Bleeding assessment tools in the diagnosis of VWD in adults and children: A systematic review and meta-analysis of test accuracy

Mohamad A. Kalot

Jorge Di Paola

et al

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs
Bleeding assessment tools in the diagnosis of VWD in adults and children: a systematic review and meta-analysis of test accuracy


1Department of Internal Medicine, State University of New York at Buffalo, Buffalo, NY; 2Department of Internal Medicine, St. Mary’s Hospital, St. Louis, MO; 3Department of Internal Medicine; 4Outcomes and Implementation Research Unit, University of Kansas Medical Center, Kansas City, KS; 5Department of Internal Medicine, Lebanese American University Medical Center, Ashrafieh, Beirut, Lebanon; 6Department of Hematology, University of Kansas Medical Center, Kansas City, KS; 7Fred & Pamela Buffett Cancer Center, Division of Oncology and Hematology, University of Nebraska Medical Center, Omaha, NE; 8Department of Internal Medicine, State University of New York, Upstate Medical University, Syracuse, NY; 9Department of Internal Medicine, University of Missouri in Kansas City, Kansas City, MO; 10Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ; 11Department of Pediatrics, Washington University in St. Louis, St. Louis, MO; 12Department of Internal Medicine, Division of Thrombosis and Hemostasis, Leiden University Medical Center, Albinusdreef, Leiden, The Netherlands; 13Infinity Way, Auburn, ME, USA; 14Department of Hematology, Auckland City Hospital, Grafton, Auckland, New Zealand; 15Versiti Blood Center of Wisconsin, Milwaukee, WI; 16Division of Hematology/Oncology, Department of Pediatrics, Medical College of Wisconsin, Wauwatosa, WI; 17Irish Centre for Vascular Biology, Royal College of Surgeons in Ireland, Dublin, Ireland; 18Aflac Cancer and Blood Disorders, Children’s Healthcare of Atlanta, Emory University, Atlanta, GA; 19Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; 20Brigham and Women’s Hospital and Harvard Medical School, Boston, MA; and 21Department of Medicine, Queen’s University, Kingston, ON, Canada

Von Willebrand disease (VWD) can be associated with significant morbidity. Patients with VWD can experience bruising, mucocutaneous bleeding, and bleeding after dental and surgical procedures. Early diagnosis and treatment are important to minimize the risk of these complications. Several bleeding assessment tools (BATs) have been used to quantify bleeding symptoms as a screening tool for VWD. We systematically reviewed diagnostic test accuracy results of BATs to screen patients for VWD. We searched Cochrane Central, MEDLINE, and EMBASE for eligible studies, reference lists of relevant reviews, registered trials, and relevant conference proceedings. Two investigators screened and abstracted data. Risk of bias was assessed using the revised tool for the quality assessment of diagnostic accuracy studies and certainty of evidence using the Grading of Recommendations Assessment, Development and Evaluation framework. We pooled estimates of sensitivity and specificity. The review included 7 cohort studies that evaluated the use of BATs to screen adult and pediatric patients for VWD. The pooled estimates for sensitivity and specificity were 75% (95% confidence interval, 66-83) and 54% (29-77), respectively. Certainty of evidence varied from moderate to high. This systematic review provides accuracy estimates for validated BATs as a screening modality for VWD. A BAT is a useful initial screening test to determine who needs specific blood testing. The pretest probability of VWD (often determined by the clinical setting/patient population), along with sensitivity and specificity estimates, will influence patient management.
Introduction

Von Willebrand factor (VWF) is a hemostatic protein that facilitates platelet adhesion and aggregation in addition to stabilizing coagulation factor VIII (FVIII). Qualitative or quantitative abnormalities in VWF can lead to von Willebrand disease (VWD). The reported prevalence of VWD is up to 1% in the general population with a symptomatic prevalence of ~1 in 1000 at the level of primary care. This prevalence may be up to 15% in women with chronic heavy menstrual bleeding, making VWD the most common inherited bleeding disorder.

