The American Journal of Surgery 215 (2018) 625–630

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

The Parkland grading scale for cholecystitis

Tarik D. Madni [a], David E. Leshikar [b], Christian T. Minshall [c], Paul A. Nakonezny [d], Canon C. Cornelius [c], Jonathan B. Imran [a], Audra T. Clark [a], Brian H. Williams [a], Alexander L. Eastman [c], Joseph P. Minei [c], Herb A. Phelan [c], Michael W. Cripps [c, *]

[a] UT Southwestern Department of Surgery, Dallas, TX, United States
[b] UC Davis, Division of Trauma and Acute Care Surgery, Sacramento, CA, United States
[c] UT Southwestern Division of Burn/Trauma/Critical Care, Dallas, TX, United States
[d] UT Southwestern Department of Clinical Sciences, Division of Biostatistics, Dallas, TX, United States

ARTICLE INFO

Article history:
Received 22 March 2017
Received in revised form
18 May 2017
Accepted 29 May 2017

Meeting submitted to: This original work was presented by poster at the 2017 annual meeting of EAST and has not been submitted or published elsewhere.

Keywords:
Gallbladder
Grading scale
Inflammation
Outcomes
Reimbursement

ABSTRACT

Background: Gallbladders (GBs) with severe inflammation have longer operative times and an increased risk for complications. We propose a grading system using intraoperative images to better stratify GB inflammation.

Methods: After reviewing the intraoperative images of GBs obtained during several hundred laparoscopic cholecystectomies, we developed a five-tiered grading system based on anatomy and inflammatory changes. Fifty intraoperative photographs were taken prior to dissection and then distributed to 11 surgeons who rated each GB's severity per the grading system. The two-way random effects Intraclass Correlation Coefficient (ICC) was used to assess the reliability among the raters.

Results: The ICC among the raters of GB severity was 0.804 (95% CI: 0.733 to 0.867; $p = 0.0001$). Nineteen GB images had greater than 82% agreement and 16 were clustered around GBs with severe inflammation (grades 3–5).

Conclusion: This study proposes a simple, reliable grading system that characterizes GB complexity based on inflammation and anatomy.

1. Introduction

Gallbladder (GB) disease affects over 20 million people in the United States,[1] making laparoscopic cholecystectomy (LC) one of the most common operations performed by general surgeons.[2] Yet, not all cholecystitis is created equal. Differences in anatomy and inflammation can wreak havoc on an otherwise straightforward operation. Increased degrees of GB inflammation have been shown to lead to more open conversion and iatrogenic injuries.[3] However, outcome comparisons, surgeon compensation, and resident case log entry all consider LC at the lowest common denominator of a straightforward “robin's egg blue” GB, which is a rare find for acute care surgeons.

Accurate and reliable stratification of the severity of GB disease requires a grading system that can be widely deployed and easily implemented. Multiple grading scales have been developed in the past to try to predict the level of difficulty for LC. Most of these scores are based on preoperative clinical findings, with few that utilize intraoperative factors.[4] These scoring systems are also complex, with multiple inputs and grades, limiting the practicality of using these scores in the operative setting. We propose that while preoperative indicators may hold some predictive value, it is not until the GB is visualized during surgery that a true determination can be made as to the severity of inflammation. To date, there are no widespread, validated grading systems utilized to stratify the intraoperative severity of GB inflammation. We believe that this is due, in part, to the inability to routinely capture high-resolution images that can be reviewed and graded by multiple

* Corresponding author. University of Texas Southwestern Medical Center, Division of Burn/Trauma/Critical Care, 5323 Harry Hines Blvd., Dallas, TX 75390-9158, United States.

E-mail addresses: tarik.madni@utsouthwestern.edu (T.D. Madni), deleshikar@ucdavis.edu (D.E. Leshikar), christian.minshall@utsouthwestern.edu (C.T. Minshall), paul.nakonezny@utsouthwestern.edu (P.A. Nakonezny), Canon.C.Cornelius@uth.tmc.edu (C.C. Cornelius), jonathan.imran@utsouthwestern.edu (J.B. Imran), audra.clark@utsouthwestern.edu (A.T. Clark), Brian.williams@utsouthwestern.edu (B.H. Williams), alexander.eastman@utsouthwestern.edu (A.L. Eastman), joseph.minei@utsouthwestern.edu (J.P. Minei), herb.phlelan@utsouthwestern.edu (H.A. Phelan), michael.cripps@utsouthwestern.edu (M.W. Cripps).

http://dx.doi.org/10.1016/j.amjsurg.2017.05.017
0002-9610/

surgeons. Recent improvements in technology that allow for image capture into the patient's electronic medical record provide the opportunity for more than one surgeon to view and evaluate the severity of disease.

