



A retrospective cohort study of risk factors for mortality among nursing homes exposed to COVID-19 in Spain

Clara Suñer^{1,15}, Dan Ouchi^{1,2,15}, Miquel Àngel Mas^{3,4}, Rosa Lopez Alarcon⁵, Mireia Massot Mesquida⁶, Núria Prat⁶, Josep Maria Bonet-Simó⁶, Marta Expósito Izquierdo⁶, Irene Garcia Sánchez⁶, Sara Rodoreda Noguerola⁶, Montserrat Teixidó Colet⁶, Joaquim Verdaguer Puigvendrelló⁶, Norma Henríquez⁶, Ramón Miralles^{2,3,4}, Eugènia Negrodo^{1,7,8}, Marc Noguera-Julian^{9,10}, Michael Marks^{11,12}, Oriol Estrada¹³, Jordi Ara¹³ and Oriol Mitjà^{1,7,10,14} ✉

Long-term care (LTC) facilities have shown remarkably high mortality rates during the coronavirus disease 2019 (COVID-19) outbreak in many countries¹, and different risk factors for mortality have been identified in this setting^{2–5}. Using facilities as the unit of analysis, we investigated multiple variables covering facility characteristics and socioeconomic characteristics of the geographic location to identify risk factors for excess mortality from a comprehensive perspective. Furthermore, we used a clustering approach to detect patterns in datasets and generate hypotheses regarding potential relationships between types of nursing homes and mortality trends. Our retrospective analysis included 167 nursing homes providing LTC to 8,716 residents during the COVID-19 outbreak in Catalonia (northeast Spain). According to multiple regression analysis, COVID-19-related and overall mortality at the facility level were significantly associated with a higher percentage of patients with complex diseases, lower scores on pandemic preparedness measures and higher population incidence of COVID-19 in the surrounding population. When grouping nursing homes into eight clusters based on common features, we found higher mortality rates in four clusters, mainly characterized by a higher proportion of residents with complex chronic conditions or advanced diseases, lower scores on pandemic preparedness, being located in rural areas and larger capacity, respectively.

Twelve months after the first outbreak of coronavirus disease 2019 (COVID-19), the global death toll associated with the pandemic amounted to more than 2 million people⁶. Residents of long-term care (LTC) facilities, such as nursing homes, are at high risk of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and also at high risk of death¹. The findings of several reports suggest that, once SARS-CoV-2 has been introduced into an LTC facility, it can spread rapidly and widely,

causing considerable impact among facility residents and local healthcare systems^{7–10}.

Large variations in COVID-19 death rates across LTC facilities have been observed. Whether high death rates are exclusively linked to vulnerability of residents, owing to advanced ages and underlying health conditions and whether structural features or geographic location of LTC facilities are a contributing factor are still not fully answered questions. Because of different policy implications of the relative influence of these two groups of characteristics, there is a need to more comprehensively understand determinants of SARS-CoV-2 spread and mortality in LTC facilities^{2,11}.

Various predictors of higher COVID-19 infection rates in nursing homes have been reported, including larger facility size, higher proportion of residents with limited income, greater percentage of African Americans (in the US), urban location and community-level rates of COVID-19 (refs. ^{8,9,12,13}). Similarly, several possible reasons for the large percentage of deaths from COVID-19 among residents in LTCs have been identified in previous studies. The first reason is individual residents' characteristics such as older age or high levels of multi-comorbidity^{3,14}. The second reason is nursing home characteristics, primarily size (that is, larger number of residents and crowding)^{4,15} and type of care provided (featured as the proportion of residents with complex or advanced diseases)^{5,10,11,14,16}. This variable can be used as a proxy to differentiate two types of senior housing centers: nursing homes that provide around-the-clock care and monitoring to residents with complex healthcare conditions, in contrast to nursing homes that provide assisted living and custodial care to residents who are mostly independent. There are also mixed facilities with both types of residents. The third reason is the level of pandemic preparedness in the nursing home. Residential settings need to follow specific guidance for managing the outbreak to prevent infections in communal living areas and shared rooms^{4,10}, to stop transmission from multiple caregivers⁷, who may work across

¹Fight AIDS and Infectious Diseases Foundation, Badalona, Spain. ²Universitat Autònoma de Barcelona, Barcelona, Spain. ³Direcció Clínica Territorial de Cronicitat Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain. ⁴Department of Geriatrics, Hospital Universitari Germans Trias i Pujol, Badalona, Spain. ⁵Direcció d'Organització i Sistemes d'Informació Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain. ⁶Direcció d'Atenció Primària Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain. ⁷Infectious Diseases Department, Hospital Universitari Germans Trias i Pujol, Badalona, Spain. ⁸Centre for Health and Social Care Research, Faculty of Medicine, University of Vic—Central University of Catalonia, Barcelona, Spain. ⁹Institut de Recerca de la Sida, IrsiCaixa, Badalona, Spain. ¹⁰Universitat de Vic—Universitat Central de Catalunya, Vic, Spain. ¹¹Clinical Research Department, London School of Hygiene & Tropical Medicine, London, UK. ¹²Hospital for Tropical Diseases, London, UK. ¹³Gerència Territorial Metropolitana Nord, Institut Català de la Salut, Barcelona, Spain. ¹⁴Lihir Medical Centre, International SOS, Lihir Island, Papua New Guinea. ¹⁵These authors contributed equally: Clara Suñer, Dan Ouchi. ✉e-mail: omitja@flsida.org

Table 1 | Characteristics of nursing homes

	Total (n = 167)
Individual residents' characteristics	
Age of residents in years	87.1 (2.1)
Percentage of male residents	26.4 (9.6)
Number of comorbidities	1.5 (0.6)
Nursing home characteristics and type of care provided	
Percentage of dependent residents ^a	82.1 (9.5)
Percentage of CCPs	46.1 (17.3)
Percentage of patients with ACD	10.5 (8.9)
Current number of residents	46.2 (29.8)
Pandemic preparedness by nursing home	
SNQ12 score	1.4 (1.7)
Percentage of residents who return home	1.4 (3.0)
Geographic location characteristics	
Mean household income in euros	36,099.6 (5,527.5)
Density of population per km ² in log ₁₀	17.9 (9.5)
Incidence rate of COVID-19 in 14-d cases per 100,000 population	900 (300)

Variables are described as mean (s.d.), computed as the mean of percentages in each facility (categorical variables) or the mean of means in each facility (continuous variables). SNQ12, number of unmet epidemic and infection-control preparedness requirements (scale of 0–12). ^aBarthel score < 50.

multiple different facilities, and to avoid shortages of healthcare resources (for example, tests and personal protective equipment)¹⁰. The fourth reason is the geographic location of nursing homes in higher-risk areas. Some authors have identified risk factors associated with characteristics of the geographic location and surrounding population, such as mean household income, population density of the area, ethnicity profile or high population incidence of COVID-19 (refs. ^{13,15,17,18}).

