



Validation of the AASLD recommendations for classification of oesophageal varices in clinical practice

Waleed Fateen^{1,2} | Krish Rangunath^{1,2} | Jonathan White^{1,2} | Amardeep Khanna^{1,2} | Marina Coletta³ | Sunil Samuel^{1,2} | Jacobo Ortiz^{1,2} | Martin James^{1,2} | Emilie Wilkes^{1,2} | Guruprasad P. Aithal^{1,2} | Indra N. Guha^{1,2} | Sarmed S. Sami⁴

¹NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK

²Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham, UK

³Gastroenterology and Endoscopy Unit, Department of Pathophysiology and Transplantation, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy

⁴Department for Targeted Intervention, Division of Surgery & Interventional Science, University College London (UCL), London, UK

Correspondence

Sarmed S. Sami, Department of Targeted Intervention, Division of Surgery & Interventional Science, University College London, 3rd floor Charles Bell House, 43-45 Foley Street, London W1W 7TS, UK.
Email: s.sami@ucl.ac.uk

Funding information

WF, KR, JW, GPA and NG are supported by the NIHR Nottingham Biomedical Research Centre (reference no: BRC-1215-20003). SSS was funded by an Olympus-Core National Endoscopy Research Fellowship Grant, Core Charity, United Kingdom (grant number: RB4803).

Handling Editor: Juan Abraides

Abstract

Background & Aims: The American Association for the Study of Liver Diseases recommends the use of a 2-grade classification system (small and large) to describe the size of oesophageal varices (OV). Data on observer agreement (OA) on this system are currently lacking. We aimed to evaluate this classification and compare it to the widely used 3-grade classification (grade 1 'small', grade 2 'medium', grade 3 'large') among operators of variable experience.

Methods: High-definition video recordings of 100 patients with cirrhosis were prospectively collected using standardised criteria. Nine observers of variable experience performed independent evaluations of the videos in random order. OV were scored using both systems. All assessments were repeated a year later by the same observers to assess intra-observer agreement.

Results: Interobserver agreement (all observers) using the 2-grade and the 3-grade system was $k = 0.71$ (95% CI: 0.64-0.78) and $k = 0.73$ (95% CI: 0.66-0.79) respectively. When using the 2-grade system, intra-observer agreement between hepatologists ($n = 3$), luminal gastroenterologists ($n = 3$) and trainee gastroenterologists ($n = 3$) was $k = 0.89$ (95% CI: 0.86-0.91), $k = 0.72$ (95% CI: 0.67-0.77), and $k = 0.74$ (95% CI: 0.67-0.8) respectively. With the 3-grade system; intra-observer agreement between the same three subgroups were $k = 0.9$ (95% CI: 0.87-0.92), $k = 0.73$ (95% CI: 0.68-0.78), $k = 0.77$ (95% CI: 0.71-0.82) respectively.

Conclusions: There was no difference in OA between the 2-grade and 3-grade classification systems. Hepatologists had significantly higher levels of consistency in grading OV. This may have implications to create alternative training models for residents and fellows in the recognition and grading of OV.

KEYWORDS

cirrhosis, endoscopy, interobserver variation, screening, varices

See Editorial on Page 754

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Liver International* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Oesophageal varices (OV) are a common finding in patients with liver disease. They occur in approximately 40% and 70% of patients with compensated and decompensated cirrhosis respectively.¹ Acute variceal bleeding is a life-threatening complication of OV with a 6-week mortality ranging between 16% and 26%.^{2,3} Guidelines recommend endoscopic surveillance of patients with known cirrhosis or portal hypertension.⁴ Index endoscopic assessments are frequently performed by endoscopists with varying levels of experience in liver disease and portal hypertension. During the procedure, the operator ascertains the location, size and appearance of the varices according to standard criteria.

The objectives of endoscopic assessment for variceal screening are two-fold. The first objective is to assess if varices are present or absent. The second objective if varices are present, is to determine whether or not they require treatment with non-selective beta blockers or endoscopic band ligation (EBL).⁴ The latter decision is primarily based on the varices size and/or presence of high-risk features.⁴ The timing of repeat procedures is also determined by the presence and size of varices. Therefore the accuracy and consistency in the classification of varices by endoscopists will have a direct effect on subsequent management.

