 **Protocol for annotating clinical documents**

In order to be able to use the data available in free text form in the patient record and to develop computerized algorithms for the recognition and processing of this information, a first step of document annotation is necessary. This file is a protocol established for the annotation of clinical documents extracted within the framework of the SwissMADE project (a first version fed iteratively afterwards). The BRAT1 tool or the GATE2 tool are used for this task

In addition to annotating the clinical documents, the annotator will also have to evaluate the causality between the antithrombotic drug(s) and the adverse event (according to evaluation criteria described in the literature3,4). The categories defined are: relationship between the drug and the adverse event **certain, probable, possible and doubtful**. For this purpose, the simplified Naranjo algorithm will be used (a validated tool for assessing the causality of adverse drug events in daily clinical practice) (Table 1). The evaluation will be done on a parallel Excel document.

Table 1 : Simplified Naranjo Algorithm5

|  | **Component** | **Score** | |
| --- | --- | --- | --- |
| **Yes** | **No/Do not know** |
| 2 | Did the adverse event appear after the suspected drug was administered? | +1 | 0 |
| 3 | Did the adverse reaction improve when the drug was discontinued or a specific antagonist was administered? | +1 | 0 |
| 4 | Did the adverse reaction reappear when the drug was readministered? | +1 | 0 |
| 5 | Are there alternative causes (other than the drug) that could on their own have caused the reaction? | 0 | +1 |

**Annotation rules**

General rules

1. In order to fulfil the different purposes of the study, five elements are considered for the annotation (defined below):
   1. Identification of the drugs involved in the event.
   2. Identification of haemorrhagic and thromboembolic (arterial and venous) complications.
   3. Identification of event markers; clinical signs, symptoms, procedures, prescription and imaging orders, biological results that indicate that an antithrombotic-related ADE may have occurred.
   4. Identification of 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.
   5. Identification of confoundings; any conditions that may explain bleeding and haemorrages other than related to the drug.
   6. *(Identification of causal markers; 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)*

Drug Annotation

For the annotation of drugs, two categories are defined: anti-thrombotic drugs and other drugs (concomitant) mentioned in the text.

|  |  |
| --- | --- |
| Tag name | Antithrombotic drug |
| Rules | 1. All anti-thrombotic drugs (e.g. Sintrom, liquemin, heparin) listed in Annex 1 will be annotated. 2. If the symbol ® is appended to the trade name (or in error to the INN), it is included in the annotation as it is considered to belong to the word. 3. Words derived from therapeutic classes will also be annotated E.g.: anticoagulation. |

|  |  |
| --- | --- |
| Tag name | Concomitant medication |
| Rules | 1. All non-antithrombotic drugs (not listed in the appendix) will also be annotated with the "concomitant medication" entity 2. If the symbol ® is appended to the trade name (or in error to the INN), it is included in the annotation as it is considered to belong to the word. 3. Words derived from therapeutic classes will also be annotated. E.g.: cortisonic |

Adverse events annotation (hemorrhagic/thromboembolic events)

For the annotation of adverse events, only haemorrhagic and thromboembolic (venous and arterial) events will be annotated. The following sub-categories are defined for these two types of events: non-severe bleeding\*, severe bleeding\*, DVT (deep vein thrombosis), PE (pulmonary embolism), TIA (transient ischemic attack), acute myocardial infarction.

|  |  |
| --- | --- |
| Tag name | Event |
| Rules | 1. The adverse event is annotated at each occurrence in the text. 2. The annotated word(s) corresponds to the largest medical concept in the text, e.g. esophageal haemorrhage, cerebellar intracerebral haemorrhage, etc. 3. When the term "drug" or "drug-induced" or "drug etiology" follows an adverse event, e.g., "drug hemorrhage", without specifying a specific molecule, this adverse event is annotated. |

Event markers will also be annotated. Clinical signs, symptoms, procedures, prescription and imaging orders, biological results that indicate that an haemorrhagic or thromboembolic (venous and arterial) event have occured will be tagged "Event marker" **(this tag will be clarified during the annotation).**

|  |  |
| --- | --- |
| Tag name | Event marker |

Risk factors annotation

Risk factors identified in the literature as "triggers" for bleeding and/or thromboembolic complications will also be annotated. A list of risk factors to be annotated is presented in Annex 2 (list established according to risk scores presented in the literature and validated by an expert group).

|  |  |
| --- | --- |
| Tag name | Risk factor |
| Rules | 1. Risk factor terms presented in Annex 3 are annotated. 2. The risk factor shall be annotated at each occurrence in the text. 3. The annotated word(s) correspond(s) to the most important medical concept in the text, e.g. lung cancer, renal failure, etc. 4. The degree of impairment is annotated (e.g. stage 3 liver failure). |