In this retrospective analysis, we investigated determinants of mortality in LCT facilities during the COVID-19 outbreak. A major component of the newness of this study comes from the use of multiple variables covering both facility characteristics (for example, size, type of care provided and preparedness) and characteristics of the geographic location (for example, household income, population density and local COVID-19 incidence). Furthermore, we used a twofold analysis to provide a more comprehensive view of determinants of mortality: first, multivariable analysis to more precisely identify the contribution of facility-specific factors and those of the communities they serve to COVID-19-related excess mortality and to overall excess mortality. Second, we used clustering approaches to detect patterns in our dataset regarding potential relationships of nursing homes grouped according to common characteristics and describing mortality trends across clusters.

Our analysis of all public and private nursing homes for LTC in the study area included 167 nursing homes providing LTC to 8,716 residents. Table 1 summarizes characteristics of the nursing homes included in the analysis. Most nursing homes had a high percentage of individuals older than 80 years (mean age at the facility level, 87.1 years). An average of 46.1% of individuals were patients with complex chronic disease (CCP), 10.5% were patients with advanced chronic disease (ACD) and 82.1% were highly dependent on daily activities. The median SNQ12 score (an ad hoc index of unmet needs among 12 essential items for pandemic preparedness and response (values 1 (highest preparedness) to 12 (lowest preparedness))) was 1.4. Individual demographic, clinical and

epidemiological characteristics of included residents are summarized in Supplementary Table 1.

Between 1 March 2020 and 1 June 2020, a total of 1,629 deaths were reported among people institutionalized in the nursing homes included in the analysis (Extended Data Fig. 1). Of these, 1,089 (66.9%) were registered as COVID-19-related deaths in the mortality registry of the Department of Health. The cause of the remaining 540 deaths could not be confirmed. Overall, the excess deaths in the analyzed nursing homes compared with deaths in the same period in the four previous years were estimated to be 971 deaths; COVID-19-related deaths accounted for 89.2% of all excess mortality. At the nursing home level, the median (interquartile range) mortality rate was 14.3 (7.6–26.1) deaths per 100 residents per 3-month study period for all-cause deaths and 3.9 (0.0–18.4) for COVID-19-related deaths.

Owing to the presence of deaths recorded as of undetermined cause, we investigated characteristics of nursing homes that may contribute to explaining mortality using two outcomes: COVID-19-related death (that is, laboratory confirmed and clinically suspected) and all-cause death. According to the multivariable analysis (Table 2), the risk of COVID-19-related deaths was higher in nursing homes with a higher percentage of complex CCPs (incidence rate ratio (IRR), 1.07; 95% confidence interval (CI), 1.04–1.11 per ten-unit increase), lower scores on pandemic preparedness measures (IRR, 1.05; 1.02–1.09 per unit increase) and geographic location in areas with a higher COVID-19 14-d incidence rate per 100,000 population (IRR, 2.39; 1.95–2.93 per 100-unit increase). Risk factors for all-cause mortality were the same as those for COVID-19-related death, and two additional factors were identified: a higher percentage of male residents (IRR, 1.07; 1.01–1.13 per ten-unit increase) and nursing homes with larger numbers of residents (IRR, 1.02; 1.01–1.03 per ten-unit increase). At the individual level, age, sex, comorbidities and underlying incidence of SARS-CoV-2 in the surrounding population were dominant risk factors for COVID-19-related death (Supplementary Table 2).

Complementary to the multivariable analysis, we conducted a cluster analysis of nursing homes grouped according to common characteristics and assessed mortality trends across clusters (Extended Data Fig. 2). The cluster analysis based on the *k*-nearest-neighbor classifier identified eight groups of nursing homes that were significantly different from each other. Fig. 1 illustrates the intensity of each characteristic (that is, the extent of the difference between the mean of a given cluster and that of the entire sample) in the resulting clusters and the contribution of each characteristic to their definition. Key characteristics of each cluster are shown in Fig. 1.

Nursing homes in cluster 1 were located in low-density populated areas with high population incidence of COVID-19 and high household income; cluster 2 had facilities with a high proportion of CCPs and patients with ACD, located in areas with low population incidence of COVID-19; cluster 3 had a low proportion of CCPs and highly dependent residents and a high proportion of male residents; cluster 4 had a higher number of residents than the median, although with a very low proportion of CCPs; nursing homes in this cluster were located in areas with low household income; cluster 5 had a low proportion of patients with ACD and dependent residents and had higher number of residents that returned home with their relatives; cluster 6 facilities were located in areas with high household income and low population incidence of COVID-19; cluster 7 had a high proportion of CCPs and patients with ACD; nursing homes in this cluster were located in densely populated areas; cluster 8 had a high SNQ12 score (indicating a very limited capacity for pandemic preparedness and response) and a high proportion of CCPs and younger residents than the median.

