medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Cardiovascular vulnerability predicts hospitalisation in primary care clinically suspected and confirmed COVID-19 patients: a model development and validation study

Florien S. van Royen[1*], Linda P.T. Joosten[1], Maarten van Smeden[2], Pauline Slottje[3], Frans H. Rutten[1], Geert-Jan Geersing[1], Sander van Doorn[1]

[1]Dept. General Practice, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[2]Dept. Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[3]Dept. General Practice, Academic network of general practice, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

*Corresponding author, e-mail: f.s.vanroyen-5@umcutrecht.nl, postal address: Universiteitsweg 100, 3584 GX, Utrecht, Netherlands

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Objectives:

Cardiovascular conditions were shown to be predictive of clinical deterioration in hospitalised patients with coronavirus disease 2019 (COVID-19). Whether this also holds for outpatients managed in primary care is yet unknown. The aim of this study was to determine the incremental value of cardiovascular vulnerability in predicting the risk of hospital referral in primary care COVID-19 outpatients.

Design:

Analysis of anonymised routine care data extracted from electronic medical records from three large Dutch primary care registries.

Setting:

Primary care.

Participants:

Consecutive adult patients seen in primary care for COVID-19 symptoms in the 'first wave' of COVID-19 infections (March 1 2020 to June 1 2020) and in the 'second wave' (June 1 2020 to April 15 2021) in the Netherlands.

Outcome measures:

A multivariable logistic regression model was fitted to predict hospital referral within 90 days after first COVID-19 consultation in primary care. Data from the 'first wave' was used for derivation (n=5,475 patients). Age, sex, the interaction between age and sex, and the number of cardiovascular conditions and/or diabetes (0, 1, or ≥2) were pre-specified as candidate predictors. This full model was (i) compared to a simple model including only age and sex and its interaction, and (ii) externally validated in COVID-19 patients during the 'second wave' (n=16,693).

Results:

The full model performed better than the simple model (likelihood ratio test $p<0.001$). Older male patients with multiple cardiovascular conditions and/or diabetes had the highest

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

predicted risk of hospital referral, reaching risks above 15-20%, whereas on average this risk was 5.1%. The temporally validated c-statistic was 0.747 (95%CI 0.729-0.764) and the model showed good calibration upon validation.

Conclusions:

For patients with COVID-19 symptoms managed in primary care, the risk of hospital referral was on average 5.1%. Older, male and cardiovascular vulnerable COVID-19 patients are more at risk for hospital referral.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Coronavirus disease 2019 (COVID-19) has led to a global pandemic ever since the first cases were described in late 2019. In order to orchestrate the flow of patients depending on the expected course of the disease, the need arose for risk profiling patients suffering most from COVID-19. Since then, a fast-growing amount of prediction models have been developed for prognosticating COVID-19 patients. However, most of these models focus on an in-hospital population, with only few prediction models focusing on the community.(1) This is unfortunate as the clinical presentation of (suspected) COVID-19 starts-off with initially mild to moderate symptoms in the first week of illness, and only in some with progression to hypoxemia for which hospital (or even ICU) admission is needed, typically occurring in the second week of illness.(2) If such deterioration occurs, the primary care physician is often the first to decide on the need and optimal timing for more impactful measures, such as intensified monitoring or ultimately referral for hospitalisation. In hospitalised COVID-19 patients, underlying cardiovascular diseases have been identified as strong predictors for further disease deterioration towards ICU admittance or death.(3–7) The incremental value of cardiovascular diseases and/or diabetes in predicting escalation of care in primary care COVID-19 patients has, however, yet to be determined.

The aim of this study was to determine the incremental value of cardiovascular vulnerability – defined by the number of cardiovascular diseases and/or type 2 diabetes mellitus – in predicting the risk of hospital referral in primary care patients with clinically suspected or confirmed COVID-19. Results from this analysis are intended to explore the incremental prognostic value of concurrent cardiovascular disease in community COVID-19 outpatients, and – if indeed present – to support clinical decision making on COVID-19 early disease management and vaccination prioritization for primary care physicians.

Methods

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Study design

This study involves an analysis of anonymised observational electronic medical record data of community people registered by the primary care physician with either confirmed or clinically suspected COVID-19. We assessed the incremental value of cardiovascular disease and/or diabetes by developing a prognostic prediction model in a cohort of patients from the 'first wave' of COVID-19 infections in the Netherlands (March 1 2020 to June 1 2020) that was temporally validated in a cohort of patients from the 'second wave' of infections in the Netherlands (June 1 2020 to April 15 2021). Where appropriate for this study, we adhered to the TRIPOD guideline for reporting prediction models.(8)

Databases

Patients were included from three similar ongoing and dynamic primary care databases run by the academic hospitals of the cities and surrounding municipalities of Utrecht and Amsterdam, containing pseudonymised medical data of approximately 850,000 patients in total: the Julius General Practitioner's Network (JGPN) University Medical Center Utrecht, the Academic Network of General Practice at VU University medical center in Amsterdam (ANH VUmc), and the Academic General Practitioner's Network at Academic Medical Center Amsterdam (AHA AMC).(9–11) Two databases (JGPN and ANH VUmc) were used to identify patients for the development of the prediction model (i.e. development cohort) and all three databases (JGPN, ANH VUmc and AHA AMC) were used to identify patients for the temporal validation (i.e. the validation cohort).