We hypothesize that a novel grading system, based solely on intraoperative images, can stratify GB inflammation. Our first specific aim was to develop a cholecystitis stratification system with a limited number of grades based on intraoperative images. Our second specific aim was to evaluate the reliability of cholecystitis severity grade assignment among a group of acute care surgeons.

2. Methods

2.1. Study setting and procedures

This study was approved by the institutional review board at the University of Texas Southwestern Medical Center. All patients at Parkland Memorial Hospital, a large, urban, tertiary referral hospital, who underwent LC by the Trauma and Acute Care Surgery service over an 8-month period between 10/2015 and 5/2016 were included in the study. Since October 2015, the standard operating procedure for LC at Parkland Memorial Hospital is to take intraoperative pictures of the GB using the laparoscope once the GB is initially visualized. These pictures were stored on the Parkland Black Diamond Video server. Intraoperative photographs were taken of the right upper abdominal structures after placement of all four laparoscopic ports. If the GB was visualized easily, it was grasped and retracted cephalad prior to taking the photograph. If severe inflammation was present which limited mobilization or the ability to visualize the GB, the picture was taken of the inflamed area. These images were referred to as the "initial view" of the GB.

Development of the grading scale was based on the concept that it should have 1) a limited number of grades, 2) be easy to remember, 3) and have consistent assignment among users. The grading scale was developed by reviewing downloaded initial view images of the GBs. We aimed for an odd-numbered grading system but felt three grades of severity (ie: none, moderate, severe) would be too broad of distinctions. Additionally, a seven-tiered grading system was thought to offer too much specificity, would be difficult to remember, and would be cumbersome to use in clinical practice. Our hypothesis was that a five-tiered grading system would give an appropriate, stepwise range for surgeons to differentiate one gallbladder from another in terms of severity.

2.2. Statistical analysis

We first determined the number of photographs and raters needed to achieve at least 80% statistical power, with a significance level of 0.05 (2-tailed), so as to detect a conservative intraclass correlation coefficient (ICC) $\leq$ 0.20. Thus, an *a priori* power analysis determined that 11 raters grading 45 GB images achieves 81% power to detect ICC as small as 0.20 with a significance level of 0.05. To err on the side of improved power, an independent reviewer selected 50 images equally representative of all five grades and placed them in random order. Afterwards, these 50 intraoperative photographs were distributed (in random order) to 11 acute care surgeons who, in turn, rated each GB's severity per the grading system. The two-way random effects ICC was used to assess the reliability (or magnitude of absolute agreement) among the 11 raters based upon the 50 GB images.[5] Demographic and clinical characteristics of the 50 patients associated with the 50 respective intraoperative photographs across the 5 image grades were described using the sample median and interquartile range (IQR) for continuous variables and the frequency and percentage for categorical variables. Statistical analysis was carried out using SAS software, version 9.4 (SAS Institute, Cary, NC). The level of significance was set at $\alpha = 0.05$ (two-tailed).

3. Results

Over 200 "initial view" images of GBs were available for analysis. A five-tiered grading scale was developed by a team of acute care surgeons that described increasing levels of GB inflammation and commonly encountered anatomic abnormalities. Once developed, an independent evaluator reviewed the 200 images and selected a cohort of approximately 30 images to test the scale among a group of acute care surgeons. None of these images were used in the final grading scale analysis. After review and discussion of these images, a final grading scale based on the initial view was developed and the results are in Table 1, with examples of each grade noted in Fig. 1.

After the 50 initial view images were randomly ordered, then each of the 11 raters graded all 50 images, for a total of 550 observations. The ICC among the 11 raters of GB severity was 0.804 (95% CI: 0.733 to 0.867; p = 0.0001; Fig. 2). The high reliability of the grading system is also demonstrated by the near-identical trend for all 11 raters, as shown in Fig. 2. Nineteen GB images had greater than 82% agreement, and 16 of those were clustered around those GBs with severe inflammation (grades 3–5). Only 3 images had less than 50% agreement and those were assigned grades 1 through 4.

