Supplementary Materials

Healthc Inform Res. 2022 April;28(2):112-122.
https://doi.org/10.4258/hir.2022.28.2.112
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Supplement A.

Supplementary Methods

Phase 1: Conversion of Source Data to OMOP-CDM

1.1 Drug exposures

In the "Drug Exposure" table of the CDM, the individual drug ingredient names in the source table were matched with the "Concept Name" column in the OMOP vocabulary (RxNorm). Some drugs were displayed in the source database as combination products, with separate ingredients represented by their generic drug names. For example, the concept of "Amoxicillin-Clavulanic Acid" did not match directly with a drug concept in the OMOP vocabulary (Figure 2). In this case, the individual active pharmaceutical components of the drug combination were traced back to the prescription order and mapped to the RxNorm code "617296 (Amoxicillin 500 MG / Clavulanate 125 MG Oral Tablet)" and mapped to OMOP Concept ID "1713694" as its source and standard Concept IDs.

1.2 Diagnosis codes

The hospital had migrated from ICD-9 to ICD-10 for recording diagnoses during the 3-year period, requiring mapping of both ICD-9 and ICD-10 codes. ICD-9 and ICD-10 diagnosis codes were mapped to the same standard Concept ID in the "Condition Occurrence" table under the OMOP-CDM. For example, the ICD-9 and ICD-10 codes for "Type 2 Diabetes Mellitus without complications" are "250.02" and "E11.9", respectively and are mapped to the same OMOP standard Concept ID "201826" (Figure 3).
Some data cleaning steps were required during the ETL process for diagnosis codes. For example, in the source database, the diagnosis code "I.255 Ischaemic cardiomyopathy" had the decimal place shifted by two places and should have been "I25.5". The mapping was rectified by moving the decimal two places to the right. Codes that had more granularity than present in the OMOP vocabulary were mapped to the parent code, such as "E14.69 Unspecified diabetes mellitus" which was mapped to the Concept ID corresponding to the parent code of "E14.6. Unspecified diabetes mellitus" in ICD-10.

1.3 Laboratory tests and investigations

In-house codes were used at source for laboratory tests. The hospital provided a mapping for a portion of the internal codes to LOINC codes (used by OMOP-CDM), which could then map to the OMOP concept IDs in the "Measurements" table. For the remaining laboratory tests, the description of the test and laboratory units were used to bridge the source data to LOINC codes using Athena[1], an open-source tool for distributing and browsing the standardized vocabularies for all instances of an OMOP CDM, as a lookup resource. Once this mapping was completed, these LOINC codes were easily converted to OMOP Concept IDs.

As an example, in the upper branch of Figure 2, the concept of "J18.9 (Pneumonia, unspecified)" in ICD-10 could be mapped to "45601123" as its OMOP source Concept ID which was mapped to the OMOP standard Concept ID "255848" (or "233604007 (Pneumonia)" in SNOMED CT). If the concept did not exist in the OMOP vocabulary, it was mapped through a manual conversion process to an OMOP concept identifier. An example of this process is shown in the lower branch of Figure 2.

[1]Athena is an OHDSI vocabularies repository managed by Odysseus Data Services, Inc. It is hosted as a web application for distributing and browsing standardized vocabularies for all instances of an OMOP-CDM.

Phase 2: Illustrative Analysis Following CDM Conversion

2.1 Sample cohort assembly and drug exposure

The OACs included for analysis were warfarin (OMOP Concept ID: 1310149) and rivaroxaban (40241331). The diagnosis codes for AF included 'Atrial Flutter' (Concept ID: 314665), 'Atrial Fibrillation' (313217), 'Atrial Arrhythmia' (4068155), and 'Atrial Fibrillation and Flutter' (4108832). The Concept IDs for both thromboembolic and bleeding events are as shown in Table S1. For patients on warfarin, the presence of an International Normalised Ratio (INR; Concept ID: 3022217) measurement was used as an additional surrogate to ascertain continued warfarin exposure.

2.2 Visualizing comparative safety, effectiveness, and utilization for benefit-risk-assessments

Bleeding or thromboembolic events could occur at any point during the observation period, as long as the criteria for drug exposure was fulfilled (described above in '*2.1 Sample cohort assembly and drug exposure*'). The list of events considered (along with their OMOP Concept IDs) are listed in Table S1. Two queries were written in Structured Query Language (SQL) to identify bleeding and thromboembolic outcomes, respectively. Each query returned the sequence of OAC exposure followed by the occurrence of the event(s), if any. If a patient had both bleeding and thromboembolic events, the earlier outcome was considered.

