


RESEARCH ARTICLE

Open Access



Bleeding in cardiac patients prescribed antithrombotic drugs: electronic health record phenotyping algorithms, incidence, trends and prognosis

Laura Pasea^{1,2}, Sheng-Chia Chung^{1,2}, Mar Pujades-Rodriguez³, Anoop D. Shah^{1,2,7}, Samantha Alvarez-Madrado⁴, Victoria Allan^{1,2}, James T. Teo⁵, Daniel Bean⁶, Reecha Sofat⁷, Richard Dobson^{1,2,6}, Amitava Banerjee^{1,2}, Riyaz S. Patel^{2,8}, Adam Timmis⁹, Spiros Denaxas^{1,2} and Harry Hemingway^{1,2,10*} 

Abstract

Background: Clinical guidelines and public health authorities lack recommendations on scalable approaches to defining and monitoring the occurrence and severity of bleeding in populations prescribed antithrombotic therapy.

Methods: We examined linked primary care, hospital admission and death registry electronic health records (CALIBER 1998–2010, England) of patients with newly diagnosed atrial fibrillation, acute myocardial infarction, unstable angina or stable angina with the aim to develop algorithms for bleeding events. Using the developed bleeding phenotypes, Kaplan-Meier plots were used to estimate the incidence of bleeding events and we used Cox regression models to assess the prognosis for all-cause mortality, atherothrombotic events and further bleeding.

Results: We present electronic health record phenotyping algorithms for bleeding based on bleeding diagnosis in primary or hospital care, symptoms, transfusion, surgical procedures and haemoglobin values. In validation of the phenotype, we estimated a positive predictive value of 0.88 (95% CI 0.64, 0.99) for hospitalised bleeding. Amongst 128,815 patients, 27,259 (21.2%) had at least 1 bleeding event, with 5-year risks of bleeding of 29.1%, 21.9%, 25.3% and 23.4% following diagnoses of atrial fibrillation, acute myocardial infarction, unstable angina and stable angina, respectively. Rates of hospitalised bleeding per 1000 patients more than doubled from 1.02 (95% CI 0.83, 1.22) in January 1998 to 2.68 (95% CI 2.49, 2.88) in December 2009 coinciding with the increased rates of antiplatelet and vitamin K antagonist prescribing. Patients with hospitalised bleeding and primary care bleeding, with or without markers of severity, were at increased risk of all-cause mortality and atherothrombotic events compared to those with no bleeding. For example, the hazard ratio for all-cause mortality was 1.98 (95% CI 1.86, 2.11) for primary care bleeding with markers of severity and 1.99 (95% CI 1.92, 2.05) for hospitalised bleeding without markers of severity, compared to patients with no bleeding.

Conclusions: Electronic health record bleeding phenotyping algorithms offer a scalable approach to monitoring bleeding in the population. Incidence of bleeding has doubled in incidence since 1998, affects one in four cardiovascular disease patients, and is associated with poor prognosis. Efforts are required to tackle this iatrogenic epidemic.

Keywords: Bleeding, Electronic health records, Phenotype, Antithrombotic therapy, Prognosis

* Correspondence: h.hemingway@ucl.ac.uk

¹Health Data Research UK, University College London, 222 Euston Road, London NW1 2DA, UK

²Institute of Health Informatics, University College London, London, UK

Full list of author information is available at the end of the article



Background

Bleeding is amongst the most common serious side effects of modern medicine, but clinicians and health systems lack basic information on how to define and monitor the occurrence and severity of bleeding in populations. Multiple clinical guidelines make recommendations for the use of antithrombotic drugs across diseases [1, 2]. Increases in the burden of common cardiovascular diseases (CVDs), new drugs (e.g. P2Y12 receptor antagonists and direct anticoagulants), implementation of long-standing trial evidence (e.g. aspirin in the secondary prevention of CVD) and prolongation (lifelong) of regimes which were initially introduced for fixed durations (e.g. dual antiplatelet therapy after acute myocardial infarction (MI)) have led to increasing antithrombotic use [3–5].

Bleeding risk stratification [3], prevention [6, 7] and management [8, 9] are mentioned in several guidelines. However, specific recommendations at the individual and population level, in specific subpopulations (e.g. with concomitant proton pump inhibitor prescription [10]), are lacking largely due to lack of data regarding the population burden (incidence, time trends and prognosis) of bleeding in people with common CVDs, time trends in incidence of bleeding of different severities with increasing antithrombotic use. Bleeding risks, often defined differently, have been described in individual diseases (atrial fibrillation (AF) [11], acute coronary syndromes [12] and stable coronary disease [13]), but there are no studies comparing the risks across common CVDs.

A central reason for these uncertainties is the lack of standardised definitions to measure bleeding occurrence and severity which are scalable across populations and different national health systems, where manual adjudication of case records (used in small numbers of bleeding events, e.g. in trials, or consented research cohorts [10, 14, 15]) is neither practical nor feasible. Consistent definitions of disease and health conditions using diverse electronic health records (EHR) across primary and hospital care can be used to make valid comparisons across countries [16–18]. Previous EHR studies of bleeding endpoints have been restricted by setting [19–21], anatomical site (e.g. upper gastrointestinal bleeding [22–24]) or data (insurance or administrative claims [25, 26]) (Additional file 1: Table S1). The efficient use of information related to bleeding (e.g. diagnosis, anatomical site, fatality, length of hospital stay, haemoglobin, transfusion, endoscopy, surgical interventions) could help to generate population estimates of bleeding occurrence and severity.

We sought to address the following questions: First, how can population-based EHR, spanning primary and hospital care, be used to define valid, replicable algorithms of bleeding occurrence and bleeding severity?

Second, what is the long-term cumulative incidence of bleeding events across patients with incident AF, acute MI and unstable and stable angina who are prescribed with different antiplatelet and anticoagulation regimens? Third, to what extent has the incidence of bleeding increased over time with the changes in antithrombotic management? Fourth, to what extent is bleeding of differing severity associated with long-term prognosis in terms of all-cause mortality, atherothrombotic events and recurrent bleeding?

We used the CALIBER [27] research platform of linked primary, hospital, myocardial ischaemia registry and mortality data. EHR phenotypes have been developed in CALIBER for acute MI [18], AF [28] and stable coronary disease [29]. Cohort studies of their associations with blood pressure [30], diabetes [31], smoking [32], socioeconomic deprivation [33], rheumatoid arthritis [34], alcohol consumption [35] and neutrophil counts [36] have supported their validity.

Methods

Linked electronic health records

We used data from the CALIBER [27] resource. CALIBER links EHR from primary care general practices (Clinical Practice Research Datalink [CPRD]), hospital admissions (Hospital Episode Statistics [HES]), myocardial ischaemia registry (Myocardial Ischaemia National Audit Project [MINAP]) and cause-specific mortality (Office for National Statistics [ONS]) data in England. The 4% sample of England's population in CPRD available for linkage is representative in terms of age, sex and overall mortality [37–39]. In CALIBER, EHR disease phenotypes [40] have been developed through collaborations between clinicians, epidemiologists and statisticians, and a number of risk factors and cardiovascular and non-cardiovascular disease endpoints have been validated for cardiovascular research [18, 27–36].

The study was approved by the Independent Scientific Advisory Committee of the Medicines and Healthcare products Regulatory Agency in the UK, protocol number 14_133.

Study population

The study population consisted of patients with CVD, i.e. those who were potential candidates for antiplatelet and/or vitamin K antagonist (VKA) therapy, in CALIBER during 1997–2010. The study period was chosen to reflect stable prescribing practice, with only warfarin and antiplatelet agents, before the introduction of multiple directly acting anticoagulants. To define this population, we used pre-existing validated disease phenotypes [<https://www.caliberresearch.org/portal>]. Patients were eligible if they were aged 18 years and above and entered the cohort at their first diagnosis of AF, acute MI, unstable angina or

stable angina in primary or hospital care records. They were followed up until death, transfer out of their primary care practice (i.e. loss to follow-up) or the date of administrative censoring (March 2010).

We analysed baseline characteristics of patients stratified by initial CVD. Using prescribing data, we summarised therapy duration (median and interquartile range days) of between cohort entry and first bleeding event. To calculate the duration, a patient's prescription was assumed to be continuous if issued within 90 days of the previous one (90 days is the longest allowed duration of prescriptions in the UK). Treatments were grouped as aspirin monotherapy, adenosine diphosphate (ADP) receptor inhibitor monotherapy, dual antiplatelet therapy (aspirin and ADP receptor inhibitor), VKA monotherapy, VKA and one antiplatelet (aspirin or ADP receptor inhibitor) and triple therapy (VKA, aspirin and ADP receptor inhibitor).