The Japanese Research Society for Portal Hypertension originally described the 3-grade classification system which is still widely used.⁵ It involves scoring varices as grade 1 (small), straight small calibre varices; grade 2 (medium), moderately enlarged, beady varices covering less than one-third of the lumen; and grade 3 (large), markedly enlarged, nodular or tumour-shaped varices occupying more than one-third of the lumen.^{6,7} The American Association for Study of Liver Disease (AASLD) proposed the 2-grade classification system. This system was originally created by the North Italian Endoscopy Club who found it to be predictive of variceal bleeding.⁸ The system was endorsed by a consensus meeting (Baveno I, 1992).⁴ It involves classifying variceal size into either small or large. The classification can be quantitative with a cut-off diameter of 5 mm as measured by an open biopsy forceps or semi quantitative using grade 1 above as small and grade 2/grade 3 as large. The quantitative approach is not widely used in clinical practice because of its challenging nature and doubtful accuracy. The technical difficulty is created by the variable degree of air insufflation, breathing pattern and peristalsis. Endoscopic examination should be performed on both minimal and maximal insufflation in order to avoid misclassification (Figure 1). Clinically, grades 2 and 3 are regarded as varices needing treatment and treated the same way.

Data on the interobserver agreement and therefore reproducibility of the 2-grade classification system are lacking. Moreover, this system has not been compared to the more widely used 3-grade approach in adults. The primary aim of this study was to compare observer agreement (OA) on variceal classification using the 2-grade vs the 3-grade system. The secondary aim was to assess the impact of observer experience on the level of agreement.

Key points

- Increased blood pressure in the abdomen can happen as a result of scarring of the liver. Blood vessels around the food pipe are fragile and often do not tolerate an increase of their pressure. This puts them at risk of rupturing into the food pipe leading to blood vomit which could be a threat to life.
- A camera test is advisable to check the state of the food pipe blood vessels. During the camera test, a careful evaluation of such blood vessels is important. According to the camera test evaluation, treatment is indicated in the form of medications to decrease the pressure or direct application of elastic bands on to the blood vessels.
- In this study we compared the observer agreement using two different grading systems for evaluation of the food pipe blood vessels. We also tested the impact of observer experience on their consistency of evaluation.
- We found that there was no significant difference between the two grading systems. However, there was a significantly higher consistency amongst experienced observers in identifying the presence and stage of such risky blood vessels. These results may have implications on training and service redesign.

2 | METHODS

2.1 | Study design

This was a prospective repeatability and reproducibility study in a tertiary referral centre (Nottingham University Hospitals NHS Trust, Nottingham, UK). All the participants provided written informed consent and this study received approval from the East Midlands Nottingham 1 research ethics committee. Endoscopy procedures were performed between 31 July 2012 and 25 February 2014 using a high definition system and videos were digitally recorded. Nine independent observers assessed OV on the video recordings. The same nine observers as well as the reference observer re-assessed the same video recordings after an interval of at least 1 year to assess for intra-observer agreement.

