The risk of bleeding with warfarin: A systematic review and performance analysis of clinical prediction rules

Karen Dahri, Peter Loewen
Vancouver Coastal Health/Providence Health Care Pharmacy Services, Vancouver, British Columbia, Canada; Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Summary
It was the objective of this article to qualitatively review and evaluate the clinical prediction rules (CPRs) available for estimating bleeding risk in patients commencing warfarin therapy. A systematic review of PubMed (1949 to December 2006), MEDLINE (1966 to December 2006), EMBASE (1980 to December 2006), Cochrane Database of Systematic Reviews (to December 2006), and International Pharmaceutical Abstracts (1970 to December 2006) was conducted. Seven studies were found that detailed CPRs used to assess risk of bleeding prior to commencing warfarin therapy. Four studies described distinct CPRs. The remaining three studies were further validations of one of the CPRs, the Outpatient Bleeding Risk Index. The Outpatient Bleeding Risk Index was classified as being of Level 2 evidence while the remaining three indices were classified as being of Level 4 evidence. In no case did the CPRs exhibit performance characteristics that would indicate “strong” ability to predict the presence of absence of major bleeding among warfarin recipients. The modified Outpatient Bleeding Risk Index exhibited moderate predictive ability for major bleeding in two studies, although pooling of all studies of this CPR did not reveal moderate or better performance. None of the CPRs identified “any bleeding” with moderate or strong predictive ability. None of the available CPRs exhibit sufficient predictive accuracy or have trials evaluating the impact of their use on patient outcomes. Hence, no existing CPR can be recommended for widespread use in practice at present.

Keywords
Bleeding risk, warfarin, clinical prediction rules

Introduction
Warfarin is a vitamin K antagonist which has been shown to be an effective anticoagulant for many different indications. Guidelines recommend its use in patients for primary and secondary prevention of venous thromboembolism (VTE), for the prevention of systemic embolism in patients with prosthetic heart valves or atrial fibrillation, for the prevention of acute myocardial infarction, and for the prevention of stroke, recurrent infarction or death in patients with acute myocardial infarction (1). Bleeding is the main adverse effect of concern when using warfarin. Estimates of bleeding rates vary widely depending on study design, with annual incidences of 0.6% for fatal bleeding, 3.0% for major bleeding and 9.6% for major or minor bleeding being reported (2). Many different factors, such as concurrent use of interacting medications, can increase a patient’s risk for bleeding (3). Risk assessment tools that are able to quantify an individual patient’s risk of bleeding while on warfarin would be a valuable addition to clinical practice by aiding clinicians in evaluating the benefits versus risks of initiating therapy. The purpose of this review was to systematically review and evaluate the performance of the clinical prediction rules (CPR; sometimes called clinical decision rules) available for evaluating bleeding risk in patients commencing warfarin therapy.

Methods
A systematic review based on the Cochrane Collaboration’s standard methodology was employed (4).

Criteria for considering studies for this review
Types of studies
We included studies that prospectively or retrospectively evaluated the ability of a CPR using only readily available clinical parameters to distinguish between patients at high and low risk of experiencing major bleeding on warfarin therapy for any indication. We excluded studies that described CPRs that were published in abstract format only, were conference proceedings, and were published in languages other than English.

Received April 23, 2007
Accepted after resubmission August 15, 2007
Prepublished online October 11, 2007
doi:10.1160/TH07–04–0297
Types of participants
Participants included adults with any condition for which warfarin was prescribed.

Types of intervention
To be included, participants had to have undergone risk stratification via a CPR that used only clinical variables readily available prior to beginning warfarin therapy. CPRs that included post-commencement findings as part of the system were excluded unless the utility of the CPR was also evaluated in the absence of this data.

Types of outcome measures
The primary outcome of interest was the ability to distinguish between patients at high and low risk of experiencing major bleeding on warfarin therapy. Predictive ability was defined in terms of likelihood ratios (LR). Authors' definitions of “major bleeding” were accepted provided they were similar to the generally accepted TIMI (Thrombolysis in Myocardial Infarction) bleeding criteria (5) and American College of Chest Physicians definition (6).

