

Jayant A. Talwalkar

Abbreviations

AUROC	Area under the receiver operating characteristic curve
AVB	Acute variceal bleeding
CSPH	Clinically significant portal hypertension
EV	Esophageal varices
HCC	Hepatocellular carcinoma
LSM	Liver stiffness measurement
LSPS	Liver spleen platelet count score
MELD	Model for End-Stage Liver Disease
NIEC	North Italian Endoscopic Club

Introduction

Previous studies have indicated that subjective estimation of risk by physicians in the absence of scientifically based risk models is inaccurate, resulting in systematic underestimation and overestimation. In turn, there continues to be a substantial need to identify individuals at risk for potentially lethal clinical events before they occur. Over the past two decades, a number of risk stratification models have been created to identify groups of patients at risk for complications of portal hypertension. The growing availability of therapies for both portal hypertension and underlying liver disease etiologies have further raised interest in developing more rigorous models

J.A. Talwalkar, MD, MPH
Division of Gastroenterology and Hepatology, Mayo Clinic,
200 First Street Southwest, Rochester, MN 55905, USA
e-mail: talwalkar.jayant@mayo.edu

using advanced methods of risk stratification [1, 2]. This chapter will discuss the evolution of methods for risk stratification model building in portal hypertension and address emerging concepts such as the incorporation of new tests into existing models, the economic impacts of risk attribution, and suggestions on how to prospectively validate consensus-driven models.

Definition of Risk Stratification

In the context of clinical medicine, risk stratification is defined as a statistical process to determine detectable characteristics associated with an increased chance of experiencing unwanted clinical outcomes. Said another way, risk stratification determines whether events in a local population are accounted for by the risk factors in that population. By identifying factors before the occurrence of an event, it may be possible to develop targeted interventions to mitigate their impact [3].

Dichotomization of Single Variables for Risk Stratification

The ability to estimate risk accurately for both individual patients and populations is a challenging concept. In clinical practice, the assessment of risk by physicians is usually based on the perception of a high or low probability for developing major clinical events over time. Furthermore, most indications for therapy are also dichotomous in nature which reinforces the decision-making process used in clinical practice [1, 2, 4]. In contrast, the syndrome of portal hypertension is a complex pathophysiological disease state where the biological and statistical basis for risk estimation certainly exceeds the limits of a dichotomous, single risk stratification variable [4, 5].

The use of variables in a dichotomous fashion is complicated by other issues. Reproducibility of variable measurement within an individual may vary by 10 % or more which is separate from the biological variation that causes additional error in measurement. Because the risk for clinical events is usually distributed across a spectrum versus being located at the extremes (high or low), a dichotomous variable alone lacks sufficient sensitivity and specificity to be a useful method of risk stratification [1, 2, 4, 6]. In general, odds ratios >15–20 are required to meaningfully affect prediction for an individual [4, 5]. However, such high odds ratios do not generally exist for individual predictors.

An example of using a single test result to assess risk comes from a recent systematic review and meta-analysis performed by Singh and colleagues [7] examining the association between quantitative liver stiffness measurement (LSM) and the future development of decompensated cirrhosis, hepatocellular carcinoma (HCC), and mortality. By pooling relevant studies for each outcome and the composite end point, a 7 % and 32 % increase in risk of liver-related event per unit of LSM was identified. The authors, however, cited heterogeneity of studies, variability in treatment and follow-up, and publication bias as potential limitations affecting precision of the results. The use of prospective cohort studies in patients at earlier stages of

chronic liver disease receiving similar treatment would be required to assess LSM as risk stratification tool for recognizing high- and low-risk patients for clinical events. Furthermore, a greater focus on assessing whether a prognostic model including measures of liver severity such as MELD is likely to provide better discriminative ability in predicting outcomes.

Clinical Prediction Models

Multivariable Models

An ideal approach should not only classify patients as high or low risk but also intermediate risk, so that the large majority of patients in a population can be assessed [4]. With the inception of risk stratification model development in most areas of clinical medicine, the predominant method used by many investigators was logistic regression analysis [1, 2, 6, 8]. In the literature, there are a multitude of publications using this approach in risk stratification of patients affected by portal hypertension. Current risk estimation systems, however, are now more commonly based on proportional hazards techniques with either Cox (semiparametric) or Weibull (parametric) approaches. In contrast to logistic regression, the proportional hazards techniques have the advantage of allowing for losses to follow-up and variable observation time among individuals within a cohort. The Cox proportional hazards method also has two additional distinct advantages : (1) no assumptions are required about the shape of the underlying survival function and (2) data is used more efficiently by allowing risk to be estimated for periods greater than the length of the study's follow-up [1, 2, 8].