Patients with VWD may experience easy bruising and bleeding, especially mucocutaneous bleeding such as epistaxis, oral cavity, and heavy menstrual bleeding as well as bleeding after childbirth and dental and surgical procedures. The clinical presentation varies greatly and the bleeding phenotype may change throughout a person’s life, leading to different management plans, depending on the type and subtype of VWD. Three types of VWD have been defined depending on the type of abnormality in VWF. Type 1 VWD occurs because of partial quantitative deficiency in VWF as a result of a defect in synthesis or increased clearance, type 2 VWD is commonly divided into 4 major qualitative variants (types 2A, 2B, 2M, and 2N), and in type 3 VWD there is an absence of VWF production.

In addition to variation in VWD management, there is limited awareness within the VWD community on the best clinical practice for screening and diagnosis. The aim of this systematic review is to determine the accuracy of bleeding assessment tools (BATs) and other nonstandard bleeding inventories as screening tests for VWD, which can be used to inform a combined strategy for diagnosis. Test accuracy estimates obtained in this systematic review were used to inform evidence-based recommendations on diagnostic strategies for the recently published clinical practice guidelines on VWD, developed by a combined effort from the American Society of Hematology, the International Society on Thrombosis and Haemostasis (ISTH), the National Hemophilia Foundation, the World Federation of Hemophilia, and the University of Kansas Medical Center. The guidelines aim to inform all stakeholders on essential issues in which there is variation or uncertainty in clinical practice and will support decision-making in the context of patients’ values and preferences.

Methods

Search strategy and data sources

We searched MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials from inception through August 2019. We also manually searched the reference lists of relevant articles and existing reviews. The search was limited to studies reporting data for accuracy of diagnostic tests. The complete search strategy is available in supplement 1. The prespecified protocol for this review is registered with PROSPERO (CRD42020147977). This review is reported in accordance with Preferred Reporting Items for Systematic reviews and Meta-Analyses for diagnostic test accuracy guidelines.

Study selection

We used the following eligibility criteria.

Studies. We included studies reporting data on diagnostic test accuracy (cohort studies, cross-sectional studies) for VWD.

Participants. Patients suspected of having VWD of any age, presenting to inpatient or outpatient settings.

Index tests for diagnosis. BATs and nonstandardized testing. We did not exclude studies based on the timing of when the index test was conducted.

Reference standards. If a reference diagnostic test was not conducted, we accepted clinical follow-up as a reference standard.

Exclusion criteria. Although studies reporting on patients with VWD as well as other bleeding disorders were eligible for inclusion, we excluded studies in which >80% of the study population included a different bleeding disorder. When possible, we extracted data separately for patients with VWD from these studies. We also excluded studies that did not provide sufficient data to determine test accuracy (sensitivity and specificity), studies only available as an abstract, studies with sample size fewer than 10 patients, and studies that used an unsuitable reference standard.

Screening and data extraction

Independent reviewers conducted title and abstract screening and full-text review in duplicate to identify eligible studies. Two reviewers completed data extraction independently and in duplicate and data were verified by a third reviewer (M.A.K.). Disagreements were resolved by discussion to reach consensus, in consultation with 2 expert clinician scientists (N.C. and P.J.). We extracted data about general study characteristics (authors, publication year, country, study design), diagnostic index test and reference standard, prevalence of VWD, and parameters to determine test accuracy (ie, sensitivity and specificity of the index test).

Risk of bias and certainty of evidence

We conducted the risk of bias assessment for diagnostic test accuracy studies using the Quality Assessment of Diagnostic Accuracy Studies-2 revised tool. We used the Grading of Recommendations Assessment, Development and Evaluation framework to assess overall certainty by evaluating the evidence for each outcome on the following domains: risk of bias, imprecision, inconsistency, indirectness, and publication bias.