Secondary outcome measures included the ability of the tools to predict mortality, minor bleeding, and erratic international normalized ratio (INR).

Search strategy for identification of studies
A systematic search of the following databases was performed: PubMed (1949 to December 2006), MEDLINE (1966 to December 2006), EMBASE (1980 to December 2006), Cochrane Database of Systematic Reviews (to December 2006), and International Pharmaceutical Abstracts (1970 to December 2006). The following search terms were used: risk assessment, assessment tools, clinical prediction tools, bleeding risk, bleeding, and warfarin. In MEDLINE, a search strategy previously identified as having 98% sensitivity for identifying CPRs was employed (7). A manual review of reference lists from retrieved articles was performed to identify any additional studies.

Methods of the review
Initially two reviewers independently evaluated the titles, abstracts, and citations of all identified articles to select those potentially meeting the inclusion criteria. Articles so selected were independently subjected to full-text review to establish whether they met the inclusion criteria and did not meet any exclusion criteria. Concordance between reviewers was measured by simple agreement. Any disagreements were resolved by discussion and consensus. Reviewers were not blinded to author or journal names.

Quality assessment
Both reviewers independently subjected the included studies to quality assessment. Three methods of quality assessment were carried out according to the methodologic standards published by Laupacis et al. (8), McGinn et al. (9), and the hierarchy of evidence for CPRs published by McGinn (9). For the methodologic standards tools, each question was assigned a score of 0 (criteria not met), 1 (criteria partially met), 2 (criteria fully met). The total score was summed to give an overall quality score.

Data extraction
Both reviewers independently read each article and abstracted data using a standardized content review form. Data extracted included: the aim of the study, design, number, type, and geographic location of sites included, sample size, duration, inclusion and exclusion criteria, patient characteristics, elements comprising the CPR, and results. Attempts were made to acquire additional information from investigators as required. Discrepancies of data extraction were resolved by consensus through review of the published report.

Analysis of CPR performance characteristics
The LR for each stratum of the CPRs was calculated using published data and the method of Peirce and Cornell (10). In this case, LRs represent the ratio of the probability of each test result (e.g. low, intermediate, high risk using CPR) in people who end up experiencing bleeding to the probability in those who do not bleed. LR is the most directly clinically applicable measure of diagnostic test performance, particularly when the tests (in this case, CPRs) produce more than two strata of results (in this case low, intermediate, or high risk of bleeding) (11). The cutoffs for performance based on LR proposed by Jaeschke et al. were used in interpreting these estimates (12, 13). Performance was considered moderate when the LR was greater than 5.0 or less than 0.20. Performance was considered strong when the LR was greater than 10.0 or less than 0.10.

For studies evaluating identical CPRs in different populations, we attempted to quantitatively combine the results in order to more precisely estimate the performance of the CPR.

Planned subgroup analysis
No subgroup analyses were prospectively planned.

Results
Description of studies
Ten studies meeting the inclusion criteria were identified (14–23). One study (14) was excluded as one of the predictors of bleeding risk used in this assessment tool was “achieved INR” which would not be available in a warfarin-naïve patient. Two studies were excluded as they assessed risk of bleeding for patients receiving anticoagulation only while in hospital and required “achieved prothrombin time (PTT)” as one of the variables (15, 16). The seven remaining studies included in this review are detailed in Table 1 (17–23).

Patient characteristics
Patient characteristics varied within the included studies. Mean age of subjects ranged from 58.4 years (19) to 80.2 years (22), although one study did not report mean age of subjects (23). Two of the studies specifically evaluated patients with VTE/pulmonary embolism (PE) (19, 21), three of the studies evaluated patients with atrial fibrillation (20, 22, 23) and two of the studies evaluated patients receiving anticoagulation for mixed indications (17, 18).