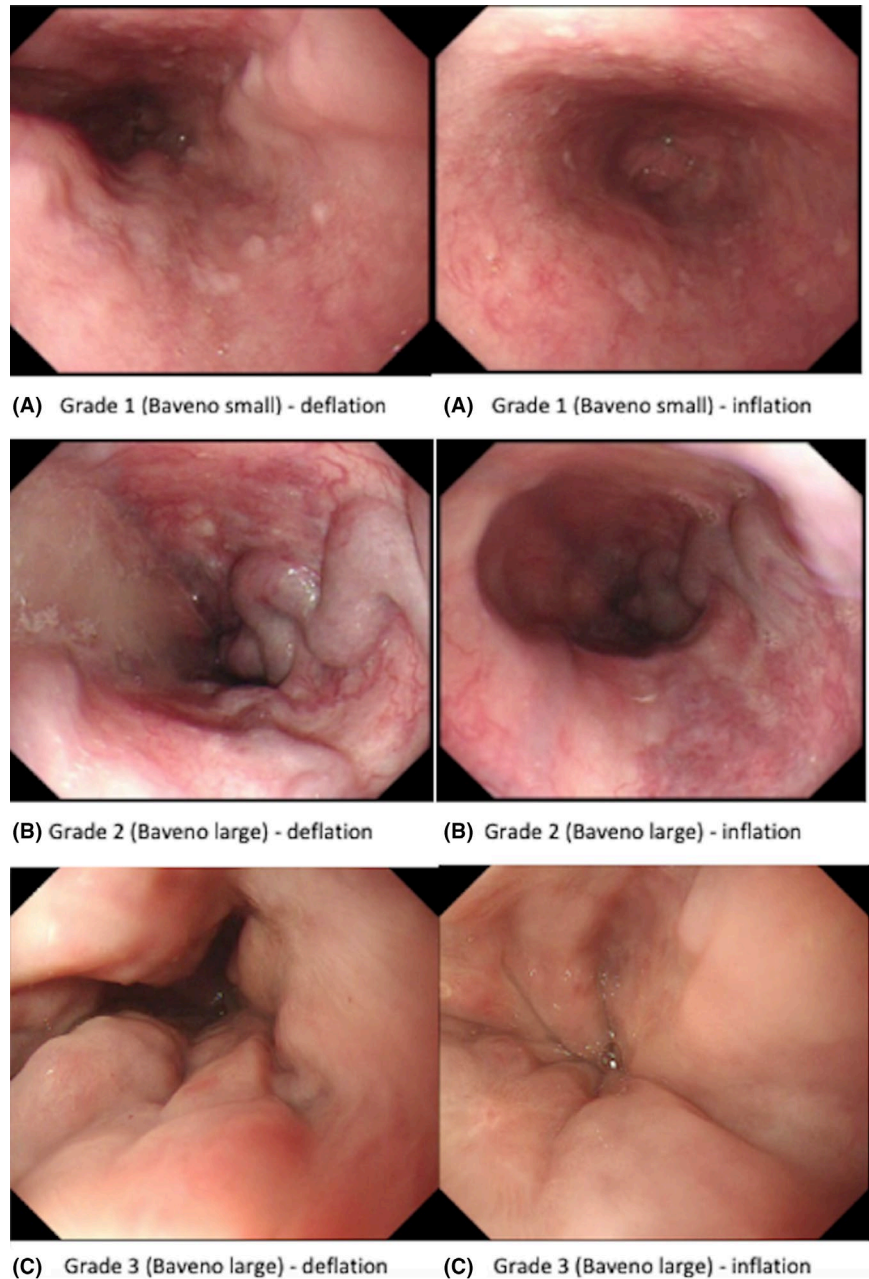
2.2 | Participants and interventions

We recruited consecutive patients with a clinical diagnosis of cirrhosis who were scheduled for a diagnostic gastroscopy as part of screening or surveillance for OV during their routine clinical care.

All procedures were performed by a single experienced endoscopist (Performed more than 1000 procedures and regularly performed EBL for at least 3 years). A 9.8-mm diameter high definition endoscope (GIF-H260; Olympus Key-Med) was used. A standardised



FIGURE 1 Corresponding images of oesophageal varices Grade 1/small (A); Grade 2/large (B) and Grade 3/large (C) during deflation (left panel) and maximum inflation (right panel)



recording protocol was used after analgesia and/or sedation (Table S1). Prior to unsedated procedures, topical pharyngeal anaesthesia was applied to the posterior pharynx (5-10 sprays, Lidocaine 10 mg/dose, Xylocaine; AstraZeneca). In case of patient preference for sedation, Midazolam (Hameln Pharmaceuticals Ltd) with or without pethidine was used.

2.3 | Rating and data collection

One-hundred anonymised video recordings from 100 patients were digitally stored (evaluation set). Nine blinded endoscopists (observers) excluding the endoscopist who recorded the procedures evaluated all the videos independent of each other and in a random order.

The observers included the following groups (a) Three hepatologists who regularly perform dedicated lists for variceal screening and EBL at our unit. (b) Three luminal gastroenterologists who performed more than 500 gastroscopies but do not perform regular variceal screening or EBL. (c) Three trainee gastroenterologists who had performed between 200 and 500 gastroscopy procedures. Each of the nine observers underwent a 30-minute training session using eight separate representative video recordings and a series of 11 still images to demonstrate the 2- and 3-grade classification systems.

The observers were asked to record their findings including: (a) presence or absence of OV. (b) Grade of OV according to the 2-grade classification system (small or large). (c) Grade of OV according to the 3-grade classification system (grade 1, 2 or 3). (d) Quality of video on a visual analogue scale of 0-10 (0 = unclear views rendering diagnosis

with confidence impossible, 10 = excellent views allowing for diagnosis with utmost confidence). Semiquantitative morphological assessment of variceal size was used by observers for both classification systems.

2.4 | Outcome measures

The primary outcomes were inter- and intra-observer agreement among the nine assessors using each of the two classification systems. Secondary outcomes were inter- and intra-observer agreement among the nine assessors stratified by level of experience (hepatologists vs luminal gastroenterologists vs trainee gastroenterologists).

