



REVIEW ARTICLE

# Critical appraisal of predictive tools to assess the difficulty of laparoscopic liver resection: a systematic review

Julie Hallet<sup>1,2,3</sup> · Patrick Pessaux<sup>4,5</sup> · Kaitlyn A. Beyfuss<sup>3</sup> · Shiva Jayaraman<sup>2,6</sup> · Pablo E. Serrano<sup>7</sup> · Guillaume Martel<sup>8</sup> · Natalie G. Coburn<sup>1,2,3</sup> · Tullio Piardi<sup>9</sup> · Alyson L. Mahar<sup>10</sup>

Received: 20 August 2018 / Accepted: 11 October 2018 / Published online: 22 October 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

**Background** Objective assessment of the difficulty of laparoscopic liver resection (LLR) preoperatively is key in improving its uptake. Difficulty scores are proposed but are not used routinely in practice. We identified and appraised predictive models to estimate LLR difficulty.

**Methods** We systematically searched the literature for tools predicting LLR difficulty. Two independent reviewers selected studies, abstracted data and assessed methodology. We evaluated tools' quality and clinical relevance using the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) guidelines.

**Results** From 1037 citations, we included 8 studies reporting on 4 predictive tools using data from 1995 to 2016 in Asia and Europe. In 4 development studies, tools were designed to predict difficulty as assigned by experts using a 10-level difficulty index, operative time, post-operative morbidity or intra-operative complications. Internal validation and performance metrics were reported in one development study. One tool was subjected to external validations in 4 studies (1 independent and geographic). Validations compared post-operative outcomes (operative time, blood loss, transfusion, major morbidity and conversion) between the risk categories. One study validated discrimination (AUROC 0.53). Calibration was not assessed.

**Conclusion** Existing tools cannot be used confidently to predict LLR difficulty. Consistent objective clinical outcomes to predict to define LLR difficulty should be established, and better-quality tools developed and validated in a wide array of populations and clinical settings, following best practices for predictive tools development and validation. This will improve risk stratification for future trials and uptake of LLR.

**Keywords** Laparoscopy · Liver · Hepatectomy · Score · Prediction

Advances in peri-operative care and surgical technique have significantly improved outcomes of liver resections, attaining mortality rates below 5% in high-volume centres [1, 2]. The benefits of minimally invasive surgery (MIS) have been highlighted in several gastrointestinal surgical procedures for the treatment of benign and malignant diseases, mainly

owing to reduced pain, blood loss, length of hospital stay and complications [3–5]. However, MIS uptake has been slower in liver surgery due to concerns regarding the technical feasibility and safety of the technique [6–10]. Twenty-five years after the first reports of LLR, only a small proportion of liver resections is performed through a minimally invasive approach [11, 12].

It is important to consider the level of complexity of different procedures when working on improving the uptake of LLR. While simpler procedures are being performed laparoscopically more commonly, complex major liver resections that are more challenging remain uncommon [9, 13–15]. The ability to objectively define the difficulty of a LLR prior to surgery would support surgeons in (1) identifying appropriate resections to perform laparoscopically at each step of the learning curve for a better ad equation of skills to surgical

Part of this work has been accepted for poster presentation at the International HPB Association Congress to be held in Geneva, Switzerland, in September 2018, and at the Canadian Surgery Forum to be held in St. John, NL, in September 2018.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00464-018-6479-3>) contains supplementary material, which is available to authorized users.

✉ Julie Hallet  
Julie.hallet@sunnybrook.ca

Extended author information available on the last page of the article

difficulty, and (2) selecting patients to benefit most from the minimally invasive approach.

Surgical scores have been proposed to rate the difficulty of LLR [16–19]. However, none of them is routinely used in practice. To be used by surgical teams, the difficulty scores must be able to accurately predict the outcomes of LLR in a diverse population of patients undergoing this procedure, and must be simple and easy to use in a preoperative clinical setting. The Morioka International Consensus on LLR recognised the usefulness of predictive tools to better select procedures according to surgeons' skills and experience, but highlighted the need for validations of existing tools prior to clinical application [13].

