BMJ Open

Variables associated to COVID-19 severity: an observational study of >13000 confirmed cases in the Basque Country, Spain

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-049066
Article Type:	Original research
Date Submitted by the Author:	14-Jan-2021
Complete List of Authors:	Vrotsou, Kalliopi; Biodonostia Instituto de Investigación Sanitaria, Grupo de AP Rotaeche, Rafael; Osakidetza-Basque Health Service, Centro de Salud de Alza Mateo-Abad, Maider; Kronikgune Machón, Mónica; Biodonostia Instituto de Investigación Sanitaria, Grupo de AP Vergara, Itziar; Biodonostia Instituto de Investigacion Sanitaria, Grupo de AP
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page:

Title:

Variables associated to COVID-19 severity: an observational study of >13000 confirmed cases in the Basque Country, Spain.

Authors:

Kalliopi Vrotsou (0000-0002-3296-3923)[1,2,3], Rafael Rotaeche[1,2,4],Maider Mateo-Abad(0000-0003-2541-8166)[2,3], Mónica Machón (0000-0001-6379-0351) [1,2,3], Itziar Vergara(0000-0001-9671-7898)[1,2,3]

Addresses:

1. Instituto de Investigación Sanitaria Biodonostia, Grupo de Atención Primaria, San Sebastián, Spain.
2. Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Baracaldo, Spain.
3. Instituto de Investigación en Servicios de Salud Kronikgune, Baracaldo, Spain.
4. Osakidetza, Centro de Salud de Alza, San Sebastián, Spain

Authors´s position:

Kalliopi Vrotsou: biostatistician.
Rafael Rotaeche: general practitioner.
Maider Mateo-Abad: biostatistician.
Monica Machón: investigator
Itziar Vergara: director of the Grupo de Atención Primaria, IIS Biodonostia

Corresponding author: Kalliopi Vrotsou,
Address: IIS Biodonostia, Paseo Dr. Beguiristain 20014, San Sebastián Spain,
Telephone number: +34 943006086
e-mail: KALLIOPI.VROTSOUKANARI@osakidetza.eus

Abstract

Objectives: To investigate which were the most relevant sociodemographic and clinical variables associated to COVID-19 severity, and uncover how their inter-relations may have affected such severity.

Design: A retrospective observational study based on electronic health record data.

Participants: Individuals ≥14 years old with a positive PCR or serology test, between the 28th of February and 31st of May, belonging to the Basque Country (Spain) public health system. Institutionalised and individuals admitted to a Hospital at Home unit were excluded from the study.

Main outcome measure: Three severity categories were established, primary care, hospital/ICU admission and exitus.

Results: A total of n=14197 cases fulfilled the inclusion criteria. Most variables presented statistically significant associations with the outcome ($p<0.0001$). The CART recursive partitioning methodology (based on n=13792) suggested that among all associations, those with, age, sex, stratification of patient health care complexity, chronic consumption of blood and blood forming organ, and nervous system drugs, as well as the total number of chronic ATC types were the most relevant. Psychosis also emerged as a potential factor.

Conclusions: Older cases are more likely to experience more severe outcomes. However, the sex, underlying health status and chronic drug consumption may interfere and alter the aging effect. Understanding the factors related to the outcome severity is of key importance when designing and promoting public health intervention plans for the COVID-19 pandemic.

Strengths and limitations of this study

- Over 13000 confirmed COVID-19, non institutionalised, ≥14 years old cases were explored
- Electronic health records data were a valuable source of information in this study
- The three-category outcome severity: primary care only, hospitalized/ICU care, and exitus was studied in a joint manner
- The CART methodology allowed exploring the big sample and the numerous variables of interest in a flexible way
- Asymptomatic cases were probably not included in this sample as during the first pandemic wave individuals with symptoms were mainly tested for the virus

Introduction

The COVID-19 disease, caused by the new coronavirus SARS-Cov-2 was initiated in December 2019 in China. On the 12th March 2020 the World Health Organization (WHO) declared it a pandemic. Its rapid expansion, along with the high death toll and the serious health aftermaths, have rendered the COVID-19 outbreak as one of the worst health crises in almost a century worldwide.

Since the first cases were detected in Spain the statistics have situated this country among the most affected in Europe, both in terms of total cases and in deaths per million people(1). International literature on COVID-19 is rapidly growing(2–7). The research conducted so far in Spain, has focused mainly on predicting the evolution of the pandemic(8), describing hospitalized individuals(9), or assessing the factors related to the risk of death(10,11). In its most recent publication, the Working group for the surveillance and control of COVID-19 in Spain, presented the factors affecting the outcomes of hospitalisation, as well as ICU admission and death for hospitalized individuals only(12). These outcomes were treated separately; each compared to cases not presenting the corresponding outcome. Therefore, studies that integrate information from different health care levels, analysing the data as a whole and applying statistical methods capable of considering the gradient of outcome severity are lacking.

The autonomous community of the Basque Country, situated in the North of Spain, has its own public health system (Osakidetza), which offers sanitary coverage to some 2.3 million people. Since 2009 Osakidetza has promoted an integrated health care model, by coordinating its different care levels and offering a more holistic approach on patient care(13). It counts with an extensive electronic health records infrastructure, where information on patient health data and episodes of care are stored. The objective of the present observational study was to describe a big series of COVID-19 infected individuals, during the first wave of the pandemic, establish their severity level, based on electronic health record data, and explore what characteristics may be associated to that severity. In particular, we were interested in understanding the structure and inter-relations

of different variables, identifying those that may have affected outcome severity the most. To this end, the Classification and Regression Trees (CART) methodology was applied. This statistical tool is suitable for exploring and uncovering complex relations, particularly useful when analysing big data sets(14).

Methods

Data source and variables

All information was extracted from the electronic health records of the Basque Country Public Health System-Osakidetza, via the Osakidetza Business Intelligence (OBI) tools. Data extraction covered the period between 28th February 2020 and 31st May 2020; corresponding to the first detected case in the Basque Country and the end of the first pandemic wave in Spain. Only individuals ≥14 years old with a COVID-19 positive Polymerase Chain Reaction (PCR) or antibody test, were included. No antigen tests were performed at that time. Cases living in residential homes and those admitted to a hospital at home unit were excluded.

The following variables were studied. Age, sex and income level derived by the pharmaceutical co-payment scheme (<18,000€, 18,000-100,000€, >100,000€). Chronic medication consumption was explored using the Anatomical Therapeutic Chemical (ATC) system at the first level. Polypharmacy, defined as the consumption of 5 or more chronic drugs, and the number of ATC types consumed were derived. Chronic pathologies based on ICD-9 codes, COVID-19 symptoms registered during consultations and flu vaccination in the year 2019. The Osakidetza stratification according to patient health care complexity was also studied. Based on a series of health data, and the use of health services during the previous year, this variable classifies individuals into four categories, ranging from less to more severe: prevention and promotion of healthy population, self-management support, disease management, and case management. Pluripathological individuals belong to the last category. This classification is renewed at the beginning of every

calendar year, for all individuals ≥14-years, registered in the Osakidetza system at least during the previous 6 months. A detailed description can be found elsewhere(15).

Given that the data were anonymous and clinical analyses could not be conducted, it was assumed that the severity of a case would be indicated by the most demanding level of medical attention received, within the study period. Four severity levels were initially identified: primary care attention only (PC), hospitalisation without intensive care unit admission (Hospital), intensive care unit admission (ICU), and Exitus. During the pandemic, several emergency ICU units were set-up within hospitals across the Basque Country. Nevertheless, this information was not reflected in the electronic health records. As a result, cases admitted to such ICUs appeared as hospital admissions. This imposed the necessity to merge Hospital and ICU admissions into one category in the current work. Cases meeting the inclusion criteria were included only once in the current analyses. The project has been approved by the ethics committee CEIm de Euskadi at 22/07/2020 (reference code: PI2020087).

Patient and public involvement

Due to the study design, no patient and public involvement was considered. Nonetheless, two of the authors are medical doctors, which has offered valuable support during this work.

Statistical analysis

Continuous variables are presented as means with standard deviations (SD), while medians and interquartile ranges (Q1, Q3) are given for discrete variables. Categorical variables are presented with frequencies and percentages (%). Comparisons were performed with the one-way ANOVA, Kruskal-Wallis and chi-square test, respectively. The Jonckheere-Terpstra and Mantel-Haenszel chi-square, both testing for a trend along the three severity groups were additionally tested(16).

CART

The CART is a non-parametric, recursive partitioning methodology. A CART tree starts with the root node, containing all the sample. At each step of the recursive process, every node may split

into offspring nodes. Nodes that split are called internal, and those that do not split any further are called terminal. The tree splits are based on the most important variables, among all candidate variables fed into the model. These splits intend to minimize the variability of the target variable within each offspring node, resulting in the most homogenous nodes, as far as the outcome variable is concerned(14). Splitting was based on the entropy criterion and each variable was allowed only once per tree branch. For a stopping rule, the number of terminal nodes, and the observations included in each of them were considered. A tree with 10 terminal nodes, each including at least 1% of the valid sample data was selected. Cost-complexity pruning was applied. Variables with significance levels p>0.010 in the three-group comparisons and those with a total frequency <1% of the valid sample were excluded from the CART stage. Missing data were omitted. Analyses were performed with the SAS software version 9.4 (Copyright (c) 2016 by SAS Institute Inc., Cary, NC, USA.). The PROC HPSPLIT function was used for tree construction.

Results

A total of n=14197 COVID-19 cases fulfilled the inclusion criteria. Of those n=9722 (68.5%) received PC only, n=3710 (26.1%) had a hospital or ICU admission (n=3630 and 80, respectively), and n=765 died (5.4%). Table 1 presents the baseline information of the sample. Overall, mean age was 53.7 (SD:17.4) years. Age increased with outcome severity. Most infected cases were females but more males were observed in the Hospital/ICU and Exitus groups. As far as the heath care complexity stratification variable was concerned the PC group presented the highest percentage of healthy individuals (36.1%), while case management was most prevalent in the Exitus group (36.6%). Based on the available information, individuals with an annual income <18.000 euros were more prevalent in the Hospital/ICU and Exitus groups, and those with higher income remained mostly in PC. Finally, the Exitus group had the highest percentage of individuals with a flu vaccination in the year 2019. This observation was consistent for cases <65 and ≥65

years of age, while percentages differed between age groups. All comparisons were statistically significant.

Chronic medication consumption data are presented in **Table 2**. Overall, the most consumed medications were those for the nervous system (38.7%), alimentary tract and metabolism (33.0%), and cardiovascular system (30.2%). With the exception of musculo-skeletal system and antiparasitic products, insecticides and repellents, the same trend was observed in all other ATC types. The consumption of alimentary tract and metabolism disorders (A), blood and blood forming organs (B), cardiovascular system (C), and nervous system diseases drugs (N) drugs exceeded 60% in the Exitus group. Both polypharmacy and the number of ATC types consumed was associated with infection severity.

As far as chronic diseases were concerned, the most prevalent condition was related to mental pathologies (**Table 3)**. In particular, 30% of the sample had received a diagnosis corresponding to the ICD-9 neurotic, personality or other nonpsychotic mental disorder. Hypertension was the next more prevalent condition (21%), followed by diseases of the blood and blood forming organs (11.5%), chronic obstructive pulmonary disease and allied conditions, as well as diseases of the esophagus, stomach and duodenum (both 10.4%). Diabetes mellitus was present in 8.5% of the sample. With the exception of neurotic, personality or other nonpsychotic conditions, that presented the same distribution along the three outcome groups, the prevalence of the most frequent pathologies increased with COVID-19 severity. A similar trend was seen in the total number of chronic diseases.

CART

The CART process indicated that age, sex, health care complexity stratification, blood and blood forming organ medication (B), as well as nervous system drugs (N) along with the frequency of ATC types consumed would be the most relevant variables in understanding the main case characteristics associated to the outcome. During this process the variable of psychoses was also flagged as important. In spite of its low prevalence (2.9%) psychoses was given a lot of weight in the older section of the population. The inclusion of this pathology resulted in a less parsimonious model; with ATC-N drugs placed in an additional tree level. Nonetheless, given that psychoses was the single variable resulting in a node with an exitus majority, and that other authors have already suggested an association between antipsychotic drugs and mortality in COVID-19 cases (10), presenting the corresponding findings was considered of relevance. Therefore, the CART process was repeated twice, first excluding and afterwards including psychoses.