Intervention characteristics
The Outpatient Bleeding Risk Index was the most frequently studied CPR. Originally developed from a retrospective review
<table>
<thead>
<tr>
<th>Study [name of CPR]</th>
<th>Sample size</th>
<th>Age</th>
<th>Length of follow-Up</th>
<th>Indication for anticoagulation</th>
<th>Major bleeding definition</th>
<th>Methodology (Level of evidence)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landefeld 1989(17) [OBRI]</td>
<td>Test = 375 Validation = 187 [prospective]</td>
<td>61 ± 14 years</td>
<td>876 patient years of outpatient warfarin therapy (mean approx 18 months)</td>
<td>Mixed indications</td>
<td>Was fatal, was life-threatening, was potentially life-threatening, led to severe blood loss, led to surgical treatment or led to moderate blood loss that was acute or subacute and was not explained by trauma or surgery.</td>
<td>Derived and validated in a split sample or a retrospective database (Level 4)</td>
<td></td>
</tr>
<tr>
<td>Beyth 1998(18) [mOBRI]</td>
<td>Test = 565 Validation = 264 [prospective] Validation = 61 ± 14 years Validation = 60 ± 6 years</td>
<td>48 months</td>
<td>Mixed indications</td>
<td>Overt bleeding that led to the loss of at least 2 units in 7 days or less, or was otherwise life-threatening (e.g. intracranial bleeding).</td>
<td>Validated in a prospective sample (with or without retrospective database validation) (level 2–3)</td>
<td>Same cohort used as in Landefeld to reassess previously developed Outpatient Bleeding Risk Index; only difference is use of entire sample.</td>
<td></td>
</tr>
<tr>
<td>Wells 2003(19) [mOBRI]</td>
<td>Validation = 222 [prospective]</td>
<td>58.4 years</td>
<td>18.5 ± 14.9 months</td>
<td>Venous thromboembolism/ pulmonary embolism</td>
<td>Loss of 2 units of blood in a 7-day period or bleeding that was otherwise life threatening.</td>
<td>Validated in a prospective sample (with or without retrospective database validation) (level 2–3)</td>
<td>Only 2 patients classified in high risk category [inadequate # of outcome events]</td>
</tr>
<tr>
<td>Aspinall 2005(20) [mOBRI]</td>
<td>Validation = 269 [prospective]</td>
<td>67.9 ± 11.4 years</td>
<td>1,308 patient years of warfarin therapy (mean approx 12 months)</td>
<td>Atrial fibrillation</td>
<td>Patient was hemodynamically unstable, required a transfusion, had an intracranial hemorrhage, or died (e.g. a GI bleed in a hypotensive patient, subdural hematoma)</td>
<td>Validated in a prospective sample (with or without retrospective database validation) (level 2–3)</td>
<td>Portion of sample had data collected retrospectively.</td>
</tr>
<tr>
<td>Kuijer 1999(21) [SBRPS]</td>
<td>Test = 241 Validation = 780 [Not clear whether prospective or retrospective]</td>
<td>Test = 63 ± 17 Validation = 60 ± 17</td>
<td>3 months</td>
<td>Venous thromboembolism/ pulmonary embolism</td>
<td>Clinically overt and associated with a decline in hemoglobin concentration of at least 20 g/l, if there was a need for transfusion of 2 units or more of red blood cells, if it was retroperitoneal or intracranial, or if it warranted permanent discontinuation of treatment.</td>
<td>Derived and validated in a split sample or a retrospective database (level 4)</td>
<td></td>
</tr>
<tr>
<td>Gage 2006(22) [HEMORRHAGES]</td>
<td>Total = 3,791; 1,694 received warfarin [Size of test vs. validation population not clear] [retrospective]</td>
<td>Warfarin = 79 years</td>
<td>3,138 patient years of follow-up (mean approx 10 months)</td>
<td>Atrial fibrillation</td>
<td>ICD-9CM codes</td>
<td>Derived and validated in a split sample or a retrospective database (level 4)</td>
<td>Genetic factors used in scoring not available. Two other factors present in &lt;1% of the sample.</td>
</tr>
<tr>
<td>Shireman 2006 (23) [name of CPR]</td>
<td>Test = 19,875 Validation = 6,470 [retrospective]</td>
<td>All ≥65 years: 88% &gt; 70 years; 43% ≥80 years</td>
<td>90 days</td>
<td>Atrial fibrillation</td>
<td>Hospitalization for “major acute bleeding” (including GI hemorrhage or intracranial hemorrhage)</td>
<td>Derived and validated in a split sample or a retrospective database (level 4)</td>
<td></td>
</tr>
</tbody>
</table>
of 565 patients beginning warfarin therapy on hospital discharge (17), this CPR has been subsequently validated in different populations (1, 18, 19). The Outpatient Bleeding Risk Index incorporated the following clinical predictors: age $\geq 65$ years; history of stroke (current or past = 1 point; both current or past = 2 points), history of gastrointestinal (GI) bleeding, serious comorbidity (renal insufficiency, recent myocardial infarction, severe anemia), and atrial fibrillation (17). One point was awarded for the presence of each predictor with subjects being classified as low (total points = 0), middle (total points = 1–2), or high-risk (total points $\geq 3$) (17). Major bleeding rates and performance characteristics are depicted in Table 2.