Electronic health record data relevant to definition of bleeding phenotypes

Within CALIBER, bleeding events were captured in primary care data (Read terms), hospital admissions administrative data (ICD-10 terms) and death registry (ICD-9 and ICD-10 terms) (Additional file 1: Table S2). The description of the terms used contained information on the anatomical site of bleeding. Hospital records indicated the diagnosis position (i.e. primary or secondary reason for hospitalisation), and the length of hospitalisation was calculated using admission and discharge dates. Procedures relevant to bleeding (transfusion, bleeding surgical interventions and endoscopy) were captured in hospitalisation records using OPCS codes. Drug prescriptions were available in primary care data, classified according to the British National Formulary (BNF) chapter. Clinical biomarkers such as haemoglobin were also captured in primary care.

Algorithmic combinations to define bleeding EHR phenotypes

The construction of the CALIBER bleeding EHR phenotype (Fig. 1) is fully explained in Additional file 1: Methods S3. In short, we applied a structured approach to phenotyping, previously demonstrated by Morley et al. [28], involving iterative steps of diagnosis code reviews, descriptive analyses and expert input. We used published trial protocols of the definition of major bleeding [14, 15, 41] to identify candidate markers of bleeding severity. We included the sub-set of markers which were available in the EHR (for example, the HES data does not record haemoglobin measurements) and evaluated the associations with short-term mortality in order to develop the severe bleeding EHR phenotype. We defined fatal bleeding as a bleeding cause of death (underlying or otherwise) in the national death registry or all-cause death within 7 days of a

bleeding record in primary or hospital care. We identified four markers of bleeding severity available within our data: (1) bleeding as a primary reason for hospitalisation combined with at least 14 days hospitalisation, (2) bleeding site (intracranial, ruptured aortic aneurysm or haemopericardium), (3) bleeding from more than one site on the same day and (4) a transfusion record in hospital care within 30 days of a bleeding record.

We classified non-fatal bleeding events as hospitalised or primary care with further markers of severity (henceforth referred to as 'hospitalised+MS' and 'primary care+MS') and hospitalised or primary care without markers of severity (referred to as 'hospitalised' and 'primary care'). For patients with no bleeding code in either primary care or hospital records, possible bleeding events may be inferred where there are records that provide evidence suggesting bleeding, for example, transfusions and low haemoglobin.

Statistical analysis

Validation of the hospitalised bleeding phenotype

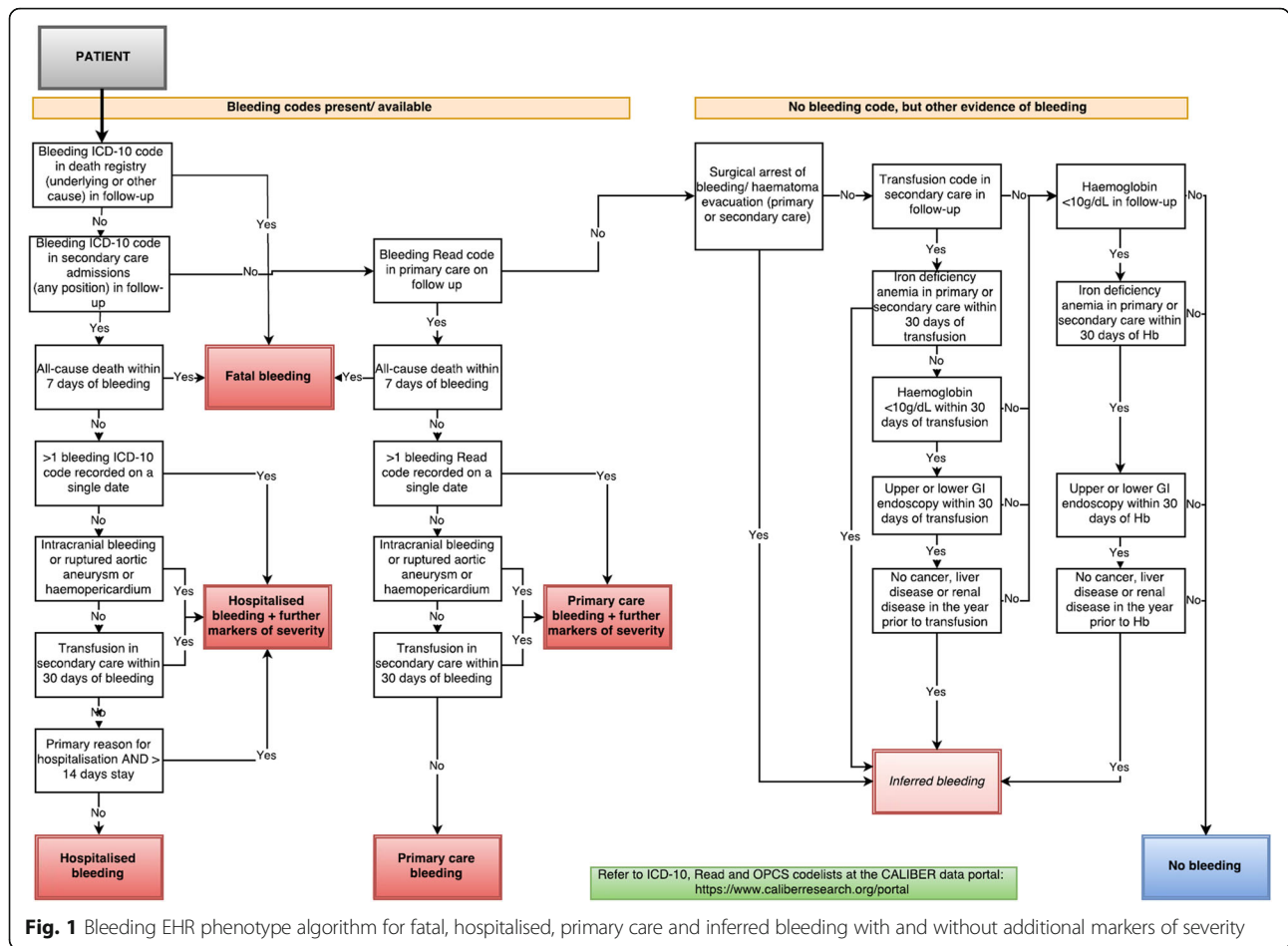
We validated the hospitalised bleeding part of the phenotype algorithm through manual case note review amongst consented patients in the SIGNUM prospective stroke cohort at 2 large NHS Trusts (University College London Hospitals NHS Foundation Trust and King's College Hospital NHS Foundation Trust). Two clinicians (blinded to the ICD-10 and OPCS-4 codes recorded) reviewed the entire hospital record (charts, referral letters, discharge letters, imaging reports) for 283 stroke patient hospital episodes. The hospital record corpus (14,364,947 words in total) was made available as single text files per patient, through the use of CogStack [42], method of enterprise-wide retrieval and extraction architecture for structured and unstructured information which integrates data across multiple EHR systems in a hospital. Bleeding assignments from the clinicians' review were compared with those from the bleeding algorithm, and we estimated the positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity using the case review data as the gold standard.

Cumulative bleeding incidence in four cardiovascular diseases

The incidence of any bleeding and fatal, hospitalised+MS or primary care+MS bleeding was assessed using Kaplan-Meier plots stratified by CVD-type AF, acute MI, unstable angina or stable angina.

The association between antithrombotic prescribing and bleeding

Cox proportional hazard models were used to estimate the hazard ratios for the association between antithrombotic therapies and first bleeding event of any severity and fatal or bleeding+MS event. Antithrombotic therapy prescriptions were included in the models as a time-dependent



variable. Possible states were no antithrombotic therapy (the reference group), aspirin, ADP receptor inhibitor, dual antiplatelet therapy, vitamin K antagonist, vitamin K antagonist and one antiplatelet (aspirin or ADP receptor inhibitor), and triple therapy. Patients were followed up until their first bleeding event of any severity and until their first fatal or bleeding+MS event. The Cox models were adjusted for age and sex.

Time trends in bleeding

We estimated the number of fatal, hospitalised+MS, primary care+MS, hospitalised and primary care bleeding events per 1000 patients at monthly intervals between 1997 and 2010. To do this, we divided the number of bleeding events recorded by the total number of patients at risk each month. Loess smoothed lines were fitted to detect changes in incidence over time. Similarly, we estimated the time trends for the number of antithrombotic prescriptions issued each month.

