Identification of risk factors for adverse drug reactions in a pharmacovigilance database

Sofia Attelind\textsuperscript{1,2} | Niclas Eriksson\textsuperscript{1,3} | Anders Sundström\textsuperscript{2} | Mia Wadelius\textsuperscript{1} | Pär Hallberg\textsuperscript{1} 

\textsuperscript{1}Department of Medical Sciences, Clinical Pharmacogenomics, Uppsala University, Uppsala, Sweden  
\textsuperscript{2}Department of Drug Safety, Swedish Medical Products Agency, Uppsala, Sweden  
\textsuperscript{3}Uppsala Clinical Research Center, Uppsala University Hospital, Uppsala, Sweden

Abstract

Introduction: In addition to identifying new safety signals, pharmacovigilance databases could be used to identify potential risk factors for adverse drug reactions (ADRs).

Objective: To evaluate whether data mining in a pharmacovigilance database can be used to identify known and possible novel risk factors for ADRs, for use in pharmacovigilance practice.

Method: Exploratory data mining was performed within the Swedish national database of spontaneously reported ADRs. Bleeding associated with direct oral anticoagulants (DOACs)-rivaroxaban, apixaban, edoxaban, and dabigatran-was used as a test model. We compared demographics, drug treatment, and clinical features between cases with bleeding (\(N = 965\)) and controls who had experienced other serious ADRs to DOACs (\(N = 511\)). Statistical analysis was performed by unadjusted and age adjusted logistic regression models, and the random forest based machine-learning method Boruta.

Results: In the logistic regression, 13 factors were significantly more common among cases of bleeding compared with controls. Eleven were labelled or previously proposed risk factors. Cardiac arrhythmia (e.g., atrial fibrillation), hypertension, mental impairment disorders (e.g., dementia), renal and urinary tract procedures, gastrointestinal ulceration and perforation, and interacting drugs remained significant after adjustment for age. In the Boruta analysis, high age, arrhythmia, hypertension, cardiac failure, thromboembolism, and pharmacodynamically interacting drugs had a larger than random association with the outcome. High age, cardiac arrhythmia, hypertension, cardiac failure, and pharmacodynamically interacting drugs had odds ratios for bleeding above one, while thromboembolism had an odds ratio below one.

Conclusions: We demonstrated that data mining within a pharmacovigilance database identifies known risk factors for DOAC bleeding, and potential risk factors such as dementia and atrial fibrillation. We propose that the method could be used in pharmacovigilance for identification of potential ADR risk factors that merit further evaluation.
1 | INTRODUCTION

A primary aim of post-marketing pharmacovigilance is to identify signals of suspected adverse drug reactions (ADRs) as early as possible. A cornerstone within this field is the spontaneous reporting of suspected ADRs from healthcare providers and other reporters to regulatory agencies and pharmaceutical companies, usually with focus on ADRs not discovered in clinical trials. Different data mining algorithms are used to find new safety signals. They are often based on disproportionality analyses that calculate the ratio between observed and expected numbers of reports for drug-event combinations, with or without mathematical modifications. There has been less focus on the identification of possible risk factors for ADRs, aside from the suspected drug based on spontaneous reports, possibly due to a perceived lack of control group. When attempted, such efforts rely either on manual reviews of case reports or on analyses stratified or adjusted for broad, often demographic, categories such as age and sex. However, since a wide range of different ADRs are reported for a drug, and since such reports stem from patients with a similar treatment indication, it should be possible to use patients with other ADRs to the same drug as controls, in order to discover potential new risk factors. Our research group has previously applied this technique to data in the pharmacovigilance database kept by the Swedish Medical Products Agency (SMPA) in data mining studies where we let reports of an ADR of interest serve as cases, and all other reports for the same drug as controls.

In the current study, we further investigate whether this methodology could identify possible risk factors for bleeding during treatment with direct oral anticoagulants (DOACs). If so, we should be able to detect known, and possibly yet unidentified risk factors.

