intensive Care during the First Phase of the pandemic in West Texas. J Prim Care Community Health. 2020;11:2150132720954687.
35. Akner G. Analysis of multimorbidity in individual elderly nursing home residents. Development of a multimorbidity matrix. Arch Gerontol Geriatr. 2009;49:413–9.
36. Zhang X, Dou Q, Zhang W, Wang C, Xie X, Yang Y, et al. Frailty as a predictor of all-cause Mortality among older nursing home residents: a systematic review and Meta-analysis. J Am Med Dir Assoc. 2019;20:657–e6634.
37. Adams C, Chamberlain A, Wang Y, Hazell M, Shah S, Holland DP, et al. The role of staff in transmission of SARS-CoV-2 in Long-Term Care facilities. Epidemiology. 2022;33:669.
38. Sugie NF, Turney K, Reiter K, Tublitz R, Kaiser D, Goodsell R, et al. Excess mortality in U.S. prisons during the COVID-19 pandemic. Sci Adv. 2023;9:eadj8104.
39. Jamart H, Van Maele L, Ferguson M, Drielsma P, Macq J, Van Durme T. La première vague de Covid-19 en Belgique et les soins primaires. Rev Med Suisse. 2020;713:2119–22.
40. Catteau L, van Loenhout J, Stouten V, Billuart M, Hubin P, Haarhuis F et al. Thematisch Verslag: Vaccinatiegraad en Epidemiologische Impact van de Covid-19-Vaccinatiecampagne in België. Gegevens tot en met 31 oktober 2021 [Internet]. Brussels, Belgium: Sciensano; 2021 p. 53. Report No.: D/2021/14.440/80. https://covid-19.sciensano.be/sites/default/files/Covid19/COVID_19_THEMATIC_REPORT_VaccineCoverageAndImpactReport_NL.pdf
41. Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. BMC Infect Dis. 2021;21:855.
42. Abdalla S, Abd-Allah F, Abdel Aziz MI. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. 2015 [cited 2024 May 17]; https://fada.birzeit.edu/handle/20.500.11889/7859
43. Farshbafnadi M, Kamali Zonouzi S, Sabahi M, Dolatshahi M, Aarabi MH, Aging. COVID-19 susceptibility, disease severity, and clinical outcomes: the role of entangled risk factors. Exp Gerontol. 2021;154:111507.
44. Griffith DM. Men and COVID-19: A Biopsychosocial Approach to Understanding Sex Differences in Mortality and Recommendations for Practice and Policy Interventions. Prev Chronic Dis [Internet]. 2020 [cited 2023 May 16];17. https://www.cdc.gov/pcd/issues/2020/20_0247.htm
45. Arceo-Gomez EO, Campos-Vazquez RM, Esquivel G, Alcaraz E, Martinez LA, Lopez NG. The income gradient in COVID-19 mortality and hospitalisation: an observational study with social security administrative records in Mexico. Lancet Reg Health Am. 2022;6:100115.
46. Bello P, Rocco L, Education. COVID-19 excess mortality. Econ Hum Biol. 2022;47:101194.
47. C DG, Ml LBNC. M. Socioeconomic inequalities in overall and COVID-19 mortality during the first outbreak peak in Emilia-Romagna Region (Northern Italy). Epidemiologia e prevenzione [Internet]. 2020 [cited 2023 Nov 21];44. https://pubmed.ncbi.nlm.nih.gov/33412821/
48. Dukhovnov D, Barbieri M. County-level socio-economic disparities in COVID-19 mortality in the USA. Int J Epidemiol. 2022;51:418–28.
49. Feng Z. Spatiotemporal pattern of COVID-19 mortality and its relationship with socioeconomic and environmental factors in England. Spat Spatiotemporal Epidemiol. 2023;45:100579.
50. Hawkins D. Social determinants of COVID-19 in Massachusetts, United States: an ecological study. J Prev Med Public Health. 2020;53:220–7.
51. van Ingen T, Akingbola S, Brown KA, Daneman N, Buchan SA, Smith BT. Neighbourhood-level risk factors of COVID-19 incidence and mortality [Internet]. medRxiv; 2021 [cited 2024 Mar 28]. p. 2021.01.27.21250618. https://www.medrxiv.org/content/https://doi.org/10.1101/2021.01.27.21250618v1
52. Alam MF, Wildman J, Rahim HA. Income inequality and its association with COVID-19 cases and deaths: a cross-country analysis in the Eastern Mediterranean region. BMJ Global Health. 2023;8:e012271.
53. Wouterse B, Geisler J, Bär M, van Doorslaer E. Has COVID-19 increased inequality in mortality by income in the Netherlands? J Epidemiol Community Health. 2023;77:244–51.
54. Albani V, Welsh CE, Brown H, Matthews FE, Bambra C. Explaining the deprivation gap in COVID-19 mortality rates: a decomposition analysis of geographical inequalities in England. Soc Sci Med. 2022;311:115319.
55. Pathirana TI, Jackson CA. Socioeconomic status and multimorbidity: a systematic review and meta-analysis. Aust N Z J Public Health. 2018;42:186–94.
56. Gold MS, Sehayek D, Gabrielli S, Zhang X, McCusker C, Ben-Shoshan M. COVID-19 and comorbidities: a systematic review and meta-analysis. Postgrad Med. 2020;132:749–55.
57. Mattos dos Santos R. Isolation, social stress, low socioeconomic status and its relationship to immune response in Covid-19 pandemic context. Brain Behav Immun - Health. 2020;7:100103.
