Research project title

Automated detection of adverse drug events from older inpatients' electronic medical records using structured data mining and natural language processing

Study Type:	Further use of biological material and health-related personal data for research in absence of informed consent in accordance with Article 34 HRA
Study Registration:	The clinical trial will be registered through clinical trial.org once the appropriate ethics committee has approved it.
Study Identifier:	**SwissMADE**
Principal Investigator:	Prof. Chantal Csajka
Sponsor:	Service of Clinical Pharmacology, CHUV Rue du Bugnon 46, 1011 Lausanne - SUISSE
Investigational Product:	Antithrombotic drugs
Protocol Version and Date	Version 2.0 dated 20.03.2018

Principal Investigator:

Clinical Pharmacology CHUV

Pr. Chantal Csajka, PhD
Service of Clinical Pharmacology
Rue du Bugnon 17
CHUV. 1005 Lausanne – CH
Phone: +41 79 556 60 46
e-mail: Chantal.Csajka@chuv.ch

Co-Investigators:

Lausanne University Hospital (CHUV)

Dr. Marie-Annick Le Pogam (MD,MPH)
Institute of Social and Preventive Medicine (IUMSP)
CHUV. 1000 Lausanne – CH
Phone : +41 21 314 89 59
e-mail : marie-Annick.Le-Pogam@chuv.ch

Prof. Bernard Burnand (MD,MPH)
Institute of Social and Preventive Medicine (IUMSP)
CHUV. 1000 Lausanne – CH
Phone : +41 21 314 72 55
e-mail : bernard.burnand@chuv.ch

PD Dr. Pierre-Olivier Lang (MD,MPH,PhD)
Geriatric and Geriatric Rehabilitation Division
CHUV. CH-1066 Epalinges
Phone : +41 21 314 62 07
e-mail : pierre-olivier.lang@unil.ch

Geneva University Hospital (HUG)

Pr. Christian Lovis (MD, MPH)
Division of Science and Medical Information
HUG. CH-1205 Genève
Phone : +41 22 372 88 83
e-mail : christian.lovis@hcuhe.ch

Dr. Nicole Vogt-Ferrier (MD)
Division of Clinical Gerontopharmacology
HUG. CH-1211 Genève
Phone : +41 22 30 565 18
e-mail : nicole-b.vogt-ferrier@hcuge.ch

University Hospital Zurich (USZ)

Dr. Patrick E. Beeler (MD)
Department of Internal Medicine
USZ. CH-8091 Zurich
Phone : +41 44 255 32 99
e-mail : Patrick.Beeler@usz.ch

Dr. Fabio Rinaldi (PhD)
Institute of Computational Linguistics
UZH. CH-8091 Zurich
Phone : +41 44 635 7132
e-mail : fabio.rinaldi@uzh.ch

Baden University Hospital (KSB)

Dr. Monika Lutters (PhD)
Pharmacy
KSB. CH-5404 Baden
Phone : +41 56 486 39 43
e-mail : monika.lutters@ksb.ch

Signature Pages

Principal Investigator:

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the project according to the protocol and the local legally applicable requirements, and in particular the data protection therefore.

Site	CHUV
Principal Investigator:	Prof. Chantal Csajka (PhD)
Date: 16/04/2018	**Signature:** Chantal Csajka

Co-Investigators:

Co-Investigator: Dr. Marie-Annick Le Pogam (MD,MPH)

Date: 26/04 | 2018 **Signature:**

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: PD Dr. Bernard Burnand (MD,MPH)

Date: **March 24, 2018** **Signature:**

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: PD Dr. Pierre-Olivier Lang (MD,MPH,PhD)

Date: 26 mai 2018

Signature:

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: Pr. Dr. Christian Lovis (MD, MPH)

Date: 24.03.2018 **Signature:**

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: Dr. Nicole Vogt-Ferrier (MD)

Date: March 23rd, 2018 **Signature:** NVogt

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: Dr. Patrick E. Beeler (MD)

Date: April 3, 2018 **Signature:** P. Z

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: Dr. Fabio Rinaldi (MD)

Date: 10.04.2018 **Signature:**

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Co-Investigator: Dr. Monika Lutters (PhD)

Date: 29.03.2018

Signature:

Kantonsspital Baden AG
Dr. Monika Lutters
Spitalapothekerin FPH
CH-5404 Baden

With my signature I have approved the research protocol version 2 (dated 20.03.2018), and I confirm hereby to conduct the study according to the protocol, current version of the Declaration of Helsinki (1964), ICH-GCP guidelines and the local legally applicable requirements.

Abbreviations

ADEs - Adverse drug events

ATC - Anatomical therapeutic chemical classification system

CDS - Clinical decision support

CFR - Code of federal regulations

CIRS - Critical Incident Reporting System

CHOP Codes - Swiss classification of surgical interventions

CTCAE - Common terminology criteria for adverse events

EC - Ethics committee

EMRs - Electronic medical records

HRA - Human Research Act

HRO – Human Research Ordinance

ICD-10-GM Codes - The International Statistical Classification Of Diseases And Related Health Problems 10th revision, German Modification

LOINC - Logical observation identifiers names and codes

MedDRA - Medical dictionary for regulatory activities

NLP - Natural language processing

PRN orders - « Pro re nata » orders

SDM - Structured data mining

SNOMED-CT - Systematic nomenclature of medical clinical terms

SwissMADE DB - SwissMADE database

Table of Contents

1. Background and rational 15
2. Objectives 16
3. Patients' eligibility criteria 16
3.1. Inclusion criteria 16
3.2. Exclusion criteria 16
4. Study procedures 17
4.1. Data 17
4.1.1. Participating centers 17
4.1.2. Source data 17
4.2. Data analysis 21
4.2.1. Identification of antithrombotic ADEs 22
4.2.2. Elaboration of algorithms based on structured data (SDM) 22
4.2.3. Elaboration of algorithms based on free texts and narratives (NLP): 23
4.2.4. ADE detection tool assessment 26
5. Data management 26
5.1. Data extraction procedures 26
5.2. Data governance 26
5.2.1. Persons responsible for data protection 28
5.2.2. Persons entitled to receive the personal data 28
5.2.3. Data storage 29
6. Scientific methodology 29
6.1. Sample size calculation for the retrospective medical record review 29
7. Ethical and regulatory aspects 31
7.1. Motivation for a proxy consent obtained from the ethics committee 31
7.2. Notification requirements 31
8. Funding, publication and declaration of interest 31
9. References 32

1. Background and rationale

Patient injury resulting from medication use [1, 2], also known as adverse drug events (ADEs), is the second most frequent complication experienced by hospitalized patients, accounting for one-third (10% to 40%) of all hospital care-related adverse events [3-5]. Between 0.2% and 65% of hospitalized patients experience at least one ADE during their stay [6-9]. Apart from increasing patient morbidity and mortality, ADEs seriously impact on hospital utilization (i.e. increased lengths of stay and readmissions) and costs [2, 4, 9-12].

Older inpatients (aged ≥ 65 years) are especially at risk of ADEs. Over 30% of older inpatients experience at least one ADE during their hospital stay and up to 70% of these events are deemed preventable [8, 13-15]. ADEs also have more severe consequences in older inpatients, inducing or worsening frailty, functional and cognitive disability, and leading to loss of autonomy, frequent and longer hospital stays, as well as an increased risk of institutionalization and death [13, 14, 16].

Cardiovascular drugs are frequently associated with ADEs in older patients, in particular the classes of antithrombotic and antihypertensive drugs. Although recommended and widely used in elderly patients at increased risk of cardiovascular events, antiplatelet and anticoagulant treatments are highly associated with bleeding complications in the older population [17-19], and are a major cause of emergency department admissions and death in elderly patients [20-22]. Antithrombotic therapy is thus like the sword of Damocles, conferring protection against thrombosis while exposing to bleeding, with severe consequences in both cases [18, 23]. A recent study indicated that bleeding events were the most common ADE (36%) in patients over 65 [24]. The drugs most frequently involved in serious ADEs were antithrombotic agents (31%). Disregarding drug interactions, contra-indications and precautions caused 20% of ADEs, and drug overdoses were present in 17% [25]. In addition, combinations of these factors and inefficacy raise particular concerns from an individual and public health point of view.

Many interventions have been conducted to improve the quality and safety of medication prescribing in older inpatients [4, 26-30]. The most recent and significant ones include the provision of evidence- based prevention tools (e.g. specific guidelines [31], lists of criteria for inappropriate medication use [30, 32]), pharmacist-based interventions (e.g. patient counselling, medication reconciliation, clinical pharmacist rounding) [29], team-based interventions (e.g. multidisciplinary geriatric teams [27]), and information and communication technologies such as computerized clinical decision support (CDS) tools within computers. Some CDS tools, called clinical event monitors, detect and monitor drug-related problems and provide feedback through alerts and reminders when certain signals regarding pharmacy orders (i.e. sudden stop orders, antidote ordering, and dose correction orders), laboratory test results or patient characteristics have been triggered [33-38].