Study population and data collection

Patients for the development cohort were included from March 1 2020 to June 1 2020 (the 'first wave' of the COVID-19 pandemic in the Netherlands). During this time period, very limited polymerase chain reaction (PCR) testing for COVID-19 was available, and moreover mainly restricted to more severe hospitalised cases. Consequently, many symptomatic patients consulting their GP were highly suggestive of COVID-19 and were not tested. We

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

therefore included all consecutive adult patients aged 18 years or older, who visited their primary care physician with confirmed COVID-19 as well as those with symptoms that were clinically suspected by the GP.

For identification of the study population and data collection, the same methods were applied in all three databases. Dutch primary care physicians record diagnoses and clinical symptoms in the electronic medical records as diagnostic codes using the International Classification of Primary Care (ICPC) coding system. The primary care physicians participating in the JGPN, AHA AMC and ANH Vumc databases are trained in and experienced with using ICPC codes. For the development cohort, COVID-19 suspected patients were identified using the ICPC codes R74 (acute upper respiratory infection), R81 (pneumonia) and R83 (other respiratory infection). At the time, primary care physicians were recommended to use R81 and R83 for indicating COVID-19 suspected and COVID-19 confirmed cases, respectively, for lack of COVID-19 specific ICPCs. Records of patients labelled with ICPC R74 (unspecified acute upper respiratory infection) were manually screened for COVID-19 suspicion in the consultation text by three (primary care) clinical scientists (FSvR, LPTJ, and SvD) and cases of doubt were discussed until agreement was reached. Patients with ICPC R74 yet without having a synonym of or reference to COVID-19 suspicion or related symptoms in the consultation text were excluded from the cohorts. Of all included patients, baseline characteristics (i.e. age; sex; relevant diagnoses; and if available Body Mass Index (BMI), oxygen saturation and C-Reactive Protein (CRP)) were collected. (History) of relevant diseases (i.e. cardiovascular disease, type 2 diabetes mellitus, hypercholesterolemia, hypertension, pulmonary diseases, cancer) were identified using ICPC. Supplementary table 1 contains all ICPC codes that were used for this study.

For the validation cohort, consecutive adult JGPN patients were included from September 1 2020 until April 15 2021, and consecutive adult AHA AMC and ANH Vumc patients were included from June 1 2020 until December 31 2020 (the ‘second wave’ of COVID-19

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infections). At this point in time, the Dutch government made PCR COVID-19 tests freely available and these were recommended for all symptomatic subjects in the Netherlands and for those who were in close contact with a confirmed COVID-19 patient. Moreover, at that time GPs were instructed to uniformly code confirmed cases in their medical records (i.e. using ICPC codes either R83 or R83.03). Thus, only confirmed COVID-19 cases (i.e. with ICPC R83 and R83.03) were included in the cohort for validation of the model.

Outcome

The primary outcome of the prediction model in this study was referral to an emergency ward for intended hospital admission. This was defined as any clinical deterioration resulting in hospital referral by the primary care physician that was recorded as such in the consultation annotation (free text) of the medical record. To capture the full spectrum of complications of COVID-19 resulting in hospitalisation, follow-up lasted 90 days after first consultation for COVID-19 suspected symptoms. To this end, all anonymised consultation texts were manually screened for any emergency hospital referral by (primary care) clinical scientists (FSvR, LPTJ, SvD, and GJG) and cases of doubt were discussed, until consensus was reached.

Candidate predictors

Based upon existing literature from hospitalised COVID-19 patients, we a-priori specified the following candidate predictors prior to the analysis phase: age, sex, the interaction between age and sex, and the number of cardiovascular diseases. The latter was defined as (history of) type 2 diabetes mellitus, heart failure, coronary artery disease, peripheral arterial disease, stroke/transient ischemic attack (TIA), venous thromboembolism (pulmonary embolism or deep venous thrombosis; VTE), and/or atrial fibrillation (AF). Presence of these diseases was based on the corresponding ICPCs (supplementary table 1) at any point before the first COVID-19 consultation in the patient's medical record. The number of cardiovascular

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

diseases were counted per patient and categorised into: no cardiovascular disease, one cardiovascular disease, or two or more cardiovascular diseases.

Sample size

The model development cohort in JGPN and ANH Vumc yielded 5,475 eligible patients with an event fraction of 0.068 (6.8%, n=373) for the primary outcome referral to the hospital. Prior to prediction analysis, the number of allowed candidate predictors was determined. Based on the proposed calculation for sample size in prediction modelling by Riley et al.(12), the maximum number of candidate predictors that can be modelled was 30 with a R^2 Cox-Snell (R^2cs) of 0.0495. As this R^2cs was estimated in absence of a known value, varying R^2cs from 0.0395 to 0.0595 yielded a minimum of 24 and a maximum of 37 candidate predictors, including interaction terms. By using the candidate predictors age, sex, the interaction between age and sex, and the number of cardiovascular diseases with three categories, the sample size of 5,475 eligible patients was deemed sufficient and large enough for model development.