Prognosis following bleeding

We used Cox proportional hazard models to estimate the hazard ratios (HR) for the association between first

bleeding events, all-cause mortality and atherothrombotic events (composite of cardiovascular death, ischaemic or unspecified stroke, or MI). Bleeding severity (hospitalised+MS, primary care+MS, hospitalised, primary care and inferred) was treated as a time-dependent variable in the models to prevent immortal time bias. The possible bleeding variable states were no bleeding (reference group), primary care, primary care+MS, hospitalised or hospitalised+MS. All patients started follow-up in the no bleeding state and changed to the relevant bleeding state at the time of their first bleeding event. Models were also adjusted for age, sex and baseline disease history (diabetes, stroke, peripheral arterial disease, cancer, renal disease, peptic ulcer, bleeding diatheses, chronic anaemia). We also explored the risk of recurrent bleeding in the subgroup of patients that had non-fatal bleeding events using Kaplan-Meier plots, following patients from the time of their first non-fatal bleeding event.

Modelling assumptions

The proportional hazards assumptions of Cox models were checked using residual and log(-log) plots. All analyses were performed using R version 3.2.

Patient involvement

No patients were involved in setting the research question and study outcome or the design and implementation of the study. There are no current plans to disseminate the results with patient groups.

Results

Study population

Our study population consisted of 128,815 patients in 224 general practices newly diagnosed with AF, acute MI, unstable angina and/or stable angina between 1997 and 2010. They were followed up for a total of 559,161 person-years, a median of 3.7 years (IQR 1.5, 6.9). The mean age was 71.5 years at cohort entry (43.8% aged ≥ 75 years), and 48.5% were women.

Patient characteristics stratified by CVD are shown in Table 1. AF patients were older than the coronary disease patients, and the majority were women. In contrast, the coronary disease patients were mostly men. The AF patients also had a higher prevalence of history of stroke, renal disease, cancer and chronic anaemia. The majority of patients in all four disease groups were prescribed at least one antithrombotic drug between cohort entry and first bleeding event or end of follow-up in those who did not bleed.

Applying the CALIBER bleeding EHR phenotype algorithm

The bleeding algorithm is shown in Fig. 1. We identified 39,804 bleeding records from 27,259 (21.2%) patients in our cohort. 59.4% of coded bleeding events were captured in primary care, 50.2% in hospital admissions and 3.8% events in death registry. Allowing a 30-day window, only 13.2% of coded bleeding events were captured in 2 or more data sources. The overlap of bleeding events between the data sources used is shown in Additional file 1: Figure S4.

We identified 1492 further possible bleeding events occurring in 1144 patients with no bleeding diagnosis recorded in primary care or hospital records through the following routes: transfusion and presence of iron deficiency anaemia diagnosis within 30 days ($n = 689$) [1]; surgical procedures to arrest bleeding or for haematoma evacuation ($n = 477$) [2]; haemoglobin < 10 g/dL, iron deficiency anaemia diagnosis and endoscopic examination within 30 days and no cancer, liver or renal disease records in the previous year ($n = 249$) [3]; transfusion, haemoglobin < 10 g/dL and endoscopic examination within 30 days and no cancer, liver or renal disease records in the year prior ($n = 77$) [4].

Validation of the hospitalised bleeding phenotype

In our validation sub-study of hospitalised bleeding in the phenotype algorithm using ICD-10 and OPCS codes, we estimated a PPV of 0.88 (95% CI 0.64, 0.99), a NPV of 0.98 (0.95, 0.99), a sensitivity of 0.71 (0.48, 0.89) and a

specificity of 0.99 (0.97, 1.00) (Additional file 1: Table S5). The ICD-10 codes that were recorded for the false-negative cases (clinicians identified bleeding in the case notes, but the algorithm did not find bleeding in the codes) are presented in Additional file 1: Table S6. The clinicians' review of free text identified seven patients with a CT scan report of haemorrhagic transformation of stroke, which did not have a bleeding as the primary cause of admission. (Additional file 1: Table S7).

Cumulative incidence of any bleeding and fatal bleeding or bleeding with markers of severity

At 5 years, 29.1% (95% CI 28.2, 29.9%) of AF patients, 21.9% (21.2, 22.5%) of MI patients, 25.3% (24.2, 26.3%) of unstable angina patients and 23.4% (23.0, 23.8%) of stable angina had bleeding of any kind (Fig. 2). Risks of fatal bleeding, hospitalised+MS or primary care+MS bleeding events at 5 years were 9.9% (9.3, 10.4%) for AF patients, 6.1% (5.8, 6.5%) for MI patients, 6.8% (6.0, 7.2%) for unstable angina patients and 5.7% (5.5, 5.9%) for stable angina.

Time trends in bleeding incidence and antithrombotic prescribing

The estimated number of hospitalised+MS bleeding events per 1000 active patients increased from 0.32 (0.24, 0.40) in January 1998 to 0.54 (0.45, 0.62) in December 2009. Contrarily, in primary care+MS, bleeding events per 1000 active patients decreased from 0.80 (95% CI 0.70, 0.91) in January 1998 to 0.34 (0.23, 0.45) in December 2009. The incidence of fatal bleeding remained steady (Fig. 3a).

There were increases in hospitalised and primary care bleeding events without markers of severity (Fig. 3b). The estimated number of hospitalised bleeding events per 1000 active patients increased from 1.02 (0.83, 1.22) in January 1998 to 2.68 (2.49, 2.88) in December 2009, and for primary care bleeding events, the increase was from 1.70 (1.44, 1.95) to 3.31 (3.06, 3.57). This corresponded to the rise of rates of prescribed antithrombotic therapies over the study period (Fig. 3c). From January 1998 to December 2009, the increase in the number of prescriptions issued per 1000 active patients for aspirin, ADP receptor inhibitor and VKA was 147.9 (95% CI 127.4, 168.3) to 465.1 (444.6, 485.6), 2.8 (0.2, 5.4) to 94.8 (92.2, 97.4) and 22.7 (19.2, 26.1) to 83.7 (80.2, 87.1), respectively.

Overall, patients prescribed with more aggressive antithrombotic therapies (dual antiplatelet therapy, vitamin K antagonists and triple therapy) had a significantly higher risk of bleeding events compared with those not prescribed antithrombotic therapies (Fig. 4). Compared with those not prescribed antithrombotic therapies, patients who were prescribed triple therapy had 3.4 (2.6, 4.4) times increased risk of any bleeding and 5.7 (3.7, 8.7) times increased risk of fatal or bleeding+MS events.