In Switzerland, CDS systems have been developed and implemented in hospital information systems [36,38]. However, current systems generate many false-positive alerts, target inappropriate prescriptions instead of clinically relevant ADEs, and do not consider the type of hospital or unit (e.g. medical, surgical), nor the patients' medical characteristics [33, 35, 39, 40]. Owing to their limitations (poor specificity, overalerting) [35, 39] and to the amount of information contained in electronic medical records (EMRs), new ADE detection and monitoring systems are currently being developed based on multiple sources of data (structured data and free texts from EMRs) and methods involving structured data mining (SDM) and natural language processing (NLP) [40-44]. These new systems' performance depends on their ability to process meaning and content from either structured or unstructured data. EMR entries that comprise drug names, doses, treatment durations and administration routes, laboratory results, diagnoses or procedure codes based on the International Classification of Diseases, Tenth Revision, German

Modification (ICD-10-GM) and the Swiss classification for surgical procedures (CHOP) can be considered as structured. Other information such as reasons for admission, patient history and conditions, nursing and medical progress notes, inpatient reports and discharge summaries are essentially available in unstructured free text or narratives.

Apart from CDS system targeting drug interactions, dosing errors, or prescription orders [25, 38, 45, 46], no ADE detection and monitoring system based on EMRs' structured data mining and natural language processing is currently available in Switzerland. Nor can "ready-made" systems from other countries be adapted as they were developed for EMRs written essentially in English [41, 42, 44].

2. Objectives

Our research hypothesis is that the automated detection of ADEs from EMRs using structured data mining and natural language processing (SDM and NLP) could significantly improve risk management and patient safety in hospitalized older inpatients with multimorbidity, frailty and polypharmacy. It could additionally provide reliable data on incidence of ADEs for health care professionals, patient safety organisations and policy-makers.

The main objective of the project is to develop and validate an electronic application for the automated detection of ADEs related to antithrombotics based on structured data and free text mining.

Secondary objectives are :

- To quantify the incidence of ADEs associated with and caused by antithrombotic drugs,
- To assess the causality, severity and preventability of detected ADEs induced by antithrombotic drugs,
- To develop strategies for the implementation of the project results to improve risk management of antithrombotic drugs in hospital settings.

3. Patients' eligibility criteria

3.1. Inclusion criteria

The project is a multicenter observational study using routinely collected health-related personal data.

The inclusion criteria are as follows:

- Patients aged ≥65
- A hospital admission between 01.01.2015-31.12.2016
- At least one antithrombotic drug prescription during each stay
- Hospital stays >24 h

3.2. Exclusion criteria

We will exclude from our study any patient for whom an explicit refusal to be involved in research projects or to give access to their personal health data is documented.

4. Study procedures

4.1. Data

4.1.1. Participating centers

Four hospitals will participate in the project:

- 2 hospitals in the French speaking part of Switzerland : Hôpitaux Universitaires de Genève (HUG) and Centre Hospitalier Universitaire Vaudois (CHUV)
- 2 in the German speaking part : Universitätsspital Zürich (USZ) and Kantonspital Baden (KSB)

4.1.2. Source data

This project will make use of health-related information routinely collected during daily practice. Relevant health-related data as defined below will be extracted by each participating hospital for all patient stays fulfilling the inclusion criteria.

For this project, we will focus on five different classes of antithrombotic drugs, i.e. heparins (unfractionated heparin and low molecular weight heparins), vitamin K antagonists (coumarin derivatives), DOACs (direct oral anticoagulants), fondaparinux (direct thrombin and factor Xa inhibitors), and antiplatelet drugs. The list of selected classes of drugs is presented in Table 1 and the list of drugs presented in Annexe 1.

Table 1: B01A Antithrombotic agents

ATC Code *	Classes of antithrombotic drugs
B01AA	Vitamin K antagonists
B01AB	Heparin group
B01AC	Platelet aggregation inhibitors excl. heparin
B01AE	Direct thrombin inhibitors
B01AF	Direct factor Xa inhibitors
B01AX05	Other antithrombotic agents : Fondaparinux

* ATC Code: Anatomical Therapeutic Chemical Classification System

Health-related data comprise: General administrative data, patient location and transfer within the hospital, pertinent clinical and laboratories biological measurements, drug-related prescriptions, diagnoses (ICD-10 codes) and surgical procedures (CHOP). All diagnostic or procedure codes will be considered a priori, and their pertinence with regard to haemorrhages or thromboses will be evaluated upon the elaboration of the algorithm. Dates will be provided with all extracted items.

A case identification number (CaseID) will be attributed to each patient's stay (from 1 to n). CaseID will be grouped by patient using a patient identification number (PID). The extracted items are presented in **Table 2** for structured data. Documents that will be extracted as free texts are presented in **Table 3**. Details of the extracted items are presented in Annexe 2.

Data will be split in a learning set and different validation sets for NLP and predictive analyses of ADEs (see 4.2)

Table 2 : Structured data extracted for the project

Data type	Extracted data	Comments (Unit)
General Administrative Data	Patient identification number	(value)
	Case identification number (admission ID, hospitalization ID, or stay ID)	(value)
	Insurance type	(category)
	Region of residence (MedStat region)	(category)
	Admission mode: admission via emergency department, planned admission, transfer, etc.	(category)
	Nationality	(category)
	Date of birth	(date)
	Gender	(category)
	Date of death (if any; enables calculation of in-hospital mortality)	(date)
Clinical measurements*	Blood pressure	(value)
	Weight	(value)
	Height	(value)
	Sum of alcohol withdrawal syndrome score	(value)
Patient location(s) and transfers*	Unit of hospitalization	(category)
	Transfers (medicine, surgery, intermediate care, intensive care)	(category)
	Date and time of admission	(date, time)
	Date and time of discharge	(date, time)
Diagnoses and procedures*	DRGs codes	(category)
	CHOP codes (with ancillary information)	(category)
	ICD-10 codes (with ranking information)	(category)
	Readmissions and reasons for readmissions (1st, 2d, 3rd, 4th, subsequent readmissions)	(category)
	Drugs coded for reimbursement	(category)

	Intensive Care Unit length of stay (in hours)	(category)
	Duration of mechanical ventilation (in hours)	(category)
	Disease severity and scores	(category)
	Nine Equivalents of Nursing Manpower use Score » (NEMS) score	(category)
Laboratory values* Laboratory results that were ordered or received within the time frame of any of the recorded/extracted stays (if available)		(value in specific unit)
	electrolytes and ions	
	Blood Ionogram (Sodium and Potassium)	(mmol/L)
	Serum lactate and bicarbonate	(mmol/L, mmol/L)
	Uric acid	(mmol/L)
	Urea	(mmol/L)
	Serum iron, Transferrin saturation and Serum Ferritin	(mmol/L , µg/L)
	enzymes	
	Serum aminotransferases	AST/ALT (UI/L)
	Serum 5'Nucleotidase activity	(UI/L)
	Creatine kinase (CK)	(UI/L)
	Gamma-glutamyltransferase (GGT)	(UI/L)
	Alkaline phosphatase (ALP)	(UI/L)
	Complete Blood count (CBC)	
	Red blood cell count	(Absolute value / mm3)
	Hemoglobin	(mmol/L)
	Hematocrit	(percentage of total blood volume)
	Mean Corpuscular Volume	(μ^3)
	White blood cell count	(Absolute value / mm3)
	Platelet count	(Absolute value / mm3)
	Reticulocyte count	(Absolute value / mm3)
	Hemostasis	
	Prothrombin time (PT)	(time)
	Activated partial thromboplastin time (APTT)	(time)
	Thrombin time (TT)	(time)

	International normalized ratio (INR)	-
	Plasma Fibrinogen	(g/L)
	Procoagulant balance	Antithrombin (g/L), Protein C and S (nmol/L), Anti-cardiolipin antibody (GPL Unit), Anti-bêta-2-glycoprotéine 1 antibody (GPL Unit)
	Individual coagulation factors	(% relative to a reference pool)
	Fibrinolysis	D-dimer (µg/L)
	Anticoagulation monitoring	Anti-Xa (% relative to a reference value), Anti-IIa (% relative to a reference value)
	Markers of coagulation	Thrombin-Antitrombin III Complex TAT (ng/ml), Fragment 1 + 2 of prothrombin (% relative to a reference value)
	Other	
	Albumin	(g/dL)
	Serum total protein	(g/dL)
	Oxygen saturation	(% relative to a reference value)
	C-reactive protein (CRP)	(mg/L)
	Myoglobin (Mb)	(µg/L)
	Troponin	(µg/L)
	Creatinine and Creatinine Clearance	(mg/L, ml/min)
	Total bilirubin, direct and indirect bilirubin	(mg/L, mg/L, mg/L)
	Glycated hemoglobin	(% relative to a reference value)
	Tumor markers available	-
Prescription/medication*	ATC code (and product ID)	(category)

All medication orders that have: (i) a planned start date ≤ discharge date AND *(ii)* a planned discontinuation date ≥ admission date	Information on dose and planned administration frequency, incl. unit (e.g. MG=milligrams)	(value, category)
	Information on administration route (e.g. intravenous administration vs. oral)	(category)
	PRN orders ("as needed": drugs available on patient's request, e.g. analgesics)	(category)
	Administrations performed (signed by nurses)	(category)