Missing data

Candidate predictors age, sex and cardiovascular disease had no missing data. Missing values in baseline of characteristics measurements of CRP, BMI and oxygen saturation level were not imputed as these determinants were not used further in predictive modelling.

Statistical analyses

Baseline characteristics were summarised using descriptive statistics with categorical variables as numbers with percentages and continuous variables as means with standard deviations or medians with interquartile ranges (IQR). A multivariable logistic regression modelling approach was used. All included patients were entered in a fixed model (full model) with the predictors age, sex, the interaction between age and sex, and the categorical number of cardiovascular diseases and/or diabetes. To determine the incremental value of

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the predictor number of cardiovascular diseases and/or diabetes, a second model (simple model) was fitted using only the predictors age, sex, and the interaction between age and sex. Age was considered as a continuous variable and was studied using a restricted cubic spline function to account for possible non-linearity with 4 knots on the percentiles 0.05, 0.35, 0.65 and 0.95.(13) Number of cardiovascular diseases and/or diabetes was taken as a dummy variable with 'no cardiovascular disease' as reference category. The incremental value of the number of cardiovascular diseases and/or diabetes was studied comparing the full and simple model's c-statistics (ΔAUC), Cox-Snell R^2cs (ΔR^2cs), and a likelihood ratio test. To determine incremental value, an alpha of 0.05 was used for the likelihood ratio test. The full model was internally validated using Harrell's bootstrapping with 100 repetitions to obtain optimism corrected estimates of the c-statistic, and R^2 and slope were calculated. For the temporal external validation of the full model, calibration and discrimination were evaluated: observed and predicted events were calculated and depicted in calibration plots and for discrimination areas under the curve (AUC/c-statistic) were calculated. Other performance measures for temporal external validation that were calculated are: calibration slope, calibration intercept, calibration in the large, R^2cs, and Brier score. Brier scores assess the overall goodness of fit of models, with smaller numbers indicating better performance. Confidence intervals for c-statistics were obtained using the Delong method. For R^2cs and Brier score confidence intervals, bootstrapping was used with repetitions set at 1000. Validation was done in the whole validation dataset as well as separately in the JGPN, ANH VUmc, and AHA AMC validation cohorts. All statistical analyses were performed in R version 4.0.3 with R base, rms, pROC, DescTools, and rmda packages.(14–18)

Ethics

This research was conducted in accordance with Dutch law and the European Union General Data Protection Regulation and according to the principles of the Declaration of Helsinki. The need for formal ethical reviewing was waived by the local medical research ethics committee of the University Medical Center Utrecht, the Netherlands as the research did not require

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

direct patient or physician involvement. The JHN, ANH VUmc and AHA AMC databases may be used for scientific purposes and contain pseudonymised routine care data from the EMRs of all patients of the participating general practices, except those patients who objected to this. Anonymized datasets were extracted from these databases by the respective data managers for the purpose of this research.

Results

Patient characteristics

Patient characteristics of the (clinically suspected and confirmed COVID-19) development cohort are described in table 1. There were 5,475 patients included in this cohort: 2,825 from JGPN and 2,650 from ANH VUmc. In JGPN, 64.6% of the patients were coded as R74, 14.7% as R81, and 21.3% as R83. In ANH VUmc, 71.5% were coded as R74, 10.7% as R81, and 19.2% as R83. Differences in patient characteristics between both datasets in the development cohort were minor. Around a quarter of patients suffered from one or more cardiovascular disease, most often type 2 diabetes and coronary artery disease.

Patient characteristics of the (confirmed COVID-19) validation cohort are described in table 2. From the total of 16,693 patients in the validation cohort 5,420 originated from JGPN, 4,989 from ANH VUmc, and 6,284 from AHA AMC. The patient characteristics in these three datasets were very similar. Around 15-20% suffered from one or more cardiovascular disease, again most often type 2 diabetes and coronary artery disease.

Model development and internal validation

All 5,475 patients in de development cohort were used for model development. 373 patients (6.8%) had the outcome hospital referral. All predefined model regression coefficients of the full and simple models with confidence intervals are shown in table 3. The apparent c-statistic of the full model was 0.693 (95%CI 0.665-0.721) and the internally validated c-

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

statistic was 0.688 (95%CI 0.660-0.716). The apparent c-statistic of the simple model was 0.681 (95%CI 0.653-0.710) and the internally validated c-statistic was 0.680 (95%CI 0.652-0.708). The full and the simple model are compared in table 4. The full model performed significantly better than the simple model (p- value for likelihood ratio test, χ^2 = 19.5, *df* = 2, *p*<0.001). Figure 1 gives a visual representation of the full model showing the predicted risks of hospital referral as a function of (increasing) age, stratified by sex and by the number of cardiovascular diseases and/or diabetes. Overall risks are higher for male patients and increase with age. Furthermore, a higher risk is observed in patients with underlying cardiovascular disease.