Therefore, the objective of this study was to systematically review the literature for studies reporting development or validation of scores predicting the difficulty of LLR.

## Methods

### Search strategy

Medline (1966–December 2017), EMBASE (1974–December 2017), Web Of Knowledge and the Scopus database (2004–December 2017) were systematically searched. No restriction was applied regarding language or the type of publication. As of December 2017, the grey literature using the OpenSIGLE database, the Trip database and Google Scholar, was queried for relevant studies. With assistance from an information specialist, the search strategy was initially developed for Medline and adapted for each data source (see Online Appendix). Keywords and MeSH (or Emtree) terms were gathered into three categories: (1) liver resection, (2) minimally invasive surgery and (3) predictive score. Medical subject headings (MeSH) did not exist for prognostication tools and so a previously developed filter for predictive models was used [20]. To increase search sensitivity, each keyword was exploded. In addition, bibliographies of all included studies were reviewed for any relevant publications.

### Study selection

This review included studies reporting on the development, validation or impact assessment of scores predicting the difficulty or outcomes of LLR (laparoscopic-assisted, totally laparoscopic or robotic). Studies including adult patients ( $\geq 18$  years old) operated for benign and malignant diseases were eligible. Studies were excluded if they reported on the prognostic impact of a single factor (unless it was updating the accuracy of an existing prognostic tool), relied on inappropriate analytic purpose (e.g. multivariate modelling not aimed at prediction, development

of novel statistical methods), were not limited to LLR or did not represent original data or research (e.g. editorial, review). Studies that included patients not fulfilling the inclusion criteria (e.g. open liver resection) were excluded if it was not possible to distinguish those patients from the larger study population. In the event of duplicate publication, we included the most relevant and the most informative study.

### Data abstraction

A standardised extraction form was developed and pilot-tested using previously published recommendations from the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines [21, 22]. The characteristics of study design, study population, prognostic factors, outcome measures and tool development (statistics modelling decisions, candidate variable selection) and validation (internal and external validation methods, measures of the model predictive accuracy) were abstracted.

### Risk of bias assessment

Considering that there is currently no published validated scoring system to assess the risk of bias of predictive tools' development and validation studies, a checklist was created using key methodological elements suggested by the TRIPOD guidelines [22]. Similar assessment has been used in prior systematic reviews of prognostic tools [23–25].

Two authors (JH and KB) selected studied, extracted data and assessed risk of bias independently. Disagreements were resolved by consensus. A single reviewer (JH) assessed the references from the grey literature and meetings' proceedings.

### Data summary

The statistics related to predictive scores' development and validation are reported in summary tables. Formal statistical evaluation of the internal or external validity was defined as assessment of discrimination and/or calibration of a score. Discrimination assesses the ability of the score to distinguish between high- and low-risk patients. Calibration assesses how closely the predicted values of the outcome match the observed outcomes in the study sample. Both these are standard in good statistical practices evaluating clinical prediction model [22, 26].