* (incl. time stamps), linked to patient ID and to case ID/admission ID

Table 3 : Free-texts and narratives extracted for the project

Extracted data	Comments (Unit)
Patient identification number (metadata)	(free-text)
Case identification number (admission ID, or hospitalization ID, or stay ID) (metadata)	(free-text)
Notes taken at admission	(free-text)
Discharge summaries and letters	(free-text)
Nurses' progress notes	(free-text)
Imaging/radiology reports	(free-text)
Specialists' (e.g. hematologist, cardiologist, angiologist, and particularly endoscopy reports) consultation notes	(free-text)
Clinical pharmacology or pharmacy service consultation notes	(free-text)
ADE / Pharmacovigilance Reports	(free-text)
CIRS Reports	(free-text)

4.2. Data analysis

All extracted data will be used to develop algorithms, which will identify ADEs. Thus, based on developed algorithms, we will be able to indicate for each hospital stay of the extracted database whether an ADE has potentially occurred. To test for the algorithms' accuracy, we will then randomly select a validation dataset from the extracted database and verify in the corresponding EMRs (gold standard) whether an ADE has truly occurred or not. As a result, algorithms will be improved according to the results of the

validation (maximization of sensitivity and specificity, as well as of positive and negative predictive values). In the end, validated algorithms will serve to identify ADEs accurately (validated outcomes). For a more detailed description of the development and validation process of algorithms, see section 4.2.4.
To identify best predictors of ADEs, we will also use predictive modelling, such as various machine learning algorithms or logistic regression models. Outcomes of interest will be the validated outcomes described above. Predictors (i.e. ADE risk factors) will be selected from structured and textual information available in extracted data. Predictive models will also undergo internal and external validation: k-fold cross-validation and accuracy assessment of predicted outcomes, respectively [47].

At last, we will study the causal relationship between ADE risk factors and validated outcomes using causal analysis [48, 49]

4.2.1. Identification of antithrombotic ADEs

In order to identify ADEs induced by antithrombotics, computational algorithms based on logical rules will be developed based on direct and indirect indicators of ADEs (Table 4). This study will focus on two types of events: hemorrhage and thromboembolic events, which will be defined according to international references [50-52]. Potential confounding factors in the causal relationship between antithrombotic drugs and related ADEs will be identified from the scientific literature (including concomitant drugs, patient characteristics and concomitant health conditions).

Table 4 : Direct and indirect ADE indicators

ADE markers	clinical signs or symptoms indicating that an antithrombotic-related ADE occurred
ADE triggers	clinical signs, symptoms, procedures, prescription and imaging orders, biological results that indicate that an antithrombotic-related ADE may have occurred
ADE confounding conditions / risk factors	conditions or factors that increase the potential for an antithrombotic-related ADE or for a spontaneous bleed or thombotic event to occur. Risk factors include patient characteristics, specific concomitant health conditions that may interact with antithrombotics, concurrent use of more than one antithrombotic
ADE causes	causal factors are risk factors that are responsible for antithrombotic-related ADEs, including inappropriate prescribing (under-, over-, mis-prescribing), interactions (drug-disease, drug- food and drug-drug interactions), inappropriate administration, and insufficient monitoring.

4.2.2. Elaboration of algorithms based on structured data (SDM)

Computational algorithms based on logical rules applied to structured data will be developed to identify ADE markers, triggers, confounding conditions /risk factors and causes.

Detection algorithms for clinical markers of ADE (i.e. hemorrhagic events or thromboembolism) and confounding clinical conditions (e.g. chronic liver or kidney disease, hypertension, diabetes, cancer,

multimorbidity) will target ICD-10-GM diagnostic codes in hospital discharge data. Regarding ADE triggers, clinical conditions (e.g. hypotension, shock, and acute kidney failure) and procedures (e.g. postoperative *control of haemorrhage*, drainage of hematoma, or surgical treatment of venous or arterial thromboembolism) will be identified from hospital discharge data by algorithms based on ICD-10-GM diagnostic codes and CHOP codes, respectively. Biological triggers of ADEs (i.e. abnormal laboratory values) will be detected by algorithms applied to laboratory results. Similarly, some algorithms based on prescription orders will search for pharmacological triggers of ADEs including sudden medication stop orders, antidote ordering, dose correction orders, under- and overdosing, misprescribing, insufficient monitoring, and drug-drug or drug-disease associations. Finally, some algorithms based on imaging orders will flag imaging triggers of ADEs (e.g. emergency head or abdominal CT-scan orders, upper endoscopy orders for haemostasis).

4.2.3. Elaboration of algorithm based on free texts and narratives (NLP):

Based on ADE markers and triggers, and confounding conditions, a set of relevant concepts will be identified in existing knowledge representation frameworks. This set of concepts will be organized in a coherent and pertinent taxonomy validated by pharmacologists, pharmacists, and geriatricians from the research team. This taxonomy will serve as a placeholder for a text-based approach of the variables of interest (e.g. clinical ADEs, laboratory or imaging results, comorbidity). For example, a structured rule for hyperkalemia could be "[LabVal: K+] $\geq$ 3.5 [unit: mmol/l]", whereas this would be expressed in a text (i.e. discharge letter, medical or nursing notes) in numerous ways, including "kaliémie élevée", "hyperkaliémie", etc. for EMRs written in French. This taxonomy will be encoded using standardized codes from existing classifications such as the Medical Dictionary for Regulatory Activities (MedDRA), the Systematic Nomenclature of Medical Clinical Terms (SNOMED-CT) and the Logical Observation Identifiers Names and Codes (LOINC) [53, 54].

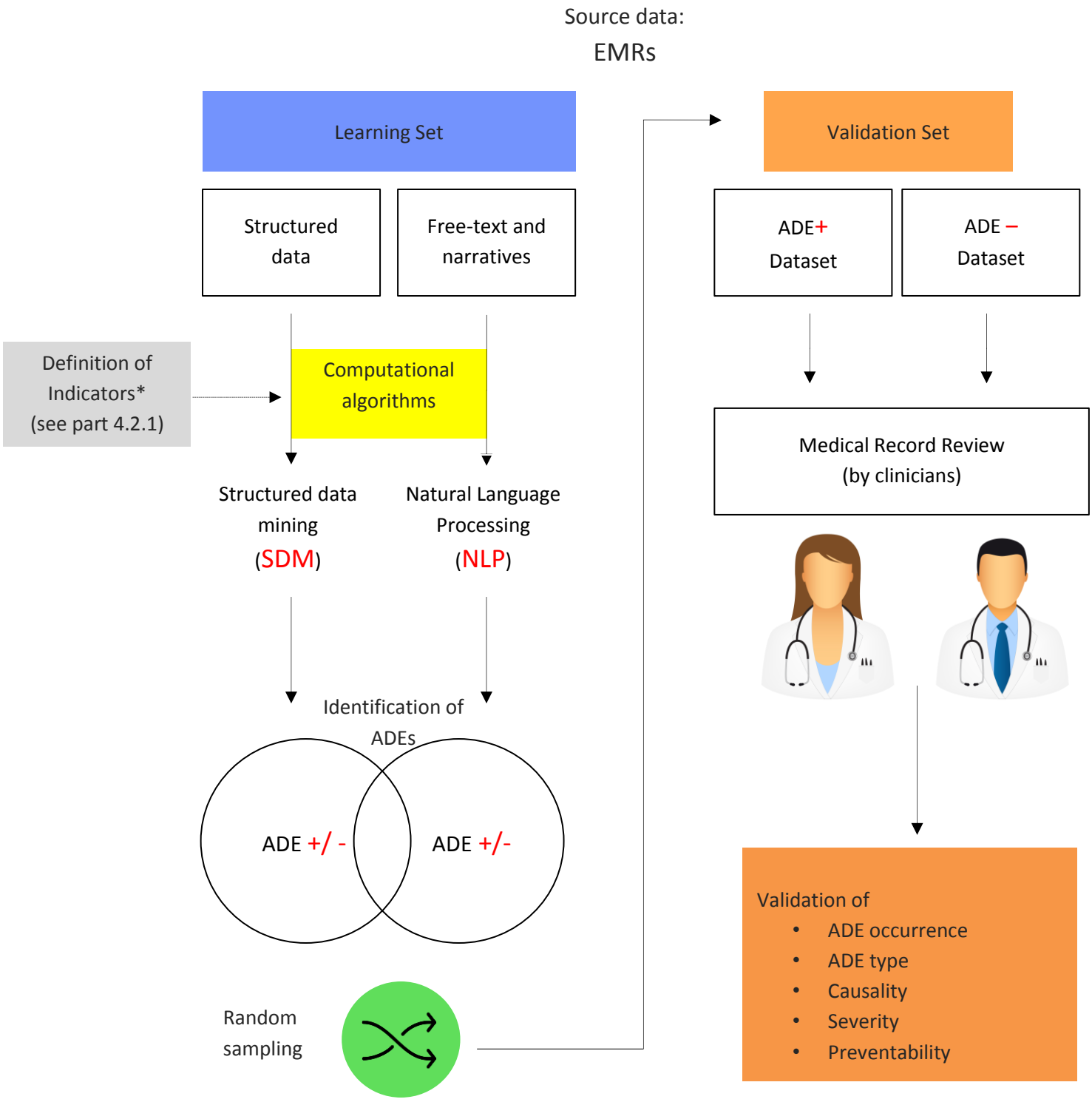
Already developed devices such as known mapping lists of free text medication orders and NLP algorithms to map drug orders to the ATC codes will help to make fast progress in terms of the inclusion of uncoded free text order entries (Table 5). The mapped ATC codes and the hierarchy of the ATC classification system will enable us to compare drugs and drug groups otherwise designated by brand names in free text orders.