The ANOVA analysis revealed significant differences in mortality (both all-cause and COVID-19-related mortality) among clusters (Fig. 2). Clusters 1, 7 and 8, particularly characterized by

Table 2 | Estimated effect of LTC facilities' characteristics on all deaths and COVID-19-related deaths at the facility level

	All-cause deaths			COVID-19-related deaths		
	Univariate analysis	Multivariable analysis		Univariate analysis	Multivariable analysis	
	IRR (95% CI)	IRR (95% CI)	P value	IRR (95% CI)	IRR (95% CI)	P value
Individual residents' characteristics						
Age of residents in years ^a	1.00 (0.97-1.02)	-	-	0.98 (0.95-1.02)	-	-
Percent of male residents	1.06 (1.00-1.12) ^e	1.07 (1.01-1.13)	0.0263	1.03 (0.95-1.12)	-	-
Number of comorbidities ^a	1.18 (1.08-1.28) ^e	-	-	1.31 (1.18-1.46) ^e	-	-
Nursing home characteristics and type of care provided						
Percent dependent residents	1.00 (0.95-1.05)	-	-	1.00 (0.93-1.07)	-	-
Percent CCPs	1.04 (1.01-1.06) ^e	1.07 (1.03-1.10)	<0.0001	1.06 (1.02-1.09) ^e	1.07 (1.04-1.11)	0.0001
Percent residents with ACD	1.06 (1.00-1.11)	-	-	1.10 (1.02-1.18) ^e	-	-
Current number of residents	1.02 (1.00-1.03) ^e	1.02 (1.01-1.03)	0.0098	0.95 (0.98-1.01)	-	-
Pandemic preparedness by nursing home						
SNQ12 ^a	1.06 (1.04-1.09) ^e	1.05 (1.03-1.08)	0.0001	1.07 (1.04-1.10) ^e	1.05 (1.02-1.09)	0.0005
Percent of residents who return home	0.95 (0.79-1.13)	-	-	0.86 (0.67-1.10)	-	-
Geographic location characteristics						
Mean household income in euros ^b	1.00 (0.99-1.00)	-	-	0.99 (0.98-1.00)	-	-
Population density, log ₁₀ (people km ⁻²) ^c	1.06 (0.89-1.27)	-	-	1.00 (0.80-1.26)	-	-
Population incidence of COVID-19 ^d	1.89 (1.63-2.20) ^e	1.88 (1.61-2.19)	<0.0001	2.25 (1.85-2.73) ^e	2.39 (1.95-2.93)	<0.0001

Dependent resident, Barthel score <50; SNQ12, number of unmet essential items for implementing preventive measures (0-12). IRRs and 95% CIs are shown for the zero-inflated Poisson model. IRR represents the estimated effect for an increase of ten units, unless otherwise indicated. Two-sided *P* values of effect size were obtained from the Wald test. ^aIRR for an increase in one unit. ^bIRR for an increase of €10,000 per annum in mean household income. ^cIRR for an increase in 1 log₁₀ (people per km²). ^dIRR for an increase in 100 cases of the COVID-19 14-d incidence rate per 100,000 population. ^e*P* value for univariate analysis <0.05. Omnibus test: all-cause death model (χ^2 statistic, 1,332; *P* value <0.0001) and COVID-19-related death model (χ^2 statistic, 2,211; *P* value <0.0001).

a geographic location with low population density, higher percentages of residents with complex or advanced chronic diseases and lower scores on the pandemic preparedness measure, respectively, had a greater COVID-19-related mortality rate than the median. Clusters 1 and 8, but not cluster 7, also reported higher all-cause mortality rates than the median, along with cluster 4, characterized by a higher number of residents.

In this analysis of 167 nursing homes, we provide a new approach to understanding differences in mortality between nursing homes during the COVID-19 outbreak. Most studies on risk factors for COVID-19 mortality in the LTC setting did not simultaneously address potential confounding factors at the resident, organizational or community level that may influence measured outcomes. Alternatively, we used multiple variables covering both facility characteristics and socioeconomic factors related to the geographic location to investigate determinants of mortality in LTC facilities using a multivariable model and a clustering approach to detect patterns of LTC facilities that are at higher risk.

Our analysis revealed that a 10% increase in the proportion of residents with complex conditions increased the mortality risk by 7%, a one-point increase in the 12-point score of unmet measures for containing SARS-CoV-2 spread increased the mortality risk by 5%, and an increase of 100 in the COVID-19 14-d incidence rate per 100,000 population increased the mortality risk by 140%.

In addition to risk factors for mortality at the facility level, we aimed to identify nursing home profiles that were at higher risk of mortality to assist policy and healthcare planning in the advent of future outbreaks. Consistent with our regression analysis, clusters with greater mortality rates than the median (clusters 1, 4, 7 and 8) were all located in neighborhoods with high incidence of COVID-19. These results align with previous studies that

reported a significant relationship between COVID-19 incidence in the catchment area and the risk of an outbreak^{12,13} and mortality¹⁸ at the facility level. The influence of the local incidence of COVID-19 on mortality underscores the paramount importance of preventing SARS-CoV-2 entry into facilities (often with new residents, staff, or visitors^{11,18,19}) and suggests that efforts to contain SARS-CoV-2 transmission in the community may also contribute to reducing numbers of deaths due to COVID-19 at local nursing homes. Results from multiple regression and cluster analyses were also consistent regarding the importance of the capacity of the nursing home for pandemic preparedness and response, as described previously^{7,10}. Although most nursing homes showed low SNQ12 scores (indicating few unmet needs for applying containment measures), facilities in cluster 8, characterized by higher SNQ12 scores (mean of five unmet items over a total of 12 essential requirements) than the median, experienced high mortality levels.

Finally, our multivariable analysis revealed a significant relationship between higher percentages of CCPs and increased mortality risk. According to local clinical guidelines, CCPs are characterized by having high clinical complexity²⁰, suggesting an increased likelihood of death in the advent of any infection or acquired disease. Similar to our findings on clinical complexity profile and vulnerability, a cohort study of US nursing home residents with COVID-19 found that impaired cognitive and physical function were independently associated with mortality³. Interestingly, cluster 2, characterized by the higher health risk of its residents, had mortality rates similar to those of clusters 4, 5 and 6, with a more favorable resident health profile. These conflicting results suggest that mechanisms driving mortality risk in nursing homes are complex and may depend on the conjunction of various factors.