2.5 | Statistical analysis and sample size calculation

Outcomes were measured using either intraclass correlation coefficient (ICC) or kappa (κ) statistic as appropriate, both of which summarise agreement within or between observers in comparison

TABLE 1 Baseline characteristics of patients and videos (n = 100)

Variable	Value
Patient	
Male gender	62%
Age (y)	60.4 (SD \pm 13.7)
Prevalence of varices (reference endoscopy)	50%
Sedation	
Midazolam (mg)	2 (IQR 0-3.13)
Pethidine (mg)	0 (IQR 0-25)
Video	
Duration (s)	49 (IQR 38-73)
VAS scores (median)	7.5 (range 5-9)
Prevalence of varices on video assessments (mode of 19 observations)	48%

Abbreviations: IQR, interquartile range; VAS, visual analogue score.

to the probability of agreement by chance. Test statistics were generated according to published methodologies as follows.⁹ For categorical data (ie varices present vs varices absent), Cohen's kappa was used in case of two observations (ie intra-observer agreement) and Fleiss kappa in case of more than two observations (ie interobserver agreement). For ordinal data (ie 2-grade staging system and 3-grade staging system), absolute agreement ICC was used, analysis of variance was performed using a two-way random effects model of individual values. The interpretation of test values was according to the guidance provided by Landis and Koch who described values <0 as 'no agreement', 0-0.20 as 'slight agreement', 0.21-0.40 as 'fair agreement', 0.41-0.60 as 'moderate agreement', 0.61-0.80 as 'substantial agreement', and 0.81-1 as 'almost perfect agreement'.¹⁰

2.6 | Sample size calculation and statistical analysis

The total number of videos required for analysis was calculated using the method suggested by Zou.¹¹ Assuming a true test-value of 0.59 (based on previous literature),¹² our nine observers were required to rate a minimum of 93 videos to yield 80% power with a confidence interval \pm 0.12 for agreement.¹²

A total of 19 assessments were performed for each video including; nine initial assessments by the observers; nine interval assessments by the observers; as well as one interval assessment by the reference endoscopist.

R-statistical computing (R version 3.4.1, Vienna, Austria) was used for analysis. The R-library was used to calculate interobserver reliability (irr). The R-libraries 'reshape2' and 'ggplot2' were used for data visualisation and 'ICC.sample.size' was used for sample size calculation.¹¹

3 | RESULTS

3.1 | Baseline characteristics

One-hundred patients were recruited to the evaluation set to allow for low quality videos or missing data. A summary of the descriptive statistics can be found in Table 1.

	2-grade			3-grade			
	0	1	2	0	1	2	3
Hepatologist	49.3	20	30.7	49.3	19	23	8.7
Interval hepatologist	47.3	23.3	29.3	47.3	21	22.3	9.3
Luminal	38.7	31.3	30	38.3	30	21	10.7
Interval luminal	41	31.3	27.7	41	29.3	22	7.6
Trainee	40	29.3	30.7	40	29.3	22.7	8
Interval trainee	51.7	22.7	25.7	51.7	22.7	17.3	8.3
Mode	48	24	28	48	24	21	7
Reference	50	34	16	50	29	13	8

TABLE 2 Prevalence of variceal stages according to different evaluations. Interval evaluations were performed at least 12 months apart. Mode marks the mode of all video evaluations performed (n = 19). Reference denotes results of the original endoscopy evaluation and report

3.2 | Prevalence of OV

During the original endoscopic evaluation the reference endoscopist graded 50%, 34% and 16% of tests ($n = 100$) as absent varices, small and large (2-grade system) respectively. Using the 3-grade system, the reference graded 50%, 29%, 13% and 8% of tests as absent varices, grade 1, grade 2 and grade 3 respectively. Agreement between the reference endoscopist and the mode of 19 video observations was $ICC = 0.76$ (95% CI: 0.67-0.83). Table 2 outlines the prevalence of scoring grades according to various evaluations performed. Absence of varices was almost identical using both grading systems. There was a consistent drop in the prevalence of grade 1 when using the 3-grade system as compared to grade 'small' using the 2-grade system, this was not statistically significant (chi-square test; $P = .3$). Figure S1 provides a summary of scores provided during all assessments performed.

3.3 | Observer agreement

3.3.1 | Interobserver agreement

Overall agreement (among all nine observers at the initial assessment) on the presence of varices, 2-grade system and 3-grade system was $k = 0.61$ ($n = 9$, 95% CI: 0.53-0.69), $k = 0.71$ ($n = 9$, 95% CI: 0.64-0.78) and $k = 0.73$ ($n = 9$, 95% CI: 0.66-0.79) respectively. Interobserver agreement amongst subgroups of observers is outlined in Table 3 and Figure 2.