We then grouped the GB images by their pathologic diagnosis of 1) cholelithiasis only, 2) acute cholecystitis, 3) chronic cholecystitis, 4) acute on chronic cholecystitis, or 5) gangrenous cholecystitis. There was a complete pathologic agreement between grade 5 and a pathologic diagnosis of gangrene, as no other grade of GB had a diagnosis of gangrene. Acute and acute on chronic cholecystitis were found in grades 3–5, while chronic cholecystitis covered the spread with at least one of each grade (Fig. 3). The complete demographic and clinical characteristics of the 50 patients associated with the 50 respective intraoperative photographs across the 5 image grades are shown in Tables 2–4.

4. Discussion

The need for a surgical scoring system to aid in the prediction of operative difficulty is imperative. Pragmatically, to be able to predict with accuracy and reliability the elevated difficulty of a LC may assist a surgeon in the decision to convert to an open operation sooner, or call for more experienced hands. Studies have demonstrated that the removal of an acutely inflamed GB can be extremely difficult and should be considered an advanced laparoscopic procedure that may require additional expertise.[6] However, the definition and stratification of inflammation remains elusive.

When caring for a patient with acute cholecystitis, there is currently a lack of a uniform grading system of disease severity.[7] In an effort to standardize disease severity, the American Association for the Surgery of Trauma has developed a grading system that grades the anatomic severity of inflammation in multiple emergency general surgery conditions. This scale ranges from 1 to 5, each grade representing an escalation from mild to severe, widespread disease.[8] While this scale has been validated in certain diseases such as colonic diverticulitis,[9] no current validation study for acute cholecystitis has been performed. In addition, the impact of each grade of each disease on surgical difficulty and patient outcomes has yet to be verified.

A number of predictive factors of LC difficulty have been verified in previous studies.[10] Preoperative factors such as male sex, diabetes mellitus, previous surgery, history of cholecystitis, white blood cell count, GB wall thickness, pericholecystic fluid, and an elevated C reactive protein have all been demonstrated to be

Table 1
Parkland grading scale for cholecystitis.

Cholecystitis Severity Grade	Description of Severity
1	Normal appearing gallbladder ("robin's egg blue") • No adhesions present • Completely normal gallbladder
2	Minor adhesions at neck, otherwise normal gallbladder • Adhesions restricted to the neck or lower of the gallbladder
3	Presence of ANY of the following: • Hyperemia, pericholecystic fluid, adhesions to the body, distended gallbladder
4	Presence of ANY of the following: • Adhesions obscuring majority of gallbladder • Grade I-III with abnormal liver anatomy, intrahepatic gallbladder, or impacted stone (Mirrizi)
5	Presence of ANY of the following: • Perforation, necrosis, inability to visualize the gallbladder due to adhesions

predictive factors for surgical difficulty and open conversion.[11–13] Multiple grading scales of LC difficulty have thus been developed in the past based on such risk factors. Most of these scoring systems, however, only utilize preoperative findings, and are derived from single institutions with limited power.[14–16]

A recent risk score was developed by Soltes et al. (2014) to predict the difficulty of LC. Based on history, physical examination, and abdominal ultrasonography measures, this risk score utilizes five levels of difficulty with significant differences in operative time, difficulty, and open conversion rates.[17] Their study, however, only assessed elective cases and did not take into consideration any intraoperative findings. Another group, Hiroto et al. (2006), proposed using the Tokyo Guidelines to stratify levels of GB inflammation, but this time it was used for acute cholecystitis.[18] Again, however, the authors used preoperative findings to grade each class.

While the previous scoring methods were based on preoperative data, the Surgical Apgar Score was developed to aid in the prediction of morbidity and mortality after a general or vascular surgery utilizing intraoperative data.[19] Using the variables of estimated blood loss, lowest mean arterial pressure, and lowest heart rate, this scale has been useful as a postoperative predictor of morbidity[20]; however, these measures compiled postoperatively are not useful for intraoperative decision-making. In an effort to predict LC difficulty, Sugure et al. (2015) developed a new grading system based on intraoperative findings. With 10 different grades, this scale has been found to be complex, as it includes multiple factors such as patient body mass index, adhesion presence, GB distention, and time to dissect out the critical view.[4] In addition to complexity, the scale has not been validated, and the quality of