Confounders annotation

To avoid any bias of confusion between the annotated adverse event and the identified antithrombotic, any condition that may cause thrombosis, bleeding and hemorrhage other than those related to antithrombotic drugs

|  |  |
| --- | --- |
| Tag name | Confounders |
| Rules | 1. Confounders terms are annotated. 2. Confounders shall be annotated at each occurrence in the text. 3. The annotated word(s) correspond(s) to the most important medical concept in the text, e.g. car accident,… |

*(Causal markers annotation)*

*Causal factors 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 specified in the text will be tagged with the name "causal marker"* ***(this tag will be clarified during the annotation).***

|  |  |
| --- | --- |
| *Tag name* | *Causal marker* |

NB: This document will be updated iteratively as the annotation work progresses.

\* Definition of clinically relevant "non-severe" hemorrhage (as defined by ISTH): A clinically relevant minor bleed is an acute or subacute clinically evident bleed that does not meet the criteria for a major bleed but does cause a clinical response in that it results in at least one of the following: admission to hospital for bleeding, or medically or surgically guided medical or surgical treatment for bleeding, or a change in antithrombotic therapy (including stopping or discontinuing the medication)

Definition of a "severe" hemorrhage (according to ISTH) : Fatal hemorrhage and/or symptomatic bleeding into a critical area or organ, such as intracranial, intraspinal, intraocular, retroperitoneal, intraarticular or pericardial, or intramuscular compartmentalized hemorrhage and/or bleeding causing a drop in hemoglobin of 2 g/dL (1.24 mmol/L) or more, or resulting in the transfusion of two or more units of whole blood or red blood cells.

**Bibliography**

1 P. Stenetorp, S. Pyysalo, G. Topi ́c, T. Ohta, S. Ananiadou, andJ. Tsujii. 2012.BRAT: a Web-based Tool for NLP-AssistedText Annotation. InProceedings of the Demonstrations atEACL 2012.

2 H. Cunningham, V. Tablan, A. Roberts, K. Bontcheva (2013) Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLoS Comput Biol 9(2): e1002854. doi:10.1371/journal.pcbi.1002854

3 Pearl, J., Causal inference in statistics: An overview. 2009: Statistics Surveys

4 Hernan, M. and J. Robins, Causal Inference Book. forthcoming., Boca Raton: Chapman & Hall/CRC. 220 pages.

5 Pharmacology Research & Perspectives, Volume: 6, Issue: 1, First published: 03 January 2018, DOI: (10.1002/prp2.373)

**Annex 1** : Liste des anti-thrombotiques à annoter

|  |  |  |  |
| --- | --- | --- | --- |
| Class | ATC | Antithrombotic | Spécialité (liste LS) |
| Vitamin K antagonists | B01AA04 | Phenprocoumon | Marcoumar® |
| B01AA07 | Acenocoumarol | Sintrom® |
| Heparins | B01AB01 | Heparin | Terme héparine toujours présent |
| B01AB02 | Antithrombin III | Kybernin P®, Atenativ® |
| B01AB04 | Dalteparin | Fragmin® |
| B01AB05 | Enoxaparin | Clexane® |
| B01AB06 | Nadroparin | Fraxiforte® |
| B01AB09 | Danaparoid | Orgaran® |
| [Platelet aggregation inhibitors](https://www.whocc.no/atc_ddd_index/?code=B01AC) | B01AC04 | [Clopidogrel](https://www.whocc.no/atc_ddd_index/?code=B01AC04&showdescription=yes) | Plavix® / Duoplavin® |
| B01AC06 | [Acetylsalicylic acid](https://www.whocc.no/atc_ddd_index/?code=B01AC06&showdescription=yes) | Aspirine® |
| B01AC09 | [Epoprostenol](https://www.whocc.no/atc_ddd_index/?code=B01AC09&showdescription=yes) | Veletri® |
| B01AC11 | [Iloprost](https://www.whocc.no/atc_ddd_index/?code=B01AC11&showdescription=yes) | Ilomedin /Ventavis® |
| B01AC13 | [Abciximab](https://www.whocc.no/atc_ddd_index/?code=B01AC13&showdescription=yes) | - |
| B01AC16 | [Eptifibatide](https://www.whocc.no/atc_ddd_index/?code=B01AC16&showdescription=yes) | Integrilin® |
| B01AC17 | [Tirofiban](https://www.whocc.no/atc_ddd_index/?code=B01AC17&showdescription=yes) | Aggrastat® |
| B01AC21 | [Treprostinil](https://www.whocc.no/atc_ddd_index/?code=B01AC21&showdescription=yes) | Remodulin® |
| B01AC22 | [Prasugrel](https://www.whocc.no/atc_ddd_index/?code=B01AC22&showdescription=yes) | Efient® |
| B01AC24 | [Ticagrelor](https://www.whocc.no/atc_ddd_index/?code=B01AC24&showdescription=yes) | Brilique® |
| B01AC25 | [Cangrelor](https://www.whocc.no/atc_ddd_index/?code=B01AC25&showdescription=yes) | Kengrexal® |
| B01AC27 | [Selexipag](https://www.whocc.no/atc_ddd_index/?code=B01AC27&showdescription=yes) | Uptravi® |
| Direct thrombin inhibitors | B01AE03 | [Argatroban](https://www.whocc.no/atc_ddd_index/?code=B01AE03&showdescription=yes) | Argatra® |
| B01AE06 | Bivalirudin | Angiox® |
| B01AE07 | [Dabigatran etexilate](https://www.whocc.no/atc_ddd_index/?code=B01AE07&showdescription=yes) | Pradaxa® |
| [Direct factor Xa inhibitors](https://www.whocc.no/atc_ddd_index/?code=B01AF) | B01AF01 | [Rivaroxaban](https://www.whocc.no/atc_ddd_index/?code=B01AF01&showdescription=yes) | Xarelto® / Xarelto vascular® |
| B01AF02 | [Apixaban](https://www.whocc.no/atc_ddd_index/?code=B01AF02&showdescription=yes) | Eliquis® |
| B01AF03 | [Edoxaban](https://www.whocc.no/atc_ddd_index/?code=B01AF03&showdescription=yes) | Lixiana® |
| Other antithrombotic agents | B01AX05 | [Fondaparinux](https://www.whocc.no/atc_ddd_index/?code=B01AX05&showdescription=yes) | Arixtra® |

