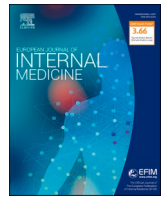




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Original article

## Incident atrial fibrillation and its risk prediction in patients developing COVID-19: A machine learning based algorithm approach

Gregory Y.H. Lip<sup>a,\*</sup>, Ash Genaidy<sup>b,\*\*</sup>, George Tran<sup>c</sup>, Patricia Marroquin<sup>b</sup>, Cara Estes<sup>b</sup><sup>a</sup> Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom<sup>b</sup> Anthem Inc., Indianapolis, IN, USA<sup>c</sup> IngenioRX, Indianapolis, IN, USA

## ARTICLE INFO

## Keywords:

Machine learning  
Main effect analysis  
Cardiovascular/non-cardiovascular multi-morbidity  
COVID-19  
Atrial fibrillation

## ABSTRACT

**Background:** The elderly multi-morbid patient is at high risk of adverse outcomes with COVID-19 complications, and in the general population, the development of incident AF is associated with worse outcomes in such patients. There is therefore the need to identify those patients with COVID-19 who are at highest risk of developing incident AF. We therefore investigated incident AF risks in a large prospective population of elderly patients with/without incident COVID-19 cases and baseline cardiovascular/non-cardiovascular multi-morbidities. We used two approaches: main effect modeling and secondly, a machine-learning (ML) approach, accounting for the complex dynamic relationships among comorbidity variables.

**Methods:** We studied a prospective elderly US cohort of 280,592 patients from medical databases in an 8-month investigation of with/without newly incident COVID19 cases. Incident AF outcomes were examined in relationship to diverse multi-morbid conditions, COVID-19 status and demographic variables, with ML accounting for the dynamic nature of changing multimorbidity risk factors.

**Results:** Multi-morbidity contributed to the onset of confirmed COVID-19 cases with cognitive impairment (OR 1.69; 95%CI 1.52–1.88), anemia (OR 1.41; 95%CI 1.32–1.50), diabetes mellitus (OR 1.35; 95%CI 1.27–1.44) and vascular disease (OR 1.30; 95%CI 1.21–1.39) having the highest associations. A main effect model (C-index value 0.718) showed that COVID-19 had the highest association with incident AF cases (OR 3.12; 95%CI 2.61–3.71), followed by congestive heart failure (1.72; 95%CI 1.50–1.96), then coronary artery disease (OR 1.43; 95%CI 1.27–1.60) and valvular disease (1.42; 95%CI 1.26–1.60). The ML algorithm demonstrated improved discriminatory validity incrementally over the statistical main effect model (training: C-index 0.729, 95%CI 0.718–0.740; validation: C-index 0.704, 95%CI 0.687–0.72). Calibration of the ML based formulation was satisfactory and better than the main-effect model. Decision curve analysis demonstrated that the clinical utility for the ML based formulation was better than the ‘treat all’ strategy and the main effect model.

**Conclusion:** COVID-19 status has major implications for incident AF in a cohort with diverse cardiovascular/non-cardiovascular multi-morbidities. Our ML approach accounting for dynamic multimorbidity changes had good prediction for new onset AF amongst incident COVID19 cases.

## 1. Introduction

Multi-morbidity is associated with adverse health outcomes and healthcare costs, especially among the elderly [1,2]. Preliminary evidence suggests that multi-morbidity is associated with confirmed COVID-19 infections notably among the elderly with several co-morbidities [3–5]. Furthermore, it has been hypothesized that COVID-19 is associated with incident atrial fibrillation (AF) but large

studies are lacking to test this hypothesis [6–8].

In light of the above, the elderly population is usually closely scrutinized due to (a) the staggering healthcare costs reaching in many countries above 70% of the national healthcare expenditures, and (b) the need to improve the quality of integrated care because of the presence of multi-morbid conditions. The elderly multi-morbid patient is at higher risk of adverse outcomes with COVID-19 complications [3–5], and in the general population, the development of incident AF is

\* Correspondence author at: Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool L7 8TX, United Kingdom.

\*\* Co-Correspondence author at: Anthem Inc., 11884 Quarterhorse Court, Cincinnati, OH 45249, USA.

E-mail addresses: [gregory.lip@liverpool.ac.uk](mailto:gregory.lip@liverpool.ac.uk) (G.Y.H. Lip), [ashraf.genaidy@anthem.com](mailto:ashraf.genaidy@anthem.com) (A. Genaidy).

