COMMENTARY

CLINICAL ENDOSCOPY

The Value of Risk Scores to Predict Clinical Outcomes in Patients with Variceal and Non-Variceal Upper Gastrointestinal Bleeding

James Yun-wong Lau

Department of Surgery, The Chinese University of Hong Kong, Shatin, Hong Kong

See "Prospective Comparison of the AIMS65 Score, Glasgow-Blatchford Score, and Rockall Score for Predicting Clinical Outcomes in Patients with Variceal and Nonvariceal Upper Gastrointestinal Bleeding" by Arunchai Chang, Chokethawee Ouejiaraphant, Keerati Akarapatima et al., on page 211-221.

I read with interest a prospective validation study by Chang et al.[1] on the use of the AIMS65 score, Glasgow-Blatchford score (GBS), and Rockall score (RS) for the prediction of clinical outcomes in patients with acute upper gastrointestinal bleeding (AUGIB). The authors studied 337 patients who demonstrated overt signs of AUGIB. They used the area under the receiving operator characteristic (AUROC), which is a performance metric that enables clinicians to discriminate between events and nonevents in patients, to determine the utility of each score type. In general, an AUROC of <0.7 is considered a sub-optimal performance. In a single hospital, the AIMS65 score, with an AUROC of 0.747, was found to be better at predicting hospital mortality than the GBS and RS, which had AUROC values of 0.671 and 0.660, respectively. The performances of the AIMS65, GBS, and RS for predictions of the need for endoscopic intervention were only moderate, with AUROCs of 0.619, 0.645, and 0.600, respectively. Similarly, these scores were not useful for predicting recurrent bleeding, with AUROCs around 0.6. In this cohort, none of the scores performed sufficiently to predict the clinical events.

The AIMS65 was useful only for predicting mortality.

Stanley et al.[2] have published the largest validation study of the AIMS65, GBS, RS, and Progetto Nazionale Emorragie Digestive (PNED) score to date. They examined these scores in 3,012 patients from six hospitals worldwide and determined that the GBS was the best predictor of death or the need for intervention (AUROC=0.86). A GBS of ≤1 is the threshold for predicting survival without intervention. The authors concluded that AUROCs for the other endpoints were <0.8 and therefore had limited clinical utility.

In the 2019 International Guidelines from the Upper Gastrointestinal Bleeding Conference Group,[3] the consensus group suggested using a GBS of ≤1 to identify those at a very low risk for outpatient management. The group could not make a recommendation regarding the use of the RS and advised against the use of the AIMS65 scores for identifying those at low risk, as 20% of high-risk patients are misclassified due to a low cut-off value. Obviously, one risk score is better at predicting outcomes than the other. The candidate predictors include low albumin level, an altered mental state, an international normalized ratio >1.5, systolic blood pressure ≤90 mmHg, and age >65 years. Our intuition is that those who score high on the AIMS65 are very sick patients with an increased risk of mortality.

We can conclude from this study (and others) that current risk scores for the management of patients with AUGIB are mostly useless; at least, none of them have been widely implemented in clinical practice. Since these scores use different chosen variables and were created for different purposes,

their performances in predicting outcomes vary. These scores require external validations, as their performances may also differ among populations. The use of a risk score is an integral part of the overall patient management algorithm; however, we do not yet have an impact study showing improved outcomes following the use of these scores. The scores are used in conjunction with our day-to-day clinical reasoning and decision-making that are based on our clinical experience and patient assessment. The RS and GBS were derived in 1994 and 2000, respectively, but the management of patients with AUGIB has evolved over the past few decades. Endoscopists are offering increasingly more advanced techniques for endoscopic hemostasis. We often use adjuvant high-dose proton pump inhibitors to prevent recurrent bleeding. We are also becoming more liberal in offering angiographic embolization in place of surgery. In contemporary practice, therefore, these scores may no longer be applicable. What are the ways forward? The British Society of Gastroenterology (BSG) recently published a multi-society consensus care bundle.[4] The BSG appears to have abandoned the use of a specific AUGIB risk score. Instead, the use of a generic early warning score (National Early Warning Score [NEWS]) is being proposed to identify patients who require critical care. The NEWS uses physiologic parameters such as respiratory and heart rates. However, the evidence for the use of NEWS is low. Should we develop a new AUGIB score? After all, the existing scoring methods are old. The RS was derived from the cohort of first National United Kingdom Audit on Gastrointestinal Bleeding, which was reported by the BSG two decades ago. The requirement for predictive model development has become more stringent in an attempt to avoid methodological mistakes. There is now a set of reporting guidelines for the derivation of predictive models (TRIPOD guidelines).[5] As I mentioned previously, any new model needs to be validated (internally and externally). Researchers should direct their efforts to assess the clinical impact of the models. Ideally, accurate predictions should be followed by timely clinical decisions, which will lead to improved clinical outcomes. Shung et al.[6] validated a machine-learning model that identified patients derived from a dataset of 1,958 individuals

from four countries with upper gastrointestinal bleeding who required hospital-based intervention. In a validation cohort from two Asian-Pacific sites, the machine-learning prediction model (area under the curve of 0.90) was superior to that of the GBS, RS, and AIMS65. Additional machine-learning models are likely being developed for specific purposes, such as identifying those at risk of further bleeding and mortality. Hopefully, these models will bring benefits to both patients and clinicians.

REFERENCES

1. Chang Arunchai, Ouejiaraphant C, Akarapatima K, Rattanasupa A, Prachayakul V. Prospective comparison of the AIMS65 score, Glasgow-Blatchford Score, and Rockall Score for predicting clinical outcomes in patients with variceal and nonvariceal upper gastrointestinal bleeding. Clin Endosc 2021;54:211-221.
2. Stanley AJ, Laine L, Dalton HR, et al. Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study. BMJ 2017;356:i6432.
3. Barkun AN, Almadi M, Kuipers E, et al. Management of novariceal upper gastrointestinal bleeding: guideline recommendations from the international consensus group. Ann Intern Med 2019;171:805-822.
4. Siau K, Hearnshaw S, Stanley AJ, et al. British Society of Gastroenterology (BSG)-led multisociety consensus care bundle for the early clinical management of acute upper gastrointestinal bleeding. Frontline Gastroenterol 2020;11:311-323.
5. Moon KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-W73.
6. Shung DK, Au B, Taylor RA, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology 2020;158:160-167.