Protocol: Evaluation of antithrombotic use and COVID-19 outcomes

Amendments

May 5th 2021: base version

May 28th 2021: v1 update

- Brought forward final time index point to May 1st 2021 to account for the time lag (min. 1 month) for data updates in the English TRE. Follow-up studies and evaluations could re-run the analysis to include other time index points in the future.
- Added clarification on management of contraindications for specific antithrombotics
- Added clarification on development of HAS-BLED phenotype
- Added time-to-event analysis (Cox regression, Kaplan Meier Curves) to analysis on COVID-19 outcomes
- Refined list of factors to analyse in factor analysis

July 2nd 2021: v2 update

- Updated question ordering
- Added DOACs vs warfarin as a new stratification for factor analysis (outcome) and covid outcomes analysis (exposure)
- Added links to codelist repository
- Updated target output exhibits across analyses
- Labile INR was removed from analysis due to a lack of data availability
- Prior major bleeding phenotype to be constructed from HES admissions rather than GDPPR due to lack of SNOMED code availability in TRE for major bleeding events
- Defined handling of incorrect AF first diagnosis dates (if all other inclusion criteria passed)
- Clarified collinearity screening process for factor analysis and removed univariable regression step
- Updated follow-up date for COVID-19 outcomes analysis to May 1st 2021 (corrected from March 31st 2021)
- "Long COVID" not included as a COVID-19 outcome due to lack of a robust phenotype broadly covering population
- Diabetes medications removed from COVID-19 outcome analysis due to high correlation with diabetes diagnosis. Diabetes kept as covariate.
- Clarified inclusion and definition of COVID-19 vaccine status
- Removed systolic blood pressure and alcohol consumption (as continuous values) and pregnancy from COVID-19 outcomes analysis due to lack of data availability
- Defined handling of missing values for BMI
- Clarified analysis steps for COVID-19 outcomes analysis

Background

Atrial fibrillation (AF) is the most common arrhythmia and significantly increases stroke risk. This risk can be effectively managed with antithrombotics (AT) [1]. In 2014, NICE updated its guidance to recommend that individuals with AF have a risk assessment for stroke and bleeding [2], and if appropriate, be offered anticoagulants (AC) rather than antiplatelets (AP) alone to reduce stroke risk [2]. This guidance helped increase the proportion of AC prescriptions in AF patients from 57.5% in 2014 to 66.9% in 2016 [3]. However, 17.4% of the at-risk population are still treated with AP alone and 15.7% are on no AT medication at all [3]. Hypotheses for this sub optimal prescribing centre around clinical over estimation of bleeding and fall risk in elderly (>85 years old) patients [3] but the potential drivers of AT medicating remain under explored at population scale.

COVID-19 has introduced another risk factor for AF patients who are at increased risk of poor outcomes if they become infected [4]. Recent observational evidence from Germany (n=6,637) suggests that regular AC use may improve outcomes for individuals hospitalised with COVID-19 [5]. However, evidence is discordant with a similar sized study in the US (n=3,772) showing no significant difference in mortality in groups on AC or AP [6]. These studies were both heavily limited by their data constraints (e.g. billing or hospital only data, moderate size, limited follow-up) which could be addressed by using the linked GP, prescriptions and hospital records available to CVD-COVID-UK.

This project will, therefore, explore three questions:

1. How many individuals in the UK with AF (and CHA_2DS_2-VASc score >= 2)[2] are not currently on antithrombotic medication?
2. What factors (e.g. location, age) are associated with a lack of antithrombotic medication in individuals with AF (and CHA_2DS_2-VASc score >= 2)?
3. Do individuals with AF (and CHA_2DS_2-VASc score >= 2) taking antithrombotic medication prior to COVID-19 infection have better outcomes (e.g. hospitalisation, mortality) than those that don't?

Addressing these questions fit directly within objective 1 of CVD-COVID-UK's scope [7] and would be relevant for health policy globally.
Moreover, in addressing these questions, this project will create the software to automatically evaluate AT use in near real-time across the whole UK population. This will provide a framework for automatic, longitudinal evaluations of NICE guidance on AT medication and generate significant efficiencies compared with replicating previous large-scale evaluations [1,3].