**Annex 2** : List of risk factors to be annotated

**I List of risk factors for hemorrhagic events** (taken from different risk scores presented in the literature, we have focused mainly on the most relevant sources with external validation.)

* Anemia1
* Chronic renal dysfunction or dialysis or acute renal dysfunction 2
* Age ≥ 75
* History of bleeding
* Hypertension3
* Chronic or acute liver dysfunction
* History of ischemic cerebrovascular accident
* INR unstable
* Chronic alcoholism
* Polymedication 4
* Cancer
* Abnormality of coagulation 5
* Genetic factors 6
* Major risk of falling, psychiatric or neurological disease
* Female gender
* Tobacco (within 2 years)
* Non-white race
* Townsend Poverty Score ( /5 units)

1 Defined as hemoglobin <13 g/dl in men and <12 g/dl in women or most recent hematocrit <30.

2 Defined as estimated glomerular filtration rate <30 mL/min or dependent on dialysis or renal transplantation or serum creatinine ≥200 mmol/L

3 Defined as diagnosed hypertension or as systolic blood pressure >160 mmHg.

4 Refers to the concomitant use of medications such as antiplatelet agents, non-steroidal anti-inflammatory drugs, etc.

5 Platelets <75,000, use of antiplatelet therapy (e.g. daily aspirin) or NSAID therapy; or blood dyscrasias.

6 CYP2C9\*2 and/or CYP2C9\*3.

Bibliography :

Fang MC et al A new risk scheme to predict warfarin-associated hemorrhage. J Amer Coll Cardiol 2011; 58: 395

Pisters R et al. A novel user-friendly score (HAS-BLED) to assess one-year risk of of major bleeding in atrial fibrillation patients: the EuroHeart study. Chest. Prepublished online March 18.2010. DOI 10: 1378/chest. 10-0134

Apostolakis S et all. Performance of the HEMORR(2)HAGES, ATRIA, and HAS-BLED bleeding risk-prediction scores in patients with atrial fibrillation undergoing anticoagulation: the AMADEUS (evaluating the use of SR34006 compared to warfarin or acenocoumarol in patients with atrial fibrillation) study. J Am Coll Cardiol. 2012 Aug 28;60(9):861-7. doi: 10.1016/j.jacc.2012.06.019

Hippisley-Cox J and Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the Qbleed scores. BMJ 2014;349:g4066 doi: 10.1136.

**II List of risk factors for venous thrombosis (DVT and PE)** (taken from different risk scores presented in the literature, we focused mainly on the most relevant sources with external validation)