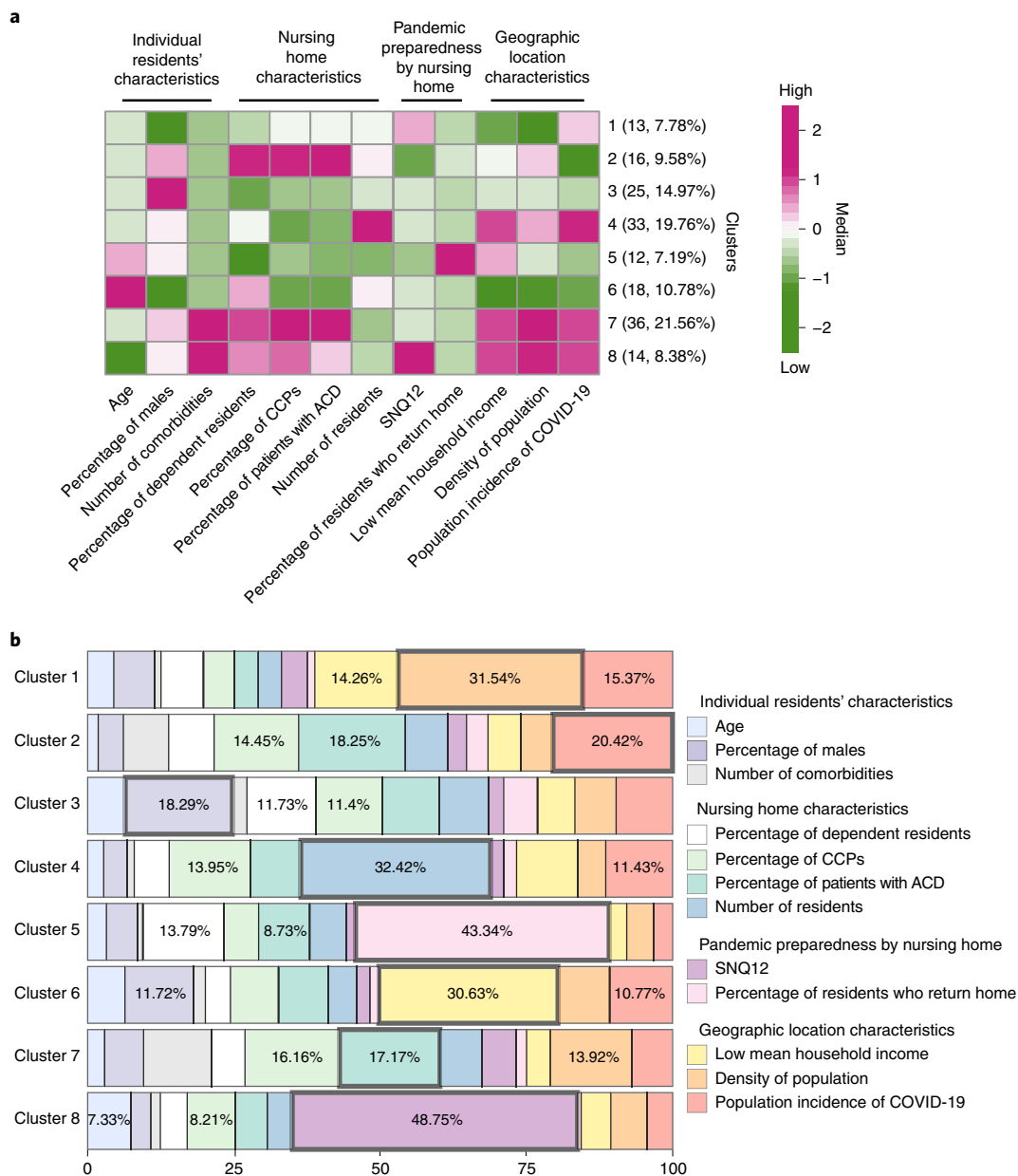


Fig. 1 | Heatmap of nursing home clusters. a, Heatmap of nursing home clusters. For each characteristic (x axis), the extent of the difference between the mean of a given cluster and the median of the entire sample is illustrated with the following color code: green tones indicate a mean of the cluster below the median of the entire sample, whereas purple tones indicate a mean of the cluster above the median of the entire sample. In both cases, more intense colors represent greater differences between the cluster and the whole sample. For each cluster, the number of nursing homes is indicated (n , percent of total). Extended Data Fig. 2 provides further details regarding the distribution of characteristics in each cluster. **b**, Contribution of variables to each cluster. Percentages are shown for the three variables with greater weight. Black squares highlight the most important variable.

Our analysis had intrinsic limitations of retrospective studies, particularly regarding data completeness. Owing to the overload of the healthcare system during the investigated period, a large number of patients who died could not be tested for SARS-CoV-2 by PCR and remained unconfirmed. Another set of information potentially relevant for our analysis that could not be collected was worker profiles in each nursing home. Unlike skilled nursing homes aimed at intermediate care or mental-health resources, which tend to be coordinated by healthcare authorities, nonspecialized nursing homes aimed at long-term stay are a case mix of organizational models. Hence, the inclusion of characteristics of the work team profile (for example, skills, resident/worker ratio and presence of physicians) might have provided interesting insights regarding the

capacity of the residence to cope with the outbreak¹⁸. In the absence of this information, we used the proportion of residents with complex or advanced diseases as a proxy to determine whether the nursing facility mostly provided intensive monitoring to residents with complex conditions or only provided assisted living to independent residents. Finally, it is noteworthy that, owing to the lack of validated scores for measuring preparedness to deal with COVID-19 outbreaks, we used an unvalidated measure of preparedness.

Our results raise important policy implications by suggesting structural factors of nursing homes and their surrounding districts that are important drivers of COVID-19-related mortality in this setting. Identification of facilities with low capacity for pandemic response located in areas with high incidence of COVID-19 could