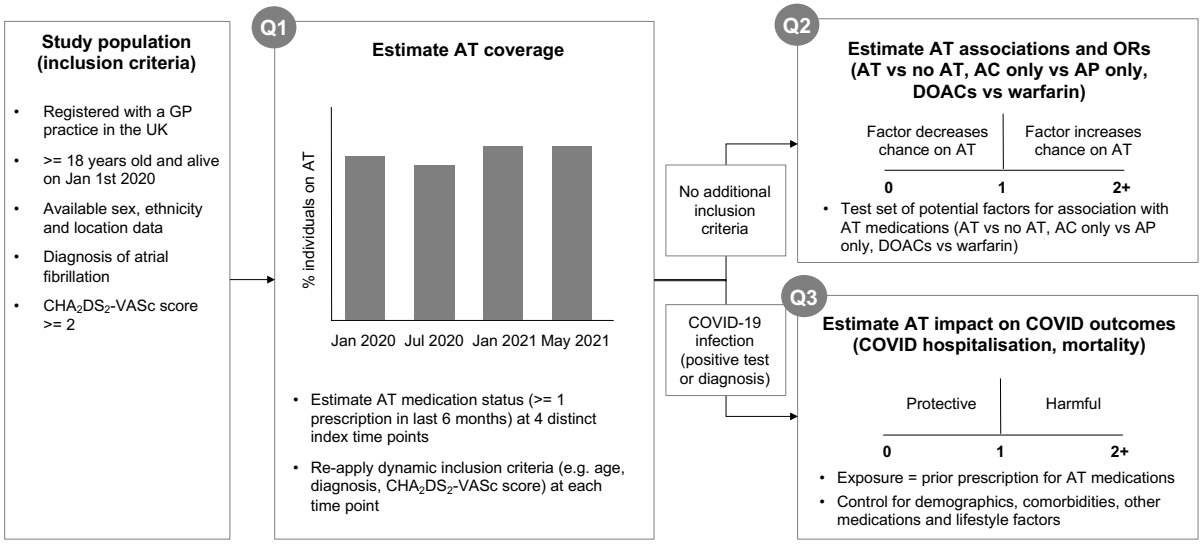
References

[1] Campbell Cowan et al, *"A 10 year study of hospitalized atrial fibrillation-related stroke in England and its association with uptake of oral anticoagulation"* - https://academic.oup.com/eurheartj/article/39/32/2975/5049393
[2] NICE guidance 2014, "Atrial fibrillation: management", - https://www.nice.org.uk/guidance/cg180
[3] Lacoin et al, *"Evolving landscape of stroke prevention in atrial fibrillation within the UK between 2012 and 2016: a cross-sectional analysis study using CPRD"* - https://bmjopen.bmj.com/content/7/9/e015363
[4] "Meta-Analysis of Atrial Fibrillation in Patients With COVID-19" - https://www.ajconline.org/article/S0002-9149(21)00056-4/fulltext
[5] Fröhlich et al, *"Impact of oral anticoagulation on clinical outcomes of COVID-19: a nationwide cohort study of hospitalized patients in Germany"* - https://link.springer.com/article/10.1007/s00392-020-01783-x

[6] Tremblay et al, *"Impact of anticoagulation prior to COVID-19 infection: a propensity score–matched cohort study"* - https://ashpublications.org/blood/article/136/1/144/458074/Impact-of-anticoagulation-prior-to-COVID-19
[7] CVD-COVID-UK - https://www.hdruk.ac.uk/projects/cvd-covid-uk-project/

Methods

Overall study design



Abbreviations: Antithrombotics (AT), anticoagulants (AC), antiplatelets (AC), direct oral anticoagulants (DOACs), odds ratios (OR)

Question 1. How many individuals in the UK with AF (and CHA2DS2-VASc score >= 2) are not currently on antithrombotic medication?

Study design

Antithrombotic coverage will be estimated at 4 distinct time index points (January 2020, July 2020, January 2021 and May 2021). The shorter final time period is due to the time lag (min. 1 month) for data updates in the English TRE. Follow-up studies and evaluations could re-run the analysis to include other time index points in the future.
The inclusion criteria for the study population (see below) will be re-applied at each time point to ensure the dynamic components (e.g. age, diagnosis, risk scores) are up to date.