<https://doi.org/10.1016/j.ejim.2021.04.023>

Received 27 March 2021; Received in revised form 20 April 2021; Accepted 22 April 2021

Available online 14 May 2021

0953-6205/© 2021 European Federation of Internal Medicine. Published by Elsevier B.V. All rights reserved.

associated with worse outcomes in such patients [9]. There is therefore the need to identify those patients with COVID-19 who are at highest risk of developing incident AF.

We therefore investigated incident AF risks in a large prospective population of elderly patients with/without incident COVID-19 cases and baseline characteristics consisting of diverse cardiovascular/non-cardiovascular multi-morbidities. Our aims were to show how multi-morbidity contributes to the onset of confirmed COVID-19 cases, and secondly, how COVID-19 had the highest association with *incident AF cases*. We used two approaches: main effect modeling as well as a machine-learning (ML) approach, accounting for the complex dynamic relationships among co-morbidity variables.

## 2. Methods

### 2.1. Cohort detailed definition and data sources

We studied a large prospective US Medicare population over an 8-month period starting April 1 2020 to determine their effects on potential new COVID-19 cases and subsequently, the incidence of new onset AF. The study population comprised of the Medicare health plan was a predominantly elderly population (i.e.  $\geq 65$  years) and individuals with disability including those in the 18–65 age span. It is financed by the US government and managed by an independent healthcare organization. In essence, the Medicare health plan consisted of Medicare Advantage and Medicare/Medicaid Plan participants and were drawn from different geographical areas across the US continent.

The co-morbid history was gathered for a two-year period prior to the start of the study and consisted of common cardiovascular and non-cardiovascular multi-morbidities. Only patients without a history of AF and COVID-19 prior to April 1, 2020 were included in the study. Furthermore, pharmacy claims were analyzed to ensure that patients do not have a history for oral anticoagulants (warfarin, direct oral anticoagulants (DOACs)) and rhythm control medications (i.e., amiodarone, disopyramide, dofetilide, dronedarone, flecainide, mexiletine, procainamide, propafenone, quinidine gluconate, quinidine sulfate) typically used in AF, with previously noted exceptions for AF cases [10].

The study cohort was gathered from medical claims databases during the April 1 2018 – Nov 30 2020 time window based on primary and secondary ICD10 codes. Each participant had to contribute at least 32 months of medical and pharmacy coverage during the study and records in the medical database (i.e., eight months for the prospective cohort investigation and 24 months of prior medical history for non-incidence AF/COVID-19 conditions). IRB approval was not required for the extraction of data from the claim databases; however, compliance with US privacy laws and Company governance is required for use of data.

### 2.2. Parameter identification and definition

At baseline (Day 0, upon entry into the study on April 1 2020), subjects *without* any history of COVID-19 and AF conditions for two years were enrolled over an 8-month period with two prospective cohorts defined as follows: (a) cohort 1 – incident (new) COVID-19 cases and followed up for incidence of AF (at least one day after occurrence of a COVID-19 case); (2) cohort 2 – non-COVID-19 cases with or without developing incident AF cases (i.e., without subsequent COVID-19 cases in later days or simultaneous COVID-19 cases on the same day).

The co-morbid history was identified on the basis of ICD10 codes (please see supplemental table S1 for ICD 10 codes), including: congestive heart failure, hypertension, diabetes mellitus, stroke, vascular disease, valvular disease, coronary artery disease, sleep apnea, chronic kidney disease, chronic obstructive pulmonary disease/bronchiectasis, major bleeding, cognitive impairment, lipid disorders, liver disease, anemia, depression, spondylosis/intervertebral discs, and osteoarthritis.

An incident COVID-19 case was determined as the first case upon

entry into the study using the US CDC code of ‘U071’. Confirmed cases of COVID-19 infections via the use of the ICD-10 code ‘‘U071’’ was recommended by the US Centers for Disease Control as of April 1 2020. In this respect, a confirmed diagnosis of COVID-19 was issued as documented by the provider, documentation of a positive COVID-19 test result, or a presumptive positive COVID-19 test result. Furthermore, ‘‘confirmation’’ does not require documentation of the type of test performed, or disease severity; the provider’s documentation that the individual has COVID-19 was sufficient for the diagnosis.