3.3.2 | Intra-observer agreement

Overall agreement (repeat observations after at least 12 months) on the presence of varices, 2-grade system and 3-grade system was $ICC = 0.72$ (95% CI: 0.69-0.75), $ICC = 0.78$ (95% CI: 0.76-0.8) and $ICC = 0.8$ (95% CI: 0.78-0.82) respectively. Intra-observer agreement amongst subgroups of observers is outlined in Table 3 and Figure 3.

3.4 | Impact of experience on observer evaluation

Hepatologists had significantly higher intra-observer agreement on the presence vs absence of varices as well as usage of 2-grade and 3-grade staging systems in comparison to the other groups. This is evidenced by non-overlapping 95% confidence intervals of ICC values as shown in Figure 3. Hepatologists had significantly higher interobserver agreement on the presence vs absence of varices in comparison to trainee gastroenterologists. Interobserver agreement otherwise was similar between the three groups Figure 2.

TABLE 3 Agreement on variceal assessments as tested by nine observers initially and subsequently repeated after an interval of at least 1 year

Presence or absence of varices		
Interobserver agreement		
Assessment	Raters	k-value (95% CI)
All observers	9	0.61 (0.53-0.69)
Hepatologists	3	0.73 (0.63-0.8)
Luminal gastroenterologists	3	0.59 (0.47-0.69)
Trainees	3	0.48 (0.33-0.61)
Intra-observer agreement		
Assessment		k-value (95% CI)
All observers		0.72 (0.69-0.75)
Hepatologists		0.84 (0.8-0.87)
Luminal gastroenterologists		0.66 (0.59-0.72)
Trainees		0.65 (0.56-0.72)
2-stage classification system		
Interobserver agreement		
Assessment	Raters	ICC-value (95% CI)
All observers	9	0.71 (0.64-0.78)
Hepatologists	3	0.75 (0.65-0.82)
Luminal gastroenterologists	3	0.7 (0.56-0.78)
Trainees	3	0.65 (0.5-0.75)
Intra-observer agreement		
Assessment		ICC-value (95% CI)
All observers		0.78 (0.76-0.8)
Hepatologists		0.89 (0.86-0.91)
Luminal gastroenterologists		0.72 (0.67-0.77)
Trainees		0.74 (0.67-0.8)
3-stage classification system		
Interobserver agreement		
Assessment	Raters	ICC-value (95% CI)
All observers	9	0.73 (0.66-0.79)
Hepatologists	3	0.77 (0.7-0.84)
Luminal gastroenterologists	3	0.73 (0.61-0.82)
Trainees	3	0.7 (0.52-0.78)
Intra-observer agreement		
Assessment		ICC-value (95% CI)
All observers		0.8 (0.78-0.82)
Hepatologists		0.9 (0.87-0.92)
Luminal gastroenterologists		0.73 (0.68-0.78)
Trainees		0.77 (0.71-0.82)

Abbreviation: CI, confidence interval.

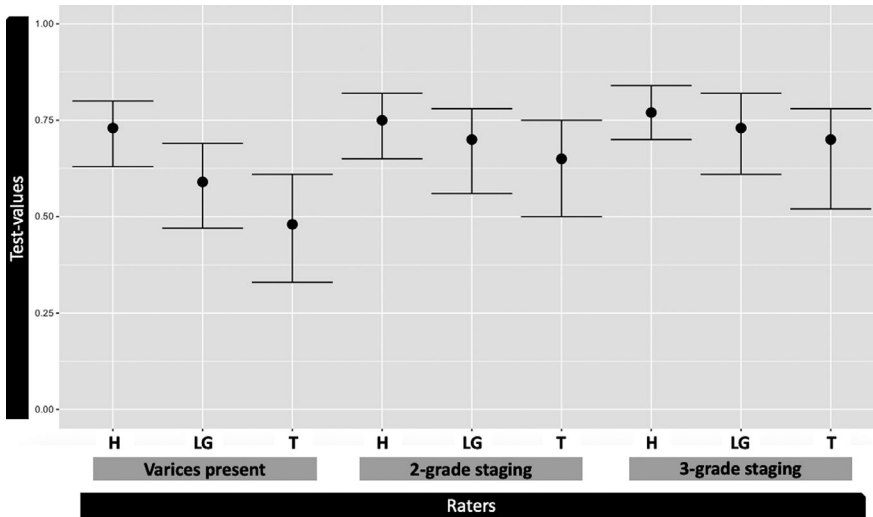


FIGURE 2 Interobserver agreement stratified by experience, Y-axis presents test-values with error bars demonstrating 95% confidence intervals. The assessments on the X-axis in order from left to right: presence of varices, size according to 2-grade classification and size according to 3-grade classification. The figure compares agreement between hepatology observers ($n = 3$) in comparison to luminal gastroenterology observers ($n = 3$) and trainee observers ($n = 3$). H, hepatologist; LG, luminal gastroenterologist; T, trainee gastroenterologist