Table 5: Examples of raw data medication orders that require a time-consuming mapping process in order to enable their consideration in the analysis.

Drug orders and free text entries (extracted raw data)	ATC code added by manual and automated mapping
Novalgin(Novaminsulfon sintetica 50% 1g/2ml	N02BB02
Novalgin1gr	N02BB02
Novalgin50%InjLsg1g/2ml	N02BB02
Novalgin i/v	N02BB02
Novaminsulfon (Inj Lös 1 g/2ml) Amp / Metamizol 500mg/ml	N02BB02
Novaminsulfon SIN injektionlösung	N02BB02
Minalgin 50% (Inj Lös 1 g/2ml) i.m./i.v. Amp / Metamizol 500mg/ml	N02BB02
Minalgin 500	N02BB02
Novaminsulfon 1g in 100ml NaCl 0,9%	N02BB02
Novalgin 1gr.	N02BB02

In the end, we will obtain for each 2015 and 2016 hospital stays of patients aged ≥65 years treated with antithrombotics a CaseID of positive ADE detected from the developped algorithms based on structured data (SDM), unstructured data (NLP), and on both types of data (SDM+NLP)(Figure 1).

Figure 1 : diagrams presenting the study procedure



ADE +/- = Hemorrhagic and thromboembolic events detected/ not detected

* defined according to scientific literature (standard pharmacological references) and clinical guidelines

4.2.4. ADE detection tool assessment

To assess the performance (sensitivity, specificity, positive and negative predictive values) of the ADE detection tool, a validation will be performed on a random sample of 600 hospital stays. ADE occurrence, type, causality, severity and preventability will be assessed by means of a patient medical record review, analysed by a team of pharmacologist, pharmacists and geriatrician from the research team (Figure 1).

To ensure that the ADE assessment and data abstraction are structured and reliable, pharmacologists and pharmacists from the research team will develop a common "ADE assessment form" both in French and German based on existing good pharmacovigilance practice rules, which will be disseminated in their original languages and English after study completion. Pairs of trained clinicians (pharmacists, pharmacologists, geriatricians) will then assess all selected medical records using this form. The causality between having taken a drug and the advent of an ADE will be assessed using existing causality assessment scales [55, 56]. The severity of the ADEs will be scored according to the Common Terminology Criteria for Adverse Events (CTCAE). Finally, an ADE will be deemed preventable if caused by a medication error that occurred during prescribing, transcribing, dispensing, administering, and monitoring or if it was due to a lack of medication adherence.

To test for the reliability of ADE assessment by pharmacists and pharmacologists, we will calculate intra- and inter-rater agreements for overall ADE occurrence, causality, severity, and preventability. For each pair of trained pharmacists and pharmacologists, the inter-rater agreement will be tested by comparing the results of the ADE assessment between members of the pair. To test for intra-rater agreement in each participating hospital, a random sample including 100 ADEs detected by SDM and NLP will be re-assessed by the assigned pair, 3 months after their first assessment. For both intra- and inter-rater agreements, the measure of agreement will be the Cohen's or uniform kappa statistic [57].

After excluding "false positive" ADEs, we will measure the cumulative incidence of "true positive" preventable ADE and non preventable ADEs for each hospital, medical units and overall.

5. Data management

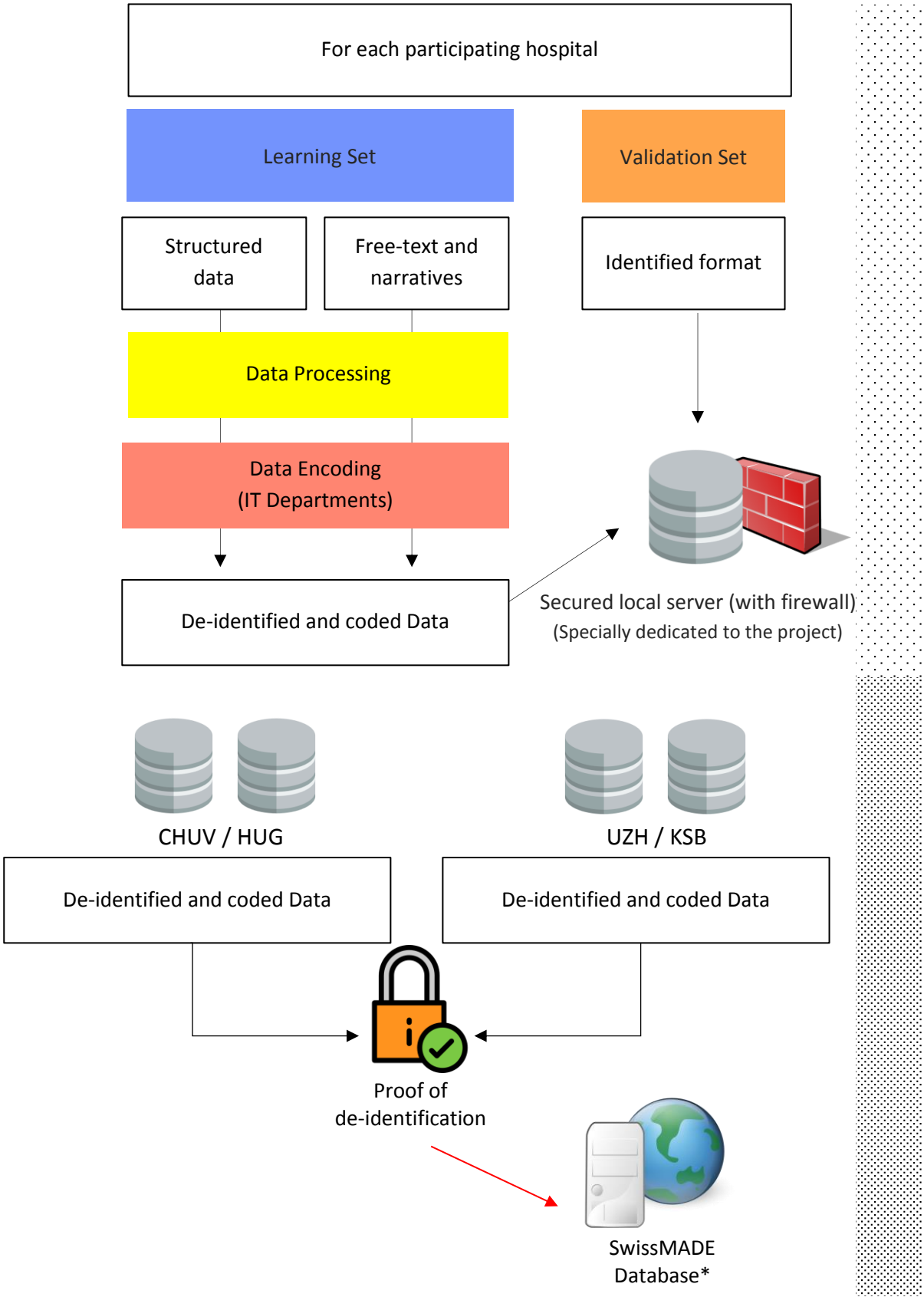
5.1. Data extraction procedures

Data will be extracted from inpatients' electronic medical records (EMRs) by the Information Technology (IT) departments of each hospital. Before being processed, structured data will be standardized in a unique common format (common data model) and unstructured data will be transformed in a machine-readable format, when necessary.
Regarding data extraction for the validation of the algorithms, individual data will be manually collected in EMRs at each hospital site and entered on a secured web-based electronic case report form that ensures traceability.

5.2. Data governance

As a consequence of the multi-site and bi-lingual (French and German) nature of the project, we will use a centralized and de-centralized data governance strategy, as presented in Figure 2.

Figure 2 : diagrams presenting the **de-centralized data governance strategy** (Step 1)



* Unless otherwise authorized, study investigators will only have access to this shared database

A two step approach will be undertaken for data processing.

Step 1. De-centralized data processsing

Raw data from the EMRs will be managed and processed by the IT team of each hospital according to established protocols. For structured data, the nominative identifiers will be coded (there are 18 identifiers to delete as described in HIPAA Privacy Rule, Code of federal regulations (CFR) - https://www.law.cornell.edu/cfr/text/45/part-164/subpart-E Title 45: Public Welfare, Subtitle A §164.514). Free text and narratives of the learning set will be de-identified and coded locally (by an encryption software) before being transferred to secured servers specially dedicated to the project within each hospital. Each hospital's IT team will be responsible for processing their own data. First data processing and analyses will be performed within each hospital.
EMRs used for the validation of the algorithms will be available in an identified format. Data will remain on a secured server in each hospital and made available to authorized investigators within each hospital.

Step 2. Centralized data processsing

Locally extracted coded data of the learning set from the German-speaking part of Swizlerland (UZH and Baden) and from the French-speaking part (CHUV and HUG) will be transferred to a centralized common SwissMADE database (SwissMADE DB). After proof of de-identification, structured items from free texts and narratives will be transferred to the SwissMADE DB.