In Sweden, the most used oral anticoagulant was for many years warfarin, a vitamin K antagonist. Since 2014, the use of DOACs-rivaroxaban, apixaban, edoxaban, and dabigatran-has increased in Sweden, and apixaban is now the leading oral anticoagulant. Atrial fibrillation is the most common indication for DOAC treatment. Most ADRs on anticoagulants reported to the SMPA pharmacovigilance database by healthcare professionals concern apixaban and warfarin, followed by rivaroxaban. Known risk factors for bleeding associated with DOACs are described in the European public assessment report (EPAR) Summary of product characteristics (SmPC). These labelled risk factors include high age, low-body weight, and reduced renal function. Additional labelled risk factors are recent surgery or invasive interventions, liver dysfunction, recent gastrointestinal ulceration or intracranial bleeding, malignancies, trauma or surgery to the brain, spine or eye, esophageal varices, vascular malformations or aneurysms, concomitant treatment with another anticoagulant, a selective serotonin reuptake inhibitor (SSRI), serotonin and noradrenaline reuptake inhibitor (SNRI), nonsteroidal anti-inflammatory drug (NSAID), an antplatelet drug, or drugs that inhibits the Cytochrome P450 enzyme 3A4 (CYP3A4) or P-glycoprotein (P-gp).

**Keywords**
- anticoagulant-induced bleedings
- direct oral anticoagulants
- pharmacovigilance database
- risk factors
- suspected adverse drug reactions

**Key Points**
1. We developed a new method for the identification of potential risk factors for adverse drug reactions in a pharmacovigilance database.
2. We used bleeding associated with direct oral anticoagulants (DOACs)-rivaroxaban, apixaban, edoxaban, and dabigatran-as a test model.
3. We identified 13 risk factors of which 11 were previously known.
4. Previously unknown potential risk factors were dementia and atrial fibrillation.
5. Implementing this method in pharmacovigilance practices could improve patient safety.

**Plain Language Summary**
Pharmacovigilance databases contain reported adverse drug reactions and information about the patient, other drugs taken at the same time as well as other diseases that could be risk factors for the adverse drug reaction. In this study, we evaluated a method for the identification of potential risk factors in a database of spontaneously reported adverse drug reactions in Sweden. Knowledge of risk factors is important to guide the selection of the best drug for an individual or patient group. Bleeding associated with drugs that prevent blood clotting, that is direct oral anticoagulants, was used as a test model. We compared demographics, treatment, and clinical characteristics between cases with bleeding and those without bleeding. We identified known risk factors for bleeding, but also a few previously unknown potential risk factors such as dementia and atrial fibrillation. We propose that the method could be used to identify potential risk factors for adverse drug reactions.
In addition to the labelled risk factors, epidemiological studies have proposed risk factors such as prior stroke/transient ischemic attack, heart failure, uncontrolled hypertension, cardiovascular disease, alcohol abuse and concomitant treatment with paracetamol.\textsuperscript{13,14}

2 | OBJECTIVES

To evaluate whether data mining in a pharmacovigilance database can be used to identify known and possible novel risk factors for ADRs, for use in pharmacovigilance practice. The study used bleeding events associated with DOACs as a test model.

3 | MATERIALS AND METHODS

3.1 | Pharmacovigilance database

This was an exploratory data mining study using case–control methodology to identify possible risk factors for bleeding during treatment with DOACs. We used data from the SMPA database of suspected ADRs reported spontaneously by healthcare professionals. This database contains reports sent to the SMPA from healthcare professionals and consumers since 1965 and on December 31, 2019, it held approximately 200,000 reports. The database uses the Medical Dictionary for Regulatory Activities (MedDRA) terminology for coding of ADR terms and diseases,\textsuperscript{15} and the World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology’s Anatomical Therapeutic Chemical (ATC) classification for the coding of drugs.\textsuperscript{16}

All spontaneously reported ADRs for DOACs (dabigatran, rivaroxaban, apixaban and edoxaban with ATC codes B01AF01, B01AF02, B01AF03, and B01AE07, respectively) received by the SMPA from healthcare professionals up until December 31, 2019, were retrieved and examined (n = 2268). Data on suspected and concomitant drugs (full ATC code), reported ADRs, age and sex of the patient, treatment indication, current and past diseases, and other relevant information from the narrative of the case were extracted. MedDRA terms were restricted to preferred terms (PT) and high level group terms (HLGT).