Data synthesis

When feasible, we combined the accuracy estimates from individual studies quantitatively (ie, pooled) for each test using OpenMetaAnalyst. We conducted a bivariate analysis for pooling sensitivity and specificity for each of the test comparisons to account for variation within and between studies. Forest plots were created for each comparison. The Breslow-Day test was used to measure the percentage of total variation across studies because of heterogeneity ($I^2$); however, the results did not influence our judgment about inconsistency from the known methodological limitations of $I^2$ in test accuracy reviews.

Diagnostic strategies for VWD are based on assessment of the pretest probability for individual patients, which provides an estimate of the expected prevalence of VWD at a population level. We calculated the absolute differences in effects for each comparison as true positives, true negatives, false positives, and false negatives. Here, we
Table 1. Characteristics of included studies

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Study design</th>
<th>No. of patients</th>
<th>Patient selection</th>
<th>Index test</th>
<th>Reference standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowman</td>
<td>2008</td>
<td>Cohort with DTA results for adults</td>
<td>217</td>
<td>Unrelated adults (age 20-88 y) recruited from primary care clinics investigated for VWD type 1 (35 male, 65 female)</td>
<td>MCMDM-1 VWD Bleeding Questionnaire. A bleeding score $\geq 4$ was considered abnormal</td>
<td>Laboratory workup including ABO blood group, VWF:Ag, VWF:RCo, and FVIII:C</td>
</tr>
<tr>
<td>Deforest</td>
<td>2015</td>
<td>Cohort with DTA results for adults</td>
<td>64</td>
<td>Adult patients (age 18-73 y) referred for the first time to a hematologist because of a problem with bleeding or bruising (11 male, 53 female)</td>
<td>Self-BAT: ISTH-BAT was converted to a laboratory workup including grade 4 reading level to produce the first version of the Self-BAT, which was then optimized to ensure agreement with the ISTH-BAT. A normal bleeding score was 0 to +5 for females and 0 to -3 for males</td>
<td>Laboratory workup including CBC, INR/PT/PTT, thrombin time, fibrinogen, ferritin, ABO blood group, VWF:Ag, VWF:RCo, FVIII:C, and VWF multimers</td>
</tr>
<tr>
<td>Philipp</td>
<td>2008</td>
<td>Cohort with DTA results for adults</td>
<td>146</td>
<td>Females (age 13-55 y) receiving a 12-page questionnaire based on the bleeding symptoms found significant in women with WWD. A screening tool was positive if 1 of 4 criteria were met: severity of heavy menstrual period, history of treatment of anemia, excess bleeding after challenges including dental surgery, surgery and delivery, family history of bleeding disorder</td>
<td>Laboratory workup including WVF:Ag and VWF:RCo</td>
<td></td>
</tr>
<tr>
<td>Bidlingmaier</td>
<td>2012</td>
<td>Cohort with DTA results for children</td>
<td>100</td>
<td>Children (age 1-17 y), 44 with a positive bleeding history, 29 referred because of an isolated APTT prolongation, and 27 because of a positive family history of bleeding</td>