One authorized person per site will be allowed access to personal data and will decide who can access to which data linked to which analysis (see 5.2.3). Remote access to the local working databases within each hospital and the SwissMADE database to authorized investigators will be made possible through a virtual Private Network (VPN).

5.2.1. Persons responsible of data protection

According to the governance plan described above, a main representative of the project within each hospital will take full responsibility of data use and protection. The responsible person per hospital is as follows:

CHUV	Pr. Chantal Csajka, PI of the project
HUG	Pr. Christian Lovis, co-investigator
UZH	Dr. Patrick E. Beeler, co-investigator
Baden	Dr. Monika Lutters, project partner

The main local representative will keep the correspondence file « coded/de-identified identifier »/« nominative identifier » in a secured place and separately from the dataset (this file will remain in each participating hospital). Access to this identifier file will be restricted to the IT team of each hospital and to the main representative of the project within each hospital.

5.2.2. Persons entitled to receive the personal data

The main representative for data protection per site will provide the authorisations to access data to other co-investigators. According to article 5 ORH, the use of personal health data is limited to those who need it to perform their tasks.

Personal data from the validation test will not be accessible by any investigators before completion of the algorithms. For validation purposes, such data will be accessible to authorized clinicians (clinical pharmacists, clinical pharmacologists and geriatricians) of the research team.

The SwissMADE database will be accessible for analyses to authorized investigators. The responsible investigator per site will make sure that all relevant information produced by each research lab will be transmitted to other groups. Extracted data might be reused for further related research projects.

5.2.3. Data storage

Each hospital's IT team will be responsible for processing their own data until transfer to the SDM and NLP working databases and the final transfer to the SwissMADE DB. The SwissMADE DB will be hosted in a secured environment backed up and linked to the academic network http://tribu.intranet.chuv/content_details.htm?cid=21945.

Since data are extracted from patients' medical records and not generated for the purpose of this study, they are submitted to the cantonal rules on health , which generally require storage for about 10 years. Any correction or deletion will be traced by the investigators in accordance with Article 5 of the law on data protection (LPD). The data will be stored until the publication of the results of the study. The correspondance file « coded identifier »/« nominative identifier » will be destroyed once the study is completed.

These measures comply with the requirements of Article 5 ORH on operational and technical measures to prevent unauthorized or inadvertent publication, modification, deletion and copying of personal health data.

5.2.4. Scientific steering committee

A scientific steering committee will be set up to allow regular discussions of the PNR74 study progress. All sub-projects ancillary to or related to the present project will be submitted to the scientific committee for approval. The scientific committee is composed of Prof. Chantal Csajka, Dr. Marie-Annick LePogam, Dr Pierre-Olivier Lang from the CHUV, Prof Christian Lovis, Dr Nicole Vogt-Ferrier from the HUG, Dr Fabio Rinaldi, Dr P. Beeler from Zürich and Dr. Monika Lütters from Baden.

6. Scientific methodology

6.1. Sample size calculation for the retrospective medical record review

The sample size was estimated to assess the performance of the SDM+NLP tool in detecting haemorrhagic adverse events. Indeed, the SDM+NLP tool is considered the critical outcome determining the feasibility of the project, and haemorrhage the most important adverse event related to antithrombotic drugs. We used a test-result based sampling method to minimize the number of medical records to be abstracted [58]. Given that CI is the cumulative incidence of ADEs detected from both structured and unstructured data ; N is the number of 2016 hospital stays of patients at risks (i.e. patients aged ≥65 years treated with antithrombotic drugs) ; P(ADE+) is the proportion of hospitals stays with an ADE detected by the SDM+NLP tool among all at-risk hospital stays (it is calculated as the number of true positive and false positive ADEs detected divided by N) ; Se is the expected sensitivity of the SDM+NLP tool (and Se 95%CI its 95% confidence interval) ; Sp is the expected specificity of the SDM+NLP tool ; PPV, its expected positive predictive value ; NPV, its negative predictive value ; we calculated sample sizes for CI ranging

from 3% to 24%, a desired Se of 80%, a 20% width for Se 95%CI, volumes of at-risk hospital stays ranging from N=2000 to 20'000 and a balanced sample of ADE+ and ADE- hospital stays (i.e. hospital stays with and without ADE, respectively) (Table 6). CI values were elicited from the literature (i.e. range for ADE cumulative incidence: 30-40% and range for proportion of haemorrhagic ADEs among ADEs: 10-40%)[13]. The values for N were estimated from annual numbers of at-risk stays in the 4 participating hospitals. We considered, in particular, selecting hospital units with a high prevalence of antithrombotic prescription (i.e. acute geriatric unit, internal medicine, cardiology, angiology, orthopaedic surgery, thoracic surgery, and cardio-vascular- surgery), which would increase the number of at-risk stays. The frequencies of 2015 at-risk stays in these selected units are 4711, 5130, 5564, 5016 for CHUV, HUG, USZ, and KSB, respectively (N=20421). We therefore made the assumption that the number of at-risk stays for the period 2015-2016 would approximate 40000. The sample size calculation was performed using STATA IC, V 14.

Table 6. Minimum sample size (all medical records, ADE + medical records and ADE – medical records) according to the ADE cumulative incidence (CI), the number of at-risk hospital stays (N), a desired Se of 80%, a 20% width for Se 95%CI, and relevant values of the proportion of positive tests [P(ADE+)]

		Sample Size: total				Sample size: ADE+				Sample size: ADE-			
		N				N				N			
CI	P(ADE+)	2'000	20'000	40'000	Undefined	2'000	20'000	40'000	Undefined	2'000	20'000	40'000	Undefined
3%	5%	all	1'840	1'831	1'957	all	141	140	150	all	1'699	1'691	1'807
	10%	all	1'941	1'935	2'016	all	247	246	257	all	1'694	1'689	1'759
	20%	all	2'000	1'997	2'055	all	381	380	392	all	1'619	1'617	1'663
	30%	all	1'979	1'977	2'018	all	475	474	484	all	1'504	1'503	1'534
5%	7%	1'169	1'088	1'084	1'208	97	90	90	100	1'072	998	994	1'108
	10%	1'186	1'128	1'124	1'217	136	130	129	140	1'050	998	995	1'077
	20%	1'207	1'173	1'172	1'231	222	216	215	226	985	957	957	1'005
	30%	1'192	1'167	1'165	1'207	280	274	274	284	912	893	891	923
10%	12%	547	524	523	613	53	51	51	60	494	473	472	553
	20%	569	552	552	610	93	91	91	100	476	461	461	510
	30%	567	555	555	596	126	123	123	132	441	432	432	464
	40%	552	541	539	573	149	146	146	155	403	395	393	418
20%	22%	245	240	240	289	31	30	30	36	214	210	210	253
	30%	254	248	247	287	47	47	46	53	207	201	201	234
	40%	249	245	245	276	62	60	60	68	187	185	185	208
	50%	237	233	233	261	72	70	70	79	165	163	163	182
24%	26%	195	192	192	235	27	26	26	32	168	166	166	203
	30%	201	196	196	234	34	33	33	39	167	163	163	195
	40%	199	196	194	226	47	46	46	53	152	150	148	173
	50%	190	186	186	214	56	55	55	63	134	131	131	151

Thus, assuming CI equal 10%, p(ADE+) equal 12%, 40'000 at-risk hospital stays, and an expected Se equal 80% with a 20% width for Se 95%CI, a random sample of at least 523 medical records (51 ADE+ medical records and 472 ADE- medical records) will be necessary to assess the SDM+NLP tool. Considering that some medical records might not be available (let's say 1%), the validation will finally require 530 medical records. We will thus abstract 15 medical records flagged as ADE+ and 120 medical records flagged as ADE- by the SDM+NLP tool during the 2015-2016 period in each of the 4 participating hospitals (135 medical records per hospital). Under these assumptions, the expected values and confidence intervals for Se, Sp, PPV and NPV should be respectively:

- Se = 80%; 95% CI Se= [68%; 88%]
- Sp= 96%; 95%CI Sp = [94%; 97%]
- PPV= 67%; 95%CI PPV = [52%; 79%]

- NPV= 98%; 95%CI NPV= [96%; 99%]

Regarding external validation of predictive models, we will compare predicted outcomes with true outcomes based on medical record screening for a small sample of 20 to 30 hospital stays for which predicted and validated outcomes diverge.

7. Ethical and regulatory aspects

The research project will be carried out in accordance with the research protocol, the Human Research Act (HRA) and the Human Research Ordinance (HRO). The protocol has been submitted to the local competent Ethics Committee (EC) and the study will be initiated after approval from CEC.

7.1. Motive for a proxy consent obtained from the ethics committee

Because of the high number of patients included in the study and the retrospective design of the study, an individual informed consent will not be sought. In this project, abiding by the procedure of a proxy consent obtained from the ethics committee is in keeping witharticle 34 of the “Loi relative à la recherche sur l’être humain ». Furthermore, obtaining individual informed consent would indeed imply all study investigators have access to nominal personal data, which is to be avoided in this retrospective research project.

Any medical records in which it is documented that the patients have expressed their refusal on the use of their personnal data for research purposes or have refused to sign the general consent (BIL) will not be included in the study.

7.2. Notification requirements

Significant changes to the protocol must be approved by the ethics committee in advance. The regular end of the research project is reported to the EC within 90 days upon completion of the project.