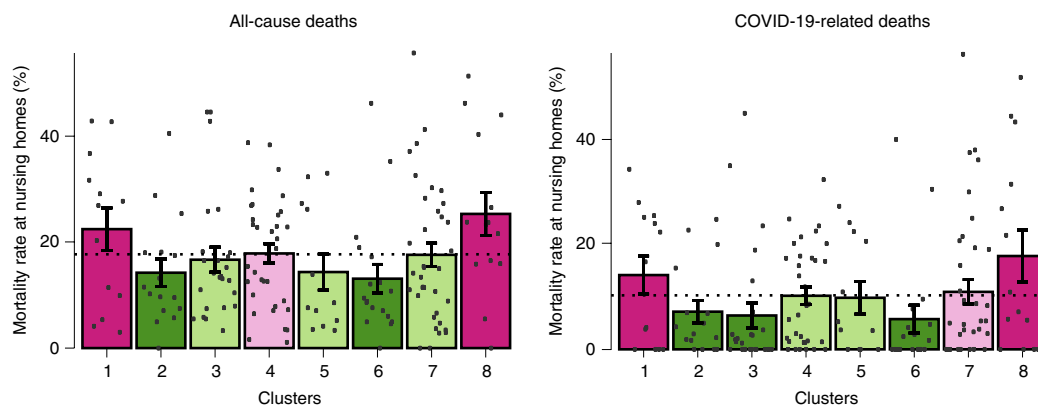


Fig. 2 | Mortality at the cluster level for all-cause deaths and COVID-19-related deaths. The dotted lines show the median mortality rate for the entire sample. Bars show the mean mortality rate of each cluster; error lines represent s.e.m. Pink and green indicate mean cluster mortality rates higher and lower than the total median, respectively. For both colors, light tones indicate that the s.d. of the cluster encompasses the overall median, whereas intense tones indicate that the whole s.d. range is above (pink) or below (green) the total median. Differences between groups were determined by one-way ANOVA. All-cause deaths, $F(7,167) = 2.83$, P value = 0.009; COVID-19-related deaths, $F(7,167) = 3.22$, P value = 0.003. Number of nursing homes per cluster (cluster (n)): 1 (13), 2 (16), 3 (25), 4 (33), 5 (12), 6 (18), 7 (36), 8 (14).

help public health officers identify facilities where preventative interventions need to be prioritized. The presence of CCPs, with high prevalence in the LTC setting^{21,22}, also increased mortality risk and should therefore be considered.

Methods

Study setting and participants. This was a retrospective cohort study of COVID-19 mortality risk factors in the residential setting in Catalonia (northeast Spain). The study included clinical, mortality and structural information corresponding to all public and private nursing homes in the administrative health region Metropolitana Nord (population of 1,986,032 people) in Barcelona, Spain, between 1 March 2020 and 1 June 2020, during the COVID-19 outbreak. Skilled nursing facilities (that is, intermediate care) and mental-health facilities were excluded from the analysis.

On 1 March 2020, the Department of Health of Catalonia launched a comprehensive disease-control program to minimize COVID-19 spread and mortality among residents in nursing homes. The containment strategy was implemented in all LTC facilities in the study area and involved 64 primary care teams that reported daily information regarding the epidemiological status of each nursing home. Primary care teams provided preventive epidemiological recommendations, including the partition of communal living areas, isolation of suspected cases and contacts and guidance on personal protective measures to nursing home workers. In the advent of a confirmed or suspected case of COVID-19, the teams also conducted systematic screening of close contacts (or all residents, in centers with high incidence) using real-time PCR with reverse transcription (RT-PCR) from nasopharyngeal swabs. No systematic screenings with PCR were conducted in nursing homes without suspected or confirmed COVID-19 cases.

The study protocol was approved by the institutional review board of the Hospital Universitari Germans Trias Pujol.

Data collection. Demographic and clinical data of residents were extracted from electronic medical records using a standardized data-collection form²⁰. Structural and organizational features of each nursing home were gathered at facility-assessment visits by the study team. The demographic and epidemiological profile of the district where the nursing home was located was retrieved from the Statistical Institute of Catalonia. Deaths were identified from the public mortality registry of the Department of Health of Catalonia in which all deaths at the country level are systematically registered, irrespective of the place of death. All data were handled according to the General Data Protection Regulation 2016/679 on data protection and privacy for all individuals within the European Union and the local regulatory framework regarding data protection. All data are available via GitHub.

Definitions. Variables regarding residents' health characteristics in each nursing home included demographic characteristics (that is, age and sex) and clinical characteristics (that is, number of comorbidities). Comorbidities were codified according to the ICD-10 system and included dementia, asthma or chronic obstructive pulmonary disease, hypertension, type 1 diabetes mellitus, type 2 diabetes mellitus, chronic kidney disease, cerebrovascular disease and cardiovascular disease.

We also recorded nursing home characteristics according to the total number of residents and type of care provided. The latter was assessed based on the proportion of residents classified by their primary care teams as residents with high dependence in activities of daily living (that is, defined as having a Barthel score < 50)²³, CCPs and patients with ACD, according to clinical guidelines of the Catalan Health Department²⁰. Catalan guidelines define CCPs based on their clinical condition (for example, multimorbidity, disability, difficult symptom control) and/or social environment (for example, lack of support from family or caregivers, isolated household). Patients with ACD are those with advanced and irreversible chronic conditions that limit their life expectancy to approximately 12 months.

Pandemic preparedness by nursing homes was characterized according to their capacity for implementing preventative measures within the facility on the basis of an ad hoc score (SNQ12 score) and organizational variables related to the percentage of residents who returned home to live with their relatives as a protective measure against COVID-19. The capacity of the nursing home for pandemic preparedness and response was assessed using an unvalidated ad hoc set of 12 essential items that yields a score called SNQ12 (sine qua non conditions for implementing the measures, Supplementary Method 1 (ref. 24)). The score indicates the number of unmet requirements, which ranges from 0 (all requirements are met) to 12 (all requirements are unmet). The requirements are related to three areas: (1) personal protective equipment, including a waterproof gown, face mask (that is, either filtering facepiece 2 or surgical), gloves and protective glasses (adequate supply, routine use and use for waste management and cleaning and disinfection), (2) surveillance and communication (routine monitoring of symptom onset by non-healthcare professionals and communication of symptoms to occupational health services) and (3) cleaning and waste management (regular hand washing before and after contact with patients with COVID-19 or their contacts, adequate laundry procedures, cleaning and disinfection of surfaces, use of an adequate disinfectant, adequate disposal of used personal protective equipment). The SNQ12 score was developed by the Catalan Ministry of Health and was distributed to medical directors of nursing homes along with the request of returning the result within the next 14 d. The number of residents returning home was recorded by LTCF with the COVIDApp²⁵. These data were also included in electronic health records.

The geographic location of the nursing home was assessed and characterized using the mean household income and density of the population in the municipality and the COVID-19 14-d incidence rate per 100,000 in the post code district (lowest administrative division) where the nursing home was located.

Excess mortality directly or indirectly due to COVID-19 was classified as COVID-19 related when individuals had a positive RT-PCR test or clinical suspicion of COVID-19. Clinical suspicion of COVID-19 was defined based on national guidelines available at the time as individuals with clinical features of acute respiratory disease of sudden onset and any severity, primarily characterized by fever, cough and shortness of breath. Other symptoms such as odynophagia, anosmia, dysgeusia, muscular pain, diarrhea, chest pain or headache could also be considered suggestive of SARS-CoV-2 infection at the physician's discretion.