Temporal external validation

Predicted risks were overall slightly higher than the observed risk (6.2% versus 4.6%) and the calibration slope was 1.36. Overall discrimination showed an AUC of 0.747 (95%CI 0.729-0.764). Performance measures based on the full validation cohort and stratified by database are shown in table 5. The overall calibration plot and the calibration plots per database separately are shown in supplementary figure 1. The hospital referral prevalence was lower in the validation datasets than in the development datasets (4.7% versus 6.8%).

Discussion

Cardiovascular vulnerability is a predictor of hospital referral in a population of 5,475 consecutive adult patients in primary care with confirmed or clinically suspected COVID-19 in the ‘first wave’ of infections in the Netherlands. This finding was confirmed by temporal validation in a population of 16,693 consecutive confirmed COVID-19 adult primary care patients in the ‘second wave’, exemplifying the robustness of our inferences. On average, in the combined data from the first and second wave (n=22,168 confirmed and clinically suspected primary care COVID-19 patients), 5.1% was referred to the hospital for considering admission. A model including the number of cardiovascular conditions and/or

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

diabetes (0, 1, or ≥2) in addition to age and sex and the interaction between age and sex, showed moderate to good performance and demonstrated consistent and good discrimination and calibration upon temporal external validation. The model showed a c-statistic of 0.747 (95%CI 0.729-0.764).

Although most COVID-19 patients experience a favourable prognosis without the need for referral for hospital care, studies on COVID-19 mainly focussed on those seen in the hospital setting. While on average the overall risk for hospital referral in this adult primary care cohort with COVID-19 was low (5.1%), it is much higher than the hospitalisation rate for other lower respiratory infections in primary care which is estimated at approximately 1% of the adult population affected.(19) In our study, age, sex and the number of concurrent cardiovascular conditions and/or diabetes predicted patients at far greater risk of hospital referral. In fact, for female patients without cardiovascular diseases or diabetes, the risk of hospital referral was well below 10% even in the eldest elderly (aged 80+). Contrastingly, in the presence of cardiovascular diseases and/or diabetes, patients experience higher risks already at younger ages, notably males. For instance, a male patient with two or more underlying cardiovascular diseases and/or diabetes, had a predicted risk of 15% already at the age of around 57 years and this predicted risk will even further increase to above 20% from the age of 80 onwards. This indicates the incremental effect of cardiovascular diseases and/or diabetes in addition to age and sex in predicting the risk for complicated COVID-19 disease trajectories in primary care patients.

Our findings overall confirm those from previous studies done in the hospital setting where age and male sex are important predictors for disease progression towards the endpoints ICU admission or death.(1,20–22) Social, behavioural, comorbidity and biological differences (ACE2 expression, sex-hormones, X-chromosome exposure) between male and female sexes all might contribute to the higher risks of COVID-19 progression observed in males, although probably not all mechanisms have been fully elucidated yet.(23,24) Also, in

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

hospitalised patients, it has been demonstrated that there is an association between cardiovascular disease and COVID-19 complicated disease trajectories, with higher prevalence of cardiovascular disease and diabetes described in those with critical illness. (3–7,25) Our study shows that this prognostically unfavourable effect is already present much earlier on in the COVID-19 disease course, at the start of symptoms in primary care.

This is in line with previous research, where this additive effect of (cardiovascular) comorbidities was also described by the 4C Mortality Score.(26) In this study, the authors demonstrated that the number of comorbidities, importantly including cardiovascular comorbidities, had a more predictive effect than taking only individual co-morbidities in predicting in-hospital mortality of COVID-19 patients.(26) Furthermore, there are two large community-based prediction studies also highlighting the importance of cardiovascular comorbidities as a predictor in the community COVID-19 population. The QCOVID model that was developed in the UK was based on data from primary care and showed a c-statistic >0.9 for the primary outcome time to death from COVID-19. The domain of that study, however, covered the whole general population *regardless* of COVID-19 diagnosis and therefore this can best be interpreted as the risk prediction of getting infected with COVID-19 *and* subsequently having complications from COVID-19. Thus, the aim of this model was to inform UK health policy and support interventions to manage COVID-19 related risks, rather than inform medical decision making during patient consultations in confirmed or clinically suspected COVID-19 cases.(27) With only 0.07% with the outcome death, and thus very low a priori chance, the c-statistic 'misleadingly' moves towards 1.0. Another similar public health based UK study in patients with and without COVID-19 identified determinants that were associated with COVID-19 related death in the OpenSAFELY primary care database by linking primary care records to reported COVID-19 related deaths. It found the most predictive clinical determinants to be increasing age, male sex, type 2 diabetes mellitus, and cardiovascular disease, similar to our findings.(28) While the domain notably differs between patients seeking primary care for COVID-19 symptoms in our study and the adult community

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

as a whole in these studies form the UK, all draw similar conclusions on the increased risk of clinical deterioration in patients with (multiple) cardiovascular disease and/or diabetes.