<td>Quantitative ISTH child bleeding score and the qualitative ITEM analysis. A bleeding score $\geq 2$ was considered abnormal</td>
<td>Laboratory workup including WVF:RCo, WVF:Ag, FVIII:C, VWF multimers</td>
</tr>
<tr>
<td>Bowman</td>
<td>2009</td>
<td>Cohort with DTA results for children</td>
<td>151</td>
<td>Children (age &lt;18 y) from the waiting room of the Childrens Outpatient Centre, the Hotel Dieu Hospital in Kingston, Ontario, investigated for VWD because of a personal history of hemorrhagic symptoms and/or a family history of VWD and/or for preoperative screening</td>
<td>PBQ: The MCMDM-1VWD Bleeding Questionnaire was modified by including pediatric-specific bleeding symptoms in the “other” category. A bleeding score $\geq 2$ was considered abnormal</td>
<td>Laboratory workup including WVF:RCo, WVF:Ag, FVIII:C, VWF multimers, genetic testing</td>
</tr>
<tr>
<td>Malec</td>
<td>2016</td>
<td>Cohort with DTA results for children</td>
<td>193</td>
<td>Children (age &lt;11 y) referred to a composite score that was considered positive when 2 of 4 criteria were positive: Tosetto bleeding score $Z_{1}$; family history of WWD or bleeding; personal history of iron deficiency anemia; and positive James early bleeding score</td>
<td>Laboratory workup including WVF:RCo and VWF:Ag</td>
<td></td>
</tr>
<tr>
<td>Marcus</td>
<td>2011</td>
<td>Cohort with DTA results for children</td>
<td>104</td>
<td>Children (age &lt;17 y) referred for evaluation of bleeding symptoms, family history of a bleeding disorder, and/or abnormal coagulation studies</td>
<td>Modified Vicenza score to include an “other” category with pediatric-specific bleeding questions. A bleeding score $\geq 2$ was considered abnormal</td>
<td>Laboratory workup including WVF:RCo and VWF:Ag</td>
</tr>
<tr>
<td>Belen, B.</td>
<td>2015</td>
<td>Case control</td>
<td>84</td>
<td>Children (age &lt;8 y) with VWD (46) and control group (32) with bleeding symptoms but had normal prothrombin time, APTT, PFA 100, VWF:Ag, VWF:RCo, and platelet function tests</td>
<td>PBQ administration. A bleeding score $\geq 2$ was considered abnormal</td>
<td>Laboratory workup including WVF:Ag, VWF:RCo, and FVIII:C</td>
</tr>
<tr>
<td>Faiz</td>
<td>2017</td>
<td>Case control</td>
<td>53</td>
<td>Women (age 14-53 y), 41 previously untested family members of VWD patients, 26 previously diagnosed VWD patients, and 27 healthy controls</td>
<td>Modified screening tool considered positive if 1 of 3 criteria were met: severity of heavy menstrual period, history of treatment of anemia, excess bleeding after challenges including dental surgery, and surgery and delivery</td>
<td>Laboratory workup including CBC, ferritin, FVIII:C, VWF:Ag, and VWF:RCo</td>
</tr>
<tr>
<td>Mittal</td>
<td>2015</td>
<td>Case control</td>
<td>1316</td>
<td>Healthy children (age &lt;18 y) without a diagnosis of a chronic medical condition presenting to a general pediatrician's office for routine or sick visits, and 35 children (21 male, 14 female) with a known diagnosis of VWD</td>
<td>PBQ. Children with total bleeding questionnaire scores $\geq 3$ were predicted to have VWD</td>
<td>Laboratory workup including WVF:Ag, VWF:RCo, and multimer analysis</td>
</tr>
</tbody>
</table>