To minimize lack of completeness, reports that had been assessed by the SMPA as nonserious were excluded, which gave a total number of 1482 individual case reports assessed as serious. Two reports of ADRs in new-borns after exposure in utero were also excluded. All reports were then categorized into two groups: bleeding or other ADRs (Supplementary Table 1). When a report included multiple ADRs and at least one of these was a bleeding event (n = 113), bleeding was selected as decisive for the choice of group. Four reports were contradictory (described both bleeding and lack of efficacy) and were excluded.

The final number of reports included in the study was 1476 (Table 1). When multiple drugs were suspected as causative by the reporter, the non-DOACs were analyzed as concomitant drugs. Doses of DOACs were converted to defined daily doses (DDD) to be made comparable, according to the definitions of the WHO Collaborating Centre for Drug Statistics Methodology.\textsuperscript{16} The DDD of rivaroxaban was 20 mg, apixaban 10 mg, edoxaban 60 mg, and dabigatran 300 mg.\textsuperscript{15}

Concomitant drugs were categorized as inhibitors of the drug metabolizing enzyme CYP3A4 and the drug transporter P-gp, based on the University of Washington Drug Interaction Database,\textsuperscript{17} or as risk drugs for pharmacodynamic interactions due to increased risk of bleeding according to the SmPCs of the DOACs\textsuperscript{12} (Supplementary Table 2).

3.2 | Statistical methods

Cases of bleeding were compared with controls, which were all other reports of serious ADRs to DOACs. The variables evaluated as potential risk factors were treatment indications, current and previous diseases, daily dose of the suspected drug, age, sex, and concomitant drugs. Only variables with at least five observations among either
<table>
<thead>
<tr>
<th>Variables</th>
<th>Bleeding/No bleeding/Controls</th>
<th>p-value</th>
<th>OR (95% CI)</th>
<th>Age adjusted p-value</th>
<th>Age adjusted OR (95% CI)</th>
<th>Boruta analysis</th>
<th>Risk factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years (mean)</td>
<td>78.1</td>
<td>70.0</td>
<td>$1.81 \times 10^{-2}$</td>
<td>$1.05^{a}(1.04-1.06)$</td>
<td>1.07</td>
<td>Included</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Concomitant diseases (MedDRA)$^{a}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias$^{e}$</td>
<td>730</td>
<td>287</td>
<td>$3.60 \times 10^{-14}$</td>
<td>2.42 (1.92-3.04)</td>
<td>1.87 $\times 10^{-5c}$</td>
<td>Included</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation$^{f}$</td>
<td>718</td>
<td>276</td>
<td>$4.08 \times 10^{-15c}$</td>
<td>2.48 (1.97-3.10)</td>
<td>6.95 $\times 10^{-6c}$</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Vascular hypertensive disorders$^{e}$</td>
<td>351</td>
<td>128</td>
<td>$1.10 \times 10^{-5c}$</td>
<td>1.71 (1.35-2.17)</td>
<td>5.48 $\times 10^{-3}$</td>
<td>Included</td>
<td>Proposed$^{14}$</td>
</tr>
<tr>
<td>Hypertension$^{f}$</td>
<td>307</td>
<td>103</td>
<td>$2.37 \times 10^{-6c}$</td>
<td>1.85 (1.43-2.39)</td>
<td>1.26 $\times 10^{-3}$</td>
<td>N/A</td>
<td>Proposed$^{14}$</td>
</tr>
<tr>
<td>Mental impairment disorders$^{e}$</td>
<td>57</td>
<td>11</td>
<td>$1.69 \times 10^{-3}$</td>
<td>2.85 (1.48-5.49)</td>
<td>4.85 $\times 10^{-2}$</td>
<td>Omitted</td>
<td></td>
</tr>
<tr>
<td>Dementia$^{f}$</td>
<td>30</td>
<td>6</td>
<td>$2.75 \times 10^{-2}$</td>
<td>2.70 (1.12-6.53)</td>
<td>0.29</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Heart failures$^{e}$</td>
<td>146</td>
<td>49</td>
<td>$3.01 \times 10^{-3}$</td>
<td>1.68 (1.19-2.37)</td>
<td>0.21</td>
<td>Included</td>
<td>Proposed$^{13,14}$</td>
</tr>
<tr>
<td>Cardiac failure$^{f}$</td>
<td>139</td>
<td>45</td>
<td>$2.16 \times 10^{-3}$</td>
<td>1.74 (1.22-2.48)</td>
<td>0.15</td>
<td>N/A</td>
<td>Proposed$^{13,14}$</td>
</tr>
<tr>
<td>Central nervous system vascular disorders$^{e}$</td>
<td>176</td>
<td>67</td>
<td>$4.56 \times 10^{-3}$</td>
<td>1.55 (1.15-2.10)</td>
<td>0.13</td>
<td>Tentative</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Renal and urinary tract therapeutic procedures$^{e}$</td>
<td>23</td>
<td>3</td>
<td>$2.13 \times 10^{-2}$</td>
<td>4.13 (1.24-13.8)</td>
<td>3.20 $\times 10^{-2}$</td>
<td>Omitted</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Gastrointestinal ulceration and perforation$^{e}$</td>
<td>20</td>
<td>2</td>
<td>$2.35 \times 10^{-2}$</td>
<td>5.39 (1.25-23.1)</td>
<td>2.36 $\times 10^{-2}$</td>
<td>Omitted</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Gastrointestinal neoplasms malignant and unspecified$^{e}$</td>
<td>22</td>
<td>3</td>
<td>$2.62 \times 10^{-2}$</td>
<td>3.95 (1.18-13.3)</td>
<td>5.93 $\times 10^{-2}$</td>
<td>Omitted</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Coronary artery disorders$^{e}$</td>
<td>129</td>
<td>50</td>
<td>$4.03 \times 10^{-2}$</td>
<td>1.43 (1.02-2.03)</td>
<td>0.43</td>
<td>Tentative</td>
<td>Proposed$^{12,14}$</td>
</tr>
<tr>
<td>Myocardial infarction$^{f}$</td>
<td>67</td>
<td>18</td>
<td>$8.44 \times 10^{-3}$</td>
<td>2.04 (1.20-3.48)</td>
<td>0.13</td>
<td>N/A</td>
<td>Proposed$^{12,14}$</td>
</tr>
<tr>
<td>Concomitant drug treatment (ATC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B01AC06 acetylsalicylic acid</td>
<td>50</td>
<td>12</td>
<td>$1.19 \times 10^{-2}$</td>
<td>2.27 (1.20-4.31)</td>
<td>2.67 $\times 10^{-2}$</td>
<td>N/A (PD interacting agents included)</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>C08CA01 amlodipine</td>
<td>65</td>
<td>19</td>
<td>$1.89 \times 10^{-2}$</td>
<td>1.87 (1.11-3.15)</td>
<td>3.52 $\times 10^{-2}$</td>
<td>N/A (CYP3A4 inhibitor tentative)</td>
<td>Labelled$^{12}$</td>
</tr>
<tr>
<td>Pharmacodynamic interacting agents</td>
<td>172</td>
<td>68</td>
<td>$2.58 \times 10^{-2}$</td>
<td>1.41 (1.04-1.91)</td>
<td>1.31 $\times 10^{-2}$</td>
<td>Included</td>
<td>Labelled$^{12}$</td>
</tr>
</tbody>
</table>