Strengths and limitations

This research contributes to the evidence-based prognostication of community COVID-19. We were able to use routine primary care databases capturing both the 'first' and 'second' wave of COVID-19 infections in the Netherlands. We used state-of-the-art methodology including external temporal validation to predict clinical deterioration in a patient population that is currently understudied. For full appreciation of our findings, however, some limitations also need to be addressed. First, the model was developed in a dataset with a low event fraction of the outcome hospital referral. Yet the number of hospital referral events did allow us to perform robust multivariable regression techniques. Second, there are limitations to using routine care registry data that could have resulted in misclassification of the study population, predictors and outcome, and most importantly it has the risk of missing values. For example, uncertainty concerning COVID-19 infection status may exist (primarily in the first wave) as COVID-19 PCR test results were not automatically linked to the primary care electronic medical records. However, the model proved its transportability in primary care patients in a different time period with satisfactory calibration and discrimination, during a time window where PCR testing was widely performed. Furthermore, the outcome hospital referral was based upon a rigorous manual extraction of medical records by pairs of researchers, albeit hospitalisation was not formally confirmed based upon linkage to hospital records. Additionally, there are differences between our development and validation population: the patients from the 'first wave' are all symptomatic patients that visited their primary care physician for symptoms suggestive of COVID-19, while the patients from the 'second wave' – due to government recommendation for individuals to get tested even in the circumstance of only mild symptoms – also include more healthy people that just informed their primary care physician of their positive COVID-19 PCR status. This could also explain the lower event fraction in the validation set (4.7% versus 6.8% in the development

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

population). Finally, the incremental value of the number of cardiovascular diseases and diabetes on prognosticating COVID-19 was assessed in different ways; although we did observe a highly significant change in the likelihood ratio test, the delta in c-statistic and R^2cs was only small to modest. Possible reasons for this include the overall low risk of hospital referral in most patients in our cohort, as well as that most patients (80.2%) in fact in our cohort did not suffer from cardiovascular diseases and/or diabetes. It has been widely acknowledged that, notably in such scenario's, a change in e.g. the c-statistic is difficult to achieve.

Clinical implications

The readily availability of the chosen primary care predictors and the clinical applicability may provide great advantages for risk profiling patient with suspected or confirmed COVID-19 in the primary care and community setting. This can have several important clinical and public health implications. First, it may be possible to identify patients that will benefit from closer monitoring and frequent follow-up at home by predicting the risk of clinical deterioration early on in the COVID-19 disease course. By intensified monitoring of higher risk patients, critical illness may be detected earlier, potentially improving prognosis. Second, risk prediction could also support advanced care planning. Informing both patients and physicians on the risk of severe illness, may help in anticipating a more stringent or more lenient management. Last, risk profiling may be used for targeting preventive measures. Vaccination strategies to *prevent* (severe) COVID-19, for instance, may prioritise those with cardiovascular disease and/or diabetes. Additionally, experimental regiments to *treat* symptomatic COVID-19 may be addressed to high-risk patients that may benefit most. Examples include for instance treatment with budesonide or colchicine; both treatment options likely benefit patients most at higher prior probability of having an adverse prognosis.(29,30) Nevertheless, in the end, risk prediction in primary care has to prove its value in daily practice at the background of changing characteristics of this challenging COVID-19 pandemic and influences of vaccination and virus mutations. We however do hope that prognostic studies, like ours, may

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

aid physicians and policy makers by making informed, evidence-based decisions and thereby improve patient outcomes.

Conclusion

In this primary care population-based study, risk of clinical deterioration leading to hospital referral after suspected or confirmed COVID-19 was on average 5.1%. This risk increased with age and was higher in males compared to females. Importantly, patients with concurrent cardiovascular disease and/or diabetes had higher predicted risks. Identifying those at risk for hospital referral could have clinical implications for COVID-19 vaccination and early disease management in primary care.

Funding statement

This work was supported by the Dutch Heart Foundation (grant number 2020T063). The Dutch Heart Foundation had no role in the design of the study, data collection, interpretation of the data, writing of the manuscript and in approving the final manuscript for publication.

Competing interests statements

The authors declare that they have no competing interests.

Author's contribution

All authors contributed to the conception and design of this study. FSVR, LPTJ, PS and SVD collected the data. FSVR, LPTJ and SVD analysed the data and all authors contributed to interpretation of the data. FSVR wrote the first draft and LPTJ, MVS, GJG, FHR, PS and SVD critically revised and approved the final manuscript.

Acknowledgements

The authors would like to thank the data managers from the JGPN, ANH VUmc and AHA AMC databases and all primary care physicians contributing data.

Data sharing statement

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The data used for this study are available from the routine primary care registries JGPN, ANH VUmc and AHA AMC. Restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data will however be available for scientific purposes from the authors upon reasonable request and with permission from the individual registries.