Table 1 Baseline characteristics of people with four common cardiac diseases

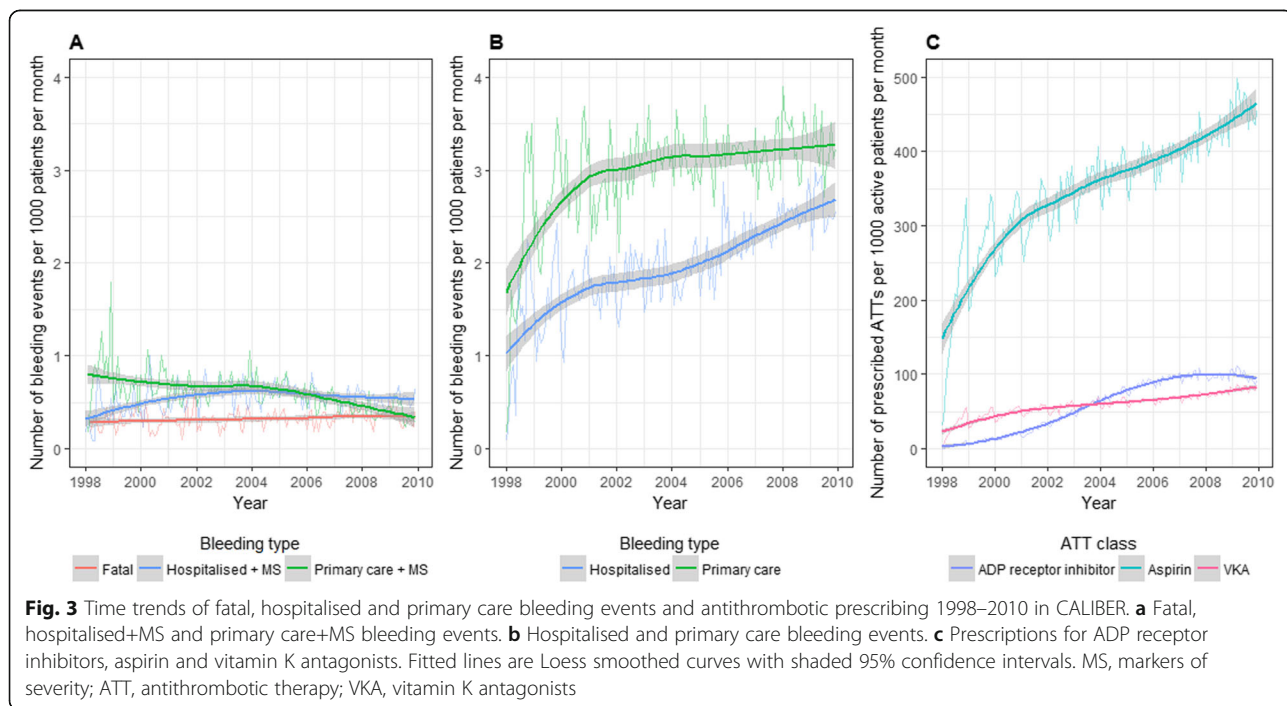
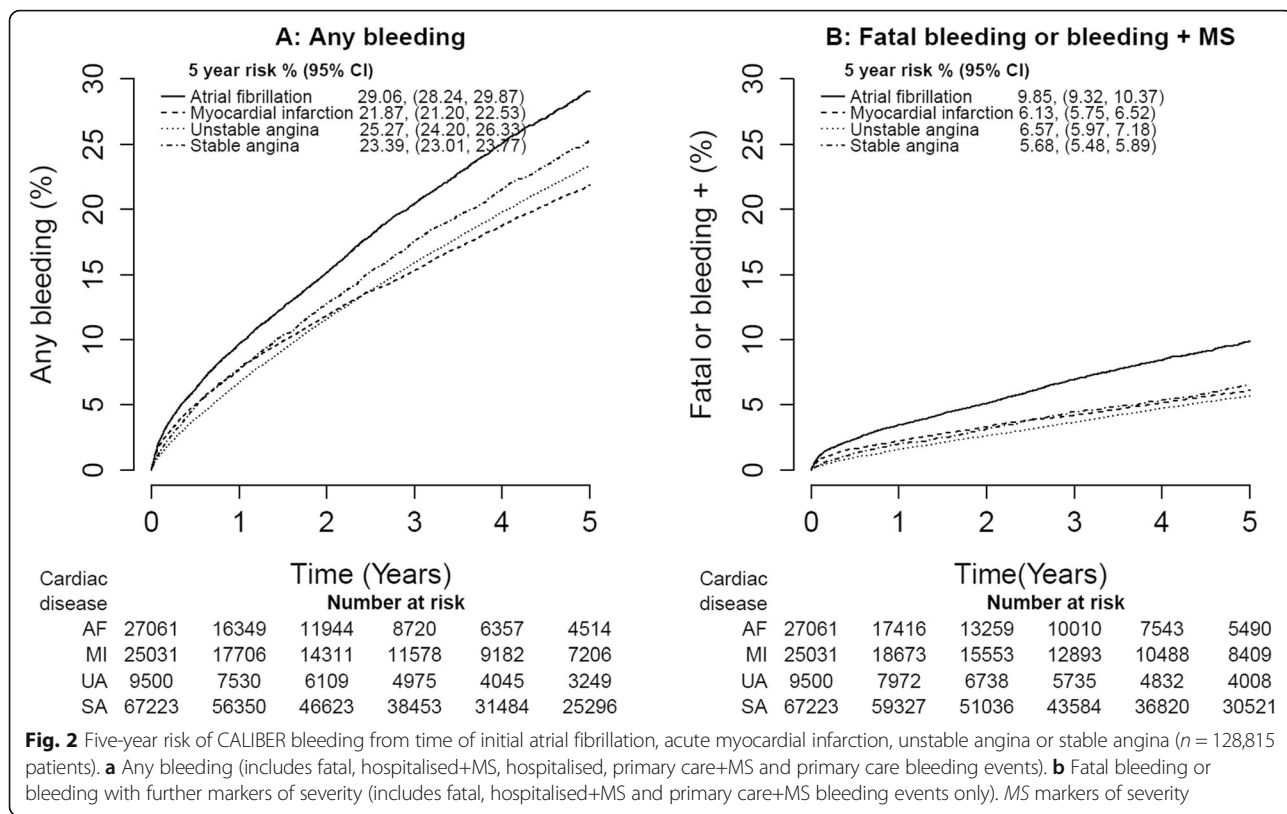
	Atrial fibrillation (n = 27,061)	Myocardial infarction (n = 25,031)	Unstable angina (n = 9500)	Stable angina (n = 67,223)
Demographics				
Age (years), mean (SD)	76.6 (12.8)	69.9 (13.5)	69.1 (13.2)	70.4 (12.3)
Age ≥ 75 years, n (%)	16,946 (62.6)	9982 (39.9)	3488 (36.7)	26,059 (38.8)
Women, n (%)	14,266 (52.7)	9206 (36.8)	4169 (43.9)	31,365 (46.7)
Highest quintile of deprivation (most deprived), n (%)	5137 (19.0)	4758 (19.1)	1947 (20.5)	13,837 (20.6)
Behaviours				
Current smoker, n (%)	2277 (10.5)	3691 (19.5)	1058 (14.0)	6229 (11.4)
History of alcohol abuse, n (%)	2627 (9.7)	2430 (9.7)	908 (9.6)	6459 (9.6)
Medical history prior to cohort entry ^a				
Type 2 diabetes, n (%)	2695 (10.0)	2922 (11.7)	1222 (12.9)	8728 (13.0)
Ischaemic or unspecified stroke, n (%)	2169 (8.0)	1462 (5.8)	558 (5.9)	3435 (5.1)
Peripheral arterial disease, n (%)	2276 (8.4)	2147 (8.6)	865 (9.1)	6199 (9.2)
Renal disease, n (%)	2570 (9.5)	1731 (6.9)	694 (7.3)	4351 (6.5)
Non-metastatic cancer, n (%)	5427 (20.1)	3158 (12.6)	1155 (12.2)	8701 (12.9)
Metastatic cancer, n (%)	526 (1.9)	209 (0.8)	74 (0.8)	520 (0.8)
Peptic ulcer, n (%)	1814 (6.7)	1713 (6.8)	753 (7.9)	5074 (7.5)
Bleeding diatheses and coagulation disorders, n (%)	312 (1.2)	175 (0.7)	77 (0.8)	534 (0.8)
Chronic anaemia, n (%)	4982 (18.4)	2808 (11.2)	1198 (12.6)	8125 (12.1)
Biomarkers at cohort entry ^b				
SBP (mmHg), mean (SD)	140 (21.8)	143 (21.2)	142 (21.2)	142 (20.5)
% Missing	29.7	33.2	25.8	21.6
Haemoglobin (g/dL), mean (SD)	12.9 (1.97)	13.5 (1.91)	13.5 (1.75)	13.6 (1.69)
% missing	56.2	65.9	62	59.1
Creatinine (mmol/L), mean (SD)	107 (59.1)	108 (56.1)	105 (55.9)	102 (46.1)
% missing	49.5	57.6	54.1	49.7
Antithrombotic therapies (n, %) and duration (median, IQR) during follow-up ^c				
Any antithrombotic therapy	16,868 (62.3)	19,950 (79.7)	7947 (83.7)	55,619 (82.7)
Aspirin monotherapy	10,787 (39.9)	16,511 (66.0)	6695 (70.5)	48,262 (71.8)
Duration (days)	382 (114, 908)	791 (267, 1742)	765 (268, 1691)	842 (305, 1752)
ADP receptor inhibitor monotherapy	1264 (4.7)	3683 (14.7)	1425 (15.0)	7351 (10.9)
Duration (days)	150 (46, 486)	94 (30, 376)	121 (42, 495)	181 (52, 652)
Dual antiplatelet therapy	1594 (5.9)	8673 (34.6)	2417 (25.4)	9539 (14.2)
Duration (days)	186 (90, 426)	349 (143, 478)	272 (98, 488)	261 (90, 476)
VKA monotherapy	7149 (26.4)	1666 (6.7)	853 (9.0)	6287 (9.4)
Duration (days)	427 (146, 1083)	216 (82, 626)	318 (110, 844)	344 (113, 938)
VKA + 1 antiplatelet	3003 (11.1)	1426 (5.7)	637 (6.7)	3892 (5.8)
Duration (days)	85 (51, 163)	106 (55, 262)	90 (54, 228)	90 (51, 214)
VKA + 2 antiplatelets	266 (1.0)	321 (1.3)	102 (1.1)	430 (0.6)
Duration (days)	68.5 (39, 93.2)	68.0 (43, 116.0)	62.5 (39, 90.0)	57.0 (35, 84.0)