An incident AF outcome was defined as occurring by at least 1 day after the development of a COVID-19 condition or upon entry into the study in the absence of any developed COVID-19 case. It was defined in terms of ICD 10 codes as reported in supplemental table S1.

The study population should not have had any history of AF or COVID-19 during the 2-year baseline period as defined in terms of ICD10 codes (see supplemental table S1) and anticoagulant/rhythm control medications for AF (see supplemental tables S2 and S3). Two demographic variables were utilized in this investigation, namely, gender and age. Age was defined as either a continuous variable or in 5 categories (18–45, 45–55, 55–65, 65–75, 75–90 years).

### 2.3. Quantitative analyses

The quantitative analysis consisted of descriptive statistics and model prediction using inferential statistics and machine learning computations. The descriptive and inferential analyses were performed using the Statistical Analysis Software (SAS) Enterprise and the ML computations were conducted using the SAS Enterprise Miner.

The descriptive analyses included identification of member counts (percent) for demographic parameters, co-morbid history, and incident COVID-19 and AF conditions (with the exception of mean (SD) for age as a continuous variable). The outcome (i.e., COVID-19 or AF) and input (i.e., co-morbid history) variables had binary representations.

### 2.4. Statistical analysis

Statistical analyses were conducted using main effects with COVID-19 or AF as an outcome, with logistic regression modeling using the SAS Enterprise software. Prediction modeling was pursued using the Enterprise SAS Miner software for complex relationships between AF as a binary outcome and comorbid history / COVID-19 status / demographic variables. All ML based modeling accounted for *dynamic changes* in risk including newly acquired risk factors, hence consisting of complex interactions among the comorbid condition history as well as incident conditions such as COVID-19 conditions. The ML based logistic regression algorithm included main effects, interaction terms and polynomial effects, with the model selection based on the stepwise procedures. Several polynomial terms were included in the ML formulation.

Model validation was based on calibration, discrimination, and clinical utility. Each model was trained on 67% of the data, with the remaining 33% data used for external validation. In this respect, the development and validation samples were extracted at random. Discriminant validity was assessed using C-indices (area under the curve) for both the development and validation samples, separately. In addition, clinical utility was assessed using decision curve analysis (DCA).

## 3. Results

### 3.1. Characteristics of cohort included in the study

The total Medicare population was 364,348 persons with 347,976 individuals contributing data to the medical databases. The final cohort included in this study consisted of 280,592 persons (mean age (SD) 72.5 (9.9) years; 58.8% female) (Table 1). There was a diversified multi-

**Table 1**

Baseline characteristics for total cohort. Values are numbers (%) unless stated otherwise/

Baseline characteristic	Total Cohort
Age group (years)	
18-45	6481 (2.3)
45-55	10094 (3.6)
55-65	27242 (9.7)
65-75	120854 (43.1)
75-90	115921 (41.3)
Age (years), mean (SD)	72.5 (9.9)
Gender	
Males	115629 (41.2)
Females	164963 (58.8)
Total	280592 (100.0)
Co-morbid history	
Congestive heart failure	18184 (6.5)
Hypertension	195769 (69.8)
Diabetes mellitus	56929 (20.3)
Stroke	15976 (5.7)
Vascular disease	33077 (11.8)
Valvular disease	28318 (10.1)
Coronary artery disease	45812 (16.3)
Chronic sleep apnea	6184 (2.2)
Chronic kidney disease	34593 (12.3)
Chronic pulmonary obstructive disease /bronchiectasis	52339 (18.7)
Major bleeding	20531 (7.3)
Cognitive impairment	9999 (3.6)
Lipid disorders	190681 (68.0)
Liver disease	27923 (10.0)
Anemia	51021 (18.2)
Depression	40121 (14.3)
Spondylosis and intervertebral discs	105504 (37.6)
Osteoarthritis	86530 (30.8)

morbid history with hypertension and lipid disorders having the highest prevalence, >65%, followed by spondylosis/ intervertebral disc and osteoarthritis, with prevalence >30%, then prevalence in the range of 15–20% for diabetes, coronary artery disease, anemia, and COPD. Other conditions had lower prevalence rates as shown in Table 1.

The incidence of AF in the new COVID-19 cases was 2.5% compared to 0.6% in the non-COVID-19 cases. The crude incidence frequency ratio for incident AF cases was 3.87 in COVID-19 cases. As shown below, this ratio drops to 3.0 after accounting for confounders.