Note: The table shows findings with p-values of <0.05 and increased odds ratios (OR) with 95% normal approximation confidence intervals (CI) in 965 cases of bleeding and 511 controls with other adverse drug reactions (ADRs). Labelled risk factor = risk factor according to the approved European Summary of product characteristics (SmPCs). Proposed risk factor = risk factor proposed in other studies (only shown when not labelled).

Abbreviations: N/A, not applicable since only higher level terms were included in the Boruta; PD, pharmacodynamic.

$^{a}$Odds ratio by increase of 1 year.

$^{b}$Concomitant diseases includes both current and past conditions as well as indications. Presented in Medical Dictionary for Regulatory Activities (MedDRA) hierarchy.

$^{c}$p-value passes the threshold for multiple tests = 1.94 $\times 10^{-4}$.

$^{d}$Predominantly (n = 136) hemorrhages and cerebrovascular accidents.

$^{e}$MedDRA high-level group term (HLGT).

$^{f}$MedDRA preferred term (PT).
cases or controls were included for categorical yes/no variables. Means and standard deviations (SD) were calculated for the daily dose of the suspected drug and patient age. The association between potential risk factors and the outcome of bleeding (dependent variable) was assessed by unadjusted and age adjusted logistic regression models, one factor at a time. A p-value of <0.05 was set as statistically significant, but since multiple tests were performed a p-value was also calculated according to the Bonferroni correction; 0.05/258 number of tests, that is, $1.94 \times 10^{-4}$ (Table 2).