Patients in each group were analysed for the type of diagnosis as well as for the use of any concurrent medications that may potentially exacerbate bleed risks, for example anti-platelets like aspirin or clopidogrel, for seven days prior to any bleeding event. Specifically, these concurrent medications were only included if the concurrent period of exposure fell within the preceding 7-day period of the event. For instance, if a patient was dispensed with aspirin for a period of three months on January 1, 2013 and the bleeding event occurred on March 1, 2013, this was considered as concurrent exposure since the dispensing period (and theoretical exposure period) includes the bleeding event date. Conversely, if another patient was dispensed with aspirin for a period of two weeks on January 1, 2013, and the event occurred on February 1, 2013, this would be excluded as concurrent exposure.

The original code from Hripcsak et al had focused their analysis on drug utilization (visualized via sunburst charts), which was insufficient for our purposes of layering on additional information on effectiveness and safety. Taking sunburst charts as inspiration, the use of a 100%, horizontally stacked, utilization-adjusted bar chart, that incorporates both drug utilization (represented by vertical bar thickness) as well as effectiveness and safety event proportions (represented by horizontal proportion within each bar) was used.

Table S1. Concept IDs and corresponding parameter descriptions for thromboembolic and bleeding events, based on the SNOMED-CT criteria

Thromboembolic event			
Concept ID	Description	Concept ID	Description
312327	Acute myocardial infarction	438447	Acute myocardial infarction of inferolateral wall
313226	Carotid artery occlusion	439296	Vertebral artery occlusion
314667	Nonpyogenic thrombosis of intracranial venous sinus	439693	True posterior myocardial infarction
315286	Chronic ischemic heart disease	440392	Retinal vascular occlusion
315296	Preinfarction syndrome	441579	Acute myocardial infarction of inferoposterior wall
315830	Prinzmetal angina	441874	Cerebral thrombosis
315832	Angina decubitus	443239	Precerebral arterial occlusion
316427	Aneurysm of coronary vessels	443454	Cerebral infarction
316437	Cerebral atherosclerosis	444406	Acute subendocardial infarction
319038	Postmyocardial infarction syndrome	4043731	Infarction - precerebral
321042	Cardiac arrest	4108356	Cerebral infarction due to embolism of cerebral arteries
321318	Angina pectoris	4110189	Cerebral infarct due to thrombosis of precerebral arteries
372924	Cerebral artery occlusion	4110190	Cerebral infarction due to embolism of precerebral arteries
373503	Transient cerebral ischemia	4110192	Cerebral infarction due to thrombosis of cerebral arteries
374055	Basilar artery syndrome	4110197	Occlusion and stenosis of multiple and bilateral cerebral arteries
374060	Acute ill-defined cerebrovascular disease	4111714	Cerebral infarction due to cerebral venous thrombosis, non-pyogenic
374371	Stenosis of precerebral artery	4111715	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
374384	Cerebral ischemia	4111716	Occlusion and stenosis of anterior cerebral artery
375557	Cerebral embolism	4111717	Occlusion and stenosis of posterior cerebral artery
376714	Vertebrobasilar artery syndrome	4112023	Occlusion and stenosis of middle cerebral artery
377001	Nonruptured cerebral aneurysm	4112024	Occlusion and stenosis of cerebellar arteries
380747	Cerebral arteritis	4112026	Sequelae of cerebral infarction
381316	Cerebrovascular accident	4120088	Cardiac arrest with successful resuscitation
381591	Cerebrovascular disease	4185117	Vertebral artery obstruction
433505	Subclavian steal syndrome	4185932	Ischemic heart disease
433783	Pulmonary artery aneurysm	4288310	Carotid artery obstruction
434056	Late effects of cerebrovascular disease	4317150	Sudden cardiac death
434376	Acute myocardial infarction of anterior wall	4334245	Retinal artery occlusion
434656	Vertebral artery syndrome	37115756	Dissection of coronary artery
436706	Acute myocardial infarction of lateral wall	40481919	Coronary atherosclerosis
437308	Basilar artery occlusion	40484167	Arteriosclerosis of artery of extremity
437540	Central retinal artery occlusion	44782775	Peripheral vascular disease associated with another disorder
438168	Aneurysm of heart	44784623	Acute coronary artery occlusion not resulting in myocardial infarction
438170	Acute myocardial infarction of inferior wall	46270031	Cerebral infarction due to occlusion of precerebral artery
438438	Acute myocardial infarction of anterolateral wall		