### 3.2. Main effect modeling

With COVID-19 as an outcome variable, the strongest associations ( $p < 0.0001$ ) were found for congestive heart failure, hypertension, diabetes, stroke, vascular disease, chronic obstructive pulmonary disease/ bronchiectasis, cognitive impairment, anemia, depression, and spondylosis / intervertebral discs (Table 2). The highest odds ratios were obtained for cognitive impairment (OR 1.69 95%CI 1.52-1.88) and anemia (1.41 95%CI 1.32-1.50), both non-cardiovascular morbidities.

With (new onset) AF as an outcome variable, main effect modeling demonstrated that the strongest associations ( $p < 0.0001$ ) were obtained with reference to COVID-19 status cases (OR 3.12 95%CI 2.61-3.710), followed by congestive heart failure (1.72 95%CI 1.50-1.96), then coronary artery disease (OR 1.43 95%CI 1.27-1.60) and valvular disease (1.42 95%CI 1.26-1.60). This was followed by chronic kidney disease, chronic obstructive pulmonary disease / bronchiectasis, anemia, lipid disorders, gender and age as a continuous variable. Females had lower risk relative to males for incident AF (OR 0.67 95%CI 0.61-0.74) (Table 3).

**Table 2**

Effects of baseline characteristics and demographic variables on COVID-19 outcomes using main effect model.

Effect	Levels	Point estimate	95% confidence interval		Pr > ChiSq
			Lower limit	Upper limit	
Congestive heart failure	(1 vs 0)	1.21	1.11	1.33	<.0001
Hypertension	(1 vs 0)	1.17	1.09	1.25	<.0001
Diabetes mellitus	(1 vs 0)	1.35	1.27	1.44	<.0001
Stroke	(1 vs 0)	1.22	1.10	1.34	<.0001
Vascular disease	(1 vs 0)	1.30	1.21	1.39	<.0001
Chronic kidney disease	(1 vs 0)	1.11	1.03	1.20	0.0049
Chronic pulmonary obstructive disease /bronchiectasis	(1 vs 0)	1.22	1.14	1.30	<.0001
Major bleeding	(1 vs 0)	1.13	1.03	1.23	0.0093
Cognitive impairment	(1 vs 0)	1.69	1.52	1.88	<.0001
Liver disease	(1 vs 0)	1.12	1.04	1.22	0.0038
Anemia	(1 vs 0)	1.41	1.32	1.50	<.0001
Depression	(1 vs 0)	1.25	1.17	1.35	<.0001
Spondylosis and intervertebral discs	(1 vs 0)	1.25	1.18	1.32	<.0001
Osteoarthritis	(1 vs 0)	1.09	1.03	1.16	0.0032
Age group	75-90	0.96	0.93	0.99	0.004
	OR 65-75				
	OR 55-65				
	OR 45-55				
	VS 18-45				

#### Note

1 - presence of condition  
0 - absence of conditions  
age groups - in years

### 3.3. Machine learning algorithm

For the training data, the c index value for the ML-based logistic regression algorithm was 0.729 (95%CI 0.718-0.740) and was incrementally higher than that obtained for the main effect model (C-index 0.718). Similar results were obtained for the externally validation cohort (0.704, 95%CI 687-0.721).

Table 4 depicts the complex relationships between the incident AF outcome and model features in terms of main effect, interactions and polynomial effects. The top three independent effects of co-morbid conditions in the main effect model (Table 3) were COVID-19 status, congestive heart failure and coronary artery disease, which were the only independent effects found in the ML based logistic regression formulation (Table 4). COVID-19 status, congestive heart failure and coronary artery disease also had interaction effects with other co-morbid conditions or demographic variables. Age was significant both as a categorical variable in interaction terms and as a continuous variable in quadratic terms.

