

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

Ash Genaidy<sup>1</sup>

<sup>1</sup>Affiliation not available

February 19, 2021

## 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. 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 complex dynamic relationships. **Methods** We studied a prospective elderly US cohort of 280592 patients from medical databases in a 8-month investigation of new 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 2.61-3.710, 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). 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 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 approach accounting for dynamic multimorbidity changes had good prediction for incident AF amongst incident COVID19 cases.

## INTRODUCTION

Multi-morbidity is associated with adverse health outcomes and healthcare costs, especially among the elderly<sup>1,2</sup>. Preliminary evidence suggests that multi-morbidity is associated with confirmed COVID-19 infections notably among the elderly with several co-morbidities<sup>3-5</sup>. Furthermore, it has been hypothesized that COVID-19 is associated with incident atrial fibrillation (AF) but large studies are lacking to test this hypothesis<sup>6,7</sup>.

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 high risk of adverse outcomes with COVID-19 complications<sup>3-5</sup>, and in the general population, the development of incident AF is associated with worse outcomes in such patients<sup>8</sup>. 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. 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.

## Methods

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 co-morbid history was gathered for two-year period prior to the start of the study and consisted of common cardiovascular and non-cardiovascular multi-morbidities. The Medicare population consisted of individuals aged 65 years and above and a smaller percentage with disability in the equal to or above the age of 18 years.

### *Cohort definition and data sources*

The study population represented the Medicare health plan (elderly population > 65 years and individuals with disability including those below 65 years) financed by the US government and managed by an independent healthcare organization. The Medicare health plan consists of Medicare Advantage and Medicare/Medicaid Plan participants, and were drawn from different geographical areas across the US continent. 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. 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.

### *Parameter identification and definition*

At baseline (Day 0), subjects *without* any history of COVID-19 and AF conditions for two years were enrolled over an 8-month period with incident (new) COVID-19 conditions occurring first and subsequently followed by incident AF outcomes by at least one day. Individuals with co-existing COVID-19 and AF cases during the same reporting day and any new AF conditions prior to incident COVID-19 cases were excluded from this study. 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.

The first incidence of COVID-19 was determined as the first case upon entry into the study using the US CDC code of ‘U072’ (designated for confirmed cases starting April 1 2020). An incident AF outcome was defined as occurring as at least 1 day after the development of a COVID-19 conditions. It was defined in terms of ICD 10 codes as reported in supplemental table S1. The incident AF cases should have had no history of conditions during the 2-year baseline period as defined in terms of ICD10 codes (see supplemental table S1) and anticoagulant/rhythm control medications (see supplemental tables S2 and S3). Two demographic variables were utilized in this investigation, namely, gender and age. Age was defined as a continuous variable or in 5 categories (18-45, 45-55, 55-65, 65-75, 75-90 years). It was calculated in reference to the first month of enrolment upon entry into the study.

### *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.

### *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 method. 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-indexes (area under the curve) for both the development and validation samples, separately. In addition, clinical utility was assessed using decision curve analysis (DCA).

## **Results**

### *Characteristics of cohort included in the study*

The total Medicare population was 364348 persons with 347976 individuals contributing data to the medical databases. The final cohort included in this study consisted of 280592 persons (mean age (SD) 72.5 (9.9) years; 58.8% female) (table 1) after applying the following exclusions: (i) AF patients in the 2-year baseline period as defined by IC10 codes and anticoagulant/rhythm control medications; and (ii) AF conditions prior to the occurrence of incident COVID-19 conditions upon entry into the study and occurring in the same reporting day for COVID-19 conditions. There was a diversified multi-morbid history with (a) hypertension and lipid disorders having the highest prevalence, >65%, followed by (b) 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.

### *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 2]

### *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 figure 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). 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.

### *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.

## **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 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 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<sup>3-5</sup>. 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 mortality when compared to AF patients without COVID-19 patients<sup>9</sup>. Our ML prediction could be incorporated into telehealth approaches to monitor patients following their COVID-19 diagnosis, for the

onset of incident AF<sup>10</sup>. Given the increasing focus on integrated care management of patients with AF<sup>11</sup>, novel ML approaches could facilitate structured management and follow-up, especially since risk profiles change in a dynamic manner over time<sup>12-14</sup>. 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<sup>15-17</sup>.

### 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.

### 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.

### Disclosures

The authors report no conflicts of interest in this work.

### 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 (Clinical research 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. *The British journal of general practice : the journal of the Royal College of General Practitioners*. 2020.
3. Mair FS, Foster HM, Nicholl BI. Multimorbidity and the COVID-19 pandemic - An urgent call to action. *J Comorb*.2020;10:2235042x20961676.
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). *Cardiovascular diabetology*. 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 medicine*. 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):827-836.
7. Stone E, Kiat H, McLachlan CS. Atrial fibrillation in COVID-19: A review of possible mechanisms. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*.

2020;34(9):11347-11354.

8. Proietti M, Marzona I, Vannini T, et al. Long-Term Relationship Between Atrial Fibrillation, Multimorbidity and Oral Anticoagulant Drug Use. *Mayo Clinic proceedings*. 2019;94(12):2427-2436.

9. Higgins JPT, Green S: Cochrane Handbook for Systematic Reviews of Interventions 5.1.0. [updated March 2011] <http://handbook.cochrane.org>.

10. 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. *Clinical cardiology*. 2020;43(11):1232-1239.

11. 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). *European heart journal*. 2020.

12. 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. *Journal of clinical medicine*. 2020;9(4).

13. 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. *Thrombosis and haemostasis*. 2018;118(4):768-777.

14. 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. *Thrombosis and haemostasis*. 2018;118(7):1296-1304.

15. 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. *European journal of internal medicine*. 2020;82:105-111.

16. Guo Y, Lane DA, Wang L, et al. Mobile Health Technology to Improve Care for Patients With Atrial Fibrillation. *Journal of the American College of Cardiology*. 2020;75(13):1523-1534.

17. Guo Y, Lane DA, Chen Y, Lip GYH, m AFAIITi. Regular Bleeding Risk Assessment Associated with Reduction in Bleeding Outcomes: The mAFA-II Randomized Trial. *The American journal of medicine*. 2020;133(10):1195-1202 e1192.

#### Hosted file

Covid\_AF\_tables\_20012021.docx available at <https://authorea.com/users/366983/articles/509826-incident-atrial-fibrillation-and-its-risk-prediction-in-patients-developing-covid-19-a-machine-learning-based-algorithm-approach>

#### Hosted file

Covid\_AF\_figures\_20012021.docx available at <https://authorea.com/users/366983/articles/509826-incident-atrial-fibrillation-and-its-risk-prediction-in-patients-developing-covid-19-a-machine-learning-based-algorithm-approach>