8. Funding, publication and declaration of interest

The results of the study will be submitted to local, national and/or international congresses and to peer reviewed journals for publication.

The authors of this protocol have no conflict of interest to declare. This study has been funded by the PNR74 project of the SNF.

9. References

1. Nebeker, J.R., P. Barach, and M.H. Samore, *Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting.* Ann Intern Med, 2004. **140**(10): p. 795-801.
2. Krahenbuhl-Melcher, A., et al., *Drug-related problems in hospitals: a review of the recent literature.* Drug Saf, 2007. **30**(5): p. 379-407.
3. de Vries, E.N., et al., *The incidence and nature of in-hospital adverse events: a systematic review.* Qual Saf Health Care, 2008. **17**(3): p. 216-23.
4. US Department of Health Human Services, *National action plan for adverse drug event prevention.* Washington, DC, 2014.
5. Halfon, P., A. Staines, and B. Burnand, *Adverse events related to hospital care: a retrospective medical records review in a Swiss hospital.* Int J Qual Health Care, 2017. **29**(4): p. 527-533.
6. Classen, D.C., et al., *'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured.* Health Aff (Millwood), 2011. **30**(4): p. 581-9.
7. Martins, A.C., F. Giordani, and S. Rozenfeld, *Adverse drug events among adult inpatients: a meta-analysis of observational studies.* J Clin Pharm Ther, 2014. **39**(6): p. 609-20.
8. Boeker, E.B., et al., *An individual patient data meta-analysis on factors associated with adverse drug events in surgical and non-surgical inpatients.* Br J Clin Pharmacol, 2015. **79**(4): p. 548-57.
9. Bouvy, J.C., M.L. De Bruin, and M.A. Koopmanschap, *Epidemiology of adverse drug reactions in Europe: a review of recent observational studies.* Drug Saf, 2015. **38**(5): p. 437-53.
10. Classen, D.C., et al., *Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality.* Jama, 1997. **277**(4): p. 301-6.
11. Hug, B.L., et al., *The costs of adverse drug events in community hospitals.* Jt Comm J Qual Patient Saf, 2012. **38**(3): p. 120-6.
12. Rottenkolber, D., J. Hasford, and J. Stausberg, *Costs of adverse drug events in German hospitals--a microcosting study.* Value Health, 2012. **15**(6): p. 868-75.
13. Klopotowska, J.E., et al., *Adverse drug events in older hospitalized patients: results and reliability of a comprehensive and structured identification strategy.* PLoS One, 2013. **8**(8): p. e71045.
14. Long, S.J., et al., *What is known about adverse events in older medical hospital inpatients? A systematic review of the literature.* Int J Qual Health Care, 2013. **25**(5): p. 542-54.
15. Morimoto, T., et al., *Incidence of adverse drug events and medication errors in Japan: the JADE study.* J Gen Intern Med, 2011. **26**(2): p. 148-53.
16. Fried, L.P., et al., *Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care.* J Gerontol A Biol Sci Med Sci, 2004. **59**(3): p. 255-63.
17. Dumbreck, S., et al., *Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines.* Bmj, 2015. **350**: p. h949.
18. Schneider, D.J. and B.E. Sobel, *Conundrums in the combined use of anticoagulants and antiplatelet drugs.* Circulation, 2007. **116**(3): p. 305-15.
19. Andreotti, F., et al., *Antithrombotic therapy in the elderly: expert position paper of the European Society of Cardiology Working Group on Thrombosis.* Eur Heart J, 2015. **36**(46): p. 3238-49.
20. Sharma, M., et al., *Efficacy and Harms of Direct Oral Anticoagulants in the Elderly for Stroke Prevention in Atrial Fibrillation and Secondary Prevention of Venous Thromboembolism: Systematic Review and Meta-Analysis.* Circulation, 2015. **132**(3): p. 194-204.
21. Lepori, V., A. Perren, and C. Marone, *[Adverse internal medicine drug effects at hospital admission].* Schweiz Med Wochenschr, 1999. **129**(24): p. 915-22.
22. Wasserfallen, J., et al., *Rate, type, and cost of adverse drug reactions in emergency department admissions.* Eur J Intern Med, 2001. **12**(5): p. 442-447.
23. May, A.E., T. Geisler, and M. Gawaz, *Individualized antithrombotic therapy in high risk patients after coronary stenting. A double-edged sword between thrombosis and bleeding.* Thromb Haemost, 2008. **99**(3): p. 487-93.
24. Kanagaratnam, L., et al., *[Serious Adverse Drug Reaction and Their Preventability in the Elderly Over 65 Years].* Therapie, 2015. **70**(5): p. 477-84.

25. Beeler, P.E., et al., *Use of an on-demand drug-drug interaction checker by prescribers and consultants: a retrospective analysis in a Swiss teaching hospital.* Drug Saf, 2013. **36**(6): p. 427-34.
26. Spinewine, A., et al., *Appropriate prescribing in elderly people: how well can it be measured and optimised?* Lancet, 2007. **370**(9582): p. 173-84.
27. Topinkova, E., et al., *Evidence-based strategies for the optimization of pharmacotherapy in older people.* Drugs Aging, 2012. **29**(6): p. 477-94.
28. Medicine, I.o., *Preventing Medication Errors: Quality Chasm Series*, ed. P. Aspden, et al. 2007, Washington, DC: The National Academies Press. 480.
29. Cresswell, K.M., et al., *Adverse drug events in the elderly.* Br Med Bull, 2007. **83**: p. 259-74.
30. Cooper, J.A., et al., *Interventions to improve the appropriate use of polypharmacy in older people: a Cochrane systematic review.* BMJ Open, 2015. **5**(12): p. e009235.
31. Onder, G., et al., *Recommendations to prescribe in complex older adults: results of the CRIteria to assess appropriate Medication use among Elderly complex patients (CRIME) project.* Drugs Aging, 2014. **31**(1): p. 33-45.
32. Renom-Guiteras, A., G. Meyer, and P.A. Thurmann, *The EU(7)-PIM list: a list of potentially inappropriate medications for older people consented by experts from seven European countries.* Eur J Clin Pharmacol, 2015. **71**(7): p. 861-75.
33. Yourman, L., J. Concato, and J.V. Agostini, *Use of computer decision support interventions to improve medication prescribing in older adults: a systematic review.* Am J Geriatr Pharmacother, 2008. **6**(2): p. 119-29.
34. Classen, D.C., et al., *Computerized surveillance of adverse drug events in hospital patients.* Jama, 1991. **266**(20): p. 2847-51.
35. Handler, S.M., et al., *A systematic review of the performance characteristics of clinical event monitor signals used to detect adverse drug events in the hospital setting.* J Am Med Inform Assoc, 2007. **14**(4): p. 451-8.
36. Harinstein, L.M., et al., *Use of an abnormal laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically Ill patients.* J Crit Care, 2012. **27**(3): p. 242-9.
37. Beeler, P.E., D.W. Bates, and B.L. Hug, *Clinical decision support systems.* Swiss Med Wkly, 2014. **144**: p. w14073.
38. Beeler, P.E., et al., *Impact of electronic reminders on venous thromboprophylaxis after admissions and transfers.* J Am Med Inform Assoc, 2014. **21**(e2): p. e297-303.
39. Fritz, D., et al., *Comparative evaluation of three clinical decision support systems: prospective screening for medication errors in 100 medical inpatients.* Eur J Clin Pharmacol, 2012. **68**(8): p. 1209-19.
40. Chazard, E., et al., *Data-mining-based detection of adverse drug events.* Stud Health Technol Inform, 2009. **150**: p. 552-6.
41. Ageno, W., et al., *Oral anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.* Chest, 2012. **141**(2 Suppl): p. e44S-88S.
42. Militaru, F.C., et al., *Pharmacogenetics aspects of oral anticoagulants therapy.* J Med Life, 2015. **8**(2): p. 171-5.
43. Robertson, L., P. Kesteven, and J.E. McCaslin, *Oral direct thrombin inhibitors or oral factor Xa inhibitors for the treatment of pulmonary embolism.* Cochrane Database Syst Rev, 2015(12): p. Cd010957.
44. Cohen, A.T., et al., *Comparison of the Novel Oral Anticoagulants Apixaban, Dabigatran, Edoxaban, and Rivaroxaban in the Initial and Long-Term Treatment and Prevention of Venous Thromboembolism: Systematic Review and Network Meta-Analysis.* PLoS One, 2015. **10**(12): p. e0144856.
45. Fattinger, K., et al., *Epidemiology of drug exposure and adverse drug reactions in two swiss departments of internal medicine.* Br J Clin Pharmacol, 2000. **49**(2): p. 158-67.
46. Hardmeier, B., et al., *Adverse drug events caused by medication errors in medical inpatients.* Swiss Med Wkly, 2004. **134**(45-46): p. 664-70.

47. Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. 2 ed. 2009: Springer-Verlag New York. 745.
48.] Pearl, J., *Causal inference in statistics: An overview*. 2009: Statistics Surveys
49. Hernan, M. and J. Robins, *Causal Inference Book*. forthcoming., Boca Raton: Chapman & Hall/CRC. 220 pages.
50. Schulman, S., et al., *Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients.* J Thromb Haemost, 2005. **3**(4): p. 692-4.
51. Kearon, C., et al., *Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH.* J Thromb Haemost, 2016. **14**(7): p. 1480-3.
52. Konstantinides, S.V., et al., *2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism.* Eur Heart J, 2014. **35**(43): p. 3033-69, 3069a-3069k.
53. Lovis, C., et al., *Medical dictionaries for patient encoding systems: a methodology.* Artif Intell Med, 1998. **14**(1-2): p. 201-14.
54. Fabry, P., et al., *Amplification of Terminologia anatomica by French language terms using Latin terms matching algorithm: a prototype for other language.* Int J Med Inform, 2006. **75**(7): p. 542-52.
55. Naranjo, C.A., et al., *A method for estimating the probability of adverse drug reactions.* Clin Pharmacol Ther, 1981. **30**(2): p. 239-45.
56. *World Health Organization. The use of the WHO-UMC system for standardized case causality assessment*. 2005, The Uppsala Monitoring Centre: Uppsala.
57. Collet, T.H., et al., *Reproducibility of diabetes quality of care indicators as reported by patients and physicians.* Eur J Public Health, 2014. **24**(6): p. 1004-9.
58. Taffe, P., et al., *Test result-based sampling: an efficient design for estimating the accuracy of patient safety indicators.* Med Decis Making, 2012. **32**(1): p. E1-12.
59. Le Pogam, M.A., et al., *Geriatric Patient Safety Indicators Based on Linked Administrative Health Data to Assess Anticoagulant-Related Thromboembolic and Hemorrhagic Adverse Events in Older Inpatients: A Study Proposal.* JMIR Res Protoc, 2017. **6**(5): p. e82.
60. (AHRQ), A.f.H.R.a.Q. *Patient Safety Indicators technical specification updates – version 7.0 (ICD 10)*. 2017 Available from: http://www.qualityindicators.ahrq.gov/Modules/PSI_TechSpec_ICD10_v70.aspx
61. LI, I., *Risk Adjustment for Measuring Health Care Outcomes*, ed. I.H.A. Press. Vol. Fourth Edition. 2013, Chicago.
62. Kim, E.J., et al., *Predicting outcomes among patients with atrial fibrillation and heart failure receiving anticoagulation with warfarin.* Thromb Haemost, 2015. **114**(1): p. 70-7.
63. Ruiz-Gimenez, N., et al., *Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry.* Thromb Haemost, 2008. **100**(1): p. 26-31.
64. Beyth, R.J., L.M. Quinn, and C.S. Landefeld, *Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin.* Am J Med, 1998. **105**(2): p. 91-9.
65. Hirsh, J., et al., *Antithrombotic and thrombolytic therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).* Chest, 2008. **133**(6 Suppl): p. 110S-112S.
66. Holbrook, A., et al., *Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.* Chest, 2012. **141**(2 Suppl): p. e152S-e184S.
67. Palareti, G., et al., *Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). Italian Study on Complications of Oral Anticoagulant Therapy.* Lancet, 1996. **348**(9025): p. 423-8.
68. Gage, B.F., et al., *Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF).* Am Heart J, 2006. **151**(3): p. 713-9.
69. Efird, L.M., et al., *Stratifying the risks of oral anticoagulation in patients with liver disease.* Circ Cardiovasc Qual Outcomes, 2014. **7**(3): p. 461-7.

70. Lee, H.H., et al., *C-reactive protein as a prognostic indicator for rebleeding in patients with nonvariceal upper gastrointestinal bleeding.* Dig Liver Dis, 2015. **47**(5): p. 378-83.
71. Donze, J., et al., *Scores to predict major bleeding risk during oral anticoagulation therapy: a prospective validation study.* Am J Med, 2012. **125**(11): p. 1095-102.
72. Zhu, W., et al., *The HAS-BLED Score for Predicting Major Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation: A Systematic Review and Meta-analysis.* Clin Cardiol, 2015. **38**(9): p. 555-61.
73. Lip, G.Y., et al., *Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) score.* J Am Coll Cardiol, 2011. **57**(2): p. 173-80.
74. Shireman, T.I., et al., *Development of a contemporary bleeding risk model for elderly warfarin recipients.* Chest, 2006. **130**(5): p. 1390-6.
75. Kuijer, P.M., et al., *Prediction of the risk of bleeding during anticoagulant treatment for venous thromboembolism.* Arch Intern Med, 1999. **159**(5): p. 457-60.
76. White, R.H., et al., *Major bleeding after hospitalization for deep-venous thrombosis.* Am J Med, 1999. **107**(5): p. 414-24.

Annexe 1: ATC Codes of the drugs concerned by the study

Class	ATC	Antithrombotic
Vitamin K antagonists	B01AA04	Phenprocoumon
	B01AA07	Acenocoumarol
Heparins	B01AB01	Heparin
	B01AB02	Antithrombin III
	B01AB04	Dalteparin
	B01AB05	Enoxaparin
	B01AB06	Nadroparin
	B01AB09	Danaparoid
Platelet aggregation inhibitors	B01AC04	Clopidogrel
	B01AC06	Acetylsalicylic acid
	B01AC09	Epoprostenol
	B01AC11	Iloprost
	B01AC13	Abciximab
	B01AC16	Eptifibatide
	B01AC17	Tirofiban
	B01AC21	Treprostinil
	B01AC22	Prasugrel
	B01AC24	Ticagrelor
	B01AC25	Cangrelor
	B01AC27	Selexipag
Direct thrombin inhibitors	B01AE03	Argatroban
	B01AE06	Bivalirudin
	B01AE07	Dabigatran etexilate
Direct factor Xa inhibitors	B01AF01	Rivaroxaban
	B01AF02	Apixaban
	B01AF03	Edoxaban
Other antithrombotic agents	B01AX05	Fondaparinux

Annexe 2: Justification of the extracted items for structured data

Data type	(Descriptive) Justification	References
General Administrative Data	Descriptive value	
Clinical measure		
Blood pressure	(To identify blood pressure disorders) Risk factor that influence hemorrhagic, thromboembolic events	(Zhu, He et al. 2015)
Weight/Height	Descriptive value and risk factor that influence hemorrhagic, thromboembolic events	(Lip, Frison et al. 2011, Bene, Dubart et al. 2014)
Sum of alcohol withdrawal syndrome score	Risk factor that influence hemorrhagic, thromboembolic events	(White, Beyth et al. 1999, Olesen, Lip et al. 2011)
Patient location(s) and transfers	Identify triggers that indicate that an antithrombotic-related ADE may have occurred	
Diagnoses and procedures	(Select hospital stays eligible for the denominator of indicators) Identify an hemorrhagic, thromboembolic event (markers)	(LI 2013, Le Pogam, Quantin et al. 2017, (AHRQ) 2017)
Laboratory values		
Measured electrolytes and ions		
Blood Ionogram: Sodium and Potassium	(Abnormal values can be a sign of a vital organ failure) Risk factor that influence hemorrhagic, thromboembolic events	(Kim, Ozonoff et al. 2015)

Lactic acid	(Abnormal values may indicate different states of shock, severe anemia or ventricular failure) Risk factor that influence hemorrhagic, thromboembolic events	(Ruiz-Gimenez, Suarez et al. 2008)
Uric acid	(To diagnose kidney disorder or to reveal hematopathies, chronic renal failure, severe hepatic insufficiency (with decreased uric acid synthesis) or signs of tumors) Risk factor that influence hemorrhagic, thromboembolic events	(Beyth, Quinn et al. 1998)
Urea	(To assess kidney function and in particular to detect renal failure. High values can also be indicative of heart damage or gastrointestinal hemorrhage) Risk factor that influence hemorrhagic, thromboembolic events	(Hirsh, Guyatt et al. 2008, Holbrook, Schulman et al. 2012)
Iron and Ferritin Level	(The determination of iron and ferritin makes the identification of anemia possible) Identify an hemorrhagic, thromboembolic event (markers) and risk factor	(Ruiz-Gimenez, Suarez et al. 2008)
Measured enzymes		
Liver transaminases	(To screen for, detect, evaluate and monitor acute and chronic liver inflammation, liver infection, liver disease and/or damage) Risk factor that influence hemorrhagic, thromboembolic events	(Palareti, Leali et al. 1996, Holbrook, Schulman et al. 2012, Kim, Ozonoff et al. 2015)
Creatine kinase	(In particular CK-MB and CK-BB- Sign of myocardial or neurological lesion) Risk factor that influence hemorrhagic, thromboembolic events	(Beyth, Quinn et al. 1998, Holbrook, Schulman et al. 2012)
Gamma-glutamyltransferase	(To evaluate for a possible liver disease or bile duct disease, sometimes to screen for or monitor alcohol abuse) Risk factor that influence hemorrhagic, thromboembolic events	(Beyth, Quinn et al. 1998)
Alkaline phosphatase	(To screen for or monitor treatment for a liver disorder) Risk factor that influence hemorrhagic, thromboembolic events	(Gage, Yan et al. 2006, Kim, Ozonoff et al. 2015)
Complete Blood count (CBC)	(The blood count can reveal a large number of pathologies: anemia, coagulation problem, viral infections or consumption of platelets) Identify an hemorrhagic, thromboembolic event and risk factors	(Schulman, Kearon et al. 2005, Hirsh, Guyatt et al. 2008, Ruiz-Gimenez, Suarez et al. 2008, Konstantinides, Torbicki et al. 2014, Kearon, Ageno et al. 2016)