Table S1. Continued

Bleeding event			
Concept ID	Description	Concept ID	Description
26727	Hematemesis	4027729	Acute duodenal ulcer with hemorrhage
28779	Bleeding esophageal varices	4049659	Subcortical hemorrhage
75043	Hemarthrosis of knee	4077201	Subarachnoid hemorrhage from basilar artery aneurysm
76784	Hemarthrosis	4077958	Subarachnoid hemorrhage from anterior communicating artery aneurysm
79864	Hematuria syndrome	4077959	Subarachnoid hemorrhage from posterior communicating artery aneurysm
192671	Gastrointestinal haemorrhage	4078446	Subarachnoid hemorrhage from middle cerebral artery aneurysm
193795	Acute gastric ulcer with hemorrhage but without obstruction	4103703	Melena
195321	Postmenopausal bleeding	4108952	Subarachnoid hemorrhage from carotid siphon and bifurcation
196442	Chronic gastric ulcer with hemorrhage and with perforation but without obstruction	4110185	Intracerebral hemorrhage, intraventricular
197018	Chronic gastric ulcer with hemorrhage but without obstruction	4110186	Intracerebral hemorrhage, multiple localized
261687	Hemoptysis	4111708	Subarachnoid hemorrhage from vertebral artery
315276	Vitreous haemorrhage	4111709	Non-traumatic subdural haemorrhage
376713	Cerebral haemorrhage	4174044	Chronic peptic ulcer with hemorrhage
432869	Hemorrhagic disorder due to circulating anticoagulants	4176892	Cortical haemorrhage
432923	Subarachnoid hemorrhage	4211001	Chronic gastric ulcer with hemorrhage
433515	Chronic gastrojejunal ulcer with haemorrhage	4218781	Cerebral hemisphere hemorrhage
434402	Acute duodenal ulcer with hemorrhage but without obstruction	4231580	Acute gastric ulcer with hemorrhage
436148	Chronic duodenal ulcer with hemorrhage but without obstruction	4232181	Chronic duodenal ulcer with hemorrhage
436430	Nontraumatic extradural hemorrhage	4294973	Chronic gastric ulcer with hemorrhage and with perforation
437312	Bleeding	4317284	Respiratory tract haemorrhage
438468	Acute gastrojejunal ulcer with hemorrhage but without obstruction	4319328	Brain stem hemorrhage
439040	Subdural hemorrhage	4326561	Cerebellar hemorrhage
439847	Intracranial hemorrhage	40483641	Hemorrhage into peritoneal cavity
443530	Hematochezia	40492969	Intraparenchymal hemorrhage of brain

Supplementary Results

Table S2. Follow-up time and event rates in the overall cohorts

	Singapore	South Korea
Median follow-up time (day)	581	2,197
Total follow-up time (day)	217,270	5,645,901
Number of bleeding events observed	89	213
Incidence rate of bleeding events per 100,000 person years	14,951	1,377
Number of thromboembolic events observed	47	283
Incidence rate of thromboembolic events per 100,000 person years	7,896	1,830

Table S3. Follow-up time and event rates in the landmark analysis performed at 3 months

	Singapore	South Korea
Median follow-up time (day)	91	91
Total follow-up time (day)	31,425	183,301
Number of bleeding events observed	25	29
Incidence rate of bleeding events per 100,000 person years	29,037	5,775
Number of thromboembolic events observed	9	36
Incidence rate of thromboembolic events per 100,000 person years	10,453	7,169