Study population

The inclusion criteria will be as follows:

- Registered with a GP practice in the UK (NHS ID present in the GDPPR dataset)
- >= 18 years old and alive on January 1st 2020
- Available sex, ethnicity and GP practice location data (based on most recent, available data across linked GDPPR, HES and ONS datasets)
- Diagnosis of atrial fibrillation (coded in GDPPR)
- CHA_2DS_2-VASc score >= 2 (built up by components coded in GDPPR)

Individuals with contraindications for individual classes of antithrombotic (e.g. direct oral anticoagulants in mitral stenosis, prosthetic mechanical valves, antiphospholipid antibody syndrome) remain included provided they are still eligible for other classes of antithrombotic (e.g. antiplatelets, warfarin). Implications on interpretation will be highlighted in the discussion section of any publications.

Antithrombotic medication status will be determined using the linked NHS BSA Dispensed Medicines dataset.
An individual will be defined as on a particular medication if they have had >=1 prescription in the previous 6 months. This is to ensure incident prescriptions are captured and also attempts to control for potentially irregular buying patterns (e.g. bulk buying) due to the pandemic.

If an individual has been on multiple medications (e.g. apixaban and aspirin) both will be counted separately and then apportioned to higher level categories (e.g. antiplatelets and anticoagulants) for aggregated analysis.

The following antithrombotic medication categories will be estimated and will represent the entire study population when summed:

- Any antithrombotic medication (anticoagulants or antiplatelets) will also be estimated and reported.
    - Anticoagulants only
    - Antiplatelets only
    - Antiplatelets and anticoagulants
- No antithrombotic medication (no coded prescription within timeframe)

Individual antithrombotic medications and sub categories will also be estimated (coded in NHS Dispensed Medicines dataset).

- Anticoagulants
    - Apixaban
    - Rivaroxaban
    - Dabigatran
    - Edoxaban
    - Warfarin
- Antiplatelets
    - Aspirin
    - Clopidogrel
    - Dipyridamole
    - Ticagrelor
    - Prasugrel

Apixaban, rivaroxaban, dabigatran and edoxaban will also be collectively categorised as DOACs.

All codelists are available for review on the CCU020 Github repository (https://github.com/BHFDSC/CCU020/tree/main/england/phenotypes).

Statistical analysis

Descriptive statistics including means, medians and proportions will be used to summarise the study population characteristics and their medication status.

Target outputs

Table (main manuscript) – study population characteristics by antithrombotic medication category (at January 1st 2020)

	Total	Any AT	AC only	AP only	AC and AP	No AT
Individuals (n / %)						
Age (yrs, +- SD)						
65-74						
>=75						
Sex (% female)						
Ethnicity						
Geographical locations						
Index of Multiple Deprivation decile						
CHA_2DS_2-VASc components (e.g. stroke history) (n / %)						
CHA_2DS_2-VASc score (e.g. 2-9) (n / %)						
HAS-BLED components (e.g. liver disease) (n / %)*						

HAS-BLED score (e.g. 0 - 9) (n / %)*						
Time since first AF diagnosis						
History of Fall						

**Note HAS-BLED may be represented by a subset of available components or replaced with the ORBIT bleeding risk score depending on ability to develop robust phenotypes for each of the individual HAS-BLED components.*

Table (supplement) - study population characteristics by antithrombotic medication category for May 1st 2021 (final time index point)

Figure (main manuscript) – Antithrombotic prescribing trends by medication category over time (stacked bar chart at each of the 4 time indices)

Figure (main manuscript) – Antithrombotic prescribing trends by individual drug over time (stacked bar chart at each of the 4 time indices)

Question 2. What factors (e.g. location, age) are associated with a lack of antithrombotic medication in individuals with AF (and CHA2DS2-VASc score >= 2)?

Study design

This question will be analysed using the study population from question 1.
Multivariable logistic regression will be used to test the association and effect size (odds ratios) of a set of potential factors on no AT medication vs any AT, AP only vs AC only and DOACs vs warfarin. Analysis will be conducted with data at the first time index point (1st January 2020) and validated for consistency on the final time index point (May 1st 2021).

Study population

As described in "Study population" for question 1.