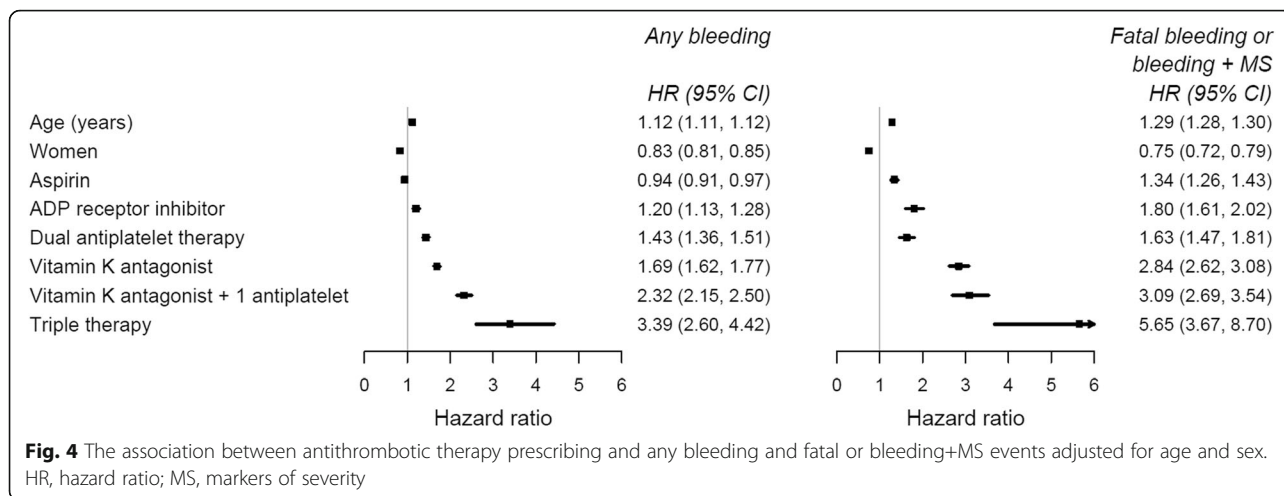
SD standard deviation, SBP systolic blood pressure, BMI body mass index, IQR interquartile range, ADP adenosine diphosphate, VKA vitamin K antagonist

^aAny record prior to cohort entry

^bNearest record to entry within 1 year prior to entry

^cBetween cohort entry and 1st bleeding event or end of follow-up





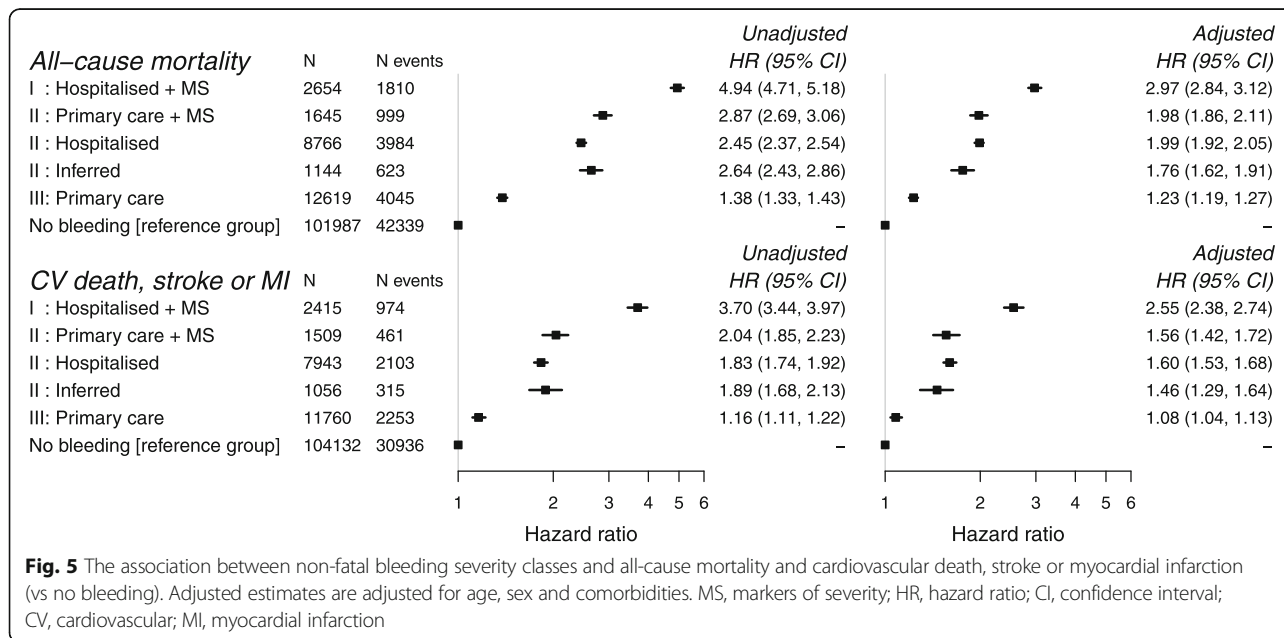
Death and atherothrombotic events following first bleeding event

Patients were at increased risk of all-cause mortality and cardiovascular death, stroke or MI following their first bleeding event, and this association was observed across all bleeding severities (Fig. 5). Based on the magnitude of relative risks for prognostic outcomes, three levels of bleeding severity were identified: The greatest prognostic risk was observed in hospitalised+MS bleeding (class I), followed by hospitalised or primary care+MS or inferred bleeding (class II). The lowest prognostic risk was associated with primary care bleeding (class III).

Compared to patients with no bleeding, the adjusted HR for all-cause mortality was 2.97 (2.84, 3.12) for

class I bleeding and 1.23 (1.19, 1.27) for class III bleeding. Similarly, the adjusted HR for cardiovascular death, stroke or MI events was 2.55 (2.38, 2.74) for class I and 1.08 (1.04, 1.13) for class III bleeding.

Risk of recurrent bleeding increased following an initial bleeding event (Additional file 1: Figure S8). The cumulative risks were greater if the initial bleeding event had further markers of severity. The 5-year recurrent event rates of any bleeding and fatal, hospitalised+MS or primary care+MS bleeding were 32.4% (31.8, 33.0), and 8.3% (7.9, 8.6), respectively. Amongst patients who initially experienced a bleeding event with markers of severity, their 5-year recurrent event rate was 37.4% (36.0, 38.8) for any bleeding and



23.1% (21.9, 24.3) for fatal, hospitalised+MS or primary care+MS bleeding.

Discussion

In a population-based study of linked primary care and hospital EHR in 128,815 patients with newly diagnosed common CVDs, we found that bleeding has doubled in incidence since 1998, affects 1 in 4 patients and is associated with poor prognosis in terms of all-cause mortality and subsequent atherothrombotic events. The phenotype algorithms made available here distinguish 3 prognostic classes of bleeding severity which may be used by health systems and public health authorities to focus efforts to tackle the growing population impact of bleeding on health outcomes.

Bleeding EHR phenotype algorithm: importance of linked electronic health records

We developed standardised and replicable EHR phenotyping algorithms for bleeding and severity measures based on available clinical information across primary and hospital care. The algorithms combine information on diagnoses, procedures, transfusion and haemoglobin. Unlike previous EHR studies which defined bleeding events using bleeding codes only, we demonstrated the depth of information readily available within linked EHR and the capability to achieve a more granular case definition by combining diagnosis terms with continuous measurements. Our results highlighted the importance of using multiple linked data sources for defining and validating the bleeding phenotype in EHR. No individual data source used in this study had complete coverage of coded bleeding diagnoses, transfusions, causes of death and other bleeding relevant data, and only 13.2% of bleeding cases were captured in multiple data sources (Additional file 1: Figure S4). Individual components of the phenotype, such as subgroups of the bleeding codes, have been validated in previous studies in CPRD [24], HES [23] and other EHR data sources [19–22, 25, 26], and our analysis of outcomes following bleeding adequately reflected expected results across levels of bleeding severity. It has been previously shown that using hospital discharge coding alone misses bleeding events compared with a manual review of case notes [10]; nonetheless, our use of multiple sources of EHR led to the estimation of a higher incidence of bleeding at 1 year than in the study with manual case note review.