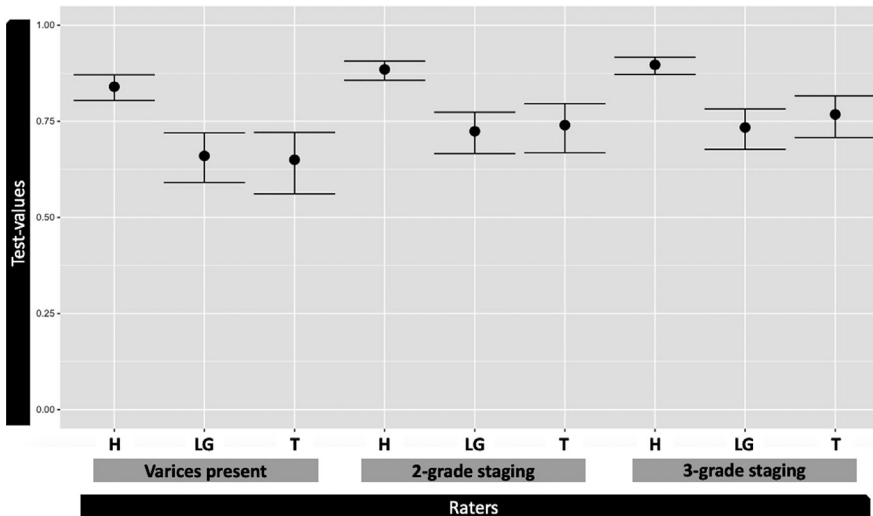


FIGURE 3 Intra-observer agreement (stratified by experience) between an initial assessment followed by an interval assessment at least 1 year. Y-axis presents test-values with error bars demonstrating 95% confidence intervals. The assessments on the X-axis in order from left to right: presence of varices, size according to 2-grade classification and size according to 3-grade classification. The figure compares agreement of hepatologists ($n = 3$) in comparison to luminal gastroenterologists ($n = 3$) and trainees ($n = 3$). H, hepatologist; LG, luminal gastroenterologist; T, trainee gastroenterologist

4 | DISCUSSION

Careful endoscopic evaluation is one of the cornerstones in management of OV. The AASLD recommends the use of a 2-grade classification system which has been previously validated as a predictor of variceal haemorrhage as opposed to the 3-grade system. Our study shows that there is no difference in both inter- and intra-observer agreement between the two systems among observers of variable experience. Hepatologists had significantly higher intra-observer agreement compared to the other two groups (Figure 3). Therefore, they may be better suited for assessing OV as they appear to be more consistent in evaluating varices over time.

The interobserver agreement of the 2-grade system was compared to the 3-grade system on a previous study. On this study, pre-specified criteria was used to grade varices using both systems. The investigators used videotape recordings from 206 patients who took part in a prophylaxis trial published in 1999 (post hoc analysis and results published in 2003 in a letter to the editor). Recordings were evaluated by four experts independently and blindly in a random

order. The kappa (k) values for agreement between the endoscopist's diagnosis and the four experts were significantly better ($k = 0.60$) for the 3-grade system compared with the 2-grade system ($k = 0.38$). To our knowledge, this is the only classification that has been directly compared to the 2-grade system.¹³ These data have significant shortcomings: firstly, it was published as a letter to the editor so no critical appraisal of the methodology is possible; secondly, it was based on video tape recording of 1990s endoscopic technology with image resolutions far less than what is currently being used in clinical practice; thirdly, recordings were reviewed by experts. Therefore, this is unlikely to reflect current practice. Other studies investigating interobserver agreement have also been published.^{12,14-19} None of the studies directly compared between the commonly used 3-grade system vs the recommended 2-grade system in adults. We found a significant improvement in intra-observer agreement with experience of the endoscopist (Figure 3). None of the Previous studies evaluated for intra-observer agreement which is an important factor in assessing the validity and reproducibility of a classification system and can help rationalise performer allocation to endoscopy lists.