Table 2: Performance characteristics of CPRs in VALIDATION groups*.

<table>
<thead>
<tr>
<th>Study (mean follow-up duration)</th>
<th># major + minor bleeds (%)</th>
<th># major bleeds (%)</th>
<th>Total in group</th>
<th>LR (95% CI) major + minor bleeding</th>
<th>LR (95% CI) major bleeding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Landefeld (17) (mean approx. 18 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>1 (2)</td>
<td>57</td>
<td>N/A</td>
<td>0.084 (0.012 – 0.59)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>N/A</td>
<td>18 (17)</td>
<td>107</td>
<td>N/A</td>
<td>0.96 (0.688 – 1.34)</td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>13 (63)</td>
<td>20</td>
<td>N/A</td>
<td>8.82 (3.82 – 20.35)</td>
</tr>
<tr>
<td><strong>Beyth (18) (48 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>2 (3)</td>
<td>80</td>
<td>N/A</td>
<td>0.19 (0.048 – 0.72)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>N/A</td>
<td>20 (12)</td>
<td>166</td>
<td>N/A</td>
<td>0.99 (0.75 – 1.32)</td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>10 (53)</td>
<td>18</td>
<td>N/A</td>
<td>9.06 (3.86 – 21.27)</td>
</tr>
<tr>
<td><strong>Wells (19) (18.5 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>7 (5)</td>
<td>0 (0)</td>
<td>128</td>
<td>0.66 (0.36 – 1.18)</td>
<td>0</td>
</tr>
<tr>
<td>Moderate</td>
<td>10 (11)</td>
<td>5 (5)</td>
<td>92</td>
<td>1.38 (0.88 – 2.16)</td>
<td>2.49 (2.11 – 2.93)</td>
</tr>
<tr>
<td>High</td>
<td>1 (50)</td>
<td>0 (0)</td>
<td>2</td>
<td>11.3 (0.74 – 173.7)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Aspinall (20) (approx. 12 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>12 (9)</td>
<td>1 (1)</td>
<td>130</td>
<td>0.99 (0.57 – 1.74)</td>
<td>0.21 (0.030 – 1.47)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>75 (8)</td>
<td>25 (3)</td>
<td>943</td>
<td>0.84 (0.73 – 0.97)</td>
<td>0.74 (0.57 – 0.96)</td>
</tr>
<tr>
<td>High</td>
<td>31 (16)</td>
<td>19 (10)</td>
<td>196</td>
<td>1.83 (1.31 – 2.56)</td>
<td>2.92 (2.02 – 4.22)</td>
</tr>
<tr>
<td><strong>Kuijer (21) (3 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>6 (4)</td>
<td>1 (0.6)</td>
<td>170</td>
<td>0.37 (0.17 – 0.79)</td>
<td>0.24 (0.035 – 1.60)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>39 (8)</td>
<td>8 (2)</td>
<td>460</td>
<td>0.93 (0.74 – 1.15)</td>
<td>0.71 (0.42 – 1.21)</td>
</tr>
<tr>
<td>High</td>
<td>26 (17)</td>
<td>10 (7)</td>
<td>150</td>
<td>2.09 (1.48 – 2.96)</td>
<td>2.86 (1.82 – 4.50)</td>
</tr>
<tr>
<td><strong>Gage (22) (approx. 10 months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>N/A</td>
<td>4 (2)</td>
<td>209</td>
<td>N/A</td>
<td>0.45 (0.17 – 1.17)</td>
</tr>
<tr>
<td>1</td>
<td>N/A</td>
<td>11 (2)</td>
<td>508</td>
<td>N/A</td>
<td>0.51 (0.29 – 0.88)</td>