Excluding psychoses

The tree generated by the CART process is depicted in **Figure 1A.** Most cases <64.7 years of age (81.4%, node 1) received mainly PC attention. In this tree branch males presented 15.3% more Hospital/ICU compared to females. Among males, those with worse health (node 8) had 19.2% more Hospital/ICU admissions, compared to the rest (node 7). The majority of males with worse baseline health status who consumed ≥3 ATC types experienced a Hospital/ICU admission.

Cases ≥64.7 years of age had mainly a Hospital/ICU outcome (52.6%), with a considerable Exitus prevalence (21.2%). Those with worse baseline health (node 6) had 4.6% more Hospital/ICU admissions and 15.8% more Exitus, compared to the rest. Cases with better baseline health status (node 5) were further split according to ATC-B consumption. Exitus for blood and blood forming organ drugs consumers was experienced in 23.2% of the cases, with the same outcome being 9.1% in non-consumers. Within this last group (node 9) the majority of females received PC

attention, while Hospital/ICU was the most prevalent outcome in males. In a similar way, among node 6 cases, males presented worse evolution than females. Finally, males consuming chronic medications for the nervous system (node 18) had 17.9% more Exitus compared to non-consumers. The 10 terminal nodes can easily be ordered less severe (i.e. <64.7 year old females, node 3) to most severe outcomes groups (i.e. ≥64.7 year old disease and case management males who consume nervous system drugs, node 10).

Including psychoses

The resulting CART model when the variable of psychoses was included in the recursive process is presented in **Figure 1B.** Psychoses, was one of the main variables of this model, and the single split variable for node 6. Inclusion of this pathology added one more level to the CART tree, with chronic nervous system drugs being a split variable for node 15. Cases with psychoses had a 50% Exitus. The ATC-N consumers presented less PC and higher Exitus compared to non-consumers. No other changes were observed compared to the Figure 1A model. In this case the most severe outcome group was was ≥64.7 year old disease and case management cases with psychosis (node 12).

Discussion

The present work has studied the sociodemographic and clinical characteristics of a big number of Spanish COVID-19 cases of the first pandemic wave. According to the information extracted from electronic health record data, the variables of age, sex, previous pathologies and chronic drug consumption may be decisive in understanding infection severity.

Both age and male sex have been flagged as important risk factors by previous COVID-19 research(2,3,7,10–12,17). The importance of age is probably undisputable, given the deterioration of the body´s immunity mechanisms and the loss of its capacity to adapt to the environment (18). The present data appears to reflect this known aging effect. In relation to the variable of sex, females presented consistently higher PC and lower Hospital/ICU in the splits where sex was present. With female Exitus being lower in two of them. Data from various European countries have highlighted that females have better COVID-19 infection outcomes than males(19). In spite of the fact that females are considered to have stronger immunity systems(20), the exact mechanisms responsible for these differences are unclear, and probably multifactorial(19). The current data, in conjunction with previous evidence call for a better understanding of the role of sex, in the current pandemic. Sex-specific analyses of future wave data should be planned. But more importantly, high quality prospective studies collecting sex-disaggregated data are needed(21).

The health care complexity stratification variable was present in both main tree arms. It should be mentioned that the way CART divided this 4-category variable into a binary one, by merging the two less severe vs. the two more severe groups was imposed by the data, not the investigators. Worse health status at the time of the infection, was associated to more hospitalizations for younger cases, and mainly to more deaths among older individuals. The inclusion of this stratification variable in the CART model is a relevant finding. Tools that stratify the general population, identifying those at greater risk, can be an asset for public health prevention programs. In the COVID-19 literature, the stratification approach has so far mainly focused on hospitalized patients(11,22,23). While one meta-analysis of in-hospital cases claimed that in COVID-19

infections underlying health conditions are even more important than age(24). Our data suggest that, at least at the local level, this very stratification variable can offer valuable information and its implementation may worth be considered when setting up public health action plans. Study of similar indicators used in other health systems would be encouraged.

As far as the drug consumption was concerned, chronic blood and blood forming organ drugs (B) and drugs for the nervous system (N), both appeared as important variables for cases ≥64.7 years of age. Cases consuming those drugs presented higher severity levels. ATC-N was the most frequent medication across all three outcome groups. ATC-B had the steepest raising in consumption from one severity level to the next. Several neurological manifestations after a COVID-19 infection have been described in the literature, with the virus perceived by certain authors as a threat for the whole nervous system(25). It is probable that individuals already suffering by chronic neurological conditions may be indeed more likely to present worse outcomes once infected(26,27). Blood related parameters like systolic and diastolic pressure, red and white cell counts, platelets, lymphocytes, among others, have been highlighted as significant predictors in different COVID-19 diagnostic models(7). An association between certain ATC-B drugs and higher odds of death in infected cases has also been observed(10). Chronic anticoagulation treatment, is referenced as protective against COVID-19 mortality by some(28), and ineffective by others(29). COVID-19 cases present a high frequency of thrombotic events, which is leading to an expansion of anticoagulation drug use when treating the disease (30). But in patients already receiving such drugs prior to infection, drug-drug interactions and infection severity should be carefully assessed before any antiviral therapy is given, or switching from oral to parenteral antithrombotic administration(31). Worse severity seen among ATC-B consumers in the current data may reflect also an increased risk for patients already under anticoagulation therapy. Poor outcomes due to therapeutic decisions and drug-drug interactions cannot be excluded either. Our continuing COVID-19 work will refine future data explorations. Obtaining for example ATC data at the second or third level, as well as information of in-patient treatments will offer more insight into these associations.

Psychoses was a relevant variable in the CART process. Antipsychotic drugs belong to the ATC-N; which is probably why allowing for the inclusion of psychoses relocated this drug type further down in the tree structure. Older patients with worse baseline health and psychoses had the highest death rate among all CART nodes. We can only hypothesize over the mechanisms that could explain such a finding. On one hand, individuals with psychotic disorders present excess mortality compared to the general population, mainly due to lifestyle choices, associated comorbidities and medication side effects(32). On the other hand, the challenging symptoms recognition and treatment management of these cases can potentially lead to a sudden health deterioration or even death(33). This could happen for example during hospital or ICU admissions. In the present sample 75% of the deaths seen in the psychoses node had been admitted to a hospital during the study period. The available information does not allow knowing whether exitus took place during the admissions, neither the in-patient treatment regime. An observational USA study of >60000 cases claimed that psychiatric disorders are a risk factor associated to higher COVID-19 diagnosis; with psychosis presenting greater risk ratios versus mood and anxiety disorders. The same study also reported an increased risk of first-time psychiatric disorders for survivors(34). Others have suggested that antipsychotics are associated to higher death rates in COVID-19 cases(10). More research in this direction is required.

The total number of chronically consumed ATC types was an important variable among cases <64.7 years of age. This variable, which could also be perceived as an indicator of the associated comorbidities, stresses even more the importance that underlying pathologies may have in determining the severity of the infection outcome(24).

In this work a surrogate outcome variable has been used. Assuming that more intensive care levels represented worse COVID-19 status is a decision also taken by previous authors(12,35–37). The available data does allow studying if admissions and deaths may have been due to other health problems.

The current study has certain limitations. The implemented information is based exclusively on electronic health record data within the previously defined dates. After that period the severity of

certain cases may have worsen. Nonetheless, the end study date corresponds to the end of the first COVID-19 wave in our area, where new infections and deaths were very low. This, in combination with the big study sample should have minimized the effect of possible outcome variations. No COVID-19 symptoms are presented. An attempt to register these symptoms was incorporated at the Osakidetza electronic records, early on after the outbreak. But, the number of symptoms and registration format evolved over the studied period; PC and hospital registrations differed; the medical staff mostly annotated symptoms in text format; while most importantly such registration was totally missing in many cases. During analysis an effort to re-code text annotations, and homogenize information from primary care and hospital data was made. In spite of that, and due to the frequency of missing values, the representativeness of the corresponding data could not be assumed. Symptoms are probably more relevant for algorithms discriminating cases from non-cases(38). Also, most likely the present data does not include information of asymptomatic cases. During the first pandemic wave no massive testings were performed in Spain. Thus, identified cases were either symptomatic, or close contacts of infected individuals. Working with health records makes recovering missing data or refining variable information a very difficult task. This was the case with the income level. Its broad categories may have obscured a more appropriate exploration. On the other hand, the high frequency of missing income level data seen in the Exitus group, is due to the "un-subscriptions" of the dead cases from the medication dispensing registry. It is important to note that the target of the Basque public health system is a health coverage based on the health needs and not the earnings of the individuals.

One of the main strength of this study is its big sample size. The consideration of three outcome groups is another advantage, which allows for a better visualization of the different severity levels of the disease. Finally, implementing the CART methodology assisted in translating a complex and multifactorial reality into an easy to follow picture. Our findings make clinical sense and are supported by previous evidence. They appear to endorse the need for public health prevention plans that consider population characteristics. At the same time, they highlight that for a

multifactorial problem to be properly treated, not only the factors affecting it, but also the inter-relations between the latter should be thoroughly studied. The COVID-19 pandemic may be a new starting point in the public health paradigm. The necessity for public health promoters to work hand-in-hand with investigators and data analysts has become indisputable, under the current circumstances. Prevention plans should be based on rigorous data and understanding of the latter. This is the only way to assure that possible re-organization and estimation of future resources can reach optimal results.

Dissemination declaration: Upon acceptance, the results will be disseminated to patient organizations, medical students or/and other interested groups or means of communication.

Data availability: The data of the current study are stored in a server of our institution. Sharing them with external investigators will be evaluated on an individual basis and will require an approval by the Osakidetza central services. The corresponding author should be contacted.

Funding source: No funding was obtained for this study.

Authors´Contributions: IV, RR and MM planned this study and obtained the permission for exploring the corresponding data by the Osakidetza central services. MMA set the filters and performed the data extraction of the electronic health record data. KV and MMA are both responsible for data cleaning and recoding. KV and MMA performed all statistical analyses. The input of IV and RR have assured a clinically meaningful perspective of all presented analyses and results. MM performed literature searches. KV drafted the first manuscript version. All authors read and contributed to the consecutive manuscript versions.

Competing interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work."

References

1. Coronavirus Update (Live): 35,732,698 Cases and 1,046,514 Deaths from COVID-19 Virus Pandemic - Worldometer [Internet]. [cited 2020 Oct 6]. Available from: https://www.worldometers.info/coronavirus/

2. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. Available from: https://isaric4c.net

3. Nachtigall I, Lenga P, Jóźwiak K, Thürmann P, Meier-Hellmann A, Kuhlen R, et al. Clinical course and factors associated with outcomes among 1904 patients hospitalized with COVID-19 in Germany: an observational study. Clin Microbiol Infect [Internet]. 2020 [cited 2020 Nov 10]; Available from: https://reader.elsevier.com/reader/sd/pii/S1198743X20304936?token=088D506A6F15F95832BA1FF82EBE9040EC1E3154088A738F47EBD33697538DEFF79877982A8A7F3857E2695415D62354

4. Hewitt J, Carter B, Vilches-Moraga A, Quinn TJ, Braude P, Verduri A, et al. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study. Lancet Public Heal [Internet]. 2020 Aug 1 [cited 2020 Nov 10];5(8):e444–51. Available from: www.thelancet.com/

5. Baqui P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. Lancet Glob Heal [Internet]. 2020 Aug 1 [cited 2020 Nov 10];8(8):e1018–26. Available from: www.thelancet.com/lancetghVol

6. Jones RC, Ho JC, Kearney H, Glibbery M, Levin DL, Kim J, et al. Evaluating Trends in

COVID-19 Research Activity in Early 2020: The Creation and Utilization of a Novel Open-Access Database. Cureus [Internet]. 2020 Aug 22 [cited 2020 Nov 10];12(8). Available from: https://pubmed.ncbi.nlm.nih.gov/32968602/

7. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ. 2020;369.