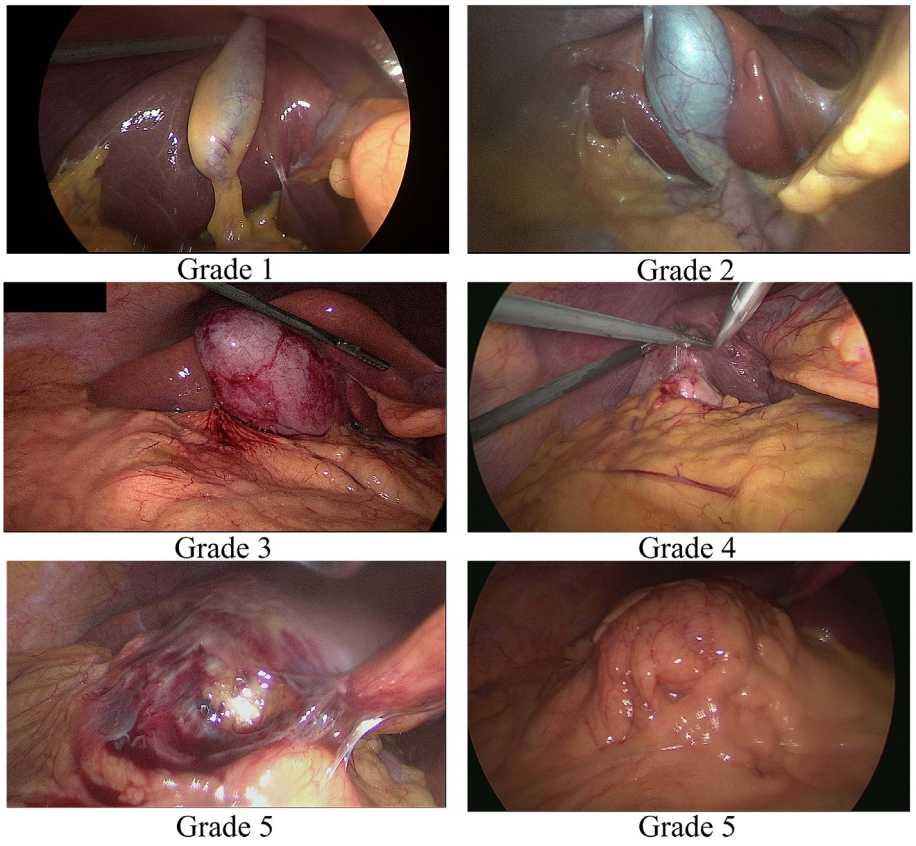

Fig. 1. Grade examples.

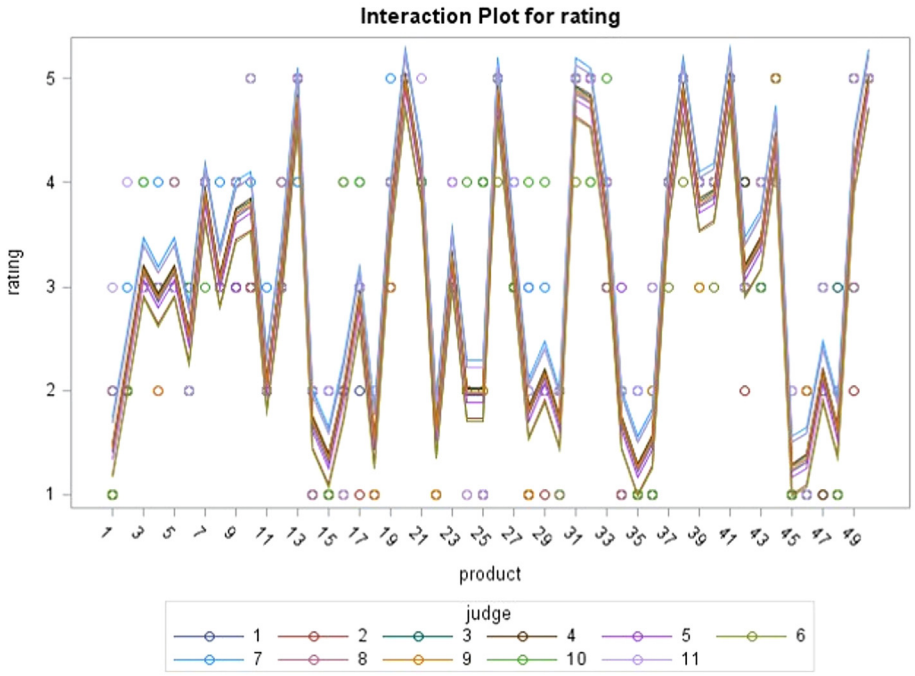

Fig. 2. Intraclass Correlation Amongst Raters. Intraoperative images of 50 gallbladders were arranged randomly. Each rater graded each image. The high reliability of the grading system is demonstrated by the near-identical trend for all 11 raters.

adhesions is not specified.