The risk stratification of patients for determining the presence of esophageal varices has been a topic of great importance in the field of portal hypertension [9]. Over time, published studies have evolved from developing models in single center cohorts to examining multiple models in several validation cohorts. Berzigotti et al. [10] recently performed a cross-sectional study using a training set of 117 patients with compensated cirrhosis to determine the predictive ability of spleen diameter, platelet count, and LSM in detecting clinically significant portal hypertension (CSPH) and esophageal varices (EV). In this study, two unique statistical models generating CSPH and EV risk scores using multivariable backward stepwise logistic regression were developed. A composite score with LSM, spleen diameter, and platelet count (LSPS) was also examined. Subsequently, the models were assessed in an independent series of 56 patients from another center. The discriminative ability of the different models was assessed by area under the receiver operating characteristic curve (AUROC) analysis. Results were noted for an LSPS score above and below 3.2 correctly classifying 85 % of patients in the training set and 75 % in the validation set that was comparable to results from the EV risk score. The authors note that all of the patients had complete test results for all measurements, and thus model performance does not account for “real-life” situations where tests provide incomplete results in some patients.

Risk Scores

Risk scores have been developed from clinical prediction models in assessing risk. Their advantage is that risk stratification is most likely to define the spectrum of risk for complications among populations with the disease of interest [4, 6]. Risk scores are commonly used in cardiovascular medicine, with the Framingham risk score as the most well-known system assessing the risk of symptomatic heart disease in asymptomatic populations. Another advantage of using risk scores is their utility in clinical practice where clinicians faced with an individual patient can reliably identify low-risk patients who do not require potentially expensive or risky therapies without compromising the quality of care [1, 2, 6].

In contrast to logistic regression and some proportional hazards models, there are relatively fewer publications in populations with portal hypertension that examine risk scores across the spectrum of disease severity. An early notable example of risk stratification system development using PH methodology is the North Italian Endoscopic Club (NIEC) prognostic model for predicting a first bleeding episode in patients with cirrhosis and esophageal varices [11]. Subsequent validation of the NIEC index in multiple independent cohorts was also performed [12].

The most prominent example of risk stratification using PH techniques is the creation of the Model for End-Stage Liver Disease (MELD) score [13, 14]. With the idea of providing risk stratification for all patients in the spectrum of disease severity related to cirrhosis, Teh and colleagues [15] studied the ability of MELD score to predict short- and medium-term risks for mortality after common surgical procedures. By multivariable analysis, only MELD score, American Society of Anesthesiologists class, and age predicted mortality at 30 and 90 days, 1 year, and long-term, independently of type or year of surgery among 772 patients with cirrhosis. Thirty-day mortality ranged from 5.7 % with MELD scores <8 to more than 50 % for patients with MELD scores >20. Given the linear relationship with mortality risk and MELD score, patients across the entire range of disease severity could be assessed with an ordinal range of MELD scores corresponding to rising time-dependent probabilities for mortality. Subsequently, multiple validation studies in separate cohorts supported the initial study's results.

Other more complicated methods also exist, including cluster analysis, tree-structured analysis, and neural networks. These methods are particularly useful for selecting the most appropriate variables when a large number of potential predictors of risk are available. However, the main problem with all of these methods is model shrinkage—their predictive ability declines sharply once the model is applied to an external dataset which limits their utility in clinical practice [1, 2, 8].

Validation, Discrimination, and Calibration of Risk Stratification Models

Internal Validation

Internal validation describes how well a constructed model performs in the dataset from which it was derived. For the most part, risk estimation systems generally perform well when assessed in this way. However, when a proportion of the same

dataset from which the model was created is used to further demonstrate validity (i.e., split-set approach), assertions of model superiority require caution as prediction is made at the exact end point in the test dataset [2, 6, 8].

External Validation

In contrast to a split-set approach, the application of a risk model in an external dataset is more appropriate for assessing external validation. In general, risk models that demonstrate similar predictive ability in different cohorts suggest that the system may have good discrimination in identifying future cases and non-cases (see below) [2, 6, 8]. Model AUROCs or c-statistic values in external validation datasets >0.7 are generally considered satisfactory. Lower values may occur when population differences in an external dataset from the testing set are known or identified after cohort comparison [8, 16].