8. Aguiar M, Ortuondo EM, Bidaurrazaga Van-Dierdonck J, Mar J, Stollenwerk N. Modelling COVID 19 in the Basque Country from introduction to control measure response. Sci Rep [Internet]. 2020 Dec 14 [cited 2020 Oct 16];10(1):17306. Available from: http://www.nature.com/articles/s41598-020-74386-1

9. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. Rev Clin Esp. 2020;

10. Poblador-Plou B, Carmona-Pírez J, Ioakeim-Skoufa I, Poncel-Falcó A, Bliek-Bueno K, Cano-Del Pozo M, et al. Baseline chronic comorbidity and mortality in laboratory-confirmed COVID-19 cases: Results from the PRECOVID study in Spain. Int J Environ Res Public Health. 2020;17(14):1–14.

11. Rubio-Rivas M, Corbella X, Mora-Luján JM, Loureiro-Amigo J, López Sampalo A, Yera Bergua C, et al. Predicting Clinical Outcome with Phenotypic Clusters in COVID-19 Pneumonia: An Analysis of 12,066 Hospitalized Patients from the Spanish Registry SEMI-COVID-19. J Clin Med. 2020;9(11):3488.

12. Working group for the surveillance and control of COVID-19 in Spain. The first wave of the COVID-19 pandemic in Spain : characterisation of cases and risk factors for severe outcomes , as at 27 April 2020. Eurosurveillance [Internet]. 2020;25(50):1–13. Available

from: http://dx.doi.org/10.2807/1560-7917.ES.2020.25.50.2001431

13. Bengoa R. Transforming health care: an approach to system-wide implementation. Int J Integr Care. 2013;13(3).

14. Zhang H, Singer BH. Recursive Partitioning and Applications. Second edi. New York: Springer Series in Statistics; 2010.

15. Orueta JF, Nuño-Solinis R, Mateos M, Vergara I, Grandes G, Esnaola S. Predictive risk modelling in the Spanish population: A cross-sectional study. BMC Health Serv Res [Internet]. 2013;13(1):1. Available from: BMC Health Services Research

16. Walker, Glenn A, 2002. Common Statistical Methods for Clinical Research with SAS Examples, Second Edition. SAS Institute. Cary, NC;

17. Wollenstein-Betech S, Cassandras CG, Paschalidis IC. Personalized Predictive Models for Symptomatic COVID-19 Patients Using Basic Preconditions: Hospitalizations, Mortality, and the Need for an ICU or Ventilator. Int J Med Inform. 2020 Aug;104258.

18. WHO Team: Social determinants of Health. World report on ageing and health [Internet]. World Health Organization. Luxembourg; 2015 [cited 2020 Dec 9]. Available from: https://www.who.int/publications/i/item/world-report-on-ageing-and-health

19. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL. Impact of sex and gender on COVID-19 outcomes in Europe. Vol. 11, Biology of Sex Differences. BioMed Central Ltd.; 2020.

20. Rider V, Martocchia A, Cava A La, Moulton VR. Sex Hormones in Acquired Immunity and Autoimmune Disease. Front Immunol | www.frontiersin.org [Internet]. 2018;9:2279. Available from: www.frontiersin.org

21. Dehingia N, Raj A. Sex differences in COVID-19 case fatality: do we know enough?

Lancet Glob Heal. 2020 Nov;

22. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ [Internet]. 2020;370:22. Available from: http://dx.doi.org/10.1136/bmj.m3339

23. Mei Y, Weinberg SE, Zhao L, Frink A, Qi C, Behdad A, et al. Risk stratification of hospitalized COVID-19 patients through comparative studies of laboratory results with influenza. EClinicalMedicine [Internet]. 2020;26:100475. Available from: https://doi.org/10.1016/j.eclinm.2020.100475

24. Starke KR, Petereit-Haack G, Schubert M, Kämpf D, Schliebner A, Hegewald J, et al. The Age-Related Risk of Severe Outcomes Due to COVID-19 Infection: A Rapid Review, Meta-Analysis, and Meta-Regression. Available from: www.mdpi.com/journal/ijerph

25. Koralnik IJ, Tyler KL. COVID-19: A Global Threat to the Nervous System. Ann Neurol. 2020;88(1):1–11.

26. Herman C, Mayer K, Sarwal A, Org N, Covid19 /. Scoping review of prevalence of neurologic comorbidities in patients hospitalized for COVID-19. 2020;

27. Romagnolo A, Balestrino R, Imbalzano G, Ciccone G, Riccardini · Franco, Carlo ·, et al. Neurological comorbidity and severity of COVID-19. J Neurol [Internet]. 123AD;1:3. Available from: https://doi.org/10.1007/s00415-020-10123-y

28. Rossi R, Coppi F, Talarico M, Boriani G. Protective role of chronic treatment with direct oral anticoagulants in elderly patients affected by interstitial pneumonia in COVID-19 era. Eur J Intern Med. 2020;77(May):158–60.

29. Schiavone M, Gasperetti A, Mancone M, Curnis A, Mascioli G, Mitacchione G, et al. Oral anticoagulation and clinical outcomes in COVID-19: An Italian multicenter experience.

Int J Cardiol. 2020;323:276–80.

30. Hanff TC, Mohareb AM, Giri J, Cohen JB, Chirinos JA. Thrombosis in COVID-19. Am J Hematol. 2020;95(12):1578–89.

31. Vivas D, Roldán V, Esteve-Pastor MA, Roldán I, Tello-Montoliu A, Ruiz-Nodar JM, et al. Recommendations on antithrombotic treatment during the COVID-19 pandemic. Position statement of the Working Group on Cardiovascular Thrombosis of the Spanish Society of Cardiology. Rev Esp Cardiol. 2020;73(9):749–57.

32. Oakley P, Kisely S, Baxter A, Harris M, Desoe J, Dziouba A, et al. Increased mortality among people with schizophrenia and other non-affective psychotic disorders in the community: A systematic review and meta-analysis. J Psychiatr Res [Internet]. 2018;102(April):245–53. Available from: https://doi.org/10.1016/j.jpsychires.2018.04.019

33. Simon L, Hashmi M, Callahan A. Neuroleptic Malignant Syndrome [Internet]. [Updated 2020 Nov 20]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-; Available from: https://www.ncbi.nlm.nih.gov/books/NBK482282/

34. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. The Lancet Psychiatry [Internet]. 2020 [cited 2020 Dec 10]; Available from: https://doi.org/10.1016/S2215-0366

35. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Heal. 2020 Aug 1;8(8):e1003–17.

36. Ghahramani S, Tabrizi R, Lankarani KB, Kashani SMA, Rezaei S, Zeidi N, et al. Laboratory

features of severe vs. non-severe COVID-19 patients in Asian populations: A systematic review and meta-analysis. Vol. 25, European Journal of Medical Research. BioMed Central; 2020.

37. Ou M, Zhu J, Ji P, Li H, Zhong Z, Li B, et al. Risk factors of severe cases with COVID-19: a meta-analysis. Epidemiol Infect. 2020 Aug 12;148:e175.

38. Zimmerman R, Nowalk MP, Bear T, Taber R, Sax T, Eng H, et al. Proposed Clinical Indicators for Efficient Screening and Testing for COVID-19 Infection from Classification and Regression Trees (CART) Analysis [Internet]. medRxiv : the preprint server for health sciences. 2020. Available from: https://doi.org/10.1101/2020.05.11.20097980

Table 1: Baseline information of the COVID-19 cases during the first wave of the pandemic

Variables	*Total (n=14197)*	Primary Care (n=9722)	Hospital/ICU (n=3710)	Exitus (n=765)	p-value
Age; mean (SD)	*53.7 (17.4)*	48.0 (14.4)	62.8 (16.1)	82.3 (10.5)	<0.0001
Sex					
Male	*5520 (38.9)*	3073 (31.6)	2031 (54.7)	416 (54.4)	<0.0001
Female	*8677 (61.1)*	6649 (68.4)	1679 (45.3)	349 (45.6)	
Health care complexity					
Missing information	*405 (2.9)*	*307 (3.2)*	*86 (2.3)*	*12 (1.6)*	
Prevention and promotion	*3878 (27.3)*	3399 (36.1)	470 (12.9)	9 (1.2)	<0.0001
Self-management support	*6821 (48.0)*	4989 (52.9)	1675 (46.2)	157 (20.8)	
Disease management	*2252 (15.9)*	891 (9.4)	1050 (28.9)	311 (41.3)	
Case management	*841 (5.9)*	136 (1.4)	429 (11.8)	276 (36.6)	
Income level					
Missing information	*854 (6.0)*	*251 (2.6)*	*130 (3.5)*	*473 (61.8)*	
<18.000 euros	*6536 (46.0)*	4297 (45.3)	2038 (56.9)	201 (68.8)	<0.0001
18.000-100.000 euros	*6670 (47.0)*	5074 (53.5)	1507 (42.0)	89 (30.4)	
>100.000 euros	*137 (1.0)*	100 (1.0)	35 (0.9)	2 (0.6)	
Flu vaccination in 2019: yes					
All vaccinated cases	*3336 (23.5)*	1322 (13.6)	1446 (39.0)	568 (74.2)	<0.0001
Vaccinated cases <65 years old	*1103 (10.1)*	814 (9.2)	265 (13.6)	24 (42.8)	<0.0001
Vaccinated cases ≥ 65 years old	*2233 (66.5)*	508 (57.7)	1181 (66.9)	544 (76.7)	<0.0001

Data are frequency (percentage), unless otherwise stated. For variables with missing information, percentages and statistical comparisons are based on valid data only. Presented p-values are based on one-way ANOVA for the variable of age and the chi-square test for the categorical variables. Cases <65 year and ≥ 65 years were n=10843 and n=3354, respectively. Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend also resulted in p<0.0001 in all comparisons.

Table 2: *Chronic medication consumption of the COVID-19 sample.*

	Total (n=14197)	**Primary Care (n=9722)**	**Hospital/ICU (n=3710)**	**Exitus (n=765)**	**p-value**
Medication (ATC type)					
Alimentary tract and metabolism (A)	*4685 (33.0)*	2234 (23.0)	1837 (49.5)	614 (80.3)	<0.0001
Blood and blood forming organs (B)	*2414 (17.0)*	889 (9.1)	1057 (28.5)	468 (61.2)	<0.0001
Cardiovascular system (C)	*4294 (30.2)*	1813 (18.6)	1893 (51.0)	588 (76.9)	<0.0001
Dermatologicals (D)	*1765 (12.4)*	1032 (10.6)	581 (15.7)	152 (19.9)	<0.0001
Genitourinary system and sex hormones (G)	*1690 (11.9)*	1050 (10.8)	505 (13.6)	135 (17.6)	<0.0001
Systemic Hormonal preparations, excluding sex hormones and insulins (H)	*1504 (10.6)*	876 (9.0)	492 (13.3)	136 (17.8)	<0.0001
Antiinfectives for systemic use (J)	*223 (1.6)*	122 (1.3)	73 (2.0)	28 (3.7)	<0.0001
Antineoplastic and immunomodulating agents (L)	*360 (2.5)*	165 (1.7)	141 (3.8)	54 (7.1)	<0.0001
Musculo-Skeletal system (M)	*3137 (22.1)*	2010 (20.7)	952 (25.7)	175 (22.9)	<0.0001
Nervous System (N)	*5494 (38.7)*	2906 (29.9)	1931 (52.0)	657 (85.9)	<0.0001
Antiparasitic products, insecticides and repellents (P)	*42 (0.3)*	24 (0.2)	15 (0.4)	3 (0.4)	0.284
Respiratory System (R)	*2603 (18.3)*	1517 (15.6)	864 (23.3)	222 (29.0)	<0.0001
Sensory Organs (S)	*863 (6.1)*	443 (4.6)	297 (8.0)	123 (16.1)	<0.0001
Various (V)	*188 (1.3)*	45 (0.5)	58 (1.6)	85 (11.1)	<0.0001
Polypharmacy: yes	*2921 (20.5)*	935 (9.6)	1357 (36.5)	629 (82.2)	<0.0001
Num. ATC types consumed: median (Q1,Q3)	*2 (0, 3)*	1 (0, 3)	3 (1, 4)	5 (3, 6)	<0.0001