As a complement to logistic regression, we used the Boruta method which is based on the machine learning algorithm Random Forest (RF). The RF analysis was based on all variables in contrast to the logistic regression analyses that were based on a maximum of two variables (i.e., age and a potential risk factor). Although models built using RF can be used for prediction, our purpose of using RF was to present the importance of the variables included in the model that is, ranking the variables according to their importance for the outcome. The RF analysis and the Boruta analysis included the variables age, sex, concomitant diseases in MedDRA HLGT, and concomitant drug treatment. For pharmacodynamically and pharmacokinetically interacting drugs, we only included the interacting group and not the individual drugs so that no drugs were included twice (229 variables). To classify which variables contributed to the RF model more than expected by chance we used the Boruta algorithm. The Boruta method performed multiple runs of RF and compared each variable’s importance to the maximum importance expected by chance, calculated on permuted copies of the original variables included in the same RF run. Variables performing worse than chance were removed from the model in a stepwise manner until all variables were either omitted, included, or the maximum number of Boruta runs was reached. Variables indistinguishable from noise after the maximum number of Boruta runs were labelled as tentative. The following Boruta settings were used: maximum number of Boruta runs: 100, variable importance mode: permutation, splitrule: maxstat (maximally selected rank statistics), number of trees: 10000 trees, mtry (randomly selected number of variables to possibly split in each node): rounded down square root of the number of variables.

Statistical calculations were done in R, version 4.2.1 for Windows. The RF and Boruta analyses were performed using the R packages Boruta and Ranger.

4 | RESULTS

The characteristics of the 1476 patients included in the analysis are shown in Table 1. The age range was 19–100 years, with a mean age

![FIGURE 1](#)  Top 15 variables out of 229 in the Boruta analysis of the variables age, sex, concomitant drug treatment, and concomitant diseases in the Medical Dictionary for Regulatory Activities (MedDRA) High Level Group Term (HLGT). The Shadow (shadowMax) variable indicates the maximum importance of the permuted variables that were included in each Boruta run.
of 78 years in the bleeding group and 70 years in the control group, and 55–50% were males. The most used DOAC was apixaban, followed by rivaroxaban, both in the case and control group. The most common indication for treatment was supraventricular arrhythmia, for example, atrial fibrillation (n = 595 for the cases and 237 for the control group), followed by peripheral thromboembolism, for example, deep vein thrombosis and prophylaxis of thrombosis. The MedDRA terms categorized as bleeding and the MedDRA terms categorized as other ADRs are shown in Supplementary Table 1.

The mean DDD was 0.89 ± 0.31 SD in the bleeding group and 0.86 ± 0.28 SD in the control group.

Table 2 shows variables with increased ORs for bleeding and p < 0.05 together with information on whether these factors are labelled in the SmPCs\textsuperscript{2,3} or have been proposed as risk factors in other studies. A total of 13 variables with p-values of <0.05 and increased ORs were found (Table 2), of which 11 were labelled or proposed risk factors (85%). High age, atrial fibrillation, and hypertension were statistically significant after correction for multiple testing.

One of the variables associated with bleeding was the main indication for DOAC treatment atrial fibrillation (MedDRA PT). Other variables were current cardiac failure (MedDRA PT) and previous myocardial infarction (MedDRA PT). Central nervous system vascular disorders (MedDRA HLGT) including current or previous stroke were associated with bleeding. Gastrointestinal ulceration and perforation (MedDRA HLGT), and malignant and unspecified gastrointestinal neoplasms (MedDRA HLGT) were also associated with bleeding. These diagnoses included gastric, duodenal, and hemorrhagic ulcer, and gastric, colon, rectal, intestinal, pancreatic, lip, and buccal cancer. The MedDRA HLGT term mental impairment disorder associated with bleeding included current dementia or vascular dementia. Renal and urinary tract therapeutic procedures (MedDRA HLGT) associated with bleeding included reports of bladder catheterization and nephrectomy. Hypertension (MedDRA PT) was reported as a current medical condition associated with bleeding.