Validation of bleeding phenotype

We provide new evidence of the validity of ICD-10 codes used in our bleeding EHR phenotype algorithm. We found a PPV of 0.88, i.e. 88% of bleeding events identified by these codes were indeed bleeding events according to the independent review of the entire hospital

record by two clinicians, blinded to the ICD-10 code assignment. The true incidence of bleeding is likely to be even higher than that detected by existing EHR phenotypes. We found that hospital codes have a sensitivity of 0.71 for detecting bleeds in the validation sub-study. Previous reports of the sensitivity of EHR ICD code-based algorithms differ in methodology and report sensitivities ranging from 0.38 [10] to 0.80 [43]. In an analysis of MI patients in a randomised trial setting, the sensitivity of a bleeding algorithm using ICD-9 codes has been shown to be as high as 0.80 when considering all diagnosis and transfusion codes [43]. The higher sensitivity may reflect the younger mean age (60 years vs > 70 years) and the greater emphasis on complete coding for billing optimisation in the USA, compared to the UK. This highlights the potential importance of assessing the context-specific validity of EHR phenotypes in different EHR systems. Upon review of the false-negative cases in our validation sub-study (Additional file 1: Table S6), none had ICD-10 or OPCS-4 codes recorded for their hospitalisation that we could reasonably include in the bleeding phenotype algorithm in order to improve the sensitivity. There have been few previous studies of the validity of ICD-10 codes in the UK against full review of hospital records, partly due to the difficulties in accessing the hospital records; our informatics approach using CogStack [42] for validation is scalable, replicable, rapid and low cost. Due to privacy restrictions in accessing primary care free-text data for research purposes, we were unable to perform a validation sub-study to assess the performance of the non-hospital bleeding in the phenotype. However, previous studies have demonstrated evidence of the accuracy and validity of primary care records and bleeding definitions [24, 44].

Ascertaining the validity of EHR phenotypes is multifaceted and may be determined by comparing the event rates and prognosis with previously published estimates [45]. Further evidence of the ability of the EHR phenotype reported here to detect bleeds comes from comparing the absolute risks that we report with studies based on manual adjudication. We found a risk of bleeding of 7% at 1-year post-MI, compared to 5.0% (based on medical claims) and 5.4% (based on physician adjudicated) [43]. Our findings were consistent with prior studies of bleeding trends over time [46], risk [43] and prognosis [23, 47, 48]. Nonetheless, efforts are required by health systems to improve the quality and completeness of data to increase the sensitivity of EHR phenotypes.

Bleeding EHR phenotype: inferring bleeding events

A previous study showed that it is appropriate to infer disease cases in EHR where diagnosis codes are absent [28]. We identified 1144 patients with no coded bleeding diagnosis present but exhibiting signs or symptoms of

bleeding, such as low haemoglobin, iron deficiency anaemia or with a recorded bleeding-related procedure, excluding cases where bleeding may not be the cause of these signs, symptoms and procedures (i.e. cancer, liver and renal diseases). This highlights the potential of looking beyond diagnosis codes in EHR to obtain more accurate estimates of bleeding in safety studies of antithrombotic use. This method requires validation, and cases identified using this method should be considered possible bleeding events and not definite.

Bleeding incidence in cardiovascular disease populations

At 5 years of follow-up, one in four patients with CVD had any bleeding event and 6.5% had fatal or severe bleeding. We provided a direct comparison of bleeding within four CVDs with varying degrees of antithrombotic use (Additional file 1: Table S9). AF had the highest bleeding 5-year rates both for any bleeding (29.1%) and fatal, hospitalised+MS or primary care+MS bleeding (9.9%). This is likely to reflect the higher use and longer duration of prescribed VKA and dual and triple therapy in AF patients. However, the incidence of bleeding in MI, unstable angina and stable angina patients was still relatively high.

Time trends in bleeding rates over the study period

So far, as we are aware, there have been no previous studies evaluating the time trends in bleeding incidence in common CVDs. In our study, we found that the rates of hospitalised bleeding per 1000 patients more than doubled from 1.02 in 1998 to 2.68 in 2009. We hypothesised that the increased use of antithrombotic therapies during this period would be associated with an increased incidence of bleeding. We indeed did identify increases in rates of hospitalised+MS, hospitalised and primary care bleeding events over time, consistent with an increase over the same time period. However, based on the results of our study, we cannot distinguish the relative contributions to the observed increase in bleeding incidence of the increasing range of available antithrombotic therapies, widening indications and changing guidelines for their use over time. Because hospitals receive reimbursement based on the ICD codes at discharge [49], it is possible that the observed increase in the rate of bleeding is partly artefactual, i.e. due to better recording over time. However, there are three lines of evidence against such an artefact: (1) we also observed increases in the rate of bleeding in an entirely separate source of data from primary care, used for clinical decision making without any financial incentives to record bleeding events; (2) this increase is consistent with previous evidence, of the increase in rates of intracerebral haemorrhage in the UK between 1981 and 2006 [46]; and (3) prescribing of antithrombotic therapies, which is

known to increase the risk of bleeding complications, has increased during the study period.

Prognosis following bleeding

These bleeding events were associated with poor outcomes suggesting an increasing burden of bleeding on healthcare systems and costs in England. Our analysis of prognosis following a non-fatal bleeding event identified three distinct levels of severity: I, hospitalised+MS; II, hospitalised, primary care+MS or inferred bleeding; and III, primary care (Fig. 5). This goes beyond the usual dichotomised classification of bleeding as either major or minor that is commonly reported. Increased bleeding severity was strongly associated with increased risks of all-cause mortality and atherothrombotic events. In particular, we found that bleeding diagnosed in primary care, without acute hospitalisation, was associated with adverse prognosis, both as class II and as class III (with and without associated markers of severity, respectively). Thus, all types of bleeding captured by the phenotype are clinically relevant. The term 'minor bleeding' may be misleading for clinicians, suggesting that no further action is required; while our study suggests that even a bleed in primary care without additional markers of severity is associated with 23% increased risk of death. Our findings are consistent with a previous study of bleeding in AF trial participants which found impaired health state utility even amongst 'minor' bleeds [48]. While we have identified associations between bleeding and prognosis, in our present analyses, we cannot claim these associations to be causal.

Limitations of EHRs

EHRs have strengths and limitations for defining bleeding. Strengths include the availability of relevant, constantly updated information, at nationally representative scale, with the opportunities for international comparison [17] and the low cost of acquiring the information. The key limitations are the lack of structured information (e.g. on bleeding severity) and inconsistency of data models in different EHR systems, which makes it difficult to combine data from multiple sites. Widespread adoption of clinically led, standardised data models such as the openEHR framework (<https://www.openehr.org/>) will help. A second limitation is that much of the information in EHR systems is in free text, which is difficult to access for research and to interpret. At a national scale, information is lacking on acute haemoglobin change, the number of units transfused and other details of bleeding to support the classification of bleeding severity. In clinical practice, these markers are used to assess bleeding severity and have high prognostic value [50]. Their addition to EHR phenotypes would be an important refinement to bleeding definitions. We showed

some evidence that haemoglobin drop might contribute to defining bleeding severity, but our data lacked haemoglobin values measured within hospital admissions. The prescribing data reported here was confined to primary care and did not include drugs prescribed during hospitalisation or over-the-counter aspirin. Therefore, the rates of prescribing reported may underestimate the true rates.

Clinical implications

Our study provides evidence of an iatrogenic epidemic, demonstrating the public health burden of increasing bleeding incidence and adverse prognosis, and suggests three clinical implications.

First, by better identifying the bleeding risks and events in EHR, the decision-making around antithrombotic therapy may be improved. It has been shown that AF patients have been prescribed oral anticoagulants despite being contraindicated due to bleeding risk, indicating that patients and clinicians may outweigh the benefits of stroke prevention over the possibility of major bleeding [51]. Furthermore, bleeding has been shown to be associated with discontinuation of warfarin [52] thus highlighting the challenge of managing benefits and harms of antithrombotic therapy. Clinicians should ensure that the decision to prescribe antithrombotic therapy is based on a personalised evaluation of both bleeding risk and atherothrombotic risk in combination with trial results [53]. Such an approach tailors drug treatment decisions to an individual's expected net benefit and is able to incorporate a patient's utility (or disutility) from bleeding and atherothrombotic events, for example, in the setting of prolonged dual antiplatelet therapy, have demonstrated the validity and feasibility (with web calculators) of such an approach using readily available clinical data [53]. Second, clinicians should be aware that patients who experience bleeding events, even those which are not hospitalised, are at particularly high risk and may warrant more intense monitoring [48]. Third, we propose that bleeding events are continually monitored and reported by organisations as part of the quality of care and outcome reporting not just in single cardiovascular diseases, but across whole health systems and whole populations. In order to do this, health systems need open and, where possible, international standards for EHR bleeding phenotypes, which will require further manual, expert refinement, in the light of system changes and ongoing evaluations of accuracy. Indeed, one general population survey of adults aged 45–75 years conducted in the USA reported antiplatelet use in 47% despite the small proportion of participants with established cardiovascular disease [54]. We have shown that the severe bleeding EHR phenotypes reported here closely match the endpoints used in trials [29]. This suggests that linked EHR can be used in ongoing reporting

to estimate the real-world impact of interventions, such as the introduction of new drugs or changes in clinical guidelines or health policy.