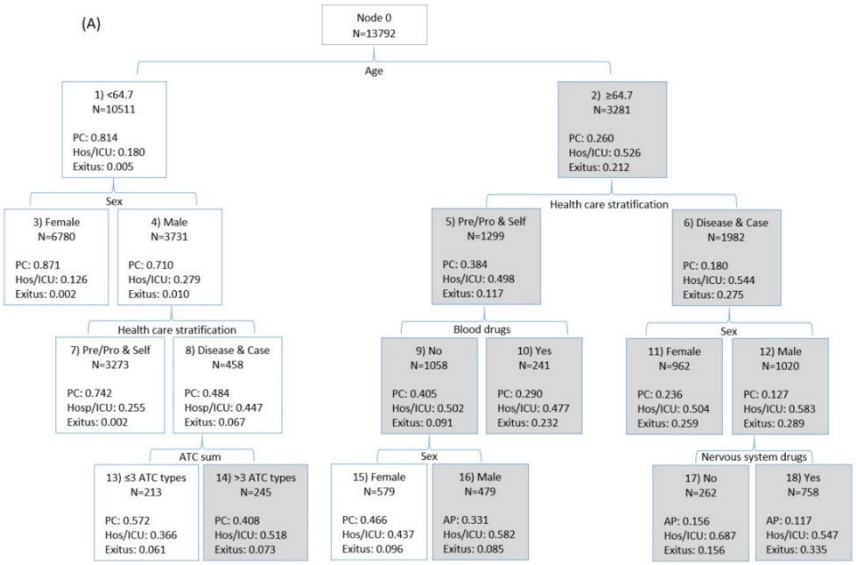
Data are frequency (percentage), unless otherwise stated. Q1, Q3: interquartile range values. ATC: Anatomical Therapeutic Chemicals. Presented p-values are based on the Chi-square test. The Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend resulted in very similar results in all comparisons.

Table 3: *Chronic diseases of the COVID-19 cases in the three outcome groups*

Disease	Total (n=14197)	Primary Care (n=9722)	Hospital/ICU (n=3710)	Exitus (n=765)	p-value
Infectious disease					
HIV infection	23 (0.2)	7 (0.1)	12 (0.3)	4 (0.5)	0.0002
Liver disease and cirrhosis	133 (0.9)	49 (0.5)	72 (1.9)	12 (1.6)	<0.0001
Malignant neoplasm	918 (6.4)	364 (3.7)	410 (11.0)	144 (18.8)	<0.0001
Endocrine diseases					
Subclinical hypothyroidism without treatment	1101 (7.8)	747 (7.7)	294 (7.9)	60 (7.8)	0.892
Diabetes Mellitus	1213 (8.5)	395 (4.1)	606 (16.3)	212 (27.7)	<0.0001
Diseases of the blood and blood-forming organs	1602 (11.3)	930 (9.6)	492 (13.3)	180 (23.5)	<0.0001
Mental disorders					
Psychoses	412 (2.9)	138 (1.4)	143 (3.9)	131 (17.1)	<0.0001
Neurotic disorders, personality disorders, and other nonpsychotic mental disorders	4258 (30.0)	2926 (30.1)	1096 (29.5)	236 (30.8)	0.712
Mental retardation	39 (0.3)	24 (0.2)	14 (0.4)	1 (0.1)	0.319
Nervous system diseases					
Hereditary and degenerative diseases of the central nervous system	419 (3.0)	145 (1.5)	149 (4.0)	125 (16.3)	<0.0001
Diseases of the circulatory system					
Hypertensive disease	2988 (21.0)	1177 (12.1)	1364 (36.8)	447 (58.4)	<0.0001
Ischemic heart disease	448 (3.2)	111 (1.1)	237 (6.4)	100 (13.1)	<0.0001
Cerebrovascular disease	611 (4.3)	189 (1.9)	266 (7.2)	156 (20.4)	<0.0001
Heart failure & Atrial fibrillation and flutter	709 (5.0)	132 (1.4)	361 (9.7)	216 (28.2)	<0.0001
Acute pulmonary heart disease & other venous embolism and thrombosis	150 (1.1)	49 (0.5)	73 (2.0)	28 (3.7)	<0.0001
Arterial embolism and thrombosis	39 (0.3)	17 (0.2)	17 (0.5)	5 (0.7)	0.002
Respiratory disease					
Chronic obstructive pulmonary disease and allied conditions	1483 (10.4)	844 (8.7)	506 (13.6)	133 (17.4)	<0.0001
Pneumonoconioses and other lung diseases due to external agents	20 (0.1)	9 (0.1)	8 (0.2)	3 (0.4)	0.038
Diseases of the Digestive system					
Diseases of esophagus, stomach and duodenum	1481 (10.4)	907 (9.3)	468 (12.6)	106 (13.9)	<0.0001
Non-infectious enteritis and colitis	643 (4.5)	500 (5.1)	121 (3.3)	22 (2.9)	<0.0001
Regional enteritis & Ulcerative Colitis	73 (0.5)	51 (0.5)	16 (0.4)	6 (0.8)	0.447
Disease of the genitourinary system					
Chronic kidney disease	398 (2.8)	87 (0.9)	188 (5.1)	123 (16.1)	<0.0001
Diseases of the skin and subcutaneous tissue					
Psoriasis	315 (2.2)	180 (1.9)	113 (3.0)	22 (2.9)	<0.0001
Diseases of the musculoskeletal system and connective tissue					
Systemic lupus erythematosus	36 (0.3)	24 (0.2)	10 (0.3)	2 (0.3)	0.972
Rheumatoid arthritis and other inflammatory polyarthropathies	125 (0.9)	59 (0.6)	55 (1.5)	11 (1.4)	<0.0001
Arthropathy associated with other disorders classified elsewhere	8 (0.1)	5 (0.1)	1 (0.0)	2 (0.3)	0.042
Multimorbidity: ≥2 chronic diseases	5326 (37.5)	2715 (27.9)	1975 (53.2)	636 (83.1)	<0.0001
Número total de enfermedades Crónicas					
Median (Q1, Q3)	1 (0, 2)	1 (0, 2)	2 (1, 3)	3 (2, 4)	<0.0001

Data are frequency (percentage) unless otherwise stated; Q1, Q3: interquartile range values. Presented p-values are based on the Chi-square test. The Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend resulted in very similar results in all comparisons.

Figure 1: CART model results for the COVID-19 outcome severity.



PC: primary care; Hos/ICU: hospital and intensive care unit. Percentages up to 3 decimal places are given. Due to rounding, node percentages may not add to 100%. Pre/Pro & Self: Prevention and promotion, and Self-management support. Disease & Case: Disease management and Case management. ATC: Anatomical Therapeutic Chemical. White nodes represent groups with a higher percentage of PC care. Grey nodes represent groups with a higher percentage of Hospital/ICU admission, and grey stripes nodes groups with a higher percentage of Exitus. Models excluding (A) and including psychosis (B) are presented.

STROBE Statement—checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Recommendation	page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3,4
Methods			
Study design	4	Present key elements of study design early in the paper	4,5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4,5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4,5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4,5*
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5, 6
		(*b*) Describe any methods used to examine subgroups and interactions	5, 6
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	5, 6
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			**page**
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	6
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table1
		(b) Indicate number of participants with missing data for each variable of interest	Table1
Outcome data	15*	Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6, 8, 9, figure1
		(*b*) Report category boundaries when continuous variables were categorized	6, figure1
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8, 9
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12, 13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Variables associated to COVID-19 severity: an observational study of non-paediatric confirmed cases from the general population of the Basque Country, Spain

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-049066.R1
Article Type:	Original research
Date Submitted by the Author:	04-Mar-2021
Complete List of Authors:	Vrotsou, Kalliopi; Biodonostia Instituto de Investigación Sanitaria, Grupo de Atención Primaria; Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC); Instituto de Investigación en Servicios de Salud Kronikgune Rotaeche, Rafael; Osakidetza-Basque Health Service, Centro de Salud de Alza Mateo-Abad, Maider; Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC); ; Instituto de Investigación en Servicios de Salud Kronikgune Machón, Mónica; Biodonostia Instituto de Investigación Sanitaria, Grupo de Atención Primaria; Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC); Instituto de Investigación en Servicios de Salud Kronikgune Vergara, Itziar; Biodonostia Instituto de Investigacion Sanitaria, Grupo de Atención Primaria; Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC); Instituto de Investigación en Servicios de Salud Kronikgune
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Global health, Health policy, Infectious diseases, Public health
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Title page:

Title:

Variables associated to COVID-19 severity: an observational study of non-paediatric confirmed cases from the general population of the Basque Country, Spain.

Authors:

Kalliopi Vrotsou (0000-0002-3296-3923)[1,2,3], Rafael Rotaeche (0000-0001-5594-3768)[1,2,4],Maider Mateo-Abad (0000-0003-2541-8166)[2,3], Mónica Machón (0000-0001-6379-0351) [1,2,3], Itziar Vergara (0000-0001-9671-7898)[1,2,3]

Addresses:

1. Instituto de Investigación Sanitaria Biodonostia, Grupo de Atención Primaria, San Sebastián, Spain.
2. Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Baracaldo, Spain.
3. Instituto de Investigación en Servicios de Salud Kronikgune, Baracaldo, Spain.
4. Osakidetza, Centro de Salud de Alza, San Sebastián, Spain

Authors´s position:

Kalliopi Vrotsou: biostatistician.
Rafael Rotaeche: general practitioner.
Maider Mateo-Abad: biostatistician.
Monica Machón: investigator
Itziar Vergara: director of the Grupo de Atención Primaria, IIS Biodonostia

Corresponding author: Kalliopi Vrotsou,
Address: IIS Biodonostia, Paseo Dr. Beguiristain 20014, San Sebastián Spain,
Telephone number: +34 943006086
e-mail: KALLIOPI.VROTSOUKANARI@osakidetza.eus

Abstract

Objectives: To investigate which were the most relevant sociodemographic and clinical variables associated to COVID-19 severity, and uncover how their inter-relations may have affected such severity.

Design: A retrospective observational study based on electronic health record data.

Participants: Individuals ≥14 years old with a positive PCR or serology test, between the 28th of February and 31st of May 2020, belonging to the Basque Country (Spain) public health system. Institutionalised and individuals admitted to a Hospital at Home unit were excluded from the study.

Main outcome measure: Three severity categories were established, primary care, hospital/ICU admission and death.

Results: A total of n=14197 cases fulfilled the inclusion criteria. Most variables presented statistically significant associations with the outcome ($p<0.0001$). The CART recursive partitioning methodology (based on n=13792) suggested that among all associations, those with, age, sex, stratification of patient health care complexity, chronic consumption of blood and blood forming organ, and nervous system drugs, as well as the total number of chronic ATC types were the most relevant. Psychosis also emerged as a potential factor.

Conclusions: Older cases are more likely to experience more severe outcomes. However, the sex, underlying health status and chronic drug consumption may interfere and alter the aging effect. Understanding the factors related to the outcome severity is of key importance when designing and promoting public health intervention plans for the COVID-19 pandemic.

Strengths and limitations of this study

- Over 13000 confirmed COVID-19, non institutionalised, ≥14 years old cases were explored
- Electronic health records data were a valuable source of information in this study
- The three-category outcome severity: primary care only, hospitalized/ICU care, and death was studied in a joint manner
- The CART methodology allowed exploring the big sample and the numerous variables of interest in a flexible way
- Information on COVID-19 symptoms was not properly registered during the first pandemic wave

Introduction

In December 2019 the new coronavirus SARS-Cov-2 initiated the COVID-19 disease in China, which soon afterwards, on March 12, was declared a pandemic by the World Health Organization (WHO). The rapid expansion of the virus, along with its high death toll and the serious health aftermaths, have rendered the COVID-19 outbreak as one of the worst health crises in almost a century worldwide.