</tr>
<tr>
<td>2</td>
<td>N/A</td>
<td>20 (4)</td>
<td>454</td>
<td>N/A</td>
<td>1.06 (0.73 – 1.54)</td>
</tr>
<tr>
<td>3</td>
<td>N/A</td>
<td>15 (6)</td>
<td>240</td>
<td>N/A</td>
<td>1.53 (0.96 – 2.43)</td>
</tr>
<tr>
<td>4</td>
<td>N/A</td>
<td>9 (8)</td>
<td>106</td>
<td>N/A</td>
<td>2.13 (1.12 – 4.03)</td>
</tr>
<tr>
<td>&lt;5</td>
<td>N/A</td>
<td>59 (4)</td>
<td>1596</td>
<td>N/A</td>
<td>0.88 (0.80 – 0.96)</td>
</tr>
<tr>
<td>$\geq$5</td>
<td>N/A</td>
<td>8 (100)</td>
<td>8</td>
<td>N/A</td>
<td>$\infty$</td>
</tr>
<tr>
<td><strong>Shireman (23) (90 days)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>35 (0.9)</td>
<td>3889</td>
<td>N/A</td>
<td>0.61 (0.47 – 0.80)</td>
</tr>
<tr>
<td>Moderate</td>
<td>N/A</td>
<td>48 (2)</td>
<td>2400</td>
<td>N/A</td>
<td>1.38 (1.13 – 1.69)</td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>12 (5.4)</td>
<td>222</td>
<td>N/A</td>
<td>3.86 (2.24 – 6.66)</td>
</tr>
<tr>
<td><strong>Pooled Byeth, Wells, Aspinall (18–20) (mOBRI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A**</td>
<td>3 (1)</td>
<td>338</td>
<td>N/A**</td>
<td>0.80 (0.06 – 0.56)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>N/A**</td>
<td>50 (4)</td>
<td>1201</td>
<td>N/A**</td>
<td>0.89 (0.74 – 1.06)</td>
</tr>
<tr>
<td>High</td>
<td>N/A**</td>
<td>29 (13)</td>
<td>216</td>
<td>N/A**</td>
<td>3.16 (2.29 – 4.37)</td>
</tr>
</tbody>
</table>

* statistically significant LRs meeting our criteria for moderate performance or better are depicted in bold. Calculations performed using http://www.healthcare.ubc.ca/calc/bayescalc2.html. **not pooled due to Beyth not reporting minor bleeding episodes.
The Outpatient Bleeding Risk Index was first modified by Beyth et al., based on a reanalysis of Landefeld’s original cohort, which resulted in atrial fibrillation being replaced by diabetes in the scoring system as atrial fibrillation no longer met the criteria for significance in the full cohort to be included in the prediction rule (18). The resulting “modified Outpatient Bleeding Risk Index” (mOBRI) was found using the “c index” to discriminate between those patients that developed major bleeding and those that did not (18). This study also compared physician classification of bleeding risk to mOBRI classification (18). Major bleeding occurred in 10% of patients that physicians classified as being at “very low” risk, 12% at “low” risk and 4% at “middle” to “very high” risk (18), suggesting superiority of the mOBRI over physician classification. Evaluation of clinical factors that contributed to bleeding episodes among the 18 high-risk patients who experienced major bleeding included supratherapeutic INR, drug-drug interactions and drug-disease interactions, all of which were deemed preventable (18).