Most grading scales which have been developed are used to predict the risk of conversion to an open cholecystectomy.[4] There is a paucity in the literature of scoring systems to predict other metrics such as hospital length of stay, iatrogenic injury, and total operative time. As acute care surgery is currently the most expensive cause of emergency hospitalization in the US, with an annual cost of over $28 billion,[21] the avoidance of additional cost is of utmost importance. Additional experience and training in laparoscopy in other surgical operations has demonstrated significant differences in length of stay, intensive care unit admission, and complication rates,[22] which can both improve both patient outcomes and lower cost. To be able to better predict the degree of difficulty of an operation, a surgeon may be able to make a more informed decision, knowing when to convert to an open operation or call for more experienced aid.

We created a grading scale of GB inflammation with an ICC of 0.804, demonstrating strong agreement amongst raters utilizing this scale. While this study was not powered to demonstrate significant differences between grades, trends are appreciated amongst the grades in Tables 3 and 4. Preoperative findings such as a thickened GB wall and pericholecystic fluid are both highest in grade 5 of disease. Operating room time, bile leak rate, length of stay, and conversion rate are all highest at grade 5 as well. In addition, Fig. 3 shows a trend in pathologic findings between grades. No grade 1 or 2 GBs were found to be acute or acute on chronic cholecystitis. Gangrenous cholecystitis was only found in GBs labeled as grade 5. We hypothesize that a such a grade-five gallbladder should have a longer operative time, increased operative difficulty, and increased post-operative complication rate compared to lower grades; however, this study was not powered to demonstrate such a significant difference. Future work will require

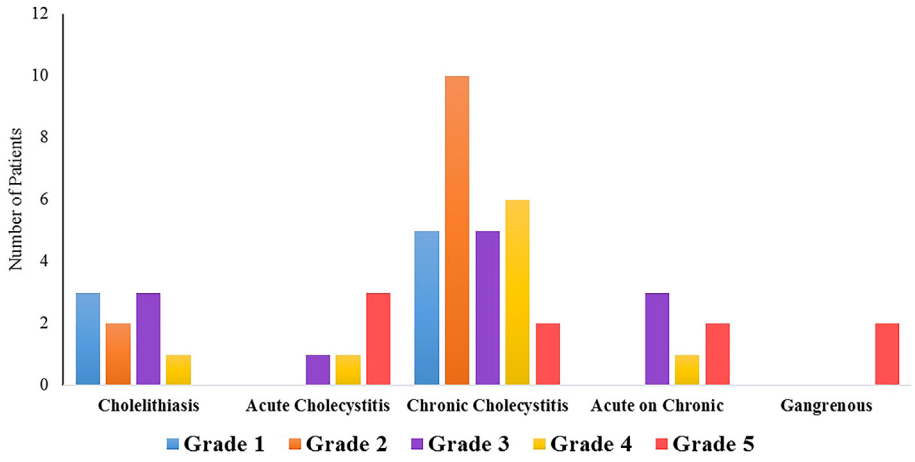

Fig. 3. Pathology diagnosis per grade.

Table 2
Demographic characteristics of the 50 patients across grades.

Demographic characteristics	Grade 1 (n = 8)	Grade 2 (n = 12)	Grade 3 (n = 12)	Grade 4 (n = 9)	Grade 5 (n = 9)
Age[a]	39 (28–43)	33 (29–45)	34 (30–41)	38 (31–43)	48 (41–51)
Female (%)	87.5	100	91.7	66.7	55.6
Hispanic (%)	62.5	83.3	83.3	77.8	77.8

[a] Median (Interquartile range).

Table 3
Preoperative clinical characteristics of the 50 patients across grades.