Since the first infections were detected in Spain the statistics have situated this country among the most affected in Europe, both in terms of total cases and in deaths per million people (1). International literature on COVID-19 is rapidly growing (2–7). The research conducted so far in Spain, has focused mainly on predicting the evolution of the pandemic (8), or the factors associated to mortality (9). Hospitalised individuals have also been described (10), and the variables related to severe outcomes in these populations have been explored (11,12). But so far, none of the previous works has considered the gradient of the COVID-19 severity by studying a multiple category outcome.

The autonomous community of the Basque Country, situated in the North of Spain, has its own public health system (Osakidetza), which offers sanitary coverage to some 2.3 million people. Since 2009 Osakidetza has promoted an integrated health care model, by coordinating its different care levels and offering a more holistic approach on patient care (13). It counts with an extensive electronic health records infrastructure, where information on patient health data and episodes of care are stored. The objective of the present observational study was to describe a big series of COVID-19 infected individuals during the first pandemic wave; establish their infection severity level, based on electronic health record data, and explore what characteristics may be associated to that severity. To this end, the Classification and Regression Trees (CART) methodology was applied. This statistical technique splits the sample into mutually exclusive sub-groups that share the same characteristics and can be particularly useful when analysing big data sets (14).

Methods

Data source and variables

All information was extracted from the electronic health records of the Basque Country Public Health System-Osakidetza, via the Osakidetza Business Intelligence (OBI) tools. Data extraction covered the period between 28th February 2020 and 31st May 2020; corresponding to the first detected case in the Basque Country and the end of the first pandemic wave in Spain. Only the health records of individuals ≥14 years old with a COVID-19 positive Polymerase Chain Reaction (PCR) or antibody test, were included, as no antigen tests were performed at that time. Data of cases living in residential homes or those admitted to a hospital at home unit were excluded.

The following variables were studied. Age, sex and income level derived by the pharmaceutical co-payment scheme (<18,000€, 18,000-100,000€, >100,000€). Chronic medication consumption was explored using the Anatomical Therapeutic Chemical (ATC) system at the first level (https://www.who.int/tools/atc-ddd-toolkit). Polypharmacy, defined as the consumption of 5 or more chronic drugs, and the number of ATC types consumed were derived. Chronic pathologies based onto the International Classification of Diseases ICD-9 codes, COVID-19 symptoms registered during consultations and flu vaccination in the year 2019 were also considered. The Osakidetza stratification according to patient health care complexity was studied. Based on a series of health data, and the use of health services during the previous year, this variable classifies individuals into four categories, ranging from less to more severe: prevention and promotion of healthy population, self-management support, disease management, and case management. Pluripathological individuals belong to the last category. This classification is renewed at the beginning of every calendar year, for all individuals ≥14-years registered in the Osakidetza system at least during the previous 6 months. A detailed description can be found elsewhere (15).

Given that the data were anonymous and clinical analyses could not be conducted, it was assumed that the severity of a case would be indicated by the most demanding level of medical attention received, within the study period. Four severity levels were initially identified: primary care

attention only (PC), hospitalisation without intensive care unit admission (Hospital), intensive care unit admission (ICU), and death. During the pandemic, several emergency ICU units were set-up within hospitals across the Basque Country. Nevertheless, this information was not reflected in the electronic health records. As a result, cases admitted to such ICUs were registered as hospital admissions. This fact imposed the necessity to merge Hospital and ICU admissions into one category in the current work. Cases meeting the inclusion criteria were considered only once in the current analyses. The project has been approved by the ethics committee CEIm de Euskadi at 22/07/2020 (reference code: PI2020087).

Patient and public involvement

Due to the study design, no patient and public involvement was considered. Nonetheless, two of the authors are medical doctors, which has offered valuable support during this work.

Statistical analysis

Continuous variables are presented as means with standard deviations (SD), while medians and interquartile ranges (Q1, Q3) are given for discrete variables. Categorical variables are presented with frequencies and percentages (%). Three-group unadjusted comparisons were performed with the one-way ANOVA, Kruskal-Wallis and chi-square test, respectively. The Jonckheere-Terpstra and Mantel-Haenszel chi-square, both testing for a trend along the three severity groups were additionally tested (16).

CART

The CART methodology is a non-parametric statistical tool, which can be very useful when handling big data sets with many variables. This statistical technique partitions the sample into smaller homogenous groups that share the same characteristics. The splitting process starts considering the whole sample that is then recursively partitioned into mutually exclusive sub-samples according to the most important variables, selected among all candidate variables. Important variables in CART are those that minimize the variability of the outcome within each sub-sample. This process results in a tree-like structure with multiple levels, which offers a visual

representation of which variables affect the outcome the most. At the same time, it allows understanding the inter-relations the indicated factors may have with one another. CART analysis is a flexible option for data sets with correlated variables, as in our case. (14,17).

The starting point of the tree structure is the root node and each split is an offspring node. Offsprings that do not split any further are called terminal. In the current analyses splitting was based on the entropy criterion and each variable was allowed only once per tree branch. For a stopping rule, the number of terminal nodes, and the observations included in each of them were considered. A tree with 10 terminal nodes, each including at least 1% of the valid sample data was selected. Cost-complexity pruning was applied. Variables with significance levels p>0.010 in the three-group comparisons and those with a total frequency <1% of the valid sample were excluded from the CART stage, while missing data were omitted (14). Analyses were performed with the SAS software version 9.4 (Copyright (c) 2016 by SAS Institute Inc., Cary, NC, USA.). The SAS proc hpsplit function was used for tree construction.

Results

A total of n=14197 COVID-19 cases fulfilled the inclusion criteria. Of n=9722 (68.5%) received PC attention only, n=3710 (26.1%) had a hospital or ICU admission (n=3630 and 80, respectively), and n=765 died (5.4%). Most cases were detected via PCR (n=8933), and this detection method was the most prevalent in all three outcome groups (PC: 51.0%, Hospital/ICU: 87.7%, Death: 93.3%). Table 1 presents the baseline information of the sample. Overall, mean age was 53.7 (SD:17.4) years, and it increased with outcome severity. Most infected cases were females, but at the same time this sex group presented lower infection severity. In particular females were more prevalent in PC (68.4%), whereas more males were observed in the Hospital/ICU and Death groups. As far as the heath care complexity stratification variable was concerned, the PC outcome group presented the highest percentage of healthy individuals (36.1%), while case management was most prevalent in the Death outcome group (36.6%). Based on the available information, individuals with an annual income <18.000 euros were more prevalent in the Hospital/ICU and Death groups, and those with higher income received mostly

PC attention. Finally, the Death group had the highest percentage of individuals with a flu vaccination in the previous year. This observation was consistent for cases <65 and ≥65 years of age, even though the corresponding percentages of the older cases were higher. All comparisons were statistically significant.

Chronic medication consumption data are presented in **Table 2**. Overall, the most consumed medications were those for the nervous system (38.7%), alimentary tract and metabolism (33.0%), and cardiovascular system (30.2%). With the exception of musculo-skeletal system and antiparasitic products, insecticides and repellents, the same trend was observed in all other ATC types. The consumption of alimentary tract and metabolism disorders (A), blood and blood forming organs (B), cardiovascular system (C), and nervous system diseases drugs (N) exceeded 60% in the Death group. Both polypharmacy and the number of ATC types consumed was associated with infection severity.

Regarding the chronic diseases, the most prevalent condition was related to mental pathologies (**Table 3)**. In particular, 30% of the sample had received a diagnosis corresponding to the ICD-9 neurotic, personality or other nonpsychotic mental disorder. Hypertension was the next more prevalent condition (21%), followed by diseases of the blood and blood forming organs (11.5%), diseases of the esophagus, stomach and duodenum (10.4%). Diabetes mellitus was present in 8.5% of the sample. With the exception of neurotic, personality or other nonpsychotic conditions, that presented the same distribution along the three outcome groups, the prevalence of the most frequent pathologies increased with COVID-19 severity. A similar trend was seen in the total number of chronic diseases. Non-infectious enteritis and colitis, and allergic asthma were the only chronic conditions presenting a descending prevalence with outcome severity, but percentage differences were low.

CART

The CART process indicated that age, sex, health care complexity stratification, the ATC categories of blood and blood forming organ medication (B), as well as nervous system drugs (N) along with the frequency of ATC types consumed would be the most relevant variables in understanding the main case characteristics associated to the outcome. During this process the variable of psychoses was also flagged as important. In spite of its low prevalence (2.9%) psychoses was given a lot of weight in the older section of the population. The inclusion of this pathology resulted in a less parsimonious model; with ATC-N drugs placed in an additional tree level. Nonetheless, given that psychoses was the single variable resulting in a node with a death majority, and that other authors have already suggested an association between antipsychotic drugs and mortality in COVID-19 cases (9), presenting the corresponding findings was considered of relevance. Therefore, the CART process was repeated twice, first excluding and afterwards including psychoses.

Excluding psychoses

The tree generated by the CART process is depicted in Figure 1. Most cases <64.7 years of age (81.4%, node 1) received mainly PC attention. In this tree branch, males presented 15.3% more Hospital/ICU compared to females. Among males, those with worse health (node 8) had 19.2% more Hospital/ICU admissions, compared to the rest (node 7). The majority of males with worse baseline health status who consumed ≥3 ATC types experienced a Hospital/ICU admission.

Cases ≥64.7 years of age had mainly a Hospital/ICU outcome (52.6%), with a considerable Death prevalence (21.2%). Those with worse baseline health (node 6) had 4.6% more Hospital/ICU admissions and 15.8% more Deaths, compared to the rest. Cases with better baseline health status (node 5) were further split according to ATC-B consumption. Death for blood and blood forming organ drugs consumers was experienced in 23.2% of the cases, with the same outcome being 9.1% in non-consumers. Within this last group (node 9) the majority of females received PC

attention, while Hospital/ICU was the most prevalent outcome in males. In a similar way, among node 6 cases, males presented worse evolution than females. Finally, males consuming chronic medications for the nervous system (node 18) had 17.9% more Deaths compared to non-consumers. The 10 terminal nodes can easily be ordered less severe (i.e. <64.7 year old females, node 3) to most severe outcomes groups (i.e. ≥64.7 year old disease and case management males who consume nervous system drugs, node 10).

Including psychoses

The resulting CART model when the variable of psychoses was included in the recursive process is presented in Figure 2. Psychoses, was one of the main variables of this model, and the single split variable for node 6. Inclusion of this pathology added one more level to the CART tree, with chronic nervous system drugs being a split variable for node 15. Cases with psychoses had a 50% Death. The ATC-N consumers presented less PC and higher Death compared to non-consumers. No other changes were observed compared to the Figure 1A model. In this case the most severe outcome group was ≥64.7 year old disease and case management cases with psychosis (node 12).

Discussion

The present work has studied the sociodemographic and clinical characteristics of a big number of Spanish COVID-19 cases of the first pandemic wave. According to the information extracted from electronic health record data, the variables of age, sex, previous pathologies and chronic drug consumption may be decisive in understanding infection severity.

Both age and male sex have been flagged as important risk factors by previous COVID-19 research (2,3,7,9,11,12,18). The importance of age is probably undisputable, given the deterioration of the body´s immunity mechanisms and the loss of its capacity to adapt to the environment (19). The present data appears to reflect this known aging effect. In relation to the variable of sex, females presented consistently higher PC and lower Hospital/ICU in the splits where sex was present. Data from various countries are suggesting that females have better COVID-19 infection outcomes than males (7,20). Females are considered to have stronger immunity systems (21). Even though the exact mechanisms responsible for these differences in the COVID-19 context are still unclear and probably multifactorial (20); certain works are hypothesizing that low androgens levels can have a protective role against this disease (22). The current data, in conjunction with previous evidence call for a better understanding of the role of sex, in the current pandemic. Sex-specific analyses of future wave data should be planned. But more importantly, high quality prospective studies collecting sex-disaggregated data are needed (23).