Statistical analysis. Continuous and categorical variables were presented as mean and s.d. (or median and interquartile range, defined by 25th and 75th percentiles) and number (%), respectively. Excess deaths were defined as the difference between

deaths reported in 2020 and the median of deaths in 2016–2019 for the same months of the year; the COVID-19 contribution to excess deaths was computed by the difference between confirmed or suspected COVID-19 deaths and all-cause mortality. In our primary analysis to determine risk factors associated with mortality, we used univariate and multivariable zero-inflated Poisson models at the facility level. Independent variables for the multivariable model were chosen using an Akaike information criterion-based backward stepwise procedure. Continuous variables were fitted assuming a linear relationship. Results were presented as the IRR and the 95% CI. In a supplementary analysis, we fitted a model to individual-level data using a random-effects logistic regression with clustering at the nursing home level (Supplementary Table 2). In a secondary analysis, we grouped nursing homes according to their characteristics using the unsupervised clustering algorithm Phenograph^{26,27}. Data were centered and scaled before running the algorithm. To validate results, we used different values of the k parameter (number of nearest neighbors) and chose one based on the mean shortest distance and the stability in the number of clusters (Supplementary Method 2). The resulting clusters were described in a heatmap that represents the intensity of each characteristic based on the difference (below or above) between the average of the given cluster and that of the overall sample. To facilitate interpretation, we used a random forest classifier to determine the weight of each variable in each cluster based on the Gini measure²⁸. The mean mortality rate of each cluster was calculated, and differences between groups were determined by one-way ANOVA. The significance threshold was set at a two-sided α value of 0.05. All analyses and plots were performed using R version 3.6 (ref. ²⁹).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data that support the findings of this study are openly available on GitHub (<https://github.com/douve/paper-code>).

Code availability

The R code for data analysis is provided as part of the replication package available on GitHub (<https://github.com/douve/paper-code>).

Received: 26 October 2020; Accepted: 17 May 2021;

Published online: 28 June 2021

References

- Trabucchi, M. & De Leo, D. Nursing homes or besieged castles: COVID-19 in northern Italy. *Lancet Psychiatry* **7**, 387–388 (2020).
- Lau-Ng, R., Caruso, L. B. & Perls, T. T. COVID-19 deaths in long-term care facilities: a critical piece of the pandemic puzzle. *J. Am. Geriatr. Soc.* **68**, 1895–1898 (2020).
- Panagiotou, O. A. et al. Risk factors associated with all-cause 30-day mortality in nursing home residents with COVID-19. *JAMA Intern. Med.* **181**, 439–448 (2021).
- Brown, K. A. et al. Association between nursing home crowding and COVID-19 infection and mortality in Ontario, Canada. *JAMA Intern. Med.* **181**, 229–236 (2021).
- Li, Y., Temkin-Greener, H., Shan, G. & Cai, X. COVID-19 infections and deaths among Connecticut nursing home residents: facility correlates. *J. Am. Geriatr. Soc.* **68**, 1899–1906 (2020).
- World Health Organization (WHO). *Coronavirus Disease (COVID-19) Situation Report – 163* https://www.who.int/docs/default-source/coronavirus/situation-reports/20200701-covid-19-sitrep-163.pdf?sfvrsn=c202f05b_2 (WHO, 2020).
- Belmin, J. et al. Coronavirus disease 2019 outcomes in French nursing homes that implemented staff confinement with residents. *JAMA Netw. Open* **3**, e2017533 (2020).
- Abrams, H. R., Loomer, L., Gandhi, A. & Grabowski, D. C. Characteristics of U.S. nursing homes with COVID-19 cases. *J. Am. Geriatr. Soc.* **68**, 1653–1656 (2020).
- Chatterjee, P., Kelly, S., Qi, M. & Werner, R. M. Characteristics and quality of US nursing homes reporting cases of coronavirus disease 2019 (COVID-19). *JAMA Netw. Open* **3**, e2016930 (2020).
- Lipsitz, L. A. et al. Stemming the tide of COVID-19 infections in Massachusetts nursing homes. *J. Am. Geriatr. Soc.* **68**, 2447–2453 (2020).
- McMichael, T. M. et al. Epidemiology of COVID-19 in a long-term care facility in King County, Washington. *N. Engl. J. Med.* **382**, 2008–2011 (2020).
- White, E. M. et al. Variation in SARS-CoV-2 prevalence in U.S. skilled nursing facilities. *J. Am. Geriatr. Soc.* **68**, 2167–2173 (2020).
- Sugg, M. M. et al. Mapping community-level determinants of COVID-19 transmission in nursing homes: a multi-scale approach. *Sci. Total Environ.* **752**, 141946 (2020).
- Liu, K., Chen, Y., Lin, R. & Han, K. Clinical features of COVID-19 in elderly patients: a comparison with young and middle-aged patients. *J. Infect.* **80**, e14–e18 (2020).
- He, M., Li, Y. & Fang, F. Is there a link between nursing home reported quality and COVID-19 cases? Evidence from California skilled nursing facilities. *J. Am. Med. Dir. Assoc.* **21**, 905–908 (2020).
- Unruh, M. A., Yun, H., Zhang, Y., Braun, R. T. & Jung, H. Y. Nursing home characteristics associated with COVID-19 deaths in Connecticut, New Jersey, and New York. *J. Am. Med. Dir. Assoc.* **21**, 1001–1003 (2020).
- Brandén, M. et al. Residential context and COVID-19 mortality among adults aged 70 years and older in Stockholm: a population-based, observational study using individual-level data. *Lancet Healthy Longev.* **1**, e80–e88 (2020).
- Gorges, R. J. & Konetzka, R. T. Staffing levels and COVID-19 cases and outbreaks in US nursing homes. *J. Am. Geriatr. Soc.* **68**, 2462–2466 (2020).
- Centers for Disease Control and Prevention (CDC). *Discontinuation of Transmission-Based Precautions and Disposition of Patients with COVID-19 in Healthcare Settings (Interim Guidance)* <https://www.cdc.gov/coronavirus/2019-ncov/hcp/disposition-hospitalized-patients.html> (CDC, 2020).
- Contel, J. C. et al. Chronic and integrated care in Catalonia. *Int. J. Integr. Care* **15**, e025 (2015).
- Ambalàs-Novellas, J., Santaegüenia, S. J., Vela, E., Clèries, M. & Contel, J. C. What lies beneath: clinical and resource-use characteristics of institutionalized older people. A retrospective, population-based cohort study in Catalonia. *BMC Geriatr.* **20**, 187 (2020).
- Hollinghurst, J. et al. The impact of COVID-19 on adjusted mortality risk in care homes for older adults in Wales, UK: a retrospective population-based cohort study for mortality in 2016–2020. *Age Ageing* **50**, 25–31 (2021).
- Seematter-Bagnoud, L. et al. Predictors of functional recovery in patients admitted to geriatric postacute rehabilitation. *Arch. Phys. Med. Rehabil.* **94**, 2373–2380 (2013).
- Catalan Health Service. *Action Plan for the Management of People in the Residential Setting and Healthcare Delivery in the Context of the COVID-19 Pandemic* <https://scientiasalut.gencat.cat/handle/11351/4847> (Catalan Health Service, 2020).
- Echeverría, P. et al. COVIDApp as an innovative strategy for the management and follow-up of COVID-19 cases in long-term care facilities in Catalonia: implementation study. *JMIR Public Health Surveill.* **6**, e21163 (2020).
- Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
- Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).