In Fig. 1, the clinical utility of main effect model and ML based logistic regression algorithm had better clinical utility in terms of net benefit than the two treatment strategies (i.e., treat all or none). In this respect, the “treat all” (i.e., provide all patients with AF-related treatments) or “treat none” (i.e., provide no treatments to any patients) interventions are two default strategies where patients are managed without the use of a model. With the above in mind, the net benefit measure reflects the number of true positives after being adjusted for any potential false positives. A model is usually useful if it has a net benefit that is better than the default strategies. As evident from Fig. 1, the developed models provide better net benefit values than the “treat

**Table 3**

Results of main effect model for incident atrial fibrillation outcome using baseline characteristics and covid status

Effect	Levels	Point estimate	95% confidence interval		Pr > ChiSq
			Lower limit	Upper limit	
Covid-19 status	(1 vs 0)	3.12	2.61	3.74	<.0001
Congestive heart failure	(1 vs 0)	1.72	1.50	1.96	<.0001
Hypertension	(1 vs 0)	1.26	1.11	1.43	0.0005
Stroke	(1 vs 0)	1.23	1.06	1.43	0.0076
Vascular disease	(1 vs 0)	1.21	1.07	1.37	0.0018
Valvular disease	(1 vs 0)	1.42	1.26	1.60	<.0001
Coronary artery disease	(1 vs 0)	1.43	1.27	1.60	<.0001
Chronic kidney disease	(1 vs 0)	1.28	1.14	1.44	<.0001
Chronic pulmonary obstructive disease /bronchictasis	(1 vs 0)	1.29	1.16	1.43	<.0001
Anemia	(1 vs 0)	1.27	1.14	1.41	<.0001
Depression	(1 vs 0)	1.16	1.02	1.32	0.0228
Lipid disorders	(1 vs 0)	0.79	0.71	0.89	<.0001
Spondylosis and intervertebral discs	(1 vs 0)	1.15	1.04	1.26	0.0045
Gender	(1 vs 0)	0.67	0.61	0.74	<.0001
Age	Interval in years	1.05	1.05	1.06	<.0001

**Note**  
1 - presence of condition  
0 - absence of conditions

all" strategy.

Above the probability threshold of 1.0%, the ML formulation provided better clinical utility than the main effect model. At a probability threshold of 1.5%, the net true positive AF events were equal to 85.5 events for the ML based logistic regression and higher than those for the main effect model (58.9 net events). In addition, the sensitivity and specificity were equal to 29.8% and 91.2%, respectively for the ML algorithm.

### 3.4. Model calibration

From calibration standpoint, the main effect model and machine learning algorithm (fig S1a and S1b) were well calibrated in the lower segment of predicted probability (0–5%). Beyond this probability range, the main effect model did not seem well calibrated due to perhaps the absence of adequate number of parameters (in other words, misspecification error in the 5% to 100% probability range) resulting in risk over-estimation. The ML based algorithm overestimated the risk beyond 5% (beyond the range of operation) but had better calibration than that obtained for the main effect model.

## 4. Discussion

In this large analysis of elderly patients free of AF and COVID-19 at baseline, but followed up for new COVID-19 cases, we developed a ML based logistic regression algorithm for predicting incident AF accounting for *dynamic changes* in risk including newly acquired risk factors. Second, DCA showed the ML based logistic regression algorithm had better clinical utility in terms of net benefit than the two treatment strategies (i.e., treat all or none).

The ML analyses demonstrated that COVID-19 status had the strongest independent association with incident AF relative to the traditional cardiovascular co-morbidities including congestive heart failure and coronary artery disease. This was also evident in the main effect

**Table 4**

Results of machine learning based algorithm for incident atrial fibrillation outcome