The RF run showed that 11 variables had estimated variable importance larger than zero (Supplementary Figure 1) and the Boruta analysis showed that six of these variables had a greater than random association with bleeding: high age, cardiac arrhythmias (i.e., atrial fibrillation), vascular hypertensive disorders (i.e., hypertension), thromboembolism, pharmacodynamic interacting agents, and heart failure (Figure 1). Of these six, five had odds ratios for bleeding above one in the logistic regression analyses and one (thromboembolism) had an odds ratio below one. In the logistic regression analyses, the odds ratio below one for thromboembolism (venous thrombosis and embolism) remained significant after correction for multiple testing (unadjusted \( p = 4.71 \times 10^{-7}, OR = 0.45 \) [95% confidence interval 0.33–0.61]). Five variables were tentatively associated with bleeding: CYP3A4 inhibitors, coronary artery disorders, sex, central nervous system vascular disorders, and pulmonary vascular disorders (pulmonary embolism). Of these five, four had odds ratios for bleeding above one in the logistic regression analyses and one, pulmonary vascular disorders, that is pulmonary embolism, had an odds ratio below one (unadjusted \( p = 2.84 \times 10^{-4}, OR = 0.53 \) [95% confidence interval 0.38–0.75]).

5 | DISCUSSION

In the current study, we have evaluated a method for identification of possible risk factors that could be used in pharmacovigilance practice, be it in initial signal assessment or in the characterization of already known ADRs. The method successfully identified several known risk factors for bleeding during DOAC treatment and two putative risk factors.

The known risk factors for bleeding during DOAC treatment identified were high age, a history of central nervous system hemorrhage, renal and urinary tract therapeutic procedures, gastrointestinal ulceration/perforation, and malignancies, as well as concomitant treatment with the CYP3A4 inhibitor amiodipine and agents interacting pharmacodynamically such as platelet aggregation inhibitors (including acetylsalicylic acid). We further identified factors that have been proposed by others, including hypertension, current heart failure, vascular disease (including ischemic heart disease) and myocardial infarction.\textsuperscript{14}

To our knowledge, a history of atrial fibrillation and dementia have not previously been associated with bleeding during treatment with DOAC and could merit investigation in further studies. However, it should be noted that dementia was not identified in the RF/Boruta analysis.

Systematic identification of risk factors defined in broad, often demographic terms may be part of regular disproportionality analyses of spontaneous ADRs. However, identification of for example, comorbidities or concomitant treatment as possible risk factors are typically done by manual review. This process is slow and risks overlooking factors that are not obvious. A structured methodology for hypothesis generation would add an important tool for the timely identification of factors that may increase the risk of ADRs. The methodology presented in this study would require several things to be in place before it could be used on a routine basis. The information in the reports need to be highly structured in the database, so that clinical factors are readily and unambiguously available. This could either be done at the time of entering a report in the pharmacovigilance database, or through quality-assured text-mining algorithms. In addition, classification of concomitant drugs as inhibitors/inducers of metabolizing enzymes or transporters could be achieved through linkage with existing databases. Finally, since any association identified with the proposed methodology is hypothetical, there is a need for readily available resources for further validation in other materials, such as registry-based epidemiological studies. These hurdles thus remain but can be overcome, with the potential for quicker identification of new risk factors. Such activities would not necessarily be limited to serious ADRs reported from healthcare professionals but could also be used for nonserious ADRs and ADRs reported from consumers.