Aspinall et al. validated Byeth’s mOBRI further in a combination retrospective and prospective study involving elderly patients (20). Subgroup analysis was conducted in patients diagnosed with atrial fibrillation within this study population (20). Using the incident rate ratio (IRR), the mOBRI was found to discriminate between the high and intermediate risk categories with respect to the rate of major bleeding (p < 0.001) (20). Patients that were classified as “high risk” had 14 times the rate of major bleeding as those classified as “low risk” (20). In the subgroup of patients with atrial fibrillation, there was a six-fold higher rate of major bleeding in patients classified as high risk compared to the other categories (20). The mOBRI was also prospectively evaluated in a population of patients receiving anticoagulation for VTE or PE in which patients were followed for a mean of 18.5 months (19). The event rates in this study were lower overall for major bleeding than previously published rates (19), and the number of patients classified as high risk was too small (n = 2) to accurately determine the risk of bleeding within this category. However, the mOBRI was able to discriminate between the low and moderate risk groups (p = 0.03) (19).

Three other CPRs for bleeding risk were identified. Kuijjer et al. developed a CPR combining age, sex and known malignancy to determine the odds ratio for bleeding in patients receiving a three-month course of anticoagulation for treatment of VTE or PE (21). The derivation sample found that the 17% of patients classified as being at low risk for a bleed experienced no bleeds while of the 20% that were classified as being high risk, 13 (26%) experienced a bleeding complication, seven of which were classified as major (21). In the validation sample 4% of low-risk subjects experienced a bleeding complication (1% major bleeding) compared with a 17% incidence of bleeding (7% major bleeding) in the high-risk category (21).

The HEMORR\HAGES CPR was derived using prior prediction rules and a review of the published literature (22). This CPR incorporated ten different characteristics — hepatic or renal disease, ethanol abuse, malignancy, age ≥ 75 years, reduced platelet count or function, uncontrolled hypertension, anemia, genetic factors (specifically CYP 2C9 single nucleotide polymorphisms), excessive fall risk and stroke — which were then validated in a retrospective patient database and compared to three other prediction schemes (22). Because of a lack of available data, genetic factors were not actually used in the study (22). Using a retrospective split-sample method (also known as “bootstrapping”) it was then compared to the mOBRI, Kuijjer CPR, and a system evidently inferred from the ELATE trial (24) in a sample of Medicare recipients with atrial fibrillation. Using the C statistic, the HEMORR\HAGES was found to have greater predictive accuracy than the other bleeding prediction schemes assessed (p < 0.001) (22).

Most recently, Shireman developed and validated a unique CPR using a large retrospective cohort of warfarin recipients at least 65 years of age with atrial fibrillation (23). A split-sample technique was used to determine an equation to compute an individual’s risk score (23). The final variables that were included in this bleeding risk model are as follows: age ≥70 years, female gender, remote bleeding event, recent bleeding event (occurring during the index hospitalization), alcohol or drug abuse, diabetes, anemia and antiplatelet drugs (23). Based on the specific characteristics, subjects were classified as being at low, moderate, or high risk of bleeding (23). In the validation sample, 0.9% of low-risk individuals, 2.0% of moderate-risk individuals, and 5.4% of high-risk individuals had major bleeding events (23). The prediction tool was able to discriminate between low-risk and high-risk subjects (p < 0.0001), between low-risk and moderate-risk subjects (p < 0.001), and between moderate-risk and high-risk subjects (p < 0.01) (23). The C statistic was used to estimate the area under the receiver-operator characteristics (ROC) curve and was reported to be 0.632, indicating a level between “no discrimination” and “acceptable discrimination” (23). Both the model by Kuijjer et al. and the mOBRI were also applied to the validation sample (23). The area under the ROC curve for the CPR by Kuijjer et al. was found to be 0.503, reflecting “no discrimination” and for the mOBRI was 0.613, reflecting a level between “no discrimination” and “acceptable discrimination” (23).