Effect	Levels	DF	Chi-Square	Pr>ChiSq	Coefficient Estimate
Intercept		1	947.09	<0.0001	-5.560000
CAD	1	1	8.55	0.0035	0.360900
CHF	1	1	41.12	<0.0001	0.357200
covid_stts	1	1	43.21	<0.0001	0.420100
ANEMIA*HYP	1 1	1	7.36	0.0067	0.115300
ANEMIA*MBldg	1 1	1	6.49	0.0108	0.113000
ANEMIA*covid_stts	1 1	1	10.27	0.0014	-0.150100
CAD*LIPDIS	1 1	1	5.82	0.0158	0.016800
CAD*age_group		4	12.79	0.0123	
	<b>1 4</b>	<b>1</b>	<b>7.87</b>	<b>0.005</b>	<b>-0.343200</b>
	<b>1 3</b>	<b>1</b>	<b>1.79</b>	<b>0.1815</b>	<b>-0.166700</b>
	<b>1 2</b>	<b>1</b>	<b>0.48</b>	<b>0.4891</b>	<b>-0.107400</b>
	<b>1 1</b>	<b>1</b>	<b>0.04</b>	<b>0.8325</b>	<b>-0.047800</b>
CHF*HYP	1 1	1	15.82	<0.0001	-0.183300
CHF*VALVD	1 1	1	5.78	0.0162	-0.093500
CKD*age_group		4	28.38	<0.0001	
	<b>1 4</b>	<b>1</b>	<b>1.91</b>	<b>0.1670</b>	<b>0.058600</b>
	<b>1 3</b>	<b>1</b>	<b>18.82</b>	<b>&lt;0.0001</b>	<b>0.256900</b>
	<b>1 2</b>	<b>1</b>	<b>3.37</b>	<b>0.0664</b>	<b>0.236100</b>
	<b>1 1</b>	<b>1</b>	<b>5.56</b>	<b>0.0184</b>	<b>-0.537000</b>
COGI*LIPDIS	1 1	1	9.72	0.0018	0.138500
COGI*SPOND	1 1	1	7.16	0.0074	-0.077900
COPD*MBldg	1 1	1	4.31	0.0378	-0.085700
COPD*covid_stts	1 1	1	4.71	0.0300	-0.092700
COPD*gndr	1 1	1	7.15	0.0075	0.081200
DEP*VALVD	1 1	1	8.55	0.0034	-0.105700
LIPDIS*MBldg	1 1	1	11.75	0.0006	-0.148600
MBldg*gndr	1 1	1	30.71	<0.0001	0.169800
STROKE*VALVD	1 1	1	7.87	0.0050	-0.110600
VD*covid_stts	1 1	1	11.71	0.0006	-0.123700
age*age		1	124.95	<0.0001	0.000304

**Note:**

CAD - coronary artery disease  
CHF - congestive heart failure  
COVID\_STTS - presence/absence of covid-19 condition  
HYP - hypertension  
CKD - chronic kidney disease  
COGI - cognitive impairment  
LIPDIS - lipid disorders  
SPOND - spondylosis and intervertebral discs  
COPD - chronic pulmonary obstructive disease /bronchictasis  
DEP - depression  
GNDR - gender (female=1, male=0)  
VALVD - valvular disease  
MBLDG - major bleeding  
VD - vascular disease  
Age - continuous variable in years  
Age group in years - 0 for 18 to 45, 1 for 45 to 55, 2 for 55 to 65, 3 for 65 60 75, 4 for 75 to 90  
Each of the comorbid conditions is labeled as 1 for presence of condition and 0 for absence

analyses. In the absence of COVID-19, the presence of congestive heart failure and coronary disease are independent cardiovascular risk factors leading to incident AF conditions; however, the presence of incident COVID-19 infection changed the importance of classic cardiovascular risk factors feeding into the development of new onset AF. There were also significant and dynamic interactions between the presence of incident COVID-19 infections and co-morbid history including anemia, chronic obstructive pulmonary disease and vascular disease.