The health care complexity stratification variable was present in both main tree arms. It should be mentioned that the way CART divided this 4-category variable into a binary one, by merging the two less severe vs. the two more severe groups was imposed by the data, not the investigators. Worse health status at the time of the infection, was associated to more hospitalizations for younger cases, and mainly to more deaths among older individuals. The inclusion of this stratification variable in the CART model is a relevant finding. Tools that stratify the general population, identifying those at greater risk, can be an asset for public health prevention programs. In the COVID-19 literature, the stratification approach has so far mainly focused on hospitalized

patients (12,24,25). While one meta-analysis of in-hospital cases claimed that in COVID-19 infections underlying health conditions are even more important than age(26). Our data suggest that, at least at the local level, this very stratification variable can offer valuable information and its implementation may worth be considered when setting up public health action plans. Study of similar indicators used in other health systems would be encouraged.

As far as the drug consumption was concerned, chronic blood and blood forming organ drugs (B) and drugs for the nervous system (N), both appeared as important variables for cases ≥64.7 years of age. Cases consuming those drugs presented higher severity levels. ATC-N was the most frequent medication across all three outcome groups. ATC-B had the steepest raising in consumption from one severity level to the next. Several neurological manifestations after a COVID-19 infection have been described in the literature, with the virus perceived by certain authors as a threat for the whole nervous system (27). It is probable that individuals already suffering by chronic neurological conditions may be indeed more likely to present worse outcomes once infected (28,29). Blood related parameters like systolic and diastolic pressure, red and white cell counts, platelets, lymphocytes, among others, have been highlighted as significant predictors in different COVID-19 diagnostic models (7). An association between certain ATC-B drugs and higher odds of death in infected cases has also been observed(9). Chronic anticoagulation treatment, is referenced as protective against COVID-19 mortality by some (30), and ineffective by others (31). COVID-19 cases present a high frequency of thrombotic events, which is leading to an expansion of anticoagulation drug use when treating the disease (32). But in patients already receiving such drugs prior to infection, drug-drug interactions and infection severity should be carefully assessed before any antiviral therapy is given, or switching from oral to parenteral antithrombotic administration (33). Worse severity seen among ATC-B consumers in the current data may reflect also an increased risk for patients already under anticoagulation therapy. Poor outcomes due to therapeutic decisions and drug-drug interactions cannot be excluded either. Our continuing COVID-19 work will refine future data explorations. Obtaining

for example ATC data at the second or third level, as well as information of in-patient treatments will offer more insight into these associations.

Psychoses was a relevant variable in the CART process. Antipsychotic drugs belong to the ATC-N medication type; which is probably why allowing for the inclusion of psychoses relocated this drug group further down in the tree structure. Older patients with worse baseline health and psychoses had the highest death rate among all CART nodes. We can only hypothesize over the mechanisms that could explain such a finding. On one hand, individuals with psychotic disorders present excess mortality compared to the general population, mainly due to lifestyle choices, associated comorbidities and medication side effects (34). On the other hand, the treatment management of these cases is challenging as alteration or abrupt cessation of their current medication could potentially lead to a sudden health deterioration or even death (35). This could happen for example during hospital and ICU admissions. In the present sample 75% of the deaths seen in the psychoses node had been admitted to a hospital during the study period. The available information does not allow knowing whether death took place during the admissions, neither the in-patient treatment regime. An observational USA study of >60000 cases claimed that psychiatric disorders are a risk factor associated to higher COVID-19 diagnosis; with psychosis presenting greater risk ratios versus mood and anxiety disorders. The same study also reported an increased risk of first-time psychiatric disorders for survivors (36). Others have suggested that antipsychotics use (9) and schizophrenia spectrum disorders (37) are associated with higher COVID-19 mortality. Even though more research in this direction is required, the available data seem to highlight the need for a close monitoring of cases with psychiatric disorders.

The total number of chronically consumed ATC types was an important variable among cases <64.7 years of age. This variable, which could also be perceived as an indicator of the associated comorbidities, stresses even more the importance that underlying pathologies may have in determining the severity of the infection outcome (26).

In this work, a surrogate outcome variable has been used. Assuming that more intensive care levels represented worse COVID-19 status is a decision also taken by previous authors (11,38–

40). The available data does not allow studying if admissions and deaths may have been due to other health problems. The female prevalence of this sample was greater than that seen in other COVID-19 publications (3,4,7), but nonetheless similar to previous studies performed in this country (9,11). In the Spanish reality, women traditionally assume the caretaker´s role for younger and older members of their families, while they also occupy more home-assisting jobs (41) and health related professions (42). All these conditions may imply higher exposure rates to the virus, which may offer a possible explanation for the sample´s sex distribution.

The current study has certain limitations. The implemented information is based exclusively on electronic health record data within the previously defined dates. After that period the severity of certain cases may have worsen. Nonetheless, the end study date corresponds to the end of the first COVID-19 wave in our area, where new infections and deaths were very low. This, in combination with the big study sample should have minimized the effect of possible outcome variations. No COVID-19 symptoms are presented. An attempt to register these symptoms was incorporated at the Osakidetza electronic records, early on after the outbreak. But, the number of symptoms and registration format evolved over the studied period; PC and hospital registrations differed; the medical staff mostly annotated symptoms in text format; while most importantly such registration was totally missing in many cases. During analysis an effort to re-code text annotations, and homogenize information from primary care and hospital data was made. In spite of that, and due to the frequency of missing values, the representativeness of the corresponding data could not be assumed. Symptoms are probably more relevant for algorithms discriminating cases from non-cases (43). During the first pandemic wave no massive population testings were performed in Spain, but at the end of that wave serology tests were administered to the health professionals and allied services of our geographic area. Thus, identified cases were either symptomatic, close contacts of cases, or individuals working in the health sector. However, the profession of the cases was not an available piece of information in this sample. Working with health records makes recovering missing data or refining variable information a very difficult task. This was also the case with the income level. Its broad categories may have obscured a more

appropriate exploration. On the other hand, the high frequency of missing income level data seen in the Death group, is due to the "un-subscriptions" of the dead cases from the medication dispensing registry. It is important to note that the target of the Basque public health system is a health coverage based on the health needs and not the earnings of the individuals.

One of the main strength of this study is its big sample size. The consideration of three outcome groups is another advantage, which allows for a better visualization of the different severity levels of the disease. Finally, implementing the CART methodology assisted in translating a complex and multifactorial reality into an easy to follow picture. Our findings make clinical sense and are supported by previous evidence. They appear to endorse the need for public health prevention plans that consider population characteristics. At the same time, they highlight that for a multifactorial problem to be properly treated, not only the factors affecting it, but also the inter-relations between the latter should be thoroughly studied. The COVID-19 pandemic may be a new starting point in the public health paradigm. The necessity for public health promoters to work hand-in-hand with investigators and data analysts has become indisputable, under the current circumstances. Prevention plans should be based on rigorous data and understanding of the latter. This is the only way to assure that possible re-organization and estimation of future resources can reach optimal results.

Dissemination declaration: Upon acceptance, the results will be disseminated to patient organizations, medical students or/and other interested groups or means of communication.

Data availability: The data of the current study are stored in a server of our institution. Sharing them with external investigators will be evaluated on an individual basis and will require an approval by the Osakidetza central services. The corresponding author should be contacted.

Funding source: No funding was obtained for this study.

Authors´Contributions: IV, RR and MM planned this study and obtained the permission for exploring the corresponding data by the Osakidetza central services. MMA set the filters and performed the data extraction of the electronic health record data. KV and MMA are both responsible for data cleaning and recoding. KV and MMA performed all statistical analyses. The input of IV and RR have assured a clinically meaningful perspective of all presented analyses and results. MM performed literature searches. KV drafted the first manuscript version. All authors read and contributed to the consecutive manuscript versions.

Competing interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work."

References

1. Coronavirus Update (Live): 109,735,851 Cases and 2,420,401 Deaths from COVID-19 Virus Pandemic - Worldometer [Internet]. [cited 2021 Feb 16]. Available from: https://www.worldometers.info/coronavirus/

2. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. Available from: https://isaric4c.net

3. Nachtigall I, Lenga P, Jóźwiak K, Thürmann P, Meier-Hellmann A, Kuhlen R, et al. Clinical course and factors associated with outcomes among 1904 patients hospitalized with COVID-19 in Germany: an observational study. Clin Microbiol Infect [Internet]. 2020 [cited 2020 Nov 10]; Available from: https://reader.elsevier.com/reader/sd/pii/S1198743X20304936?token=088D506A6F15F95832BA1FF82EBE9040EC1E3154088A738F47EBD33697538DEFF79877982A8A7F3857E2695415D62354

4. Hewitt J, Carter B, Vilches-Moraga A, Quinn TJ, Braude P, Verduri A, et al. The effect of frailty on survival in patients with COVID-19 (COPE): a multicentre, European, observational cohort study. Lancet Public Heal [Internet]. 2020 Aug 1 [cited 2020 Nov 10];5(8):e444–51. Available from: www.thelancet.com/

5. Baqui P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. Lancet Glob Heal [Internet]. 2020 Aug 1 [cited 2020 Nov 10];8(8):e1018–26. Available from: www.thelancet.com/lancetghVol

6. Jones RC, Ho JC, Kearney H, Glibbery M, Levin DL, Kim J, et al. Evaluating Trends in COVID-19 Research Activity in Early 2020: The Creation and Utilization of a Novel Open-Access Database. Cureus [Internet]. 2020 Aug 22 [cited 2020 Nov 10];12(8). Available from: https://pubmed.ncbi.nlm.nih.gov/32968602/

7. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ. 2020;369.

8. Aguiar M, Ortuondo EM, Bidaurrazaga Van-Dierdonck J, Mar J, Stollenwerk N. Modelling COVID 19 in the Basque Country from introduction to control measure response. Sci Rep [Internet]. 2020 Dec 14 [cited 2020 Oct 16];10(1):17306. Available from: http://www.nature.com/articles/s41598-020-74386-1

9. Poblador-Plou B, Carmona-Pírez J, Ioakeim-Skoufa I, Poncel-Falcó A, Bliek-Bueno K, Cano-Del Pozo M, et al. Baseline chronic comorbidity and mortality in laboratory-confirmed COVID-19 cases: Results from the PRECOVID study in Spain. Int J Environ Res Public Health. 2020;17(14):1–14.

10. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. Rev Clin Esp. 2020;

11. Working group for the surveillance and control of COVID-19 in Spain. The first wave of the COVID-19 pandemic in Spain : characterisation of cases and risk factors for severe outcomes , as at 27 April 2020. Eurosurveillance [Internet]. 2020;25(50):1–13. Available from: http://dx.doi.org/10.2807/1560-7917.ES.2020.25.50.2001431

12. Rubio-Rivas M, Corbella X, Mora-Luján JM, Loureiro-Amigo J, López Sampalo A, Yera Bergua C, et al. Predicting Clinical Outcome with Phenotypic Clusters in COVID-19

Pneumonia: An Analysis of 12,066 Hospitalized Patients from the Spanish Registry SEMI-COVID-19. J Clin Med. 2020;9(11):3488.

13. Bengoa R. Transforming health care: an approach to system-wide implementation. Int J Integr Care. 2013;13(3).

14. Zhang H, Singer BH. Recursive Partitioning and Applications. Second edi. New York: Springer Series in Statistics; 2010.

15. Orueta JF, Nuño-Solinis R, Mateos M, Vergara I, Grandes G, Esnaola S. Predictive risk modelling in the Spanish population: A cross-sectional study. BMC Health Serv Res [Internet]. 2013;13(1):1. Available from: BMC Health Services Research

16. Walker, Glenn A, 2002. Common Statistical Methods for Clinical Research with SAS Examples, Second Edition. SAS Institute. Cary, NC;

17. Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and Regression Tree Analysis in Public Health: Methodological Review and Comparison with Logistic Regression. Ann Behav Med. 2003;26(3):172–81.

18. Wollenstein-Betech S, Cassandras CG, Paschalidis IC. Personalized Predictive Models for Symptomatic COVID-19 Patients Using Basic Preconditions: Hospitalizations, Mortality, and the Need for an ICU or Ventilator. Int J Med Inform. 2020 Aug;104258.