References

1. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369: m1328.
2. Gandhi R, Lynch J, del Rio C. Mild or Moderate Covid-19. *N Engl J Med*. 2020;383(18):1757-1766.
3. Linschoten M, Peters S, van Smeden M, et al. Cardiac complications in patients hospitalised with COVID-19. *Eur Heart J Acute Cardiovasc Care.* 2020;9(8):817-823.
4. Li B, Yang J, Zhao F, et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res in Cardiol.* 2020;109(5):531-538.
5. Bello-Chavolla O, Bahena-López J, Antonio-Villa N, et al. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. *J Clin Endocrinol Metab*. 2020;105(8):2752-2761.
6. Docherty A, Harrison E, Green C, et al. Features of 20□133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ.* 2020;369: m1985.
7. Javanmardi F, Keshavarzi A, Akbari A, et al. Prevalence of underlying diseases in died cases of COVID-19: A systematic review and meta-analysis. *PLoS One.* 2020;15(10):e0241265.
8. Moons K, Altman D, Reitsma J, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Elaboration. *Ann Intern Med.* 2015;162(1):W1-73.

9. Julius General Practitioner's Network (JGPN). Https://juliuscentrum.umcutrecht.nl/nl/julius-huisartsen-netwerk. Accessed 07 May 2021.

10. Academisch Huisartsennetwerk AMC (AHA). Https://huisartsgeneeskunde-amc.maglr.com/amsterdam_umc_huisartsgeneeskunde_amc_jaarbeeld_2018/academisch-huisartsennetwerk-amc. Accessed 07 May 2021.

11. Academisch Netwerk Huisartsgeneeskunde VU (ANH). Https://www.vumc.nl/anh/database-anh.htm. Accessed 07 May 2021.

12. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* 2020;368:1–12.

13. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. [Internet]. 2nd edition. Springer. 2015. Accessed 11 May 2021. Available via: https://link.springer.com/content/pdf/10.1007/978-3-319-19425-7.pdf.

14. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available via: https://www.R-project.org/.

15. Harrell Jr FE. rms: Regression Modeling Strategies. R package version 6.2-0. 2021. Available via: https://CRAN.R-project.org/package=rms.

16. Xavier R, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. MC Bioinforma. 2011;12:77.

17. Andri Signorell et mult. al. DescTools: Tools for descriptive statistics. R package version 0.99.41. 2021.

18. Brown M. rmda: Risk Model Decision Analysis. R package version 1.6. 2018, Available via: https://CRAN.R-project.org/package=rmda.

19. Little P, Stuart B, Smith S, et al. Antibiotic prescription strategies and adverse outcome

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

for uncomplicated lower respiratory tract infections: Prospective cough complication cohort (3C) study. *BMJ.* 2017;357.

20. Liu K, Chen Y, Lin R, et al. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection.* 2020;80(6):e14-e18.

21. Liu W, Tao Z, Wang L, et al. Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. *Chin Med J.* 2020;133(9):1032-1038.

22. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):497–506.

23. Haitao T, Vermunt J, Abeykoon J, et al. COVID-19 and Sex Differences. *Mayo Clin Proc.* 2020;95(10):2189-2203.

24. Gebhard C, Regitz-Zagrosek V, Neuhauser H, et al. Impact of sex and gender on COVID-19 outcomes in Europe. *Biol of Sex Differ.* 2020;11(1).

25. Nishiga M, Wang D, Han Y, et al. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nat Rev Cardiol.* 2020;17(9):543-558.

26. Knight SR, Ho A, Pius R, 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.* 2020;370: m3339.

27. Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *BMJ.* 2020;371 :m3731.

28. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020;584(7821):430–436.

29. Ramakrishnan S, Nicolau DV, Langford B, et al. Inhaled budesonide in the treatment of early COVID-19 (STOIC): a phase 2, open-label, randomised controlled trial. *Lancet Respir Med.* 2021;19(21):1–10.

30. Tardif J, Bouabdallaoui N, Allier PLL, et al. Efficacy of Colchicine in Non-Hospitalized

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Patients with COVID-19. MedRxiv 21250494 [Preprint]. January 27, 2021 [cited 2021 May] https://doi.org/10.1101/2021.01.26.21250494.

Table 1 - Baseline characteristics of the development cohort of 5,475 community patients with confirmed or clinically suspected COVID-19 from the 'first wave' in the Netherlands

Characteristics	**JGPN (n=2,825)**	**ANH VUmc (n=2,650)**
Age in years (IQR)	48 (34-62)	49 (36-62)
Male sex	1144 (40.1%)	1068 (40.3%)
Cardiovascular and metabolic diseases		
Heart failure	109 (3.9%)	80 (3.0%)
Coronary artery disease	230 (8.1%)	189 (7.1%)
Atrial fibrillation	139 (4.9%)	177 (6.7%)
Peripheral arterial disease	49 (1.7%)	37 (1.4%)
History of stroke/TIA	136 (4.8%)	117 (4.4%)
History of VTE	123 (4.4%)	100 (3.8%)
Type 2 diabetes mellitus	300 (10.6%)	262 (9.9%)
0 CVD/Diabetes	2132 (75.5%)	2053 (77.5%)
1 CVD/Diabetes	440 (15.6%)	386 (14.6%)
≥2 CVD/Diabetes	253 (9.0%)	211 (8.0%)
Other comorbidities		
History of any cancer	198 (7.0%)	207 (7.8%)
Hypertension	721 (25.5%)	616 (23.2%)
Hypercholesterolemia	261 (9.2%)	331 (12.5%)
All pulmonary disease	575 (20.4%)	432 (16.3%)
COPD	234 (8.3%)	155 (5.8%)
Asthma	434 (15.4%)	315 (11.9%)
BMI in kg/m^2 (IQR)	27 (24-31) (n=1,091)	27 (24-31) (n=1,072)
Median oxygen saturation in % (IQR)	98 (95-98) (n=191)	98 (96-99) (n=87)
CRP in mg/L (IQR)	6 (2-23) (n=480)	3 (1-9) (n=646)
Hospital referrals	185 (6.5%)	188 (7.1%)