In the main effect model, cardiovascular and non-cardiovascular

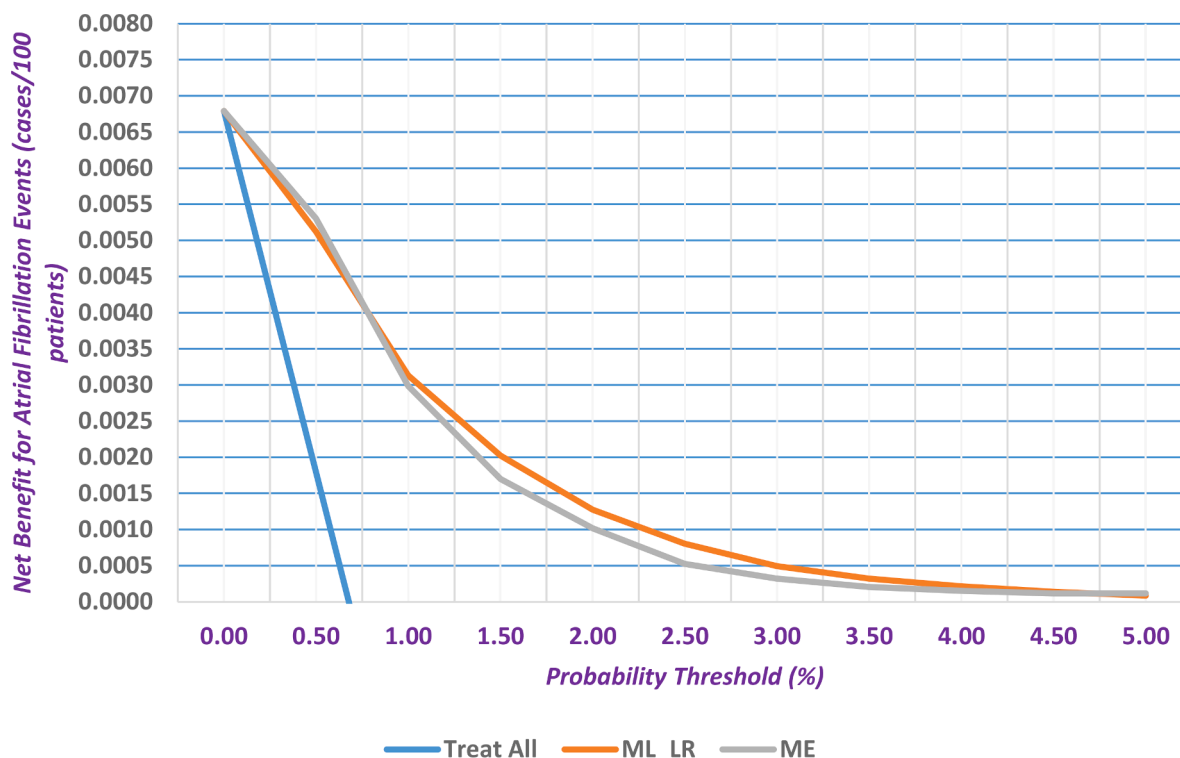


Fig. 1. Decision curve analysis for main effect model (ME), machine learning based logistic regression formulation (ML\_LR) and treat all strategy.

multi-morbidities had significant roles in the spectrum of AF disease complexity in addition to the emergent COVID-19 as a risk factor. As expected, multi-morbidity played an important role in increasing the risk of COVID-19 infection [3–5]. Demographic variables continued to demonstrate their importance as risk factors associated with the incidence of AF. Age implicated its effects in non-linear terms using both (a) quadratic effects when modelling age as a continuous variables; and (b) interactive terms (with coronary artery disease and chronic kidney disease) upon the use of age as a categorical variable. Gender showed its influence in interactive terms with the co-morbid history (chronic obstructive pulmonary disease, major bleeding).

Our findings are important given the worse prognosis amongst COVID-19 patients with AF, with a higher risk of thrombosis and mortality when compared to AF patients without COVID-19 patients [11]. Our ML prediction could be incorporated into telehealth approaches to monitor patients following their COVID-19 diagnosis, for the onset of incident AF [12]; an important consideration given that many COVID-19 infections could be asymptomatic [13]. Given the increasing focus on integrated care management of patients with AF [14], and the need for thromboprophylaxis in such patients [15], novel ML approaches could facilitate structured management and follow-up, especially since risk profiles change in a dynamic manner over time [16–18]. Such a structured approach to holistic AF care, including proactive risk evaluation, has been shown to be associated with improved clinical outcomes, especially with a reduction in hospitalisations and bleeding events [19–21].

#### 4.1. Limitations

Our study is limited by its observational design and shorter follow-up period. As with observational cohorts the possibility of residual confounding remains. One should keep in mind the potential bias emerging due to healthcare services concentrating on the treatment of COVID-19 cases and possibly leading to the cancellation of routine services, such as office visits for established chronic conditions. This extent of possible bias is not known but should be kept in mind. Additional research would

be required to assess the implications of these results on integrated care management for such AF patients [8]. Finally, the results of this study are only applicable for the incident AF cases for which the prior history of anticoagulants and heart rhythm control were applied as exclusion criteria for this purpose. Therefore, the effects of prior use of anticoagulants cannot be ascertained from this study.

## 5. Conclusions

COVID-19 status had the strongest independent association with incident AF, compared to the traditional cardiovascular co-morbidities including congestive heart failure and coronary artery disease. An ML approach elicited the complex dynamic relations which lead to the incidence of AF and in general showed better performance than the statistical main-effect model in terms of discriminatory validity, clinical utility as well as model calibration.

#### Data availability

Data are available as presented in the paper. According to US laws and corporate agreements, our own approvals to use the Anthem and Ingenio-Rx data sources for the current study do not allow us to distribute or make patient data directly available to other parties.