Our study had several strengths. The infrastructure included the use of modern, high definition endoscopes and recording performed by a single reference endoscopist experienced in the assessment and treatment of varices on a dedicated portal hypertension list each week. All videos were prospectively recorded using the same standardised protocol set a priori (Table S1) in order to further reduce operator bias. The use of video rather than still images enables more realistic and unbiased views for observers representative of real-life practice. Finally, we evaluated for interval intra-observer agreement to assess which observers are likely to be more consistent with their own decision-making over time and therefore may be more suited for undertaking these procedures in clinical practice.

This study also has some limitations. The observers did not have access to the clinical context or the complete endoscopic examination of the cases as they would do in real-life for practical reasons. This is the case for the vast majority of OA studies. Knowledge of gastric appearances, including presence of portal hypertensive gastropathy and/or fundal varices, can inform judgement of OV in some cases of uncertainty. Secondly, real-life agreement test-values maybe lower than our study as a single endoscopist recorded the procedure under standardised protocol, which is outlined in Table S1. Which is never the case in real-life.

These data support the use of both the 3-grade and 2-grade classification systems; the latter has been validated as a predictor of variceal haemorrhage and is recommended by the AASLD clinical practice guidelines.⁷ The subjective nature of differentiating (a) no varices from small varices and (b) small varices from non-small varices remains a challenge. We found that usage of the 2-stage classification system does not decrease such discrepancies. It is important to ensure that varices needing treatment are acted on if present and correct classification is the key. Altering classification systems may not provide the answer. This study highlights the value of an experienced operator. Our study suggests that hepatologists, who all perform dedicated varices lists at our centre, were more consistent as evidenced by the significantly higher intra-observer agreement on both the presence vs absence as well as the grade of OV (Figure 3). This is unlikely to be the case for hepatologists who do not perform regular variceal screening.

Whether or not there is a 'glass ceiling' with subjective human classification systems, ie possibly no classification system will improve OA significantly, remains a matter for debate. While having dedicated lists for OV screening performed by hepatologists in tertiary centres may improve consistency in diagnosis and staging of OV as well as treatment decisions, this may reduce the experience of general endoscopists in other settings. Utilisation of better technologies to stratify for cirrhosis and OV including non-invasive markers and MRI may have a future role.²⁰ More recently non-invasive tools such as Baveno criteria,^{21,22} liver stiffness to spleen/platelet score²³ and platelet-spleen ratio score²⁴ have been validated as good negative markers to predict the absence of OV. Such non-invasive tests may help minimise the subjective nature of human classification systems and reduce the overall work-load of variceal screening endoscopy. This will enable the evaluation of selected cases on dedicated lists for the assessment of portal hypertension.

This prospective investigation of the inter- and intra-observer agreement among nine observers with 100 videos revealed substantial agreement using both the 2-grade and 3-grade classification systems. This provides validity for using the 2-stage system which has been validated as a predictor of variceal haemorrhage and is recommended by the AASLD. Hepatologists had significantly higher levels of consistency in identifying both the presence and stage of OV. This may have implications to create alternative training models for residents and fellows in the recognition and grading of OV.

DISCLOSURE

The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

CONFLICT OF INTEREST

There are no known personal or financial conflicts of interest to report.

AUTHOR CONTRIBUTIONS

Dr. W. Fateen contributed to the acquisition of data; statistical analysis; interpretation of data; and drafted the manuscript. Dr. J. White and Dr. A. Khanna contributed to the acquisition of data. Dr. M. Coletta, Dr. S. Samuel, Dr. J. Ortiz and Dr. E. Wilkes contributed to acquisition of data. Dr. M. James contributed to acquisition of data; critical revision of the manuscript for important intellectual content. Prof. G.P. Aithal, Prof. K. Ragnath and Dr. I.N. Guha contributed to acquisition of data; critical revision of the manuscript for important intellectual content. Dr. S.S. Sami contributed to this study concept and design; acquisition of data; analysis and interpretation of data; critical revision of the manuscript for important intellectual content; study supervision; and accepts full responsibility for the conduct of this study.