Statistical analysis

The list of factors below will be tested for association:

- Age (normalised to units of standard deviation (Z-score))
- Sex
- Ethnicity
- Geographical location (based on registered GP practice)
- Index of multiple deprivation decile
- A binary variable for each unique disease component of the CHA_2DS_2-VASc and HAS-BLED scores

    - Congestive heart failure
    - Hypertension history
    - Stroke
    - Vascular disease
    - Diabetes
    - Uncontrolled hypertension (>160 mmHg systolic)
    - Renal disease
    - Liver disease
    - Prior major bleeding (coded in HES due to no data available in GDPPR)
    - Hazardous alcohol use
- Time since first AF diagnosis
    - <2 years
    - 2-5 years
    - >5 years
- History of fall

Descriptive statistics including means, medians and proportions will be used to summarise the study population's characteristics.

Where the date of first AF diagnosis used to calculate time since first AF diagnosis is demonstrably incorrect (e.g. before date of birth), the date of first AF diagnosis will be imputed with the mean date from the cohort.

To test association with AT medications, each factor will be tested for collinearity with the other pre-selected factors using a correlation matrix (Pearson's coefficient, r). Any factors presenting high collinearity ($r > 0.5$) with factors outside of their own sub-categories will not be included in the multivariable logistic regression model which will be used to estimate the reported effect size (odds ratios) for each factor.

Target outputs

Figure (main manuscript) – Forest plot of odds ratios for all input variables for AT vs no AT (January 1st 2020)

Figure (supplement) – Correlation matrix for pre-selected variables

Figure (supplement) – Forest plot of odds ratios for all input variables for AT vs no AT (May 1st 2021)

Figure (supplement) – Forest plot of odds ratios for all input variables for AC vs AP (January 1st 2020)

Figure (supplement) – Forest plot of odds ratios for all input variables for DOACs vs warfarin (January 1st 2020)

R shiny app to include tables with ORs, CI ranges and p-values for all comparison groups.

Question 3. Do individuals with AF (and CHA2DS2-VASc score >= 2) taking antithrombotic medication prior to COVID-19 infection have better outcomes (e.g. hospitalisation, mortality) than those that don't?

Study design

This question will be analysed using a retrospective cohort study design with individuals that meet the inclusion criteria described in "Study population" under question 1 on January 1st 2020. It will follow-up eligible individuals until May 1st 2021 and compare outcomes from COVID-19 for those that took AT prior to COVID-19 infection with those that didn't whilst controlling for measurable confounders (see "Exposure, outcomes and covariates"). Three regression analyses will be used to test the association and effect size of AT on COVID-19 outcomes (see "Statistical analysis").

Study population

Individuals meeting inclusion criteria in question 1, plus:

- COVID-19 infection will be defined as either a positive test (Pillar 1 and Pillar 2 polymerase chain reaction swab tests) or a coded diagnosis in primary or secondary care (see Thygesen et al **[add reference when available]** for further details and codelists) between January 1st 2020 and May 1st 2021

Exposure, outcomes and covariates

Exposure

The primary exposure criteria will be individuals with a prior prescription for any AT on January 1st 2020.
A prior prescription will be defined as in question 1 (>=1 prescription in the previous 6 months).
Exposure will also be tested for AC only vs AP only and DOACs vs warfarin (as in factor analysis – Question 2).

Outcomes

Two binary outcomes will be measured:

- COVID-19 hospitalisation - any hospital admission with a recorded COVID-19 diagnosis
- COVID-19 death - individuals with a COVID-19 diagnosis on their death certificate, a registered death within 28 days of their first recorded COVID-19 event or a discharge destination denoting death after a COVID-19 hospitalisation

Refer to Thygesen et al **[add reference when available]** for further details and codelists

Covariates

Although confounding will be reduced by restricting the cohort to people with AF and a CHA_2DS_2-VASc score >=2, other covariates will be included to control for other potential confounding factors.