#### Declaration of Competing Interest

The authors report no conflicts of interest in this work.

#### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejim.2021.04.023.

## References

- [1] Stirland LE, González-Saavedra L, Mullin DS, Ritchie CW, Muniz-Terrera G, Russ TC. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. *BMJ (Clin Res Ed)* 2020;368:m160.
- [2] Soley-Bori M, Ashworth M, Bisquera A, et al. Impact of multimorbidity on healthcare costs and utilisation: a systematic review of the UK literature. *Br J Gener Pract* 2020.
- [3] Mair FS, Foster HM, Nicholl BI. Multimorbidity and the COVID-19 pandemic - an urgent call to action. *J Comorb* 2020;10:2235042×20961676.
- [4] Maddaloni E, D'Onofrio L, Alessandri F, et al. Cardiometabolic multimorbidity is associated with a worse Covid-19 prognosis than individual cardiometabolic risk factors: a multicentre retrospective study (CoViDiab II). *Cardiovas Diabetol* 2020; 19(1):164.
- [5] Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: a federated electronic medical record analysis. *PLoS Med* 2020;17(9):e1003321.
- [6] Wang Y, Wang Z, Tse G, et al. Cardiac arrhythmias in patients with COVID-19. *J Arrhythm* 2020;36(5):827–36.
- [7] Stone E, Kiat H, McLachlan CS. Atrial fibrillation in COVID-19: a review of possible mechanisms. *FASEB J* 2020;34(9):11347–54.
- [8] Angeli F, Spanevello A, De Ponti R, et al. Electrocardiographic features of patients with COVID-19 pneumonia. *Eur J Intern Med* 2020;78:101–6.
- [9] Proietti M, Marzona I, Vannini T, et al. Long-term relationship between atrial fibrillation, multimorbidity and oral anticoagulant drug use. *Mayo Clin Proc* 2019; 94(12):2427–36.
- [10] Lip GY, Genaidy A, Tran G, Marroquin P, Estes C, Sloop S. Improving stroke risk prediction in the general population: common clinical rules, a new multimorbid index and machine learning based algorithms. *Thromb Haemost* 2021.
- [11] Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions* 5.1.0. March 2011 [updated], <http://handbook.cochrane.org>.
- [12] Hermans ANL, van der Velden RMJ, Gawalko M, et al. On-demand mobile health infrastructures to allow comprehensive remote atrial fibrillation and risk factor management through teleconsultation. *Clin Cardiol* 2020;43(11):1232–9.
- [13] Milani GP, Montomoli E, Bollati V, et al. SARS-CoV-2 infection among asymptomatic homebound subjects in Milan, Italy. *Eur J Intern Med* 2020;78: 161–3.
- [14] Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2020.
- [15] 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. *European J Intern Med* 2020;77:158–60.
- [16] Fauchier L, Bodin A, Bisson A, et al. Incident comorbidities, aging and the risk of stroke in 608,108 patients with atrial fibrillation: a nationwide analysis. *J Clin Med* 2020;9(4).
- [17] Chao TF, Lip GYH, Lin YJ, et al. Incident risk factors and major bleeding in patients with atrial fibrillation treated with oral anticoagulants: a comparison of baseline, follow-up and delta HAS-BLED scores with an approach focused on modifiable bleeding risk factors. *Thromb Haemost* 2018;118(4):768–77.
- [18] Yoon M, Yang PS, Jang E, et al. Dynamic changes of CHA2DS2-VASc score and the risk of ischaemic stroke in asian patients with atrial fibrillation: a nationwide cohort study. *Thromb Haemost* 2018;118(7):1296–304.
- [19] Guo Y, Guo J, Shi X, et al. Mobile health technology-supported atrial fibrillation screening and integrated care: a report from the mAFA-II trial long-term extension cohort. *Eur J Intern Med* 2020;82:105–11.
- [20] Guo Y, Lane DA, Wang L, et al. Mobile health technology to improve care for patients with atrial fibrillation. *J Am Coll Cardiol* 2020;75(13):1523–34.
- [21] Guo Y, Lane DA, Chen Y, Lip GYH. m AFAII. Regular bleeding risk assessment associated with reduction in bleeding outcomes: the mAFA-II randomized trial. *Am J Med* 2020;133(10):1195–202. e1192.