* Age ≥ 65
* Presence of acquired or genetic thrombophilia 1
* Recent TVP/EP (≤ 3, 6 or 12 months depending on the study)
* Active cancer 2
* Paralysis of the lower members: hemiparesis, hemiplegia, paraplegia
* Presence of cardiovascular disease: chronic heart failure, congenital heart disease, history of myocardial infarction, hypertension.
* Recent stroke (≤ 3, 6 or 12 months?)
* Presence of chronic lung disease: chronic respiratory failure, BPCO
* Presence of inflammatory disease: inflammatory disease of the digestive tract (Crohn's disease, ulcerative colitis), rheumatoid arthritis
* Diabetes
* Obesity 3
* Varicose veins of the lower limbs
* Post-phlebitic disease
* Hormone therapy (contraception)
* Chronic Kidney Disease: nephrotic syndrome, moderate to end-stage chronic kidney disease, dialysis, kidney transplantation
* Anti-cancer treatment: tamoxifen, thalidomide, etc.
* Myeloproliferative syndromes: polycythemia/polyglobulism, essential thrombocythemia
* Presence of a central venous catheter or transvenous pacemaker
* Recent high-risk thromboembolic surgery (≤ 3 months)
* Recent Surgery
* Severe infection/recent sepsis (≤ 1 or 3 months depending on studies)
* Recent hospitalization (≤ 3 months)
* Institutionalization 4
* Transfusions ≥ 4 units of whole blood or packed red blood cells within 72 hours prior to orthopaedic surgery
* Dehydration
* Active smoking (discussed) 5
* Antipsychotics treatments
* Immobilization from any cause (e.g. neurological disease with paresis, bed rest, from any cause, immobilization during a long plane trip etc.).
* Pollution

1Thrombophilia (inherited or acquired thrombophilia), personal or familial. -Genetics: common: in Caucasians: Leiden factor mutation, (Arg506Gln), prothrombin mutation (G20210A). Rare: deficiencies in protein C, protein S or antithrombin; -Acquired: antiphospholipid antibody syndrome (venous thrombotic events, arterial, or repeated miscarriages in association with the presence of one or more of the following abnormalities: lupus anticoagulant, antibody, anticardiolipin, β2 GP1)

2 Neoplasia +/- chemotherapy accounts for 20% of venous thrombosis in the community. Particularly high risk for the following cancers: pancreas, lymphoma, leukaemia, digestive tumours, malignant brain tumours. An additional risk is noted in oncology patients receiving immunosuppressive or cytotoxic therapy.

3 RR (relative risk) 2.3, 95% CI (95% confidence interval): 1.7-3.2; RR 2.7 if BMI >40 kg/m2

4 Hospitalization and nursing home residence (accounts for 60% of venous thrombosis in the community). The incidence of DVT is equal between hospitalization in general surgery and in medical clinics.

5 RR 1.5, 95% CI 0.95-1.5 on average; RR at 4.3 95%, CI 5.7-7.1 for > 20 APU in young smokers compared with young non-smokers. For smokers taking contraception the RR is 8.8; 95% CI 5.7-13, compared with women who do not smoke and do not take contraception.

Bibliography :

Barbar S et al Risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua Prediction Score. J Thromb and Haemost. 2010; 8: 2450.

J Hippisley-Cox C. Coupland Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism : Prospective cohort study. BMJ 2011 (343)

Caprini JA, Arcelus JI, Hasty JH, Tamhane AC, Fabrega F. Clinical assessment of venous thromboembolic risk in surgical patients. Semin Thromb Hemost. 1991;17 Suppl 3:304-12.

Laporte S et al. Clinical predictors for fatal pulmonary embolism in 15'520 patients with venous thromboembolism. Circulation 2008; 117: 1711

Heitz JA et al. Risk factors for deep vein thrombosis and pulmonary embolism. Arch Intern Med 2000; 160: 809.

Righini M, Bounameaux H. Venous thrombosis: risk factors and management. Herz.

2007 Feb;32(1):27-34. Review. PubMed PMID: 17323032.

1. **Liste des facteurs de risque pour les événements thromboemboliques artérielles sous anticoagulant**

* Hypertension1
* Age ≥ 65
* Diabetes
* Stroke history (ischemic stroke, transient ischemic attack or systemic embolism)
* Vascular disease
* Female gender
* Fibrillation or Atrial Flutter
* Presence or replacement of a cardiac valve prosthesis
* Smoking
* Obesity
* Dyslipidemia
* Thrombophilia
* Hypothyroidism
* Goute disease
* Chronic Respiratory Disease
* Chronic renal dysfunction or dialysis
* Recent acute infection
* Recent severe infection
* Chronic Infection
* Cancer
* Myeloproliferative syndrome
* Autoimmune inflammatory diseases
* Female gender
* Recent Hospitalization
* Institutionalization

1 Defined as diagnosed hypertension

Bibliography :

Lang S, Ederhy S, Boyer-Chatene L, Dufaitre G, Di Angelantonio E, Meuleman C, et al. CHADS2 Score predicts the occurrence of death, stroke and heart failure in patients with atrial fibrillation: a single-center cohort study. J Am Coll Cardiol. 2010;55(10 s1):A8.E79-A8.E79.

Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest. 2010 Feb;137(2):263-72.

**Annex 3** : Relationship between ADE and medication error