Acknowledgements

We thank G. Carot-Sans for providing medical writing support during the preparation of the manuscript. Funding sources were the crowdfunding campaign YoMeCorono (<https://www.yomecorono.com/>) and the Generalitat de Catalunya. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

C.S., D.O., M.A.M., R.L.A. and O.M. designed the study. D.O., M.A.M., R.L.A., E.N., M.M.M., N.P., J.M.B.-S., M.E.I., I.G.S., S.R.N., R.M., M.T.C., J.V.P., N.H. and J.A. collected data. C.S., D.O., M.M., M.N.-J. and O.M. analyzed data. C.S., D.O., M.M., M.N.-J., M.A.M., M.M.M., R.M. and O.M. interpreted data. C.S., D.O. and O.M. wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43587-021-00079-7>.

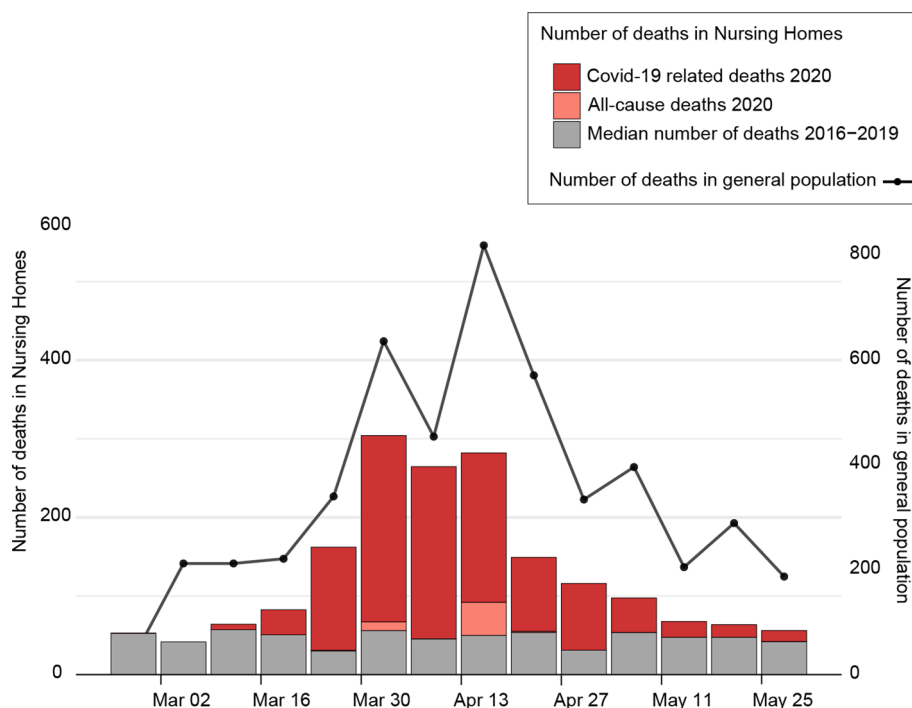
Correspondence and requests for materials should be addressed to O.M.

Peer review information *Nature Aging* thanks Thomas Perls and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

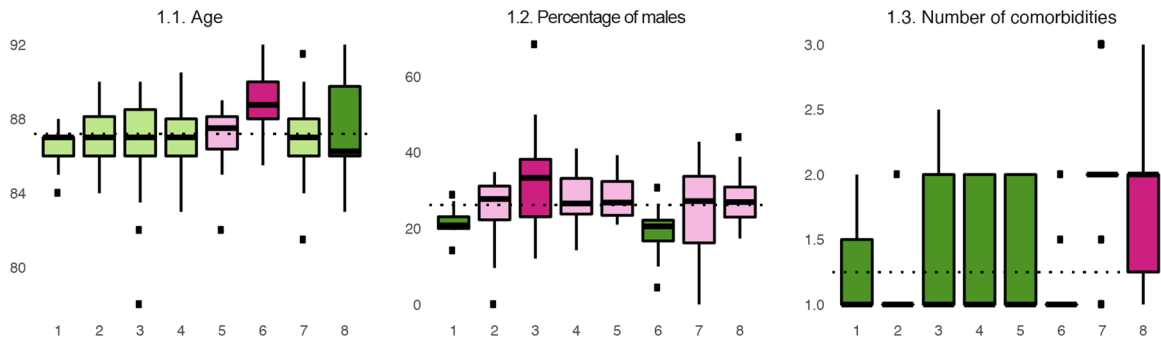
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature America, Inc. 2021

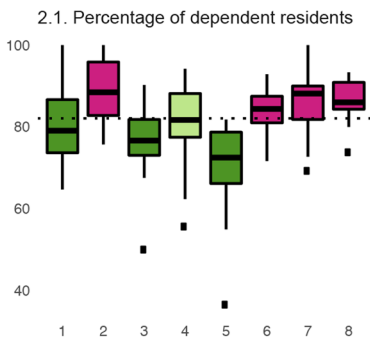


Extended Data Fig. 1 | Excess mortality of 2020 relative to the average of the past four years (2016–2019). Bars show the number of weekly deaths reported in 2020 in all nursing homes included in the analysis over the study period. Deaths reported in 2020 have been classified as Covid-19 confirmed and unconfirmed, which include deaths of individuals with suspected Covid-19 diagnosis. The median number of deaths for the same weeks in the previous 4 years (2016–2019) is shown in grey. The continuous line shows the death toll attributed to Covid-19 in the general population of the catchment area.