58. Bonde JPE, Begtrup LM, Jensen JH, Flachs EM, Schlünssen V, Kolstad HA, et al. Occupational risk of SARS-CoV-2 infection: a nationwide register-based study of the Danish workforce during the COVID-19 pandemic, 2020–2021. Occup Environ Med. 2023;80:202–8.
59. Chen Y-H, Riley AR, Duchowny KA, Aschmann HE, Chen R, Kiang MV, et al. COVID-19 mortality and excess mortality among working-age residents in California, USA, by occupational sector: a longitudinal cohort analysis of mortality surveillance data. Lancet Public Health. 2022;7:e744–53.
60. Brandily P, Brébion C, Briole S, Khoury LA, Poorly Understood. Disease? The Unequal Distribution of Excess Mortality Due to COVID-19 Across French Municipalities [Internet]. 2020 [cited 2023 Jun 26]. https://hal.science/halshs-02895908
61. Capasso A, Kim S, Ali SH, Jones AM, DiClemente RJ, Tozan Y. Employment conditions as barriers to the adoption of COVID-19 mitigation measures: how the COVID-19 pandemic may be deepening health disparities among low-income earners and essential workers in the United States. BMC Public Health. 2022;22:870.
62. Kaplan RM, Fang Z, Kirby J. Educational attainment and health outcomes: data from the Medical expenditures Panel Survey. Health Psychol. 2017;36:598–608.
63. Pförtner T-K, Dohle S, Hower KI. Trends in educational disparities in preventive behaviours, risk perception, perceived effectiveness and trust in the first year of the COVID-19 pandemic in Germany. BMC Public Health. 2022;22:903.
64. Ahmad K, Erqou S, Shah N, Nazir U, Morrison AR, Choudhary G, et al. Association of poor housing conditions with COVID-19 incidence and mortality across US counties. PLoS ONE. 2020;15:e0241327.
65. Anson J. The migrant mortality advantage: a 70 Month follow-up of the Brussels Population. Eur J Popul. 2004;20:191–218.
66. Darmon N, Khlat M. An overview of the health status of migrants in France, in relation to their dietary practices. Public Health Nutr. 2001;4:163–72.
67. Deboosere P, Gadeyne S. La sous-mortalité des immigrés adultes en Belgique: une réalité attestée par les recensements et les registres. Popul (French Edition). 2005;60:765–811.
68. Ayoubkhani D, Nafilyan V, White C, Goldblatt P, Gaughan C, Blackwell L, et al. Ethnic-minority groups in England and Wales-factors associated with the size and timing of elevated COVID-19 mortality: a retrospective cohort study linking census and death records. Int J Epidemiol. 2021;49:1951–62.
69. Mackey K, Ayers CK, Kondo KK, Saha S, Advani SM, Young S, et al. Racial and ethnic disparities in COVID-19-Related infections, hospitalizations, and deaths: a systematic review. Ann Intern Med. 2021;174:362–73.
70. Irizar P, Pan D, Kapadia D, Bécares L, Sze S, Taylor H et al. Ethnic inequalities in COVID-19 infection, hospitalisation, intensive care admission, and death: a global systematic review and meta-analysis of over 200 million study participants. eClinicalMedicine [Internet]. 2023 [cited 2023 May 4];57. https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(23)00054-8/fulltext#secsectitle0120
71. Hayward SE, Deal A, Cheng C, Crawshaw AF, Orcutt M, Vandrevala TF et al. Clinical outcomes and risk factors for COVID-19 among migrant populations in high-income countries: a systematic review [Internet]. medRxiv; 2020 [cited 2024 Mar 29]. p. 2020.12.21.20248475. https://www.medrxiv.org/content/https://doi.org/10.1101/2020.12.21.20248475v1

72. Kjøllesdal M, Skyrud K, Gele A, Arnesen T, Kløvstad H, Diaz E, et al. The correlation between socioeconomic factors and COVID-19 among immigrants in Norway: a register-based study. Scand J Public Health. 2022;50:52–60.
73. Jaljaa A, Caminada S, Tosti ME, D'Angelo F, Angelozzi A, Isonne C, et al. Risk of SARS-CoV-2 infection in migrants and ethnic minorities compared with the general population in the European WHO region during the first year of the pandemic: a systematic review. BMC Public Health. 2022;22:143.
74. Kang SJ, Hyung JA, Han H-R. Health literacy and health care experiences of migrant workers during the COVID-19 pandemic: a qualitative study. BMC Public Health. 2022;22:2053.
75. Knights F, Carter J, Deal A, Crawshaw AF, Hayward SE, Jones L, et al. Impact of COVID-19 on migrants' access to primary care and implications for vaccine roll-out: a national qualitative study. Br J Gen Pract. 2021;71:e583–95.
76. Is health a right. for all? An umbrella review of the barriers to health care access faced by migrants: Ethnicity & Health: Vol 20, No 5 - Get Access [Internet]. [cited 2024 Mar 29]. https://www.tandfonline.com/doi/full/https://doi.org/10.1080/13557858.2014.946473
77. OECD, Belgique. Profils de santé par pays 2021 [Internet]. Paris: Organisation for Economic Co-operation and Development; 2022 [cited 2024 Apr 16]. https://www.oecd-ilibrary.org/social-issues-migration-health/belgique-profils-de-sante-par-pays-2021_17c34ec7-fr
78. Cavillot L, van Loenhout JAF, Devleesschauwer B, Wyndham-Thomas C, Oyen HV, Ghattas J et al. Sociodemographic and socioeconomic disparities in COVID-19 vaccine uptake in Belgium: a nationwide record linkage study. J Epidemiol Community Health [Internet]. 2023 [cited 2024 Jan 15]; https://jech.bmj.com/content/early/2023/12/26/jech-2023-220751

## Publisher's note