19. WHO Team: Social determinants of Health. World report on ageing and health [Internet]. World Health Organization. Luxembourg; 2015 [cited 2020 Dec 9]. Available from: https://www.who.int/publications/i/item/world-report-on-ageing-and-health

20. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL. Impact of sex and gender on COVID-19 outcomes in Europe. Vol. 11, Biology of Sex Differences. BioMed Central Ltd.; 2020.

21. Rider V, Martocchia A, Cava A La, Moulton VR. Sex Hormones in Acquired Immunity and Autoimmune Disease. Front Immunol | www.frontiersin.org [Internet]. 2018;9:2279. Available from: www.frontiersin.org

22. Mohamed MS, Moulin TC, Helgi, Schiöth B. Sex differences in COVID-19: the role of androgens in disease severity and progression. Endocrine [Internet]. 2020;71:3–8. Available from: https://doi.org/10.1007/s12020-020-02536-6

23. Dehingia N, Raj A. Sex differences in COVID-19 case fatality: do we know enough? Lancet Glob Heal. 2020 Nov;

24. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ [Internet]. 2020;370:22. Available from: http://dx.doi.org/10.1136/bmj.m3339

25. Mei Y, Weinberg SE, Zhao L, Frink A, Qi C, Behdad A, et al. Risk stratification of hospitalized COVID-19 patients through comparative studies of laboratory results with influenza. EClinicalMedicine [Internet]. 2020;26:100475. Available from: https://doi.org/10.1016/j.eclinm.2020.100475

26. Starke KR, Petereit-Haack G, Schubert M, Kämpf D, Schliebner A, Hegewald J, et al. The Age-Related Risk of Severe Outcomes Due to COVID-19 Infection: A Rapid Review, Meta-Analysis, and Meta-Regression. Available from: www.mdpi.com/journal/ijerph

27. Koralnik IJ, Tyler KL. COVID-19: A Global Threat to the Nervous System. Ann Neurol. 2020;88(1):1–11.

28. Herman C, Mayer K, Sarwal A, Org N, Covid19 /. Scoping review of prevalence of neurologic comorbidities in patients hospitalized for COVID-19. 2020;

29. Romagnolo A, Balestrino R, Imbalzano G, Ciccone G, Riccardini · Franco, Carlo ·, et al.

Neurological comorbidity and severity of COVID-19. J Neurol [Internet]. 123AD;1:3. Available from: https://doi.org/10.1007/s00415-020-10123-y

30. Rossi R, Coppi F, Talarico M, Boriani G. Protective role of chronic treatment with direct oral anticoagulants in elderly patients affected by interstitial pneumonia in COVID-19 era. Eur J Intern Med. 2020;77(May):158–60.

31. Schiavone M, Gasperetti A, Mancone M, Curnis A, Mascioli G, Mitacchione G, et al. Oral anticoagulation and clinical outcomes in COVID-19: An Italian multicenter experience. Int J Cardiol. 2020;323:276–80.

32. Hanff TC, Mohareb AM, Giri J, Cohen JB, Chirinos JA. Thrombosis in COVID-19. Am J Hematol. 2020;95(12):1578–89.

33. Vivas D, Roldán V, Esteve-Pastor MA, Roldán I, Tello-Montoliu A, Ruiz-Nodar JM, et al. Recommendations on antithrombotic treatment during the COVID-19 pandemic. Position statement of the Working Group on Cardiovascular Thrombosis of the Spanish Society of Cardiology. Rev Esp Cardiol. 2020;73(9):749–57.

34. Oakley P, Kisely S, Baxter A, Harris M, Desoe J, Dziouba A, et al. Increased mortality among people with schizophrenia and other non-affective psychotic disorders in the community: A systematic review and meta-analysis. J Psychiatr Res [Internet]. 2018;102(April):245–53. Available from: https://doi.org/10.1016/j.jpsychires.2018.04.019

35. Simon L, Hashmi M, Callahan A. Neuroleptic Malignant Syndrome [Internet]. [Updated 2020 Nov 20]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-; Available from: https://www.ncbi.nlm.nih.gov/books/NBK482282/

36. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in

the USA. The Lancet Psychiatry [Internet]. 2020 [cited 2020 Dec 10]; Available from: https://doi.org/10.1016/S2215-0366

37. Nemani K, Li C, Olfson M, Blessing EM, Razavian N, Chen ; Ji, et al. Association of Psychiatric Disorders With Mortality Among Patients With COVID-19 Multimedia Supplemental content. JAMA Psychiatry [Internet]. 2021; Available from: https://jamanetwork.com/

38. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Heal. 2020 Aug 1;8(8):e1003–17.

39. Ghahramani S, Tabrizi R, Lankarani KB, Kashani SMA, Rezaei S, Zeidi N, et al. Laboratory features of severe vs. non-severe COVID-19 patients in Asian populations: A systematic review and meta-analysis. Vol. 25, European Journal of Medical Research. BioMed Central; 2020.

40. Ou M, Zhu J, Ji P, Li H, Zhong Z, Li B, et al. Risk factors of severe cases with COVID-19: a meta-analysis. Epidemiol Infect. 2020 Aug 12;148:e175.

41. Hellgren Z, Serrano I. Transnationalism and Financial Crisis: The Hampered Migration Projects of Female Domestic Workers in Spain. Available from: www.mdpi.com/journal/socsci

42. Lázaro-Pérez C, Ángel Martínez-López J, Gómez-Galán J, López-Meneses E. Anxiety About the Risk of Death of Their Patients in Health Professionals in Spain: Analysis at the Peak of the COVID-19 Pandemic. Available from: www.mdpi.com/journal/ijerph

43. Zimmerman R, Nowalk MP, Bear T, Taber R, Sax T, Eng H, et al. Proposed Clinical Indicators for Efficient Screening and Testing for COVID-19 Infection from Classification

and Regression Trees (CART) Analysis [Internet]. medRxiv : the preprint server for health sciences. 2020. Available from: https://doi.org/10.1101/2020.05.11.20097980

Table 1: Baseline information of the COVID-19 cases during the first wave of the pandemic

Variables	*Total (n=14197)*	Primary Care (n=9722)	Hospital/ICU (n=3710)	Death (n=765)	p-value
Age; mean (SD)	*53.7 (17.4)*	48.0 (14.4)	62.8 (16.1)	82.3 (10.5)	<0.0001
Sex					
Male	*5520 (38.9)*	3073 (31.6)	2031 (54.7)	416 (54.4)	<0.0001
Female	*8677 (61.1)*	6649 (68.4)	1679 (45.3)	349 (45.6)	
Health care complexity					
Missing information	*405 (2.9)*	*307 (3.2)*	*86 (2.3)*	*12 (1.6)*	
Prevention and promotion	*3878 (27.3)*	3399 (36.1)	470 (12.9)	9 (1.2)	<0.0001
Self-management support	*6821 (48.0)*	4989 (52.9)	1675 (46.2)	157 (20.8)	
Disease management	*2252 (15.9)*	891 (9.4)	1050 (28.9)	311 (41.3)	
Case management	*841 (5.9)*	136 (1.4)	429 (11.8)	276 (36.6)	
Income level					
Missing information	*854 (6.0)*	*251 (2.6)*	*130 (3.5)*	*473 (61.8)*	
<18.000 euros	*6536 (46.0)*	4297 (45.3)	2038 (56.9)	201 (68.8)	<0.0001
18.000-100.000 euros	*6670 (47.0)*	5074 (53.5)	1507 (42.0)	89 (30.4)	
>100.000 euros	*137 (1.0)*	100 (1.0)	35 (0.9)	2 (0.6)	
Flu vaccination in 2019: yes					
All vaccinated cases	*3336 (23.5)*	1322 (13.6)	1446 (39.0)	568 (74.2)	<0.0001
Vaccinated cases <65 years old	*1103 (10.1)*	814 (9.2)	265 (13.6)	24 (42.8)	<0.0001
Vaccinated cases ≥ 65 years old	*2233 (66.5)*	508 (57.7)	1181 (66.9)	544 (76.7)	<0.0001

Data are frequency (percentage), unless otherwise stated. For variables with missing information, percentages and statistical comparisons are based on valid data only. Presented p-values are based on one-way ANOVA for the variable of age and the chi-square test for the categorical variables. Cases <65 year and ≥ 65 years were n=10843 and n=3354, respectively. Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend also resulted in p<0.0001 in all comparisons.

Table 2: *Chronic medication consumption of the COVID-19 sample.*

	Total (n=14197)	**Primary Care (n=9722)**	**Hospital/ICU (n=3710)**	**Death (n=765)**	**p-value**
Medication (ATC type)					
Alimentary tract and metabolism (A)	*4685 (33.0)*	2234 (23.0)	1837 (49.5)	614 (80.3)	<0.0001
Blood and blood forming organs (B)	*2414 (17.0)*	889 (9.1)	1057 (28.5)	468 (61.2)	<0.0001
Cardiovascular system (C)	*4294 (30.2)*	1813 (18.6)	1893 (51.0)	588 (76.9)	<0.0001
Dermatologicals (D)	*1765 (12.4)*	1032 (10.6)	581 (15.7)	152 (19.9)	<0.0001
Genitourinary system and sex hormones (G)	*1690 (11.9)*	1050 (10.8)	505 (13.6)	135 (17.6)	<0.0001
Systemic Hormonal preparations, excluding sex hormones and insulins (H)	*1504 (10.6)*	876 (9.0)	492 (13.3)	136 (17.8)	<0.0001
Antiinfectives for systemic use (J)	*223 (1.6)*	122 (1.3)	73 (2.0)	28 (3.7)	<0.0001
Antineoplastic and immunomodulating agents (L)	*360 (2.5)*	165 (1.7)	141 (3.8)	54 (7.1)	<0.0001
Musculo-Skeletal system (M)	*3137 (22.1)*	2010 (20.7)	952 (25.7)	175 (22.9)	<0.0001
Nervous System (N)	*5494 (38.7)*	2906 (29.9)	1931 (52.0)	657 (85.9)	<0.0001
Antiparasitic products, insecticides and repellents (P)	*42 (0.3)*	24 (0.2)	15 (0.4)	3 (0.4)	0.284
Respiratory System (R)	*2603 (18.3)*	1517 (15.6)	864 (23.3)	222 (29.0)	<0.0001
Sensory Organs (S)	*863 (6.1)*	443 (4.6)	297 (8.0)	123 (16.1)	<0.0001
Various (V)	*188 (1.3)*	45 (0.5)	58 (1.6)	85 (11.1)	<0.0001
Polypharmacy: yes	*2921 (20.5)*	935 (9.6)	1357 (36.5)	629 (82.2)	<0.0001
Num. ATC types consumed: median (Q1,Q3)	*2 (0, 3)*	1 (0, 3)	3 (1, 4)	5 (3, 6)	<0.0001

Data are frequency (percentage), unless otherwise stated. Q1, Q3: interquartile range values. ATC: Anatomical Therapeutic Chemicals. Presented p-values are based on the Chi-square test. The Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend resulted in very similar results in all comparisons.