APTT, activated partial thromboplastin time; CBC, complete blood count; DTA, diagnostic test accuracy; INR, international normalized ratio; ITEM, Test Question Analysis; PBQ, Pediatric Bleeding Questionnaire; PFA 100, Platelet Function Assay; PT, prothrombin time; PTT, partial thromboplastin time; UMDNJ, University of Medicine and Dentistry of New Jersey.
present the results for the low, intermediate, and high pretest probability groups.

Results

Description of studies

The initial search retrieved 5693 nonduplicate studies, of which 669 were included for full-text review. Following full-text review, we identified 106 studies eligible for data abstraction, of which 13 answered the questions addressed in this systematic review. A list of excluded studies is provided in supplement 2. Reasons for exclusion at full-text review were ineligible study design, study population, or diagnostic test, sample size <10 patients, and not enough information to determine diagnostic test accuracy for VWD (Figure 1).

All the included studies reported on the use of BATs in VWD. Table 1 summarizes general characteristics of included studies, as

Figure 1. Study flow diagram for included studies.
Use of BATs as a screening tool for VWD

We pooled test accuracy of BATs when used as a screening tool for VWD from 7 cohort studies, including 112 participants. Studies used laboratory testing (VWF:Ag, VWF:RCo, FVIII:C) as a reference standard for confirming VWD, with some studies also including historic clinical diagnosis. The pooled estimates for sensitivity and specificity were 75% (95% confidence interval, 66-83) and 54% (29-77), respectively (high certainty in the sensitivity results and moderate certainty in the specificity results). Figure 2 shows the forest plot displaying the sensitivity and specificity from individual studies and the pooled estimates for BAT when used as a screening tool for VWD. The complete risk of bias assessment for individual studies is included in supplement 3.

BATs results were illustrated for 1000 patients from a low prevalence population undergoing the test (3% prevalence, which is typically seen in patients investigated for VWD because of a personal history of abnormal laboratory test [e.g., increased partial thromboplastin time]), intermediate prevalence (20% prevalence which is typically seen in patients investigated for VWD because of a personal history of bleeding symptoms [e.g., mucocutaneous bleeding]), and high prevalence (50% prevalence, which is typically seen in patients investigated for VWD as a first-degree relative for a patient with VWD); absolute differences indicate a low (<20%) proportion of false negative. Overall, the test was shown to be highly sensitive and moderately specific and the certainty of evidence was moderate to high. Table 2 shows Grading of Recommendations Assessment, Development and Evaluation test accuracy evidence summary for BAT when used as a screening test for VWD. The interactive summary of findings can be accessed using the following link: https://gdt.gradepro.org/presentations/#/isof/isof_c5b33e22-a646-4654-9f09-b820aff36c5c-1569520689536?k=eump67.

Discussion

This review presents pooled estimates of test accuracy for commonly available BATs used as a screening test for VWD. Importantly, the certainty of evidence was moderate to high. BATs had a sensitivity and specificity of 75% (95% confidence interval, 66-83) and 54% (29-77), respectively. These are overall results that include men and children; if adult women are evaluated separately, the sensitivity is much higher (100% in some studies). The benefit of using BATs is to identify patients suspected of having VWD who may otherwise be missed without this tool in clinic. Additionally, using a BAT will allow for the quantification of bleeding symptoms in patients. However, recommendations on whether to use BATs as a screening tool in clinical practice

**Figure 2.** Forest plots for sensitivity and specificity for individual studies and the pooled estimates of BAT when used as a screening tool for VWD.
**Table 2. GRADE test accuracy evidence summary for BAT when used as a screening test for VWD**

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>0.75 (95% CI 0.66-0.83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>0.54 (95% CI 0.29-0.77)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>3%</th>
<th>20%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect per 1000 patients tested</td>
<td>23 (20-25)</td>
<td>150 (132-165)</td>
<td>376 (331-413)</td>
</tr>
<tr>
<td>Test accuracy CoE</td>
<td>HIGH</td>
<td>MODERATE</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. of studies (No. of patients)</th>
<th>Study design</th>
<th>Risk of bias</th>
<th>Indirectness</th>
<th>Inconsistency</th>
<th>Imprecision</th>
<th>Publication bias</th>
<th>Pretest probability of 3%</th>
<th>Pretest probability of 20%</th>
<th>Pretest probability of 50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (patients with suspected patients)</td>
<td>7 studies, 112 patients</td>
<td>Cross-sectional (cohort type accuracy study)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>None</td>
<td>23 (20-25)</td>
<td>150 (132-165)</td>
<td>376 (331-413)</td>
</tr>
</tbody>
</table>

False negatives (patients incorrectly classified as not having suspected patients)

| True negatives (patients without suspected patients) | 7 studies, 863 patients | Cross-sectional (cohort type accuracy study) | Not serious | Not serious | Serious§ | Not serious | None | 523 (284-744) | 431 (234-614) | 270 (147-384) |