Future research

International standards for the EHR definition of bleeding occurrence and severity using available national and regional clinical records and based on the approach described here should be developed. Transparent reporting of EHR phenotype algorithms is required in order to make bleeding research more replicable and to compare the incidence and prognosis of bleeding of different severities in different countries and across different health systems [17]. This is important to understand the extent to which, if any, newer antithrombotic agents such as direct oral anticoagulants and ticagrelor are halting the trend of increased incidence of bleeding or reducing the severity of bleeding events. The method validation of disease code-based EHR phenotypes against the full hospital record reported here is scalable to other diseases and other hospitals.

Conclusion

Bleeding is a major public health problem; it is common in patients with CVD, the incidence of hospitalisation for bleeding is increasing, and it is associated with high mortality. The comprehensive and reproducible bleeding EHR phenotype with three levels of severity that we have developed is informative in mortality, risk of fatal or non-fatal atherothrombotic events, and recurrent bleeding. It can be used and further developed in EHR studies of bleeding outcomes or antithrombotic safety.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12916-019-1438-y>.

Additional file 1. Supplementary methods, figures and tables.

Abbreviations

ADP: Adenosine diphosphate; AF: Atrial fibrillation; CPRD: Clinical Practice Research Datalink; CVD: Cardiovascular disease; EHR: Electronic health records; HES: Hospital Episode Statistics; ICD: International Classification of Diseases; MI: Myocardial infarction; MINAP: Myocardial Ischaemia National Audit Project; MS: Markers of severity; NPV: Negative predictive value; ONS: Office for National Statistics; OPCS: Office of Population Censuses and Surveys Classification of Interventions and Procedures; PPV: Positive predictive value; VKA: Vitamin K antagonist

Acknowledgements

Not applicable.

Authors' contributions

LP, SCC, MPR, AS, SAM, VA, JTT, DB, RS, RD, AB, RSP, SD and HH contributed to the idea and design of the study. LP, JTT and DB extracted and prepared the data for analysis. LP performed the analysis. LP and HH drafted the manuscript, with revisions by SCC, MPR, AS, SAM, VA, JTT, DB, RS, RD, AB, RSP and SD. LP guarantees the quality and accuracy of the results presented. All authors read and approved the final manuscript.

Funding

The investigators on this study were supported by multiple funding sources including the Medical Research Council Prognosis Research Strategy (PROGRESS) Partnership (HH, grant G0902393/99558) and Medical Research Council Population Health Scientist Fellowship (S-CC: grant MR/M015084/1) and by awards to establish the Farr Institute of Health Informatics Research, London and Scotland from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, NIHR, National Institute for Social Care and Health Research, and Wellcome Trust (LP, S-CC, MPR). LP was supported by an AstraZeneca PhD studentship. The views expressed in this paper do not necessarily represent the views of the funding bodies. LP had full access to the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors had final responsibility for the decision to submit for publication.

HH is a National Institute for Health Research (NIHR) senior investigator. His work is supported by (1) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council; Economic and Social Research Council; Department of Health and Social Care (England); Chief Scientist Office of the Scottish Government Health and Social Care Directorates; Health and Social Care Research and Development Division (Welsh Government); Public Health Agency (Northern Ireland); British Heart Foundation; and Wellcome Trust. (2) The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC. (3) The National Institute for Health Research University College London Hospitals Biomedical Research Centre.

ADS is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

This study represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and the NIHR Biomedical Research Centre at University College London Hospitals. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Availability of data and materials

Access to the data for authorised researchers is provided within the UCL data safe haven (<https://www.ucl.ac.uk/isd/itforslms/services/handling-sens-data>) for researchers who have undergone data safe haven and information governance training. Linked CALIBER data (primary care data, Hospital Episode Statistics and Office for National Statistics mortality data) were obtained from the Clinical Practice Research Datalink (www.cprd.com). Access to data is only available once approval has been obtained through the individual constituent entities controlling access to the data. The phenotype algorithms described in this paper are freely available via the CALIBER website at www.caliberresearch.org, and the CALIBER data portal is available for consultation online at <http://www.caliberresearch.org>. The data are available under licence from CPRD. The phenotyping algorithms for bleeding and all EHR phenotypes used in this study are openly available at <https://www.caliberresearch.org/portal>.

Ethics approval and consent to participate

Anonymized primary care EHRs were obtained via the CPRD which has broad ethical approval for purely observational research using linked primary/secondary care data for supporting medical purposes that are in the interests of patients and the wider public. Linkages with hospital EHR were performed by NHS Digital, the statutory body in England responsible for providing core healthcare information technology and curating many of the national datasets. The study was approved by the Independent Scientific Advisory Committee of the Medicines and Health care products Regulatory Agency in the UK, protocol number 14_133.

Consent for publication

Not applicable.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation for the submitted work, no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work. JTT has received research grant funding from Pfizer-BMS (manufacturer of a direct oral anticoagulant) for a completed clinical trial on atrial fibrillation in stroke. AB has been an advisory panel member for Boehringer Ingelheim and Novo Nordisk in the last 3 years.

Author details

¹Health Data Research UK, University College London, 222 Euston Road, London NW1 2DA, UK. ²Institute of Health Informatics, University College London, London, UK. ³Leeds Institute of Health Sciences, University of Leeds, Leeds, UK. ⁴Health Data Research UK Scotland, Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK. ⁵Department of Stroke and Neurology, King's College Hospital NHS Foundation Trust, London, UK. ⁶Department of Biostatistics and Health Informatics, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, UK. ⁷Department of Clinical Pharmacology, University College London Hospital NHS Foundation Trust, London, UK. ⁸Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK. ⁹Bart's Heart Centre, Queen Mary University London, London, UK. ¹⁰The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK.

Received: 11 March 2019 Accepted: 1 October 2019

Published online: 20 November 2019

References

- Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(7):2160–236.
- Smith SC Jr, Allen J, Blair SN, Bonow RO, Brass LM, Fonarow GC, et al. AHA/ACC guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease: 2006 update: endorsed by the National Heart, Lung, and Blood Institute. *Circulation*. 2006;113(19):2363–72.
- Roffi M, Patrono C, Collet JP, Mueller C, Valgimigli M, Andreotti F, et al. 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2016;37(3):267–315.
- Bonaca MP, Bhatt DL, Cohen M, Steg PG, Storey RF, Jensen EC, et al. Long-term use of ticagrelor in patients with prior myocardial infarction. *N Engl J Med*. 2015;372(19):1791–800.
- Levine GN, Bates ER, Bittl JA, Brindis RG, Fihn SD, Fleisher LA, et al. 2016 ACC/AHA guideline focused update on duration of dual antiplatelet therapy in patients with coronary artery disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention, 2011 ACCF/AHA guideline for coronary artery bypass graft surgery, 2012 ACC/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease, 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction, 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes, and 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery. *Circulation*. 2016;134(10):e123–55.
- Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS the task force for the management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC endorsed by the European Stroke Organisation (ESO). *Eur J Cardiothorac Surg*. 2016;50(5):e1–e88.
- Lip GYH, Banerjee A, Boriani G, Chiang CE, Fargo R, Freedman B, et al. Antithrombotic therapy for atrial fibrillation: CHEST guideline and expert panel report. *Chest*. 2018;154(5):1121–201.