Hemostasis assessment	Identify an hemorrhagic, thromboembolic event and risk factors	(Schulman, Kearon et al. 2005, Konstantinides, Torbicki et al. 2014, Kearon, Ageno et al. 2016)
Other measured biological values available		
Albumine	(To screen for and help diagnose a liver disorder or kidney disease) Risk factor that influence hemorrhagic, thromboembolic events	(Hirsh, Guyatt et al. 2008, Efird, Mishkin et al. 2014)
C-reactive protein	(To identify the presence of inflammation) Risk factor that influence hemorrhagic, thromboembolic events	(Hirsh, Guyatt et al. 2008, Lee, Park et al. 2015)
Myoglobin	(To determine various cardiovascular disorders) Risk factor that influence hemorrhagic, thromboembolic events	(Gage, Yan et al. 2006, Hirsh, Guyatt et al. 2008)
Troponin	(May be a sign of myocardial infarction, pulmonary embolism or myocarditis) Risk factor that influence hemorrhagic, thromboembolic events	(Gage, Yan et al. 2006, Ruiz-Gimenez, Suarez et al. 2008)
Creatinine and Creatinine Clearance	(To help diagnose kidney disease) Risk factor that influence hemorrhagic, thromboembolic events	(Ruiz-Gimenez, Suarez et al. 2008, Donze, Rodondi et al. 2012)
Bilirubin	(To screen for or monitor liver disorders or hemolytic anemia) Risk factor that influence hemorrhagic, thromboembolic events	(Ruiz-Gimenez, Suarez et al. 2008, Zhu, He et al. 2015)
Glycated hemoglobin	(To monitor a person's diabetes) Risk factor that influence hemorrhagic, thromboembolic events	(Shireman, Mahnken et al. 2006, Lip, Frison et al. 2011)
Tumor markers avalaible	Risk factor that influence hemorrhagic, thromboembolic events	(Kuijer, Hutten et al. 1999, White, Beyth et al. 1999, Gage, Yan et al. 2006)
Prescription/medication	Identify an hemorrhagic, thromboembolic event (markers), risk factors, triggers and causal factors	(Holbrook, Schulman et al. 2012)

Bibliography related to the justification of extracted items

Agency for Healthcare Research and Quality (AHRQ). Patient Safety Indicators technical specification updates – version 7.0 (ICD 10). 2017 Sept. URL: http://www.qualityindicators.ahrq.gov/Modules/PSI_TechSpec_ICD10_v70.aspx Bene, J., A. E. Dubart, C.

Senis, M. Auffret, J. Caron and S. Gautier (2014). "Risk factors associated with a thrombotic or bleeding event in patients treated with vitamin K antagonists." J Mal Vasc 39(4): 248-255.

Beyth, R. J., L. M. Quinn and C. S. Landefeld (1998). "Prospective evaluation of an index for predicting the risk of major bleeding in outpatients treated with warfarin." Am J Med 105(2): 91-99.

Donze, J., N. Rodondi, G. Waeber, P. Monney, J. Cornuz and D. Aujesky (2012). "Scores to predict major bleeding risk during oral anticoagulation therapy: a prospective validation study." Am J Med 125(11): 1095-1102.

Efird, L. M., D. S. Mishkin, D. R. Berlowitz, A. S. Ash, E. M. Hylek, A. Ozonoff, J. I. Reisman, S. Zhao, G. K. Jasuja and A. J. Rose (2014). "Stratifying the risks of oral anticoagulation in patients with liver disease." Circ Cardiovasc Qual Outcomes 7(3): 461-467.

Gage, B. F., Y. Yan, P. E. Milligan, A. D. Waterman, R. Culverhouse, M. W. Rich and M. J. Radford (2006). "Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF)." Am Heart J 151(3): 713-719.

Hirsh, J., G. Guyatt, G. W. Albers, R. Harrington and H. J. Schunemann (2008). "Antithrombotic and thrombolytic therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition)." Chest 133(6 Suppl): 110S-112S.

Holbrook, A., S. Schulman, D. M. Witt, P. O. Vandvik, J. Fish, M. J. Kovacs, P. J. Svensson, D. L. Veenstra, M. Crowther and G. H. Guyatt (2012). "Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines." Chest 141(2 Suppl): e152S-e184S.

Kearon, C., W. Ageno, S. C. Cannegieter, B. Cosmi, G. J. Geersing, P. A. Kyrle, A. Subcommittees on Control of, Predictive and D. Diagnostic Variables in Thrombotic (2016). "Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH." J Thromb Haemost 14(7): 1480-1483.

Kim, E. J., A. Ozonoff, E. M. Hylek, D. R. Berlowitz, A. S. Ash, D. R. Miller, S. Zhao, J. I. Reisman, G. K. Jasuja and A. J. Rose (2015). "Predicting outcomes among patients with atrial fibrillation and heart failure receiving anticoagulation with warfarin." Thromb Haemost 114(1): 70-77.

Konstantinides, S. V., A. Torbicki, G. Agnelli, N. Danchin, D. Fitzmaurice, N. Galie, J. S. Gibbs, M. V. Huisman, M. Humbert, N. Kucher, I. Lang, M. Lankeit, J. Lekakis, C. Maack, E. Mayer, N. Meneveau, A. Perrier, P. Pruszczyk, L. H. Rasmussen, T. H. Schindler, P. Svitil, A. Vonk Noordegraaf, J. L. Zamorano, M. Zompatori, D. Task Force for the and C. Management of Acute Pulmonary Embolism of the European Society of (2014). "2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism." Eur Heart J 35(43): 3033-3069, 3069a-3069k.

Kuijer, P. M., B. A. Hutten, M. H. Prins and H. R. Buller (1999). "Prediction of the risk of bleeding during anticoagulant treatment for venous thromboembolism." Arch Intern Med 159(5): 457-460.

Le Pogam, M. A., C. Quantin, O. Reich, P. Tuppin, A. Fagot-Campagna, F. Paccaud, I. Peytremann-Bridevaux and B. Burnand (2017). "Geriatric Patient Safety Indicators Based on Linked Administrative Health Data to Assess Anticoagulant-Related Thromboembolic and Hemorrhagic Adverse Events in Older Inpatients: A Study Proposal." JMIR Res Protoc 6(5): e82.

Lee, H. H., J. M. Park, S. W. Lee, S. H. Kang, C. H. Lim, Y. K. Cho, B. I. Lee, I. S. Lee, S. W. Kim and M. G. Choi (2015). "C-reactive protein as a prognostic indicator for rebleeding in patients with nonvariceal upper gastrointestinal bleeding." Dig Liver Dis 47(5): 378-383.

Lezzoni LI. (2013) Risk adjustment for measuring Healthcare Outcomes, 4th edn. Chicago, IL: Health Administration Press.

Lip, G. Y., L. Frison, J. L. Halperin and D. A. Lane (2011). "Comparative validation of a novel risk score for predicting bleeding risk in anticoagulated patients with atrial fibrillation: the HAS-BLED (Hypertension, Abnormal Renal/Liver Function, Stroke, Bleeding History or Predisposition, Labile INR, Elderly, Drugs/Alcohol Concomitantly) score." J Am Coll Cardiol 57(2): 173-180.

Olesen, J. B., G. Y. Lip, P. R. Hansen, J. Lindhardsen, O. Ahlehoff, C. Andersson, P. Weeke, M. L. Hansen, G. H. Gislason and C. Torp-Pedersen (2011). "Bleeding risk in 'real world' patients with atrial fibrillation: comparison of two established bleeding prediction schemes in a nationwide cohort." J Thromb Haemost 9(8): 1460-1467.

Palareti, G., N. Leali, S. Coccheri, M. Poggi, C. Manotti, A. D'Angelo, V. Pengo, N. Erba, M. Moia, N. Ciavarella, G. Devoto, M. Berrettini and S. Musolesi (1996). "Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). Italian Study on Complications of Oral Anticoagulant Therapy." Lancet 348(9025): 423-428.

Ruiz-Gimenez, N., C. Suarez, R. Gonzalez, J. A. Nieto, J. A. Todoli, A. L. Samperiz, M. Monreal and R. Investigators (2008). "Predictive variables for major bleeding events in patients presenting with documented acute venous thromboembolism. Findings from the RIETE Registry." Thromb Haemost 100(1): 26-31.

Schulman, S., C. Kearon, S. Subcommittee on Control of Anticoagulation of the, T. Standardization Committee of the International Society on and Haemostasis (2005). "Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients." J Thromb Haemost 3(4): 692-694.

Shireman, T. I., J. D. Mahnken, P. A. Howard, T. F. Kresowik, Q. Hou and E. F. Ellerbeck (2006). "Development of a contemporary bleeding risk model for elderly warfarin recipients." Chest 130(5): 1390-1396.

White, R. H., R. J. Beyth, H. Zhou and P. S. Romano (1999). "Major bleeding after hospitalization for deep-venous thrombosis." Am J Med 107(5): 414-424.

Zhu, W., W. He, L. Guo, X. Wang and K. Hong (2015). "The HAS-BLED Score for Predicting Major Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation: A Systematic Review and Meta-analysis." Clin Cardiol 38(9): 555-561.