## Results

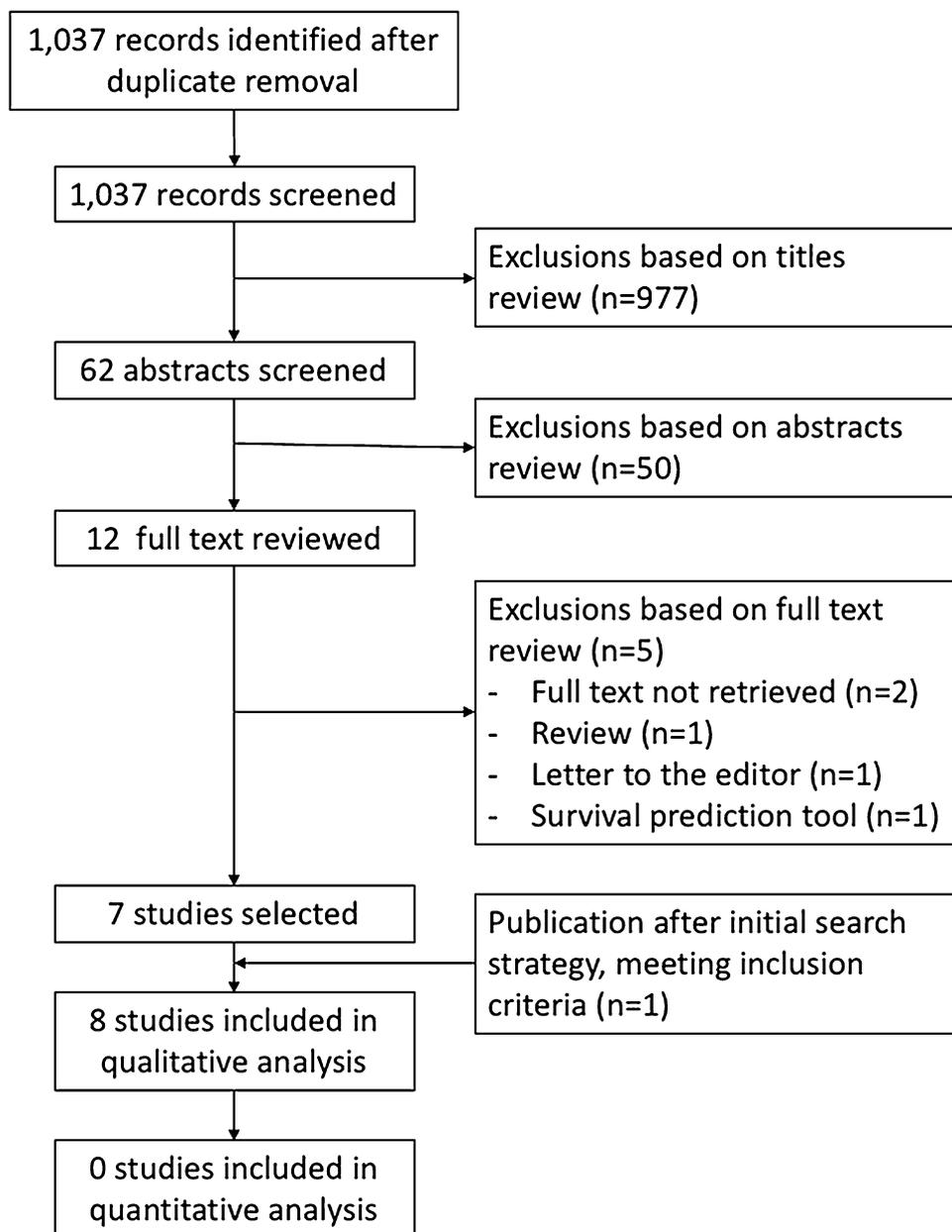
### Systematic search

The search results are presented in Fig. 1. Seven articles were initially included. One additional article was included later because it was published immediately after the initial search strategy was performed. Four tools predicting the difficulty of LLR were identified: the Difficulty Scoring System, the Hasegawa score, the Difficulty

of Laparoscopic Liver Resections Classification and the Halls tool (Table 1) [16, 27, 28]. Four articles contained solely an external validation of one of the identified scores [29–32]. We did not identify studies assessing the effectiveness or implementation of tools in clinical practice.

All tools were presented as risk scores. One tool provided the statistical equation, including coefficients and the intercept, used to create the risk score [16]. Publication of the equation is recommended by the TRIPOD guidelines in order to allow adequate external validation of the tool with appropriate statistical methods [22].