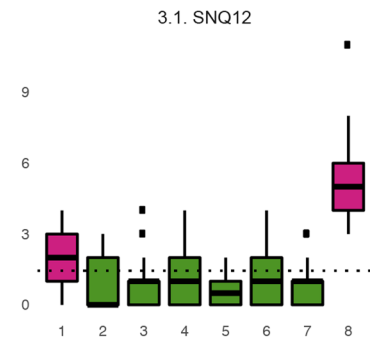
1. Individual residents' characteristics



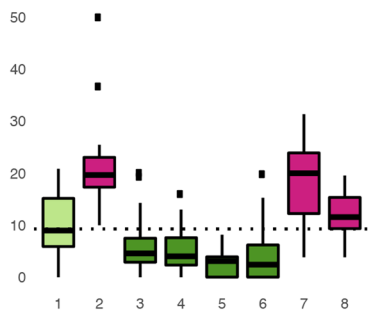
2. Nursing home characteristics



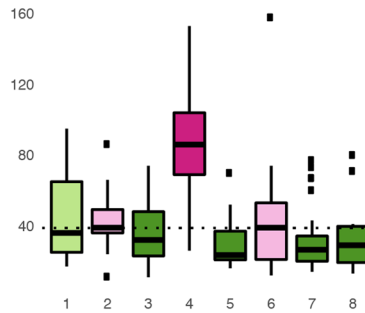
3. Pandemic preparedness by nursing home



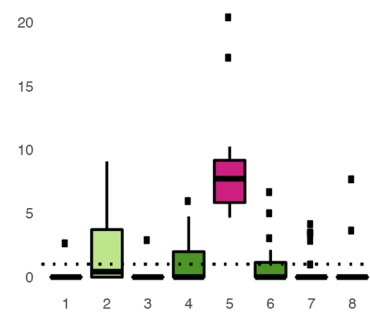
2.3. Percentage of ACD patients



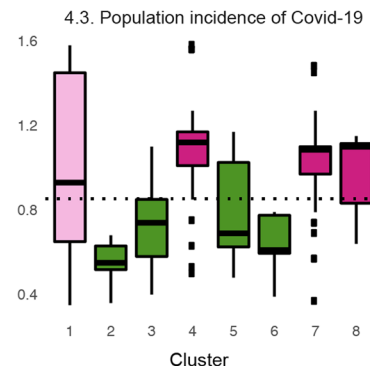
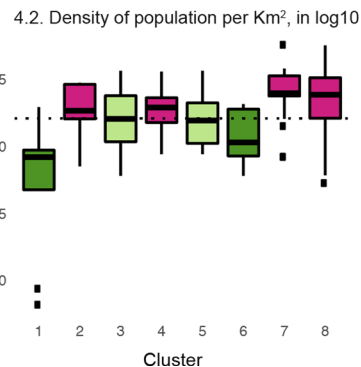
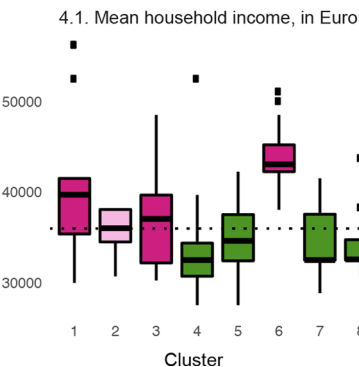
2.2. Percentage of CCPs



3.2. Percentage of residents who return home



4. Geographic location characteristics



Extended Data Fig. 2 | Characteristics among nursing home clusters. For each cluster (x-axis), box plots display the five-number summary of a given characteristic (minimum, first quartile, median, third quartile, and maximum). Black points represent outliers. The median of the entire sample is shown with a dashed line. The extent of the difference between the mean of a given cluster and the median of the entire sample is illustrated with the following colour code: green tones indicate a mean of the cluster below the median of the entire sample, whereas purple tones indicate a mean of the cluster above the median of the entire sample. In both cases, more intense colours represent greater differences between the cluster and the whole sample.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|--------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Retrospective qualitative cohort study.
Research sample	The study included clinical, mortality, and structural information corresponding to all public and private nursing homes, and their residents, in the administrative health region Metropolitana Nord (population 1,986,032 people) in Barcelona, Spain. Our analysis included 167 nursing homes providing long-term care to 8,716 residents. 2333 were males and the mean age of the study cohort was 85.7. This study sample was representative of elderly people living in nursing homes during the first months of the Covid-19 pandemic.
Sampling strategy	The sample size comprised of the entire population living in all public and private nursing homes in the administrative health region Metropolitana Nord in Barcelona, Spain. As such, no sample calculation was performed.
Data collection	Demographic and clinical data of residents were extracted from electronic medical records. Data was collected using the standardized data collection procedures from the Catalan health electronic registry. The structural and organizational features of each nursing home were gathered at facility assessment visits by the study team. The demographic and epidemiological profile of the nursing home district was retrieved from the Statistical Institute of Catalonia. Deaths were identified from the Mortality Registry of the Department of Health of Catalonia. Blinding was not relevant to our study since it is based on health registry data that is already collected and stored, and no allocation to experimental groups was performed.
Timing	Data was collected between March 1 and June 1, 2020, during the Covid-19 outbreak.
Data exclusions	The following exclusion criteria were pre-specified: skilled nursing facilities (i.e., intermediate care) and mental health facilities. The rationale behind this exclusion was to focus in older people that are long-term care residents of nursing homes.
Non-participation	As a retrospective study, all residents were included and non-participation is not applicable.
Randomization	This was an observational retrospective study and, therefore, no randomization was used.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See above
Recruitment	No recruitment was performed. The data is based on Catalan health registries and, as such, common biases of self-selection or loss to follow-up are avoided.
Ethics oversight	The study protocol was approved by the institutional review board of Hospital Germans Trias Pujol, which waived the

Ethics oversight

obtention of informed consent. All data were collected from administrative databases and therefore there was no need for patient recruitment.

Note that full information on the approval of the study protocol must also be provided in the manuscript.