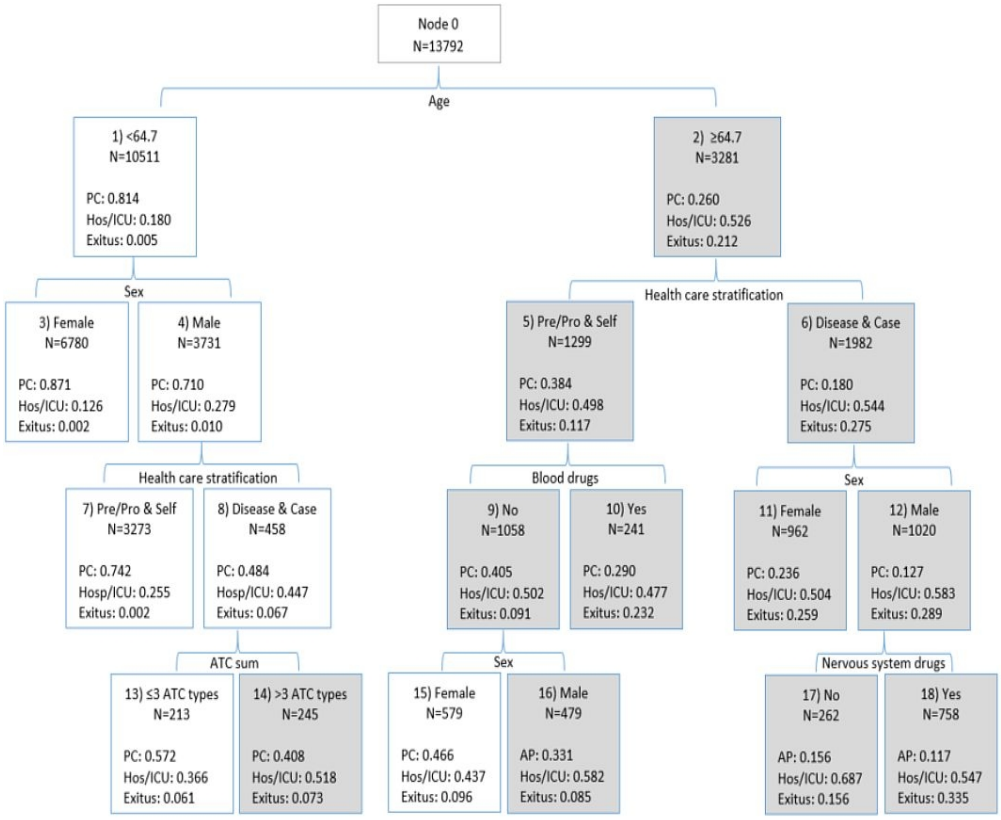
Table 3: *Chronic diseases of the COVID-19 cases in the three outcome groups*

Disease	Total (n=14197)	Primary Care (n=9722)	Hospital/ICU (n=3710)	Death (n=765)	p-value
Infectious disease					
HIV infection	23 (0.2)	7 (0.1)	12 (0.3)	4 (0.5)	0.0002
Liver disease and cirrhosis	133 (0.9)	49 (0.5)	72 (1.9)	12 (1.6)	<0.0001
Malignant neoplasm	918 (6.4)	364 (3.7)	410 (11.0)	144 (18.8)	<0.0001
Endocrine diseases					
Subclinical hypothyroidism without treatment	1101 (7.8)	747 (7.7)	294 (7.9)	60 (7.8)	0.892
Diabetes Mellitus	1213 (8.5)	395 (4.1)	606 (16.3)	212 (27.7)	<0.0001
Diseases of the blood and blood-forming organs	1602 (11.3)	930 (9.6)	492 (13.3)	180 (23.5)	<0.0001
Mental disorders					
Psychoses	412 (2.9)	138 (1.4)	143 (3.9)	131 (17.1)	<0.0001
Neurotic disorders, personality disorders, and other nonpsychotic mental disorders	4258 (30.0)	2926 (30.1)	1096 (29.5)	236 (30.8)	0.712
Mental retardation	39 (0.3)	24 (0.2)	14 (0.4)	1 (0.1)	0.319
Nervous system diseases					
Dementia	126 (0.8)	20 (0.2)	32 (0.8)	74 (9.6)	<0.0001
Other hereditary and degenerative diseases of the central nervous system	307 (2.1)	127 (1.3)	122 (3.2)	58 (7.5)	<0.0001
Diseases of the circulatory system					
Hypertensive disease	2988 (21.0)	1177 (12.1)	1364 (36.8)	447 (58.4)	<0.0001
Ischemic heart disease	448 (3.2)	111 (1.1)	237 (6.4)	100 (13.1)	<0.0001
Cerebrovascular disease	611 (4.3)	189 (1.9)	266 (7.2)	156 (20.4)	<0.0001
Heart failure & Atrial fibrillation and flutter	709 (5.0)	132 (1.4)	361 (9.7)	216 (28.2)	<0.0001
Acute pulmonary heart disease & other venous embolism and thrombosis	150 (1.1)	49 (0.5)	73 (2.0)	28 (3.7)	<0.0001
Arterial embolism and thrombosis	39 (0.3)	17 (0.2)	17 (0.5)	5 (0.7)	0.002
Respiratory disease					
Allergic asthma	354 (2.4)	258 (2.6)	88 (2.3)	8 (1.0)	0.019
Chronic obstructive pulmonary disease and allied conditions (excl. allergic asthma)	1190 (8.3)	630 (6.4)	432 (11.6)	128 (16.7)	<0.0001
Pneumonoconioses and other lung diseases due to external agents	20 (0.1)	9 (0.1)	8 (0.2)	3 (0.4)	0.038
Diseases of the Digestive system					
Diseases of esophagus, stomach and duodenum	1481 (10.4)	907 (9.3)	468 (12.6)	106 (13.9)	<0.0001
Non-infectious enteritis and colitis	643 (4.5)	500 (5.1)	121 (3.3)	22 (2.9)	<0.0001
Regional enteritis & Ulcerative Colitis	73 (0.5)	51 (0.5)	16 (0.4)	6 (0.8)	0.447
Disease of the genitourinary system					
Chronic kidney disease	398 (2.8)	87 (0.9)	188 (5.1)	123 (16.1)	<0.0001
Diseases of the skin and subcutaneous tissue					
Psoriasis	315 (2.2)	180 (1.9)	113 (3.0)	22 (2.9)	<0.0001
Diseases of the musculoskeletal system and connective tissue					
Systemic lupus erythematosus	36 (0.3)	24 (0.2)	10 (0.3)	2 (0.3)	0.972
Rheumatoid arthritis and other inflammatory polyarthropathies	125 (0.9)	59 (0.6)	55 (1.5)	11 (1.4)	<0.0001
Arthropathy associated with other disorders classified elsewhere	8 (0.1)	5 (0.1)	1 (0.0)	2 (0.3)	0.042
Multimorbidity: ≥2 chronic diseases	5326 (37.5)	2715 (27.9)	1975 (53.2)	636 (83.1)	<0.0001
Total number of chronic diseases					
Median (Q1, Q3)	1 (0, 2)	1 (0, 2)	2 (1, 3)	3 (2, 4)	<0.0001

Data are frequency (percentage) unless otherwise stated; Q1, Q3: interquartile range values. Presented p-values are based on the Chi-square test. The Jonckheere-Terpstra and Mantel-Haenszel chi-square test for trend resulted in very similar results in all comparisons.

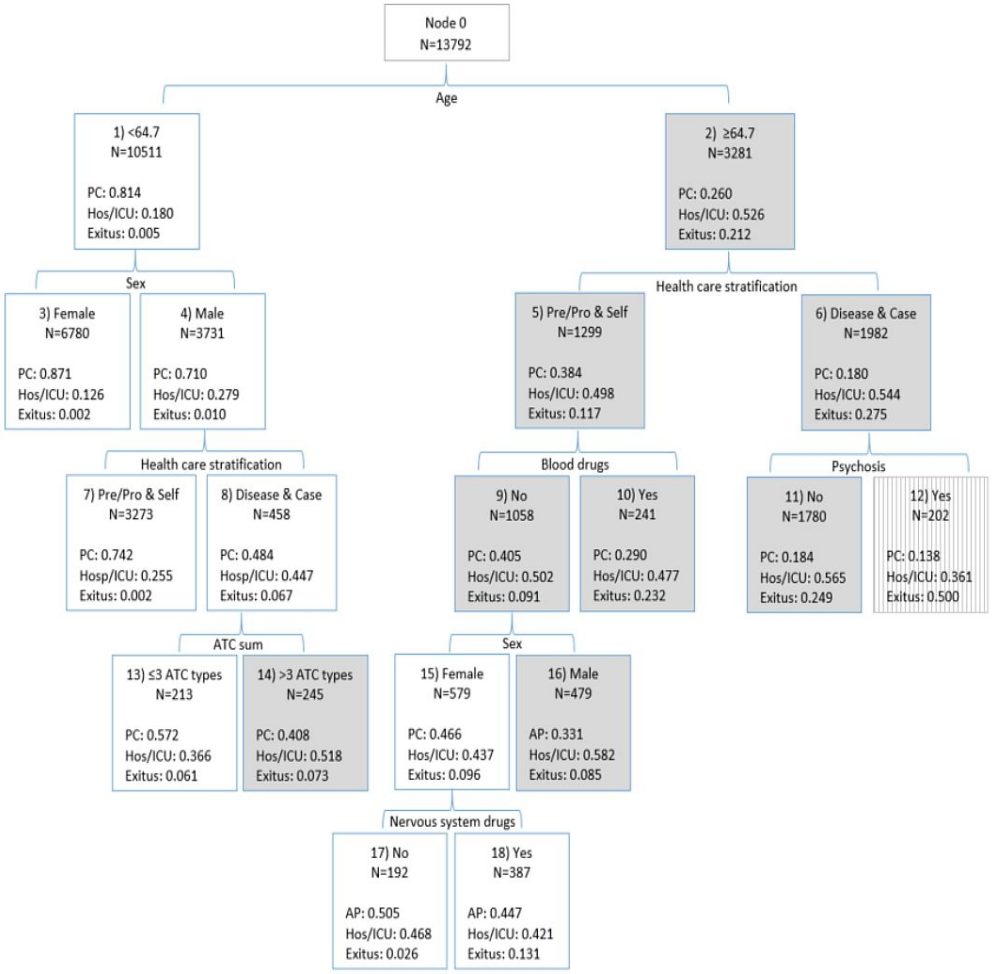
Fig 1: CART model without psychosis. *PC: primary care; Hos/ICU: hospital and intensive care unit. Percentages up to 3 decimal places are given. Due to rounding, node percentages may not add to 100%. Pre/Pro & Self: Prevention and promotion, and Self-management support. Disease & Case: Disease management and Case management. ATC: Anatomical Therapeutic Chemical. White nodes represent groups with a higher percentage of PC care. Grey nodes represent groups with a higher percentage of Hospital/ICU admission.*

Fig 2: CART model with psychosis*. PC: primary care; Hos/ICU: hospital and intensive care unit. Percentages up to 3 decimal places are given. Due to rounding, node percentages may not add to 100%. Pre/Pro & Self: Prevention and promotion, and Self-management support. Disease & Case: Disease management and Case management. ATC: Anatomical Therapeutic Chemical. White nodes represent groups with a higher percentage of PC care. Grey nodes represent groups with a higher percentage of Hospital/ICU admission, and grey stripes nodes groups with a higher percentage of Death.*



CART model without psychosis. PC: primary care; Hos/ICU: hospital and intensive care unit. Percentages up to 3 decimal places are given. Due to rounding, node percentages may not add to 100%. Pre/Pro & Self: Prevention and promotion, and Self-management support. Disease & Case: Disease management and Case management. ATC: Anatomical Therapeutic Chemical. White nodes represent groups with a higher percentage of PC care. Grey nodes represent groups with a higher percentage of Hospital/ICU admission.

90x90mm (300 x 300 DPI)



CART model with psychosis. PC: primary care; Hos/ICU: hospital and intensive care unit. Percentages up to 3 decimal places are given. Due to rounding, node percentages may not add to 100%. Pre/Pro & Self: Prevention and promotion, and Self-management support. Disease & Case: Disease management and Case management. ATC: Anatomical Therapeutic Chemical. White nodes represent groups with a higher percentage of PC care. Grey nodes represent groups with a higher percentage of Hospital/ICU admission, and grey stripes nodes groups with a higher percentage of Death.

90x90mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of *cross-sectional studies*

	Item No	Recommendation	page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3,4
Methods			
Study design	4	Present key elements of study design early in the paper	4,5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4,5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4,5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4,5*
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5, 6
		(*b*) Describe any methods used to examine subgroups and interactions	5, 6
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	5, 6
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			**page**
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	6
		(b) Give reasons for non-participation at each stage	6
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table1
		(b) Indicate number of participants with missing data for each variable of interest	Table1
Outcome data	15*	Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6, 8, 9, figure1
		(*b*) Report category boundaries when continuous variables were categorized	6, figure1
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8, 9
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12, 13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.