**Fig. 1** Flow diagram of study selection



**Table 1** Characteristics of included studies and predictive tools

| Author                | Tool studied   | Country | Data source                | Type of practice                          | Population  | Data collection | Sample size | Validation or Development | Number of events                   | Follow-up duration | Primary outcome predicted by the tool  |
|-----------------------|--|---------|----------------------------|---|---|-----------------|-------------|---------------------------|------------------------------------|--------------------|--|
| Ban et al. [16]       | Difficulty Scoring System                                  | Japan   | Retrospective cohort study | Tertiary centres<br>Multi-institutional   | Pure LLR for neoplasm   | 2011–2014       | 90          | Development               | NR                                 | NR                 | Difficulty index (assessed by experts)   |
| Hasegawa et al. [27]  | Hasegawa model   | Japan   | Retrospective cohort study | Tertiary centre<br>Single institution     | Pure LLR for neoplasm<br>Exclusions: hybrid, hand-assisted, concurrent procedure  | 2006–2016       | 187         | Development               | NR                                 | NR                 | Operative time   |
| Kawaguchi et al. [28] | Difficulty of Laparoscopic Liver Resections Classification | France  | Retrospective cohort study | Tertiary centre<br>Single institution     | Pure LLR for benign and malignant neoplasm<br>Exclusions: cyst fenestration, repeat hepatectomy, multiple resections (> 4), vascular or biliary reconstruction and extrahepatic resection | 1995–2015       | 452         | Development               | Morbidity: n=114<br>Mortality: n=5 | 90 days            | Post-operative morbidity and mortality   |
| Halls et al. [18]     | Halls model  | Europe  | Retrospective cohort study | Tertiary centres<br>Multiple institutions | LLR for benign and malignant neoplasm<br>Exclusions: vascular reconstruction, hepaticojejunostomy, multi-visceral resection, hilar cholangiocarcinoma                                     | 1997–2016       | 2856        | Development               | 541                                | 90 days            | Intra-operative complication: Satava classification, including elevated blood loss, injury to adjacent structures and conversion |

Table 1 (continued)

| Author                 | Tool studied              | Country | Data source                | Type of practice                                      | Population   | Data collection | Sample size | Validation or Development                      | Number of events | Follow-up duration | Primary outcome predicted by the tool    |
|------------------------|---------------------------|---------|----------------------------|---|--|-----------------|-------------|--|------------------|--------------------|--|
| Uchida et al. [29]     | Difficulty Scoring System | Japan   | Retrospective cohort study | Tertiary centre<br>Single institution                 | All LLRs at the institution  | 2010–2015       | 78          | Validation of <i>Difficulty Scoring System</i> | NR               | NR                 | Operative time<br>Blood loss             |
| Im et al. [30]         | Difficulty Scoring System | Korea   | Retrospective cohort study | Tertiary centre<br>Single institution                 | Pure LLR for cancer  | 2003–2015       | 90          | Validation of <i>Difficulty Scoring System</i> | NR               | NR                 | Not specified (multiple outcomes tested) |
| Tanaka et al. [31]     | Difficulty Scoring System | Japan   | Retrospective cohort study | Community and tertiary centres<br>Multi-institutional | Pure LLR for cancer<br>Exclusions: non-cancer, multiple lesions, concurrent procedure, missing data for validation | 2010–2014       | 2199        | Validation of <i>Difficulty Scoring System</i> | NR               | NR                 | Not specified (multiple outcomes tested) |
| Periyasamy et al. [32] | Difficulty Scoring System | Korea   | Retrospective cohort study | Tertiary centre<br>Single institution                 | Pure LLR for hepatocellular carcinoma  | 2004–2015       | 301         | Validation of <i>Difficulty Scoring System</i> | NR               | NR                 | Not specified (multiple outcomes tested) |

LLR laparoscopic liver resection, NR not reported

\*Studies have tested the tool against different outcomes; primary is the outcome the tool was developed to predict, secondary are additional outcomes for which the developed tool was tested

## Score development methods

Details of the tools' development and validation studies are presented in Table 2. None of the included studies relied on prospectively collected data aimed at creating a clinical prediction tool. Four studies used data from Japan (57%), 2 from Korea, one from France and one from multiple European institutions. Data related to patients undergoing LLR from time periods covering 2003 to 2016 in most studies, except for two studies that used data going back to 1995 and 1997. Three studies included patients from multiple institutions, two reporting on tool development and one on tool validation. The cohort sample sizes ranged from 78 to 2856. The number of events were reported in one out of 4 development studies. Two studies reported the follow-up duration (extrapolated from description of morbidity being assessed at 90 days following surgery).