ORCID

Waleed Fateen  <https://orcid.org/0000-0001-9989-7169>

Guruprasad P. Aithal  <https://orcid.org/0000-0003-3924-4830>

REFERENCES

1. Kovalak M, Lake J, Mattek N, Eisen G, Lieberman D, Zaman A. Endoscopic screening for varices in cirrhotic patients: data from a national endoscopic database. *Gastrointest Endosc*. 2007;65:82-88.
2. Reverter E, Tandon P, Augustin S, et al. A MELD-based model to determine risk of mortality among patients with acute variceal bleeding. *Gastroenterology*. 2014;146:412-419.e3.
3. Fortune BE, Garcia-Tsao G, Ciarleglio M, et al. Child-Turcotte-Pugh Class is best at stratifying risk in variceal hemorrhage. *J Clin Gastroenterol*. 2017;51:446-453.
4. Garcia-Tsao G, Abraldes JG, Berzigotti A, Bosch J. Portal hypertensive bleeding in cirrhosis: risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology*. 2017;65:310-335.
5. Inokuchi K, Inokuchi K. The general rules for recording endoscopic findings on esophageal varices. *Jpn J Surg*. 1980;10:84-87.
6. Idezuki Y. General rules for recording endoscopic findings of esophago-gastric varices (1991). Japanese Society for Portal Hypertension. *World J Surg*. 1995;19(3):420-422; discussion 423.

7. Garcia-Tsao G, Sanyal AJ, Grace ND, Carey W; Practice Guidelines Committee of the American Association for the Study of Liver Diseases, Practice Parameters Committee of the American College of Gastroenterology. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. *Hepatology*. 2007;46:922-938.
8. North Italian Endoscopic Club for the Study and Treatment of Esophageal Varices. Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices. *N Engl J Med*. 1988;319:983-989.
9. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8:23-34.
10. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*. 1977;33:363-374.
11. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med*. 2012;31:3972-3981.
12. Calès P, Zabotto B, Meskens C, et al. Gastroesophageal endoscopic features in cirrhosis. Observer variability, interassociations, and relationship to hepatic dysfunction. *Gastroenterology*. 1990;98:156-162.
13. Calès P, Oberti F, Bernard-Chabert B, Payen J-L. Evaluation of Baveno recommendations for grading esophageal varices. *J Hepatol*. 2003;39:658-659.
14. Ho C, Jr M, Mg B. A comparison of the radiologic and esophagoscopy diagnosis of esophageal varices. *N Engl J Med*. 1961;265:160-164.
15. Ho C, Hw S, Observer BM. Observer variation in the endoscopic diagnosis of esophageal varices. A prospective investigation of the diagnostic validity of esophagoscopy. *N Engl J Med*. 1965;272:830-834.
16. Calès P, Buscail L, Bretagne JF, et al. Interobserver and intercenter agreement of gastro-esophageal endoscopic signs in cirrhosis. Results of a prospective multicenter study. *Gastroenterol Clin Biol*. 1989;13:967-973.
17. Bendtsen F, Skovgaard LT, Sørensen TI, Matzen P. Agreement among multiple observers on endoscopic diagnosis of esophageal varices before bleeding. *Hepatology*. 1990;11:341-347.
18. Pungpapong S, Keaveny A, Raimondo M, et al. Accuracy and interobserver agreement of small-caliber vs. conventional esophagogastroduodenoscopy for evaluating esophageal varices. *Endoscopy*. 2007;39:673-680.
19. D'Antiga L, Betalli P, De Angelis P, et al. Interobserver agreement on endoscopic classification of oesophageal varices in children. *J Pediatr Gastroenterol Nutr*. 2015;61:176-181.
20. Palaniyappan N, Cox E, Bradley C, et al. Non-invasive assessment of portal hypertension using quantitative magnetic resonance imaging. *J Hepatol*. 2016;65:1131-1139.
21. de Franchis R, Faculty BVI. Expanding consensus in portal hypertension: report of the Baveno VI Consensus Workshop: stratifying risk and individualizing care for portal hypertension. *J Hepatol*. 2015;63:743-752.
22. Maurice JB, Brodtkin E, Arnold F, et al. Validation of the Baveno VI criteria to identify low risk cirrhotic patients not requiring endoscopic surveillance for varices. *J Hepatol*. 2016;65:899-905.
23. Kim BK, Han K-H, Park JY, et al. A liver stiffness measurement-based, noninvasive prediction model for high-risk esophageal varices in B-viral liver cirrhosis. *Am J Gastroenterol*. 2010;105:1382-1390.
24. Colli A, Gana JC, Yap J, et al. Platelet count, spleen length, and platelet count-to-spleen length ratio for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis. In: Colli A, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2017.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Fateen W, Ragunath K, White J, et al. Validation of the AASLD recommendations for classification of oesophageal varices in clinical practice. *Liver Int*. 2020;40:905-912. <https://doi.org/10.1111/liv.14310>