Discrimination

Several measures exist to assess the overall pattern of risk stratification model performance including sensitivity, specificity, AUROC, c-statistic, and clinical likelihood ratios [6, 8]. Although used for assessing diagnostic test performance, AUROC has increased in use for assessing the discrimination ability of a risk model (i.e., how well the model can identify future cases with clinical events and non-cases). AUROC technique using threshold cut points provide sensitivity and specificity parameters which are better understood by physicians [16]. In turn, reporting the sensitivity and specificity at threshold cut points for distinguishing high from low risk is helpful. It is generally accepted that AUCROCs and c-statistic values ≥ 0.80 denote excellent discrimination [6, 8, 16, 17].

Calibration

Risk prediction models also require a high degree of calibration to fulfill the goals of internal and external validation. Calibration is defined by how well the predicted event rates correspond to the observed events. Models which can discriminate well but have marginal ability for calibration usually result in misclassifying high- and low-risk persons for clinical events [1, 2, 16, 17]. Risk estimation systems can also change how well calibrated they are based on different baseline rates for the event in question in different geographic regions. Methods to assess reclassification after modification of risk stratification models have been developed and are now beginning to be used more frequently in emerging literature. Of note, a system with perfect calibration will have a lower value of discrimination (between 0.8 and 0.9) as they are linked concepts [2, 8, 16, 17].

Despite mortality rates as high as 20 % following acute variceal bleeding (AVB), existing risk stratification models have seldom been used to determine prognosis given their lack of external validation. Recently, Reverter et al. [18] examined

multiple techniques to assess advanced performance metrics of risk stratification models for acute variceal bleeding (AVB). Among 178 patients with cirrhosis and esophageal AVB who received standard therapy from 2007 to 2010, several risk models including MELD and a modified version of MELD were assessed for the ability to predict mortality within 6 weeks of AVB presentation. In addition to discrimination and calibration assessment, the models were further examined in separate cohorts from Canada and Spain. With an observed 6-week mortality frequency of 16 %, MELD was the best model in terms of discrimination. Following recalibration by adding the use of a logistic regression model, a MELD score of 11 was associated with a 5 % risk of mortality (i.e., low-risk group), while a MELD score of 19 was associated with a 20 % mortality rate (i.e., high risk). The MELD-based model showed excellent discrimination (AUC 0.87) in both external cohorts, while calibration was excellent in the Canadian cohort. Overprediction of mortality risk in high MELD score patients within the Spain cohort suggested less robust calibration.

Integrating Current Tests into Existing Risk Stratification Models

Several novel markers for risk stratification have undergone evaluation as tools to assess prognosis in patients with portal hypertension. Elastography imaging has received the most attention recently, with serum fibrosis markers and genomic polymorphism analyses also examined as potential tests. As discussed earlier, no single test is likely to provide adequate risk stratification [1, 2, 4, 5]. In contrast, studies have been conducted to improve risk estimation through the incorporation of new risk factors into existing models. However, improving a model's AUC from 0.80 to 0.90 by adding a new marker requires the novel test result to have an independent odds ratio >3 which is highly uncommon given significant correlations with 1 or more risk factors for portal hypertension. Conversely, the absence of improved discrimination (as measured by the AUC or c-statistic) suggests the novel marker is unlikely to be useful as a screening test [5, 8, 16]. Additional challenges exist based on the strong correlation among parameters that address the same physiology. Choosing which tests to combine has also not been standardized to date [1, 2].

Asrani and colleagues [19] examined the contribution of LSM by magnetic resonance elastography in identifying patients at increased risk for hepatic decompensation among patients with cirrhosis. Among 430 subjects with varying stages of cirrhosis, the mean LSM value was independently associated with decompensated cirrhosis after adjustment for MELD score, age, gender, albumin, and platelet count at baseline. However, the odds ratio for LSM was only 1.13. In the follow-up cohort, the hazard rate of hepatic decompensation was 1.42 per unit increase in LSM over time. However, for subjects with compensated disease and mean LSM values >5.8 kPa (equivalent to roughly 18 kPa by transient elastography), the hazard rate of hepatic decompensation was 4.96 compared to an individual with compensated cirrhosis and lower mean LSM values. This study highlights the limitation of LSM alone for risk stratifying all patients with compensated cirrhosis.

Contemporary Issues in Risk Stratification Modeling

Competing Risks

The presence of competing risks for death in patients with portal hypertension modifies the relationship between risk stratification models and mortality. It is clear that many risk factors for portal hypertension are also significantly associated with death due to other liver-related causes such as hepatocellular carcinoma. Current risk stratification strategies do not typically account for competing risks, which limits some of their discrimination and calibration utilities. Risk stratification models examining short-term mortality risk will also be compromised when applied to populations with longer life expectancies [1, 2]. In the Asrani study, cause-specific Cox PH analysis adjusting for competing risks was utilized to determine the association between elevated LSM and development of decompensation [19].