Preoperative measures	Grade 1 (n = 8)	Grade 2 (n = 12)	Grade 3 (n = 12)	Grade 4 (n = 9)	Grade 5 (n = 9)
% GBW (US)	12.5	0	16.7	11.1	44.4
% PCF (US)	0	0	8.33	0	22.2
WBC[a]	8.7 (7–9)	9.4 (8–11)	9.3 (7–11)	8.8 (8–11)	11.8 (9–14)
AST[a]	85 (20–757)	23.5 (20–36)	49 (23–77)	37 (26–190)	23 (20–44)
ALT[a]	59.5 (20–659)	24 (17–33)	46 (25–75)	52 (26–172)	26 (14–66)
ALP[a]	114.5 (108–149)	84.5 (80–120)	81.5 (101–104)	99 (96–122)	87 (78–94)
TB[a]	0.5 (0.3–1.8)	0.4 (0.2–0.7)	0.5 (0.2–0.8)	0.5 (0.4–1.1)	0.5 (0.4–1)

% GBW, percent with thickened gallbladder wall; US, ultrasound; % PCF, percent with pericholecystic fluid; WBC, white blood cell count;, AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; TB, total bilirubin.
[a] Median (Interquartile range).

Table 4
Perioperative clinical characteristics of the 50 patients across grades.

Perioperative measures	Grade 1 (n = 8)	Grade 2 (n = 12)	Grade 3 (n = 12)	Grade 4 (n = 9)	Grade 5 (n = 9)
OR Time (min)[a]	64.5 (52–82)	62.5 (55–74)	73.5 (52–92)	73.0 (72–77)	93.0 (72–100)
% OC	0	0	0	0	11.1
LOS (days)[a]	1.8 (1–2)	1.8 (1–3)	1.5 (1–2)	2 (2–5)	2.5 (1–3)
Bile Leak %	0	0	0	0	11.1

OR, operating room; min, minutes; OC, open conversion, LOS, length of stay.
[a] Median (Interquartile range).

additional patients to appropriately power a study that can evaluate potential differences in postoperative complications and outcomes between grades.

In addition to the prediction of the difficulty of an LC, our grading system may also help change the billing practices for surgeons. Current procedural terminology codes are used to communicate uniform information about procedures among different medical providers.[23] Although a "22" modifier can be added to represent increased procedural time or technical difficulty, there is no reliable metric that can corroborate that effort. Other operations, such as a skin grafting, have one code for a certain area covered, and then additional codes can be added as the wound covered increases in size. These extra codes are meant to acknowledge the additional work put forth for the given operation, thus justifying increased compensation. A system with additional codes could also pertain to GBs. A LC is also one of the few laparoscopic cases that is designated as "basic" for resident case logs that are needed to graduate from residency. However, some of the more difficult, acutely inflamed GBs are often extremely difficult and should be considered an "advanced" case.[6] Our grading scale could be utilized to standardize the description of GB inflammation and would help to delineate difficult from straightforward cases. If complexity of case could be documented, then difficult GBs may be logged as "advanced" laparoscopic cases. Another justification for a standardized scale is that operating room time utilization is often based on the average time needed to complete the index case. Correlating grades of cholecystitis on outpatient and inpatients in an institution could allow for a more accurate representation of operating room time needs.

A few limitations exist within our study. First, the ICC analysis was based on a retrospective review of still, initial view, intraoperative GB images. A live, intraoperative image may demonstrate different qualities and thus could affect the overall grade. Second, this study was derived from a single institution. While 11 different surgeons graded each image, they are all part of the same burn, trauma, and critical care division, which could have affected responses that would have been different at a separate institution.

Future work will focus on further reliability assessment and validation of our grading scale and evaluate its relationship with such outcomes as operative time, complication rates, and length of stay. We aim to see if differences in such outcomes can be seen between grades or groups of grades with a larger number of patients. In this sense, our five-tiered grading system is our starting point, and we aim to refine this scale if needed over time with additional data. Finally, a multicenter study is required to further evaluate the interrater reliability amongst reviewers from different centers and regions.

5. Conclusion

This study proposes a simple, reliable grading system that characterizes GB complexity based on inflammation and anatomy. The classification system meets the *a priori* requirements of being simple and easily reproducible among surgeons with the highest level of agreement in those with severe inflammation. This study lays the groundwork to determine if it can be used for risk stratification and to predict patient outcomes, increase resident complex laparoscopic case entry, and improve changes in surgeon reimbursement.

Disclosure information

The authors have no conflicts of interest to disclose.