False positives (patients incorrectly classified as having suspected patients)

| True positives | 7 studies, 112 patients | Cross-sectional (cohort type accuracy study) | Not serious | Not serious | Not serious | Not serious | None | 23 (20-25) | 150 (132-165) | 376 (331-413) |

The majority of included studies were judged to be low risk of bias for test and reference standard interpretation. Although there was unclear reporting regarding flow and timing in some studies, the certainty of evidence was generally not downgraded for risk of bias. The patient selection risk of bias was low in 7 cohort studies and high in 6 case control studies that were not included to calculate the pooled estimate.

CoE, certainty of evidence.

*Typically seen in patients investigated for VWD because of a personal history of abnormal laboratory test (e.g., increased APTT).
†Typically seen in patients investigated for VWD because of a personal history of bleeding symptoms (e.g., mucocutaneous bleeding).
§Typically seen in in patients investigated for VWD as a first-degree relative for a patient with VWD.

The point estimates of specificity are not homogeneous, which was not explained by a priori determined analysis (e.g., based on difference in risk of bias of the studies), and can be due to differences in the setting and disease prevalence.
The pooled sensitivity and specificity estimates of the tests from this review apply only when the test is performed alone; however, BATs can be used as part of different diagnostic strategies to inform clinical decision-making.

**Conclusion**

This comprehensive systematic review is the first to synthesize and evaluate the accuracy of BATs as a screening tool for the diagnosis of VWD in adults and children. Estimates of sensitivity and specificity from this review were used to inform evidence-based recommendations for a clinical practice guideline. Prevalence or pretest probability of VWD in a population is essential to consider when making clinical decisions about relying on the BAT results to rule in or rule out VWD diagnosis.

**Acknowledgments**

The systematic review team acknowledges Jenny Castano, members of the American Society of Hematology (ASH), the International Society on Thrombosis and Haemostasis, the National Hemophilia Foundation, the World Federation of Hemophilia, and the von Willebrand disease (VWD) diagnosis guidelines panel members for their assistance and administrative support.

This systematic review was conducted to support the development of the ASH 2020 guidelines for diagnosis and management of VWD. The entire guideline development process was funded by ASH. Through the Outcomes and Implementation Research Unit at the University of Kansas Medical Center, some researchers received salary or grant support and others participated to fulfill requirements of an academic degree or program or volunteered their time.

**Authorship**

Contribution: R.A.M., M.A.K., and N.H. contributed to study design, study selection, data extraction, statistical analysis, and interpretation of results; M.A.K., N.H., O.A., O.D., A.E.A, S.T., B.M., A.B.D., and A.Q. contributed to study selection and data extraction; M.A.K. and R.A.M. contributed to drafting the report; B.A., J.D.P., J.C.J.E., V.J.-P., C.M., R.M., J.S.O., R.S., P.D.J., N.T.C., and V.F. contributed to the interpretation of results, and critical revision of the report; and all authors approved the final version of the manuscript.

Conflict-of-interest disclosure: P.D.J. and J.C.J.E. receive research funding from CSL Behring, Bayer, and Takeda. The remaining authors declare no competing financial interests.

**ORCID profiles:** M.A.K., 0000-0002-6581-4561; A.B.D., 0000-0002-7671-9965; O.A., 0000-0002-8740-9989; R.B.-P., 0000-0002-6010-9900; N.T.C., 0000-0003-4100-7826; P.D.J., 0000-0003-4649-9014; J.S.O., 0000-0003-0309-3313; R.A.M., 0000-0002-2901-0875.

**Correspondence:** Reem A. Mustafa, Division of Nephrology and Hypertension, Department of Medicine, 3901 Rainbow Blvd, University of Kansas Medical Center, Kansas City, KS 66160; e-mail: rmustafa@kumc.edu.

**References**