8. Halvorsen S, Storey RF, Rocca B, Sibbing D, Ten Berg J, Grove EL, et al. Management of antithrombotic therapy after bleeding in patients with coronary artery disease and/or atrial fibrillation: expert consensus paper of the European Society of Cardiology Working Group on Thrombosis. *Eur Heart J*. 2017;38(19):1455–62.
9. Niessner A, Tamargo J, Morais J, Koller L, Wassmann S, Husted SE, et al. Reversal strategies for non-vitamin K antagonist oral anticoagulants: a critical appraisal of available evidence and recommendations for clinical management—a joint position paper of the European Society of Cardiology Working Group on Cardiovascular Pharmacotherapy and European Society of Cardiology Working Group on Thrombosis. *Eur Heart J*. 2017;38(22):1710–6.
10. Li L, Geraghty OC, Mehta Z, Rothwell PM. Age-specific risks, severity, time course, and outcome of bleeding on long-term antiplatelet treatment after vascular events: a population-based cohort study. *Lancet*. 2017;390(10093):490–9.
11. Hansen ML, Sorensen R, Clausen MT, Fog-Petersen ML, Raunso J, Gadsboll N, et al. Risk of bleeding with single, dual, or triple therapy with warfarin, aspirin, and clopidogrel in patients with atrial fibrillation. *Arch Intern Med*. 2010;170(16):1433–41.
12. Sorensen R, Hansen ML, Abildstrom SZ, Hvelplund A, Andersson C, Jorgensen C, et al. Risk of bleeding in patients with acute myocardial infarction treated with different combinations of aspirin, clopidogrel, and vitamin K antagonists in Denmark: a retrospective analysis of nationwide registry data. *Lancet*. 2009;374(9706):1967–74.
13. Hamon M, Lemesle G, Tricot O, Meurice T, Deneve M, Dujardin X, et al. Incidence, source, determinants, and prognostic impact of major bleeding in outpatients with stable coronary artery disease. *J Am Coll Cardiol*. 2014;64(14):1430–6.
14. Mehran R, Rao SV, Bhatt DL, Gibson CM, Caixeta A, Eikelboom J, et al. Standardized bleeding definitions for cardiovascular clinical trials: a consensus report from the Bleeding Academic Research Consortium. *Circulation*. 2011;123(23):2736–47.
15. Wiviott SD, Antman EM, Gibson CM, Montalescot G, Riesenmeyer J, Weerakkody G, et al. Evaluation of prasugrel compared with clopidogrel in patients with acute coronary syndromes: design and rationale for the TRial to assess Improvement in Therapeutic Outcomes by optimizing platelet Inhibition with prasugrel Thrombolysis In Myocardial Infarction 38 (TRITON-TIMI 38). *Am Heart J*. 2006;152(4):627–35.
16. Chung SC, Gedeberg R, Nicholas O, James S, Jeppsson A, Wolfe C, et al. Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK. *Lancet*. 2014;383(9925):1305–12.
17. Rapsomaniki E, Thureson M, Yang E, Blin P, Hunt P, Chung S-C, et al. Using big data from health records from four countries to evaluate chronic disease outcomes: a study in 114 364 patients after myocardial infarction. *Eur Heart J Qual Care Clin Outcomes*. 2016;2(3):172–83.
18. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ*. 2013;346:f2350.
19. Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. *Thromb Res*. 2006;118(2):253–62.
20. Cunningham A, Stein CM, Chung CP, Daugherty JR, Smalley WE, Ray WA. An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiol Drug Saf*. 2011;20(6):560–6.
21. Friberg L, Skeppholm M. Usefulness of health registers for detection of bleeding events in outcome studies. *Thromb Haemost*. 2016;116(6):1131–9.
22. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiol*. 1996;7(1):101–4.
23. Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res*. 2012;12(1):392.
24. de Abajo FJ, Rodriguez LA, Montero D. Association between selective serotonin reuptake inhibitors and upper gastrointestinal bleeding: population based case-control study. *BMJ*. 1999;319(7217):1106–9.
25. Wahl PM, Rodgers K, Schneeweiss S, Gage BF, Butler J, Wilmer C, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf*. 2010;19(6):596–603.
26. Valkhoff VE, Coloma PM, Masclee GM, Gini R, Innocenti F, Lapi F, et al. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. *J Clin Epidemiol*. 2014;67(8):921–31.
27. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol*. 2012;41(6):1625–38.
28. Morley KI, Wallace J, Denaxas SC, Hunter RJ, Patel RS, Perel P, et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. *PLoS One*. 2014;9(11):e110900.
29. Timmis A, Rapsomaniki E, Chung SC, Pujades-Rodriguez M, Moayyeri A, Stogiannis D, et al. Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations. *BMJ*. 2016;353:i3163.
30. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet*. 2014;383(9932):1899–911.
31. Shah AD, Langenberg C, Rapsomaniki E, Denaxas S, Pujades-Rodriguez M, Gale CP, et al. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1.9 million people. *Lancet Diabetes Endocrinol*. 2015;3(2):105–13.
32. Pujades-Rodriguez M, George J, Shah AD, Rapsomaniki E, Denaxas S, West R, et al. Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1 937 360 people in England: lifetime risks and implications for risk prediction. *Int J Epidemiol*. 2015;44(1):129–41.
33. Pujades-Rodriguez M, Timmis A, Stogiannis D, Rapsomaniki E, Denaxas S, Shah A, et al. Socioeconomic deprivation and the incidence of 12 cardiovascular diseases in 1.9 million women and men: implications for risk prediction and prevention. *PLoS one*. 2014;9(8):e104671.
34. Pujades-Rodriguez M, Duyx B, Thomas SL, Stogiannis D, Rahman A, Smeeth L, et al. Rheumatoid arthritis and incidence of twelve initial presentations of cardiovascular disease: a population record-linkage cohort study in England. *PLoS One*. 2016;11(3):e0151245.
35. Bell S, Daskalopoulou M, Rapsomaniki E, George J, Britton A, Bobak M, et al. Association between clinically recorded alcohol consumption and initial presentation of 12 cardiovascular diseases: population based cohort study using linked health records. *BMJ*. 2017;356:j909.
36. Shah AD, Denaxas S, Nicholas O, Hingorani AD, Hemingway H. Neutrophil counts and initial presentation of 12 cardiovascular diseases: a CALIBER cohort study. *J Am Coll Cardiol*. 2017;69(9):1160–9.
37. Gallagher AM, Puri S, Staa TV. Linkage of the General Practice Research Database (GPRD) with other data sources. *Pharmacoepidemiol Drug Saf*. 2011;20(5):230–364.
38. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)*. 2014;36(4):684–92.
39. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827–36.
40. National Institutes of Health. Electronic health records-based phenotyping 2014 [Available from: <http://rethinkingclinicaltrials.org/resources/ehr-phenotyping/>].
41. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antithrombotic medicinal products in non-surgical patients. *J Thromb Haemost*. 2005;3(4):692–4.
42. Jackson R, Kartoglu I, Stringer C, Gorrell G, Roberts A, Song X, et al. CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. *BMC Med Inform Decis Mak*. 2018;18(1):47.
43. Guimaraes PO, Krishnamoorthy A, Kaltenbach LA, Anstrom KJ, Effron MB, Mark DB, et al. Accuracy of medical claims for identifying cardiovascular and bleeding events after myocardial infarction: a secondary analysis of the TRANSLATE-ACS study. *JAMA Cardiol*. 2017;2(7):750–7.
44. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol*. 2010;69(1):4–14.
45. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc*. 2019. <https://doi.org/10.1093/jamia/ocz105>.
46. Lovelock CE, Molyneux AJ, Rothwell PM. Change in incidence and aetiology of intracerebral haemorrhage in Oxfordshire, UK, between 1981 and 2006: a population-based study. *Lancet Neurol*. 2007;6(6):487–93.

47. Boggon R, van Staa TP, Timmis A, Hemingway H, Ray KK, Begg A, et al. Clopidogrel discontinuation after acute coronary syndromes: frequency, predictors and associations with death and myocardial infarction—a hospital registry-primary care linked cohort (MINAP-GPRD). *Eur Heart J*. 2011;32(19):2376–86.
48. Wang K, Li H, Kwong WJ, Antman EM, Ruff CT, Giugliano RP, et al. Impact of spontaneous extracranial bleeding events on health state utility in patients with atrial fibrillation: results from the ENGAGE AF-TIMI 48 Trial. *J Am Heart Assoc*. 2017;6(8):e006703.
49. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)*. 2012;34(1):138–48.
50. Kikkert WJ, van Geloven N, van der Laan MH, Vis MM, Baan J Jr, Koch KT, et al. The prognostic value of bleeding academic research consortium (BARC)-defined bleeding complications in ST-segment elevation myocardial infarction: a comparison with the TIMI (Thrombolysis In Myocardial Infarction), GUSTO (Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries), and ISTH (International Society on Thrombosis and Haemostasis) bleeding classifications. *J Am Coll Cardiol*. 2014;63(18):1866–75.
51. O'Brien EC, Holmes DN, Ansell JE, Allen LA, Hylek E, Kowey PR, et al. Physician practices regarding contraindications to oral anticoagulation in atrial fibrillation: findings from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF) registry. *Am Heart J*. 2014;167(4):601–9 e1.
52. O'Brien EC, Simon DN, Allen LA, Singer DE, Fonarow GC, Kowey PR, et al. Reasons for warfarin discontinuation in the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF). *Am Heart J*. 2014;168(4):487–94.
53. Pasea L, Chung SC, Pujades-Rodriguez M, Moayyeri A, Denaxas S, Fox KAA, et al. Personalising the decision for prolonged dual antiplatelet therapy: development, validation and potential impact of prognostic models for cardiovascular events and bleeding in myocardial infarction survivors. *Eur Heart J*. 2017;38(14):1048–55.
54. Williams CD, Chan AT, Elman MR, Kristensen AH, Miser WF, Pignone MP, et al. Aspirin use among adults in the U.S.: results of a national survey. *Am J Prev Med*. 2015;48(5):501–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