**Methodological quality of included studies**

Table 3 lists the scores for the quality assessment completed by the reviewers for the included studies. According to the predefined criteria, the study by Beyth et al. was found to have the highest quality score, and the study by Gage et al. was found to have the lowest with scores of 21 and 14 out of a possible total score of 34, respectively. Many of the methodological criteria were found to be lacking consistently across all of the studies. All of the included studies failed to assess both outcome and predictive variables, with the exception of the Beyth study that had blind assessment of the outcome variables only. Describing the results of a clinical prediction rule using methods such as sensitivity, specificity, and likelihood ratios was only attempted in the study by Kuijjer. All of the studies also failed to evaluate the reproducibility of the predictive variables and the rule itself, to describe a course of action to follow with the results of the rule, and to prospectively measure the clinical use of the CPR.

**Levels of evidence**

Table 1 depicts the design and population characteristics of the included studies and their level of evidence. Three of the included studies derived a CPR that was validated in either a split sample or a retrospective database (21–23). One of the CPRs was derived and subsequently validated in a prospective sample (17).
The remaining three studies validated this CPR in various patient populations (18–20). The CPRs were validated in either a patient population with one primary condition (VTE/PE, atrial fibrillation) or in mixed populations. Based on the evidence hierarchy by McGinn et al. (9), the mOBRI would be classified as having Level 2 evidence as it has been validated in different populations prospectively and has demonstrated reproducibility. Evidence that its application has changed clinician prescribing behavior with favorable outcomes is not yet available. The Kuijer, HEMORR$_2$HAGES, and Shireman CPRs are classified as Level 4 evidence and require significant further evaluation before being widely adopted as they consist only of derivation and validation in retrospective databases and split samples (21–23).

**Predictive ability of included studies**

The performance characteristics of the included studies are described in Table 2.

**Discussion**

**Quality**

Based on the comprehensive quality rating system used, we consider the overall quality of the available studies to be poor.

**Performance**

Using the pre-specified thresholds for clinical usefulness based on LR estimates (LR>10, >5 or <0.2, <0.1), in no case did the included trials demonstrate that any stratum of the studied CPRs were associated with a LR sufficiently large or small to strongly predict major bleeding or its absence. A possible exception would be a HEMORR$_2$HAGES score of ≥5 strongly predicting major bleeding, but this estimate is based on a retrospective cohort of only eight patients.

Landefeld (OBRI) and Byeth (mOBRI) demonstrated moderate predictive ability for their respective CPRs of both “low” (0.08 and 0.19 respectively) and “high” (LR 8.8 and 9.1 respectively) risk stratification. Available evidence for the performance of Kuijer’s CPR and the individual strata of HEMORR$_2$HAGES CPR are not encouraging.
Pooling of the three trials evaluating the mOBRI revealed that neither high, intermediate, nor low major bleeding risk categories had sufficient predictive ability to be considered moderate or better (LR > 0.5 or < 0.2).

Despite its large sample size and high precision of LR estimates, Shireman’s CPR did not yield LRs of sufficient magnitude to be considered of even moderate usefulness in practice.

None of the studied CPRs demonstrated the ability to predict “any bleeding” (i.e. major or minor bleeding) with moderate or strong ability.

Several explanations for the lack of clinically relevant predictive ability of the studied CPRs are possible. In studies with short follow-up periods and/or subjects at low inherent risk of bleeding, the incidence of bleeding was not high enough to be likely to result in LRs that would be categorized as moderately or strongly predictive. One study (23) overcame the sample size issue but was of such short duration that, despite enrolling what were expected to be higher-risk subjects, proved with a high degree of precision that their CPR simply lacked the ability to predict bleeding to a clinically meaningful degree. A longer follow-up period using this tool or larger sample sizes using the mOBRI might reveal clinically useful predictive ability, but the data available suggests this is not promising. It is possible that readily available patient factors, no matter how they are combined, simply do not contain sufficient information to meaningfully differentiate between levels of bleeding risk.