References

1. Everhart JE, Khare M, Hill M, Maurer KR. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology*. 1999;117(3):632–639.
2. Csikesz NG, Singla A, Murphy MM, Tseng JF, Shah SA. Surgeon volume metrics in laparoscopic cholecystectomy. *Dig Dis Sci*. 2010;55(8):2398–2405.
3. Hussain A. Difficult laparoscopic cholecystectomy: current evidence and strategies of management. *Surg Laparosc Endosc Percutaneous Tech*. 2011;21(4):211–217.
4. Sugrue M, Sahebally SM, Ansaloni L, Zielinski MD. Grading operative findings at laparoscopic cholecystectomy- a new scoring system. *World J Emerg Surg WJES*. 2015;10:14.
5. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant methods Psychol*. 2012;8(1):23–34.
6. Schwaitzberg Steven D, Scott Daniel J, Jones Daniel B, et al. Threefold increased bile duct injury rate is associated with less surgeon experience in an insurance claims database. *Surg Endosc*. 2014;28(11):3068–3073.
7. Crandall Marie L, Agarwal Suresh, Muskat Peter, et al. Application of a uniform anatomic grading system to measure disease severity in eight emergency general surgical illnesses. *J Trauma Acute Care Surg*. 2014;77(5):705–708.
8. Shafi S, Aboutanos M, Brown CV, et al. Measuring anatomic severity of disease in emergency general surgery. *J trauma Acute care Surg*. 2014;76(3):884–887.
9. Savage SA, Klekar CS, Priest EL, et al. Validating a new grading scale for emergency general surgery diseases. *J Surg Res*. 2015;196(2):264–269.
10. Rosen M, Brody F, Ponsky J. Predictive factors for conversion of laparoscopic cholecystectomy. *Am J Surg*. 2002;184(3):254–258.
11. Nidoni R, Udachan TV, Sasnur P, et al. Predicting difficult laparoscopic cholecystectomy based on clinicoradiological assessment. *J Clin Diagn Res JCDR*. 2015;9(12):Pc09–Pc12.
12. Kabul Gurbulak E, Gurbulak B, Akgun IE, et al. Prediction of the grade of acute cholecystitis by plasma level of C-reactive protein. *Iran Red Crescent Med J*. 2015;17(4):e28091.
13. Yang TF, Guo L, Wang Q. Evaluation of preoperative risk factor for converting laparoscopic to open cholecystectomy: a meta-analysis. *Hepato Gastroenterol*. 2014;61(132):958–965.
14. Agrawal N, Singh S, Khichy S. Preoperative prediction of difficult laparoscopic cholecystectomy: a scoring method. *Niger J Surg Off Publ Niger Surg Res Soc*. 2015 Jul-Dec;21(2):130–133.
15. Gupta N, Ranjan G, Arora MP, et al. Validation of a scoring system to predict difficult laparoscopic cholecystectomy. *Int J Surg Lond Engl*. 2013;11(9):1002–1006.
16. Randhawa JS, Pujahari AK. Preoperative prediction of difficult lap chole: a scoring method. *Indian J Surg*. 2009;71(4):198–201.
17. Soltes M, Radoňak J. A risk score to predict the difficulty of elective laparoscopic cholecystectomy. *Wideochirurgia i inne Tech maloinwazyjne = Videosurgery Miniinvasive Tech*. 2014;9(4):608–612.
18. Hirota M, Takada T, Kawarada Y, et al. Diagnostic criteria and severity assessment of acute cholecystitis: Tokyo Guidelines. *J Hepato Biliary Pancreatic Surg*. 2007;14(1):78–82.
19. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg*. 2007;204(2):201–208.
20. Chandra A, Mangam S, Marzouk D. A review of risk scoring systems utilised in patients undergoing gastrointestinal surgery. *J Gastrointest Surg Off J Soc Surg Alimentary Tract*. 2009;13(8):1529–1538.
21. Ogola GO, Gale SC, Haider A, Shafi S. The financial burden of emergency general surgery: national estimates 2010 to 2060. *J trauma Acute care Surg*. 2015;79(3):444–448.
22. Hsu GP, Morton JM, Jin L, Safadi BY, Satterwhite TS, Curet MJ. Laparoscopic Roux-en-Y gastric bypass: differences in outcome between attendings and assistants of different training backgrounds. *Obes Surg*. 2005;15(8):1104–1110.
23. Leslie-Mazwi TM, Bello JA, Tu R, et al. Current procedural terminology: history, structure, and relationship to valuation for the neuroradiologist. *AJNR Am J Neuroradiol*. 2016;37(11):1972–1976.