Dynamic Risk Profiling

Most risk stratification models incorporate variables as static entities when in most cases they are actually dynamic in nature. Continuous risk markers, such as liver stiffness, can vary within individuals when measured at different times or when repeated over time. Thus, the timing of risk assessment is important [1, 2]. Temporal variations in portal pressure including time of day [20, 21], season [22], and relationship with exertion [23, 24] have all been documented which affect timing of measurement as well. Finally, the frequency with which risk should be assessed is unknown because the duration of the predictive value of a test is rarely studied.

Economic Implications of Risk Stratification

One of the stated goals of risk stratification is to identify all individuals at high risk for major clinical events and to pursue treatments, when available, to prevent these events. However, for a randomized trial, this may require screening and evaluating 10–20 times as many patients to identify the 5–10 % of patients who are at high risk. Screening costs, therefore, could outpace costs of the study and its interventions and thus may prevent conduct of the study. Well-designed studies to improve risk stratification models could also incur costs that may be prohibitive as well. In clinical practice, a key goal of risk stratification is to identify those patients at low risk for clinical events who would not benefit from an expensive or invasive therapy. Notably, if an alternative therapy of equivalent efficacy and lower cost becomes available, the performance of risk stratification models could change. From a health economics perspective, recognizing high-risk groups that do not benefit from interventions due to competing risks (in addition to low-risk patients) also decreases the overall costs and increases effectiveness [1, 2, 25].

Unmet Needs

Despite advances in the approaches and techniques for developing risk stratification models over time, a number of unmet needs in this field remain. The majority of study designs used for model building are retrospective in nature given that the frequency of events is already known. In contrast, the conduct of prospective studies (observational or interventional) examining the efficacy of risk stratification models in predicting events would confirm excellent performance that is defined retrospectively. Developing consensus on strategies and evaluation plans for incorporating new tests into existing risk models is also needed as current approaches are nonsystematic. Assessing the robustness of risk stratification models in selected populations with portal hypertension is also needed with specific attention to the elderly, racial and ethnic minorities, populations with multiple comorbidities, and those residing in different geographic areas [1, 2, 6, 25].

Future Pathway for Risk Stratification

Recommendations have been proposed elsewhere [1, 2, 25] that define a pathway for improving the development and application of risk stratification models, which are relevant for populations with portal hypertension:

1. Establishing baseline risk models composed of important, readily available clinical variables for common patient groups
2. Generating a consensus list of currently available risk stratification techniques that should be assessed for improving performance of baseline model
3. Thorough evaluation of the added prognostic utility of novel risk markers, including assessment of interactions, discrimination, calibration, model fit, and reclassification
4. Evaluation of optimized risk stratification approaches in randomized clinical trials
5. Creation of a full and transparent process for promoting clinical trials supported by all stakeholders

Conclusion

Developing and validating risk stratification models in populations with portal hypertension remains a daunting process. While locating a simple algorithm or test for predicting mortality or major clinical events is ideal, this will not be realistic given that no single test result can adequately represent the pathophysiologic complexity of portal hypertension. As methodologies for risk model development have moved from logistic regression analysis to proportional hazards techniques, an increased emphasis on developing risk scores including patients at intermediate risk of adverse clinical events will improve the relevance of predictive models. Notably, the MELD score has been able to serve in this

capacity to date as compared to more traditional but categorical systems like Child-Pugh classification. There also need to be additional refinements which account for the dynamic nature of clinical variables and the knowledge of competing risks that can influence the risk for major clinical events. As risk stratification models are being developed using advanced statistical techniques in cooperation with biostatisticians, these strategies should be considered for testing in prospective randomized clinical trials to establish their utility and also to identify models where new tests can be incorporated to determine if risk stratification improves.