**Applicability to practice**

Using the dual criteria of levels of evidence and quantitative performance characteristics (LR), none of the identified CPRs are supported by sufficient evidence to recommend their adoption in clinical practice. The mOBRI is the most developed in terms of quantity of supporting evidence; however, it suffers from underwhelming predictive performance and its widespread uptake into clinical practice could only be recommended once its impact on clinical decision-making had been evaluated. The other CPRs require further validation studies before subjecting them to trials of their clinical impact (17, 21–23).

The HEMORR² HAGES CPR shows some promise by being shown to have better predictive ability (using the c statistic, not the more clinically relevant LR method) (22), although neither the individual score strata nor the grouped scores < 5 produced useful LR performance and its clinical applicability could be hampered by its relative complexity.

For an assessment tool to become commonly used in clinical practice ease of use is important. The clinical variables that comprise the tools should be readily available to the clinician at the point of care. The Outpatient Bleeding Risk Index and the scoring system by Kuijjer involve a small number of easily ascertained variables (17, 21). The HEMORR² HAGES CPR, involves numerous factors, some of which are not easily accessible or reliably assessable (e.g. platelet function, ethanol abuse, CYP 2C9 single nucleotide polymorphisms, excessive fall risk), limiting its clinical applicability (22). Shireman’s CPR, while requiring knowledge of eight easily ascertained factors, demands that users perform fractional calculations and have non-integer cutoff values on hand for interpretation (23). This could impede its utility in practice even if its performance characteristics were more convincing.

Other limitations of the studies in this review include different definitions of major and minor bleeding and use of retrospective databases to validate the tools.

Warfarin remains a recommended therapy for the management of many different disease states that require anticoagulation. The risk of bleeding has been found to be a deterrent to prescribing warfarin in patients where a benefit of therapy is clearly indicated. A retrospective cohort study that evaluated patients with atrial fibrillation found that while 86% of the patients could be stratified as being at high risk for a stroke, only 55% of these patients were receiving warfarin (25). Perceived bleeding risk was found to be one of the negative predictors of warfarin use in this study (25). Others have found similar results (26–28). This perception of bleeding risk may not accurately reflect a patient’s true risk. In one of the validation studies, the Outpatient Bleeding Risk Index compared index scores with physicians’ personal predictions of the probability of major bleeding and found inaccurate estimates on the physician’s part as to the actual risk of bleeding (18).

CPRs could aid decision-making by providing an accurate assessment of a patient’s risk of bleeding. The quantification of the risk of bleeding could then be balanced against the risk of developing thromboembolic events. Using an assessment tool, patients identified as being at a higher risk of bleeding who choose to commence anticoagulation therapy could potentially be monitored more closely along with the minimization of any modifiable risk factors. Unfortunately, the evidence to date is insufficient to recommend regular use of the currently available bleeding risk indices in clinical practice.

**Future research needs**

Future research should focus on developing more accurate CPRs and evaluating the impact on patient outcomes of the available CPRs for bleeding risk. Clinician attitudes towards the use of CPRs to determine bleeding risk in patients being considered for warfarin should also be further studied.

**Conclusion**

Bleeding risk CPRs could provide clinicians with valuable individualized patient information to aid in decision-making prior to initiation of oral anticoagulation therapy. Unfortunately, none of the available CPRs exhibit sufficient predictive accuracy or have trials evaluating the impact of their use on patient outcomes. Hence, no existing CPR can be recommended for widespread use in practice at present.

While at present the true risk may not be reliably estimated, risk factors that are known to increase an individual patient’s risk of bleeding should be evaluated and minimized if possible. Further prospective trials are required to develop a CPR that can be reliably employed in clinical practice.
References