The full list of the covariates to be included (based on status on January 1st 2020) in multivariable logistic regression analyses is below:

- Demographics
    - Age (normalised to units of standard deviation (Z-score))
    - Sex
    - Ethnicity
    - Socio-economic position (measured by Index of Multiple Deprivation rank)
- Comorbidities
    - A binary variable for each unique disease component of the CHA_2DS_2-VASc and HAS-BLED scores
        - Congestive heart failure

        - Hypertension history
        - Stroke
        - Vascular disease
        - Diabetes
        - Uncontrolled hypertension (>160 mmHg systolic)
        - Renal disease
        - Liver disease
        - Prior major bleeding (coded in HES due to no data available in GDPPR)
        - Hazardous alcohol use
- Other medications (>= 1 prescription in previous 6 months, BNF category in brackets)
    - Cardiovascular
        - Anti-hypertensives (2.5)
        - Lipid-regulating drugs (2.12)
    - Other
        - Proton pump inhibitors (1.3.5)
        - NSAID (10.1.1)
        - Corticosteroids (1.5.2, 6.3.2)
        - Other immunosuppressants (1.5.1, 1.5.3, 8.2)
    - COVID-19 Vaccination status – defined as at least one vaccine prior to COVID-19 infection
- Other lifestyle risk factors and relevant biomarkers
    - Body mass index (most recent recorded value within past 10 years, whilst age>18 years in GDPPR, normalised to units of standard deviation (Z-score))
    - Smoking status (ever smoker – based on status codes in GDPPR)

Statistical analysis

Descriptive statistics including means, medians and proportions will be used to summarise the study population's characteristics and outcomes.

Where BMI (the only continuous variable with possible missing data) is not available, the BMI value will be imputed with the mean BMI value from the cohort.

Three multivariable regression analyses will be used for each of the 2 outcomes (COVID-19 death, COVID-19 hospitalisation) across the 3 AT medication exposure groups (any AT vs no AT, AC only vs AP only, DOACs vs warfarin) with the listed covariates included as input variables.

Firstly, multivariable logistic regression will be used with the two outcomes as binary events (1 for COVID-19 death, 1 for COVID-19 hospitalisation) and the pre-selected covariates ("Covariates") as input variables after passing a collinearity screen as described in the factor analysis ("Question 2").

Secondly, to address residual confounding not captured in the covariates, a new multivariable logistic regression model will be constructed on the binary outcomes adjusted for propensity to be on AT medication. The propensity score will be estimated using multivariable logistic regression on AT medication usage vs no AT medication with the same set of covariates used in step 1. The estimated propensity score will then be included in the adjusted model as an additional covariate (normalised to units of standard deviation (Z-score)).

Thirdly, to capture the dimension of time, a multivariable Cox regression model will be constructed with outcomes transformed to days to event. All covariates including the adjusted propensity score will be included.

As a sensitivity analysis to evaluate the impact of different stages of the pandemic, the above analysis steps will then be repeated for the time period Jan 1st 2020 – Dec 1st 2020, prior to the introduction of vaccines and the peak of second wave.

Target outputs

Table (main manuscript) – study population characteristics by antithrombotic medication category for COVID-19 cohort only (at January 1st 2020) – *same format as earlier table ("Question 1") with new variables added*

Figure (main manuscript) – comparison of AT medication exposures (AT vs no AT, AC vs AP and DOACs vs warfarin) on COVID-19 outcomes (follow up to May 1st 2021) using propensity score adjusted multivariable logistic regression

Figure (supplement) - correlation matrix for pre-selected variables in COVID-19 outcome analysis

Figure (supplement) - comparison of AT medication exposures (AT vs no AT, AC vs AP and DOACs vs warfarin) on COVID-19 outcomes (follow up to May 1st 2021) using Cox regression

Figure (supplement) - comparison of AT medication exposures (AT vs no AT, AC vs AP and DOACs vs warfarin) on COVID-19 outcomes (follow up to December 1st 2020) using propensity score adjusted multivariable logistic regression

Figure (supplement) - comparison of AT medication exposures (AT vs no AT, AC vs AP and DOACs vs warfarin) on COVID-19 outcomes (follow up to December 1st 2020) using Cox regression

R shiny app to include tables with ORs, CI ranges and p-values for all comparison groups.

Target outputs across questions

The end-to-end code for the analysis pipeline and summary charts will be made available on the BHF CVD-COVID trusted research environment and Github with key findings summarised for dissemination in an open-access published paper.

The project will work closely with medications and methodology projects led by Reecha Softa (CCU014), Spiros Denaxas (CCU013) and Angela Wood (CCU005) to create reproducible methods and algorithmic phenotypes for CHA2DS2-VASc and HAS-BLED scores that can support automated evaluations of AT medication