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

BMI = Body Mass Index; CRP = C-reactive protein; CVD = cardiovascular disease; IQR = interquartile rage; TIA = transient ischemic attack; VTE=venous thromboembolism.

Table 2 - Baseline characteristics of the validation cohort of 16,693 community patients with confirmed COVID-19

Characteristics	**JGPN (n=5,420)**	**ANH VUmc (n=4,989)**	**AHA AMC (n=6,284)**
Age in years (IQR)	43 (30-56)	47 (34-59)	49 (36-60)
Male sex	2400 (44.3%)	2121 (42.5%)	2462 (39.2%)
Cardiovascular diseases			
Heart failure	86 (1.6%)	64 (1.3%)	101 (1.6%)
Coronary artery disease	262 (4.8%)	270 (5.4%)	346 (5.5%)
Atrial fibrillation	122 (2.3%)	139 (2.8%)	176 (2.8%)
Peripheral arterial disease	39 (0.7%)	44 (0.9%)	39 (0.6%)
History of stroke/TIA	123 (2.3%)	151 (3.0%)	229 (3.6%)
History of VTE	124 (2.3%)	134 (2.7%)	226 (3.6%)
Type 2 diabetes mellitus	476 (8.8%)	495 (9.9%)	715 (11.4%)
0 CVD/Diabetes	4566 (84.2%)	4059 (81.4%)	4964 (79.0%)
1 CVD/Diabetes	586 (10.8%)	663 (13.3%)	955 (15.2%)
≥2 CVD/Diabetes	268 (4.9%)	267 (5.4%)	365 (4.8%)
Other comorbidities			
History of any cancer	207 (3.8%)	331 (6.6%)	304 (4.8%)
Hypertension	883 (16.3%)	1,009 (20.2%)	1,486 (23.6%)
Hypercholesterolemia	309 (5.7%)	493 (9.9%)	527 (8.4%)
All pulmonary diseases	648 (12.0%)	610 (12.2%)	867 (13.8%)
COPD	174 (3.2%)	186 (3.7%)	184 (2.9%)
Asthma	534 (9.9%)	477 (9.0%)	726 (11.6%)
BMI in kg/m^2 (IQR)	28 (24-32) (n=1,685)	27 (24-32) (n=1,823)	29 (25-33) (n=2,178)
Median oxygen saturation in % (IQR)	98 (96-98) (n=176)	98 (97-99) (n=134)	98 (97-99) (n=75)
CRP in mg/L (IQR)	3 (2-12) (n=545)	2 (1-5) (n=936)	3 (1-10) (n=1,055)
Hospital referrals	219 (4.0%)	187 (3.7%)	357 (5.7%)

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

BMI = Body Mass Index; CRP = C-reactive protein; CVD = cardiovascular disease; IQR = interquartile range; TIA = transient ischemic attack; VTE=venous thromboembolism.

Table 3 - Model development and internal validation using logistic regression

	Full model		**Simple model**	
Predictor	**Regression coefficient**	**95% confidence interval**	**Regression coefficient**	**95% confidence interval**
Intercept	-4.110	-6.338; -2.149	-3.988	-6.228; -2.038
Age	0.026	-0.034; 0.092	0.022	-0.037; 0.088
Age′	0.045	-0.170; 0.251	0.085	-0.127; 0.289
Age′′	-0.172	-0.719; 0.389	-0.268	-0.809; 0.286
Female sex	0.324	-2.358; 3.105	0.214	-2.450; 2.974
Interaction sex and age	-0.019	-0.102; 0.062	-0.016	-0.099; 0.065
Interaction sex and age′	0.032	-0.247; 0.314	0.015	-0.262; 0.296
Interaction sex and age′′	-0.056	-0.805; 0.686	-0.014	-0.758; 0.726
CVD 1	0.564	0.280; 0.844		
CVD ≥2	0.636	0.282; 0.983		
Apparent c-statistic	0.693	0.665; 0.721	0.681	0.653; 0.710
R^2cs	0.030	0.021; 0.039	0.026	0.018; 0.035
Internal validation c-statistic	0.688	0.660; 0.716	0.680	0.652; 0.708
Internal validation R^2	0.070		0.061	
Internal validation slope	0.965		0.957	

Regression coefficients with 95% confidence intervals, c-statistic, and internal validation performance measures of the full model with predictors age, sex and number of cardiovascular diseases and the simple model with only age, sex and the interaction

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

between age and sex as predictors. Age was divided into three subgroups (shown as age, age' and age'') using restricted cubic spline function to account for non-linearity. CVD = cardiovascular disease; R^2cs = R^2 Cox-Snell.