References

1. Goldberger JJ, Cain ME, Hohnloser SH et al (2008) American Heart Association/American College of Cardiology Foundation/Heart Rhythm Society scientific statement on noninvasive risk stratification techniques for identifying patients at risk for sudden cardiac death: a scientific statement from the American Heart Association Council on Clinical Cardiology Committee on Electrocardiography and Arrhythmias and Council on Epidemiology and Prevention. *Circulation* 118(14):1497–1518
2. Goldberger JJ, Buxton AE, Cain M, Costantini O, Exner DV, Knight BP, Lloyd-Jones D, Kadish AH, Lee B, Moss A, Myerburg R, Olgin J, Passman R, Rosenbaum D, Stevenson W, Zareba W, Zipes DP (2011) Risk stratification for arrhythmic sudden cardiac death: identifying the roadblocks. *Circulation* 123(21):2423–2430
3. Miller CC et al (2001) Risk stratification: a practical guide for clinicians. Cambridge University Press, Cambridge
4. Gehl A, Haas D, Fuster V (2005) Primary prophylaxis with the implantable cardioverter-defibrillator: the need for improved risk stratification. *JAMA* 294(8):958–960
5. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 159:882–890
6. Cooney MT, Dudina AL, Graham IM (2009) Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. *J Am Coll Cardiol* 54:1209–1227
7. Singh S, Fujii LL, Murad MH, Wang Z, Asrani SK, Ehman RL, Kamath PS, Talwalkar JA (2013) Liver stiffness is associated with risk of decompensation, liver cancer, and death in patients with chronic liver diseases: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 11(12):1573–1584
8. Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21:128–138
9. de Franchis R (2006) Noninvasive diagnosis of esophageal varices: is it feasible? *Am J Gastroenterol* 101:2520–2522
10. Berzigotti A, Seijo S, Arena U, Abraldes JG, Vizzutti F, García-Pagán JC, Pinzani M, Bosch J (2013) Elastography, spleen size, and platelet count identify portal hypertension in patients with compensated cirrhosis. *Gastroenterology* 144(1):102–111
11. The North Italian Endoscopic Club for the Study and Treatment of Esophageal Varices (1988) Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices. *N Engl J Med* 319:983–989
12. Merkel C, Zoli M, Siringo S, van Buuren H, Magalotti D, Angeli P, Sacerdoti D, Bolondi L, Gatta A (2000) Prognostic indicators of risk for first variceal bleeding in cirrhosis: a multi-center study in 711 patients to validate and improve the North Italian Endoscopic Club (NIEC) index. *Am J Gastroenterol* 95(10):2915–2920

13. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, ter Borg PC (2000) A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology* 31:864–871
14. Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL et al (2001) A model to predict survival in patients with end-stage liver disease. *Hepatology* 33:464–470
15. Teh SH, Nagorney DM, Stevens SR, Offord KP, Therneau TM, Plevak DJ, Talwalkar JA, Kim WR, Kamath PS (2007) Risk factors for mortality after surgery in patients with cirrhosis. *Gastroenterology* 132(4):1261–1269
16. Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935
17. Diamond GA (1992) What price for perfection?: calibration and discrimination of clinical prediction models. *J Clin Epidemiol* 45:85–89
18. Reverter E, Tandon P, Augustin S, Turon F, Casu S, Bastiampillai R, Keough A, Llop E, González A, Seijo S, Berzigotti A, Ma M, Genescà J, Bosch J, García-Pagán JC, Abraldes JG (2014) A MELD-based model to determine risk of mortality among patients with acute variceal bleeding. *Gastroenterology* 146(2):412–419
19. Asrani SK, Talwalkar JA, Kamath PS, Shah VH, Saracino G, Jennings L, Gross JB, Venkatesh S, Ehman RL (2014) Role of magnetic resonance elastography in compensated and decompensated liver disease. *J Hepatol* 60(5):934–939
20. Merican I, Sprengers D, McCormick PA, Minoli G, McIntyre N, Burroughs AK (1993) Diurnal pattern of variceal bleeding in cirrhotic patients. *J Hepatol* 19(1):15–22
21. García-Pagán JC, Feu F, Castells A, Luca A, Hermida RC, Rivera F, Bosch J, Rodés J (1994) Circadian variations of portal pressure and variceal hemorrhage in patients with cirrhosis. *Hepatology* 19(3):595–601
22. Sato G, Matsutani S, Maruyama H, Saisho H, Fukuzawa T, Mizumoto H, Burioka N (1998) Chronobiological analysis of seasonal variations of variceal hemorrhage. *Hepatology* 28(3):893–895
23. García-Pagán JC, Santos C, Barberá JA, Luca A, Roca J, Rodriguez-Roisin R, Bosch J, Rodés J (1996) Physical exercise increases portal pressure in patients with cirrhosis and portal hypertension. *Gastroenterology* 111(5):1300–6.25
24. Bandi JC, García-Pagán JC, Escorsell A, François E, Moitinho E, Rodés J, Bosch J (1998) Effects of propranolol on the hepatic hemodynamic response to physical exercise in patients with cirrhosis. *Hepatology* 28(3):677–8226
25. Goldberger JJ (2010) The coin toss: implications for risk stratification for sudden cardiac death. *Am Heart J* 160:3–7