There are several limitations to the present study. First, the study population represents only those reported by healthcare professionals as having experienced a suspected ADR to a DOAC, and not a random selection of patients treated with a DOAC. Although this may limit the representativeness, all patients are from the same population of treated patients and the results do not indicate that spurious associations were common. Several known risk factors were identified,
demonstrating that the method is valid for hypothesis generation. Second, reports of serious and life-threatening ADRs could be biased towards better completeness, thereby presenting the possibility that these reports contain a more thorough description of patient characteristics. Although we cannot completely rule out this possibility, we attempted to minimize the problem by excluding reports of nonserious ADRs. Third, some factors identified as possible risk factors could be associated with others, making them guilty by association. Fourth, with the chosen study design, we cannot assess statistical interaction terms between the drug of interest and patient characteristics that increase the risk of the studied reaction per se. ADR reports are based on suspicion of an association between a reaction and a treatment, and some factors may be identified as risk factors for an ADR when they are actually risk factors for that reaction. If frequent, a clinical characteristic underlying the reaction, for example bleeding, could be mistaken for a risk factor for an ADR. This is more likely when an association is weak, and further emphasizes the need for validation studies using other data and different methodologies. It is important to stress that statistically significant associations do not prove causality but should be viewed as hypothesis generating. Fifth, we cannot exclude the possibility that health care professionals were prone to report ADRs associated with known risk factors, such as those that are labelled or have been suggested in previous studies. This potential selection bias could have influenced the results so that they were biased towards the identification of already established risk factors. There is also a risk for false negative findings, in particular if a clinical characteristic is rare, which can present problems with statistical power. This is illustrated by three labelled risk factors (Renal and urinary tract therapeutic procedures, Gastrointestinal ulceration and perforation, Gastrointestinal neoplasms malignant and unspecified) that were found to be significant in the logistic regression, but not in the Boruta analysis. Sixth, it is possible that risk factors may differ between individual DOACs, as has been suggested by others, and this could have been further explored through, for example, analyses stratified by individual drug. However, this would have presented an issue with statistical power.

Logistic regression analyses were performed for one variable at a time. Adjustment was only made for age, which is a risk factor for most diseases. The reason not to adjust for several risk factors in one larger model, was to reduce the risk of missing an interesting, but numerically weak association. In an attempt to address this, we used the Boruta analysis, based on RF analysis, where all potential risk factors are analyzed at once including quite complex interactions. However, these results should be interpreted with caution, and each finding should be regarded as a potential risk factor that needs further investigation in other studies.

6 \ CONCLUSIONS

The current study evaluated the use of a data mining method within a pharmacovigilance database of spontaneously reported ADRs to identify possible risk factors for such reactions, using bleeding associated with DOACs as a test model. A total of 13 clinical factors were identified, of which 11 were either labelled in the product information or previously proposed risk factors. We show that this method identified both known and possible novel risk factors for further investigation in other studies.

Two factors not previously described as increasing the risk of bleeding were identified as suggestive risk factors: atrial fibrillation and dementia. However, the risk of DOAC-induced bleeding is known to be lower than the risk of stroke and systemic embolism in untreated atrial fibrillation, and therefore there is a net clinical benefit of DOAC treatment in patients with atrial fibrillation.

In summary, we detected several known risk factors for DOAC-induced bleeding, as well as two new potential risk factors, by data mining in a pharmacovigilance database. We propose that pharmacovigilance registries could be used for the identification of potential risk factors that merit further study. Implementing this method in pharmacovigilance practices could contribute to improved patient safety by enabling timely identification and management of potential drug-related risks.

AUTHOR CONTRIBUTIONS

All authors contributed to the conceptualization of this work and the design of the study. Sofia Attelind and Niclas Eriksson performed the data analysis, and all authors interpreted the results. Sofia Attelind drafted the first draft of the manuscript. All authors read, revised, edited, and approved the final version of the manuscript.

FUNDING INFORMATION

Funding from the Swedish Brain Foundation (FO2020-0234), Swedish Heart-Lung Foundation (20200777), STROKE-Riksförbundet, Selander’s Foundation, and Clinical Research Support at Uppsala University Hospital (ALF). Funding of open access by Uppsala University.

CONFLICT OF INTEREST STATEMENT

Sofia Attelind and Anders Sundström are employed at the Swedish Medical Products Agency, SE-751 03 Uppsala, Sweden. The views expressed in this paper are the personal views of the authors and not necessarily the views of the Government agency.

Niclas Eriksson reports institutional research grant from Bristol-Myers Squibb/Pfizer. Mia Wadelius and Pär Hallberg declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the current study are not publicly available due to the European General Data Protection Regulation, which requires us to protect the identity of participants, but datasets are partly available from the corresponding author on reasonable request.

ETHICS STATEMENT

This study was approved by the Swedish Ethical Review Authority (2019-05523).
ORTCID
Sofia Attelind https://orcid.org/0000-0002-7631-7376
Niclas Eriksson https://orcid.org/0000-0002-2152-4343
Anders Sundström https://orcid.org/0000-0003-2337-3371
Mia Wadelius https://orcid.org/0000-0002-6368-2622
Pär Hallberg https://orcid.org/0000-0003-3465-3280

REFERENCES

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.