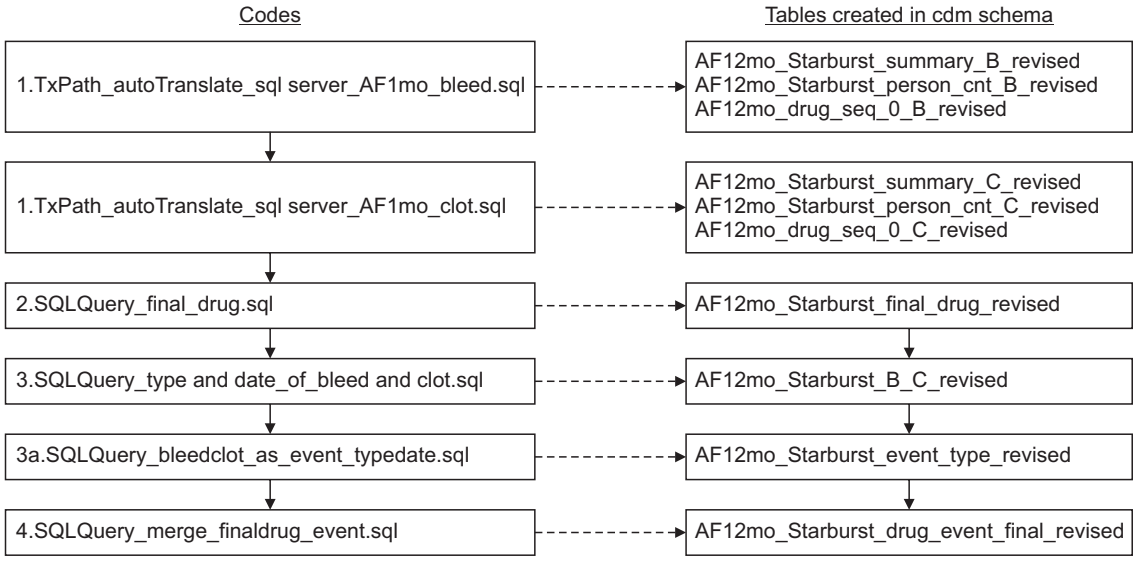
Supplementary Codes

SQL and R codes used to generate cohorts, analysis, and visualisations are attached separately.

Supplement B.

SQL scripts

1. Pre-requisites: cdm stored in MSSQL database
2. These scripts assume that the common data model tables are stored in an MSSQL database "cdm" under the schema "dbo". Hence before running the codes, please replace "cdm.dbo" with the name of the schema where your cdm database resides.
3. Please run the codes with starting with 1,2,3,3a and 4 in sequence to generate the tables in cdm schema, as follows:



4. SQL files with an "A" in front of the title are used to extract the data related to demographics, concurrent medications and duration for the patients in the cohort. This is for further analysis of the characteristics of patients, and not needed for generating the relative utilisation, safety and effectiveness chart.

Code	Data generated
A.SQL Query_concur_med_bleed.sql	Table of last concurrent medications (aspirin, NSAIDs, and antiplatelets) taken by patients in the cohort who experienced bleeding
A.SQLQuery_bleeds_and_clots.sql	Dates of earliest bleed and clot events of patients in the cohort
A.SQLQuery_demographics_bleed.sql	Demographic data (year of birth, ethnicity, gender and race)
A.SQLQuery_diagnosis.sql	Atrial fibrillation diagnoses codes of patients in the cohort.
A.SQLQuery_duration_bleed.sql	Time from taking the anticoagulant to bleeding event - Includes patients who have experienced bleeding, including those who had a clot first - drug_concept_id gives the concept_id of the anticoagulant time_to_bleed refers to the time between the start of the drug era and the condition occurrence (i.e. bleeding) - RowNum refers to the sequence of the drug era, in descending order. To get info on the earliest drug era (which may or may not be the same as the final drug used by the patient), filter for RowNum = 1
A.SQLQuery_duration_clot.sql	Time from taking the anticoagulant to clot event - Includes patients who have experienced clots, including those who had a bleed first - drug_concept_id gives the concept_id of the anticoagulant - time_to_bleed refers to the time between the start of the drug era and the condition occurrence (i.e. bleeding) - RowNum refers to the sequence of the drug era, in descending order. To get info on the earliest drug era (which may or may not be the same as the final drug used by the patient), filter for RowNum = 1

R codes

1. Pre-requisites: R and RStudio with the following libraries installed: ggplot2, readxl, stringr
2. The final table created in step 3 of the SQL section, AF12mo_Starburst_drug_event_final_revised, will have 3 columns: “drug”, “number” and “event” and should be copied and saved as an excel file called bleed_clots_first_occurrence.xlsx.
3. In line 1 of the R-code “Relative_safety_effectiveness_plot.R”, change the working directory to the directory containing bleed_clots_first_occurrence.xlsx
4. In line 43 of the R-code, change the number in the title to the number of patients in the final cohort (refer to AF12mo_Starburst_summary_B_revised – “Number of persons in final qualifying cohort”).
5. Running the R code will generate the Relative effectiveness, safety and utilisation chart.
6. To create multiple charts in one diagram, use the other R code “Relative_safety_effectiveness_plot_multiple.R”