Transfusion and Apheresis Science xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Transfusion and Apheresis Science

journal homepage: www.elsevier.com/locate/transci

Bleeding by the numbers: The utility and the limitations of bleeding scores, bleeding prediction tools, and bleeding case definitions

Emily K. Rimmer, Donald S. Houston*

Section of Hematology/Oncology, Department of Internal Medicine, Max Rady College of Medicine, University of Manitoba, Canada

A B S T R A C T

A patient's history of bleeding, whether spontaneous or in response to challenges, provides important information about both the likelihood of that patient having a biochemically-defined hemostatic defect, and that patient's risk of future bleeding. Other variables including age, comorbidities and medications influence these probabilities. Scoring systems have been devised in an effort to make the estimates quantitative in specific populations. An example of a *bleeding score* is the MCMDM1-VWD questionnaire, which was developed to predict the likelihood of a patient having von Willebrand disease. It sums standardized details of the bleeding history, weighted by severity. The HAS-BLED score typifies *bleeding prediction tools*, developed to predict bleeding during anticoagulant therapy. Although prior bleeding is one item in this score, other comorbidities like hypertension or a history of stroke count for more. A third and related concept is that of *bleeding case definitions*, which are critical to standardize the reporting of outcomes in trials of antithrombotic agents, and which have entrenched the recognition of different severities of bleeding. We advocate that future efforts should blend some of these features. Information about comorbidities and medication use could refine the interpretation of bleeding events in a bleeding score. So could the introduction of a denominator reflecting the number and duration of challenges to which the patient has been exposed when bleeding might have been expected. More detailed information about the type, frequency and severity of prior bleeding could improve the prognostic power of bleeding prediction tools. More detailed history-based scores might ultimately supersede biochemical testing in many cases.

1. Quantitation of bleeding history

It is an axiom of clinical medicine that diagnosis begins with the history. Yet history-taking has remained more art than science. Although it may be computed intuitively, and rarely with any conscious reference to the concepts expounded by Reverend Bayes, it is the clinician's estimation of the pre-test probability of disease based upon the history that directs further investigation in almost every clinical encounter. It makes sense, then, to refine those pre-test probabilities by making the information gathered from history as quantitative as possible.

Bleeding disorders are a domain that is particularly challenging. All of us experience bleeding, yet bleeding disorders are uncommon. Tools to help discern what constitutes abnormal bleeding are therefore of substantial potential utility to clinicians.

Fortunately, the assessment of bleeding disorders is a domain in which efforts to make history taking quantitative have born fruit. *Bleeding scores* have been constructed to tally specific items of the medical history, with a view to determining whether a patient is likely or unlikely to have a bleeding disorder such as von Willebrand disease. Note that the scores in use to date have been designed and validated mainly for congenital bleeding disorders, and especially for von Willebrand disease. A relevant example of a *bleeding score* is the Molecular and Clinical Markers for the Diagnosis and Management of Type 1 von Willebrand disease (MCMDM-1VWD) bleeding assessment tool [1].

A related development has been the derivation of *bleeding prediction tools*. These have generally been developed to predict the risk of bleeding (especially of major bleeding) in patients who are started on antithrombotic therapy. The best developed are for anticoagulation of patients with atrial fibrillation, where there is a straightforward desire to estimate as precisely as possible whether the benefit of anticoagulation (reduction of risk of stroke) exceeds the risk of serious bleeding from the anticoagulant. A widely used example is the HAS-BLED score [2] which, despite the acronym, estimates whether bleeding will occur, not whether it has occurred. Although initially developed to predict diagnoses, bleeding scores such as the MCMDM-1VWD can also be used to predict the risk of clinical bleeding.