Table 4 – Comparing the full and the simple model

ΔAUC	0.012
ΔR^2cs	0.004
Likelihood ratio, χ^2	19.531, df=2, *p* = 5.740e-05

ΔAUC, ΔR^2cs and Deviance (likelihood ratio) in comparing model with and model without number of cardiovascular diseases (0, 1 or ≥2) modelled. ΔAUC is calculated by subtracting model 2 unadjusted c-statistic from model 1 unadjusted c-statistic. ΔR^2cs is calculated by subtracting model 2 R^2cs from model 1 R^2cs.

Table 5 - performance temporal validation

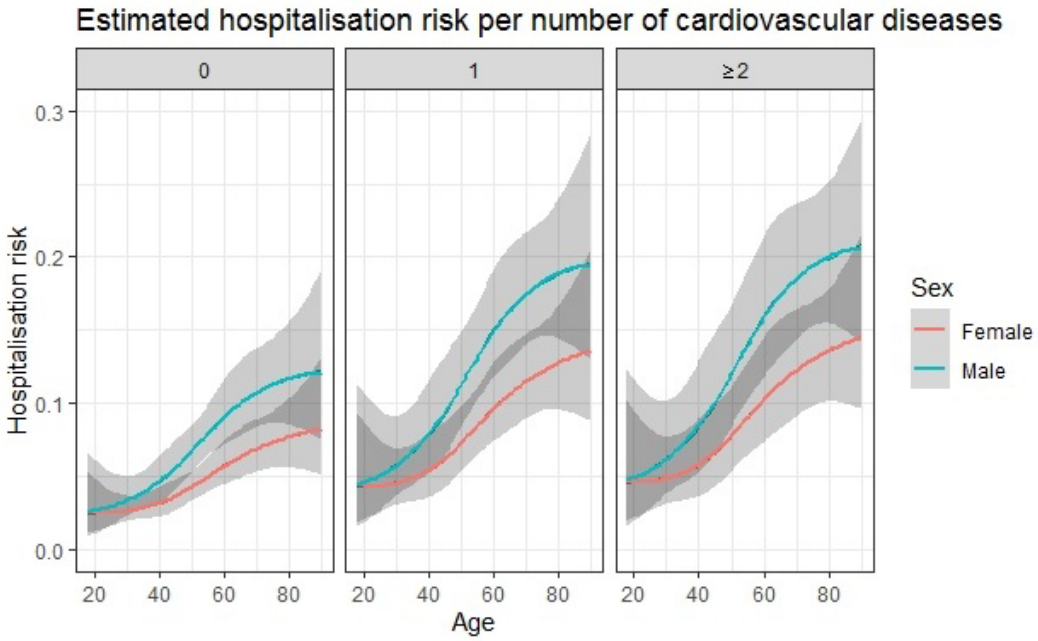
	Total validation population (n=16,693)	**Total patients JGPN (n=5,420)**	**Total patients AHA AMC (n=6,284)**	**Total patients ANH Vumc (n=4,989)**
Area under the curve	0.747 (CI 0.729; 0.764)	0.780 (CI 0.750; 0.810)	0.718 (CI 0.691; 0.745)	0.751 (CI 0.714; 0.789)
Calibration slope	1.358 (CI 1.246; 1.471)	1.510 (CI 1.306;1.718)	1.218 (CI 1.050; 1.387)	1.415 (1.190; 1.645)
Calibration intercept	-0.333 (CI -0.408; -0.261)	-0.396 (CI -0.535; -0.262)	-0.143 (CI -0.253; -0.036)	-0.561 (CI -0.712; -0.416)
Calibration in the large: Mean observed risk Mean predicted risk	 0.046 0.062	 0.040 0.058	 0.057 0.065	 0.038 0.063
R^2cs	0.035 (CI 0.029;	0.040 (CI 0.030;	0.033 (CI 0.024;	0.032 (CI 0.021;

medRxiv preprint doi: https://doi.org/10.1101/2021.05.12.21257075; this version posted June 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	0.041)	0.050)	0.042)	0.042)
Brier score	0.042 (CI 0.039; 0.045)	0.037 (CI 0.033; 0.041)	0.052 (CI 0.047; 0.056)	0.035 (CI 0.031; 0.040)

Temporal validation performance measures with 95% confidence intervals (CI). R^2cs = R^2 Cox-Snell.

Figure 1 – Predicted risk of hospitalisation



Plotted risk of hospital referral in (clinically suspected and confirmed) COVID-19 primary care patients at a certain age and stratified by sex and by the number of cardiovascular diseases and/or diabetes (i.e. 0, 1 or ≥2). Confidence intervals are shown in grey.