A third concept that is linked to the previous two is what we might call *bleeding case definitions*. Not all bleeding is clinically overt; gastrointestinal bleeding in particular can be occult. In the context of clinical trials of antithrombotic drugs, in particular, it is essential to be able to accurately measure the number of patients who suffered bleeding, and the severity of the bleeding. These can depend on clinical context. An example is the Thrombosis in Myocardial Infarction (TIMI) trial definition of major bleeding: a rather extreme hemoglobin drop of 50 g/L,

* Corresponding author at: CancerCare Manitoba, 675 McDermot Avenue, Winnipeg MB R3E 0V9, Canada.
E-mail address: houston@cc.umanitoba.ca (D.S. Houston).

https://doi.org/10.1016/j.transci.2018.07.004

1473-0502/

or intracranial bleeding [3]. A more moderate consensus case definition for major bleeding in trials of antithrombotic therapies was developed by the International Society on Thrombosis and Hemostasis (ISTH): a 20 g/L drop in hemoglobin, transfusion of 2 or more units of red cells, bleeding into a critical site, or fatal bleeding [4]. The ISTH has subsequently promulgated a consensus definition of clinically relevant non-major bleeding [5].

In a *bleeding score*, the patient's own history of bleeding is used to estimate his or her likelihood of having a bleeding disorder, and therefore (indirectly) the likelihood of future bleeding. With a *bleeding prediction tool*, a variety of patient characteristics are used to estimate the probability of bleeding events occurring after an iatrogenic pharmacologically-induced bleeding diathesis is imposed. A *bleeding case definition* is used to determine whether bleeding happened, and how bad it was.

2. Bleeding scores

The MCMDM-1VWD bleeding score is a good example of the effort to quantitate the bleeding history. It has been developed and externally validated as a screening tool for von Willebrand disease [1]. It was derived from the Vicenza bleeding score developed by Rodeghiero and colleagues [7]. It has been largely supplanted by the ISTH bleeding assessment tool (ISTH-BAT) [8] and more recently a self-administered version (Self-BAT) [9], but these are all closely related, and we will focus on the MCMDM-1VWD as it has the most supporting data. The development and application of these various tools has been recently reviewed [10,11].

There is substantial need for an aid to assess pre-test probability for von Willebrand disease (vWD). vWD is the most common inherited bleeding disorder, affecting up to 1% of the population, depending on the definition used [12]. The majority of affected individuals lie in the shoulder of the very broad bell curve that represents von Willebrand factor (vWF) plasma levels in the population. The diagnosis of vWD is problematic because measured vWF levels can fluctuate substantially, and because there are many other factors that influence primary hemostatic function other than the vWF level, including platelet count, platelet reactivity, concentrations of other clotting factors, and hematocrit, some of which themselves are quite widely variable in the population. There is no simple cut-off below which people have bleeding symptoms. This contrasts with hemophilia. A large proportion of patients with hemophilia have severe disease with a factor VIII or IX level below 0.01 IU/ml. Those with severe disease all bleed, and measurement of a single plasma level yields an unambiguous diagnosis. At a first approximation, a Factor VIII or IX level tells you most of what you need to know about bleeding risk in hemophilia, in strong contrast to the clinical ambiguity surrounding mild to moderately reduced vWF levels.

The MCMDM-1VWD bleeding questionnaire is designed to be administered by an expert, usually a hematologist. It itemizes 12 bleeding symptoms, each of which is scored from 0 to 4 in severity: epistaxis, bruising, bleeding from minor wounds, oral cavity bleeding, gastrointestinal bleeding, bleeding with dental extractions, bleeding with surgery, menorrhagia, post-partum hemorrhage, muscle hematomas, hemarthroses, and CNS bleeding.

For example, occasional brief nosebleeds lasting less than 10 min are awarded 0 points, in recognition that minor nosebleeds are part of normal life. Epistaxis lasting more than 10 min is awarded 1 point, while a nosebleed that required transfusion nets 4 points. Points can also be deducted if a patient has tolerated certain hemostatic challenges (i.e. scored as −1); for example, having undergone childbirth twice without excessive bleeding, which was felt to reassure against the likelihood of a congenital bleeding disorder. The negative points were removed in the ISTH-BAT version.

In an effort to simplify the MCMDM-1VWD questionnaire, a condensed version was developed and tested on a population of 217 patients being investigated for possible vWD, 42 with known vWD, and 100 normal controls. The sensitivity was 100%, specificity 87%, PPV 20%, and NPV 100%, with area under the ROC curve of 0.96 for laboratory-confirmed von Willebrand disease [13]. These results were prospectively validated in two studies. The first recruited 215 consecutively-referred patients seen at two specialty bleeding-disorders clinics, in Italy and the Netherlands [14]. A positive diagnosis of a 'mild bleeding disorder' (platelet function defect, vWD, Factor XI deficiency, or mild hemophilia) was made on laboratory testing in 56. The sensitivity of the bleeding score > 3 was 41%, and specificity 81%. Negative predictive value (i.e. for a score of 2 or less) was calculated to be 99.3% if the prevalence of mild bleeding disorder was 1% of those questioned, and would still be 84.5% in a selected population among whom 20% would be confirmed to have a mild bleeding disorder. The area under the ROC curve was 0.63. A second study evaluated 30 women presenting with menorrhagia for which there was no evident gynecological or endocrinological cause [15]. The MCMDM-1VWD score had a sensitivity of 85%, specificity of 90%, positive predictive value of 89% and negative predictive value of 86% for a laboratory-confirmed bleeding disorder. The ROC analysis showed an area under the curve of 0.910.

To address the special case of children, a specific Pediatric Bleeding Questionnaire (PBQ) has been developed, in which the MCMDM-1VWD questionnaire is augmented by including pediatric-specific bleeding symptoms including umbilical stump bleeding, cephalohematoma, post-circumcision bleeding, post-venipuncture bleeding, and macroscopic hematuria [16,17]. With this modified score the upper limit of normal is 1 point in children. The score returned sensitivity of 83%, specificity of 79%, positive predictive value of 0.14 and negative predictive value of 0.99, with the area under the ROC curve of 0.88(16).

A consensus working committee of the ISTH has modified the MCMDM-1VWD questionnaire, making a number of clarifications and minor modifications, resulting in the ISTH-bleeding assessment tool (ISTH-BAT) [8]. The ISTH-BAT has also been prospectively validated in adults and children, with normal values defined: for children 0–2 points, for men 0–3 points, and for women 0–5 points [18].

Thus, the development of quantifiable bleeding scores has been a significant advance. The strengths of such bleeding scores are that a diversity of bleeding symptoms is represented, relatively reproducible definitions are provided, variable severity of bleeding is incorporated, and prospective validation has been undertaken. Despite their strengths, there are a number of limitations to current bleeding scores. For each limitation, we suggest potential modifications that may further improve their utility.

3. Limitations of bleeding scores

First, the MCMDM-1VWD and its various derivatives were developed by expert consensus, having strong face validity, but without a rigorous methodology for determining which bleeding symptoms should be included. Indeed, the score includes several items (post-partum hemorrhage, oral bleeding and GI bleeding) that had non-significant odds ratios for a diagnosis of von Willebrand disease [1]. In contrast, an example of a more formal derivation of a clinical prediction rule is the HERDOO2 rule developed by Rodger and colleagues for the prediction of risk of recurrent venous thrombosis after stopping anticoagulation for a first unprovoked event [19]. This model was constructed by selecting items that independently contributed to risk via multivariable conditional logistic regression.

Second, to make the scores readily calculable, the points for each item are given integer values, and different bleeding manifestations (e.g. nosebleeds and menorrhagia) are awarded equal values. Ideally, point values could be fractional, and more closely proportional to the actual coefficient for each term in the multivariable model. In fact, some components of the bleeding score have a higher likelihood of contributing to the diagnosis [20,21]; these items presumably would be weighted more. Moreover, interaction terms could be included in the