The methodology for the included studies is summarised in Table 3. Three out of 4 tools identified predictive factor for inclusion using p-value-based variable selection (univariate and/or multivariate analyses). The fourth tool relied solely on a priori identification of clinically relevant variables.

Two studies reported the frequency and/or handling of missing data. Two tools predicting continuous outcomes used linear regression models to develop their predictive tool, one used logistic regression for categorical outcome and the fourth did not report the statistic model used. None of the studies reported verifying the statistical assumptions

of the models. One tool provided the statistical equation used to create the risk score [16].

## Populations, predictive factors and outcomes

The populations addressed by each predictive tool are detailed in Table 1. All tools were developed to assess difficulty of LLR in patients undergoing pure LLR for neoplasm. The predictive factors included in the tools and the outcomes being predicted are presented in Table 3. There was significant heterogeneity in both the predictive factors included in the tools and the outcomes being assessed.

Up to 5 factors were included in a single tool. A total of 10 predictive factors were included across all tools. Extent of liver resection and tumour location within the liver were common to three tools.

The Difficulty Scoring System tool was developed to predict a difficulty index whereby experts had assigned a score from 1 to 10 to specific types of LLR and which had been tested for inter- and intra-rater reliability in prior studies. The Hasegawa tool indicated being developed to predict operative time. The Difficulty of LLR Classification tool's development study did not explicitly mention the outcome to be predicted, but it is understood it was developed to predict 90-day post-operative morbidity defined by the Clavien–Dindo classification and mortality [33]. The Halls tool was developed to predict intra-operative complications defined by the Satava classification [34]. Two tools then tested the developed tool against additional outcomes. The Difficulty Scoring System development study assessed the

**Table 2** Methodological evaluation of the included studies

| Methodological criterion              | Ban et al. [16]  | Hasegawa et al. [27]                              | Kawaguchi et al. [28]                                | Halls et al. [18]   |
|---------------------------------------|--|---|--|---|
| Tool tested                           | Difficulty Scoring System  | Hasegawa model                                    | Difficulty of LLR Classification                     | Halls model   |
| Predictive factor selection method    | Literature-based/clinical reasoning<br>Screening using univariable/multivariable methods | Screening using univariable/multivariable methods | Literature-based/clinical reasoning—defined a priori | Literature-based/clinical reasoning—confirmed with survey of experts in the field |
| Missing data                          |  |   |  |   |
| Number of patients                    | 2  | NR  | NR   | NR  |
| Methods for handling                  | Exclusion  | NR  | NR   | Exclusion   |
| Handling continuous predictors*       | Linear<br>Dichotomised/categorised   | Linear  | Dichotomised/categorised                             | Dichotomised/categorised  |
| Analysis                              |  |   |  |   |
| Analytic model used                   | Linear regression  | Linear regression                                 | NR   | Binary logistic   |
| Model formula reported                | Yes  | Yes   | NR   | NR  |
| Statistical model assumptions checked | NR   | NR  | NR   | NR  |

\*Studies may have used multiple methods

LLR laparoscopic liver resection, NA not applicable, NR not reported

**Table 3** Predictive factors included in and outcomes predicted by the models reported in this review

| Tool tested                                     | Ban et al. [16]           | Hasegawa et al. [27] | Kawaguchi et al. [28]            | Halls et al. [18] | Uchida et al. [29]        | Im et al. [30]            | Tanaka et al. [31]        | Periyasamy et al. [32]    |
|---|---------------------------|----------------------|----------------------------------|-------------------|---------------------------|---------------------------|---------------------------|---------------------------|
|   | Difficulty Scoring System | Hasegawa model       | Difficulty of LLR Classification | Halls model       | Difficulty Scoring System | Difficulty Scoring System | Difficulty Scoring System | Difficulty Scoring System |
| <i>Type of variables</i>                        |                           |                      |                                  |                   |                           |                           |                           |                           |
| Demographics                                    | –                         | –                    | –                                | –                 | NA (validations)          |                           |                           |                           |
| Physical exam                                   | –                         | X                    | –                                | –                 |                           |                           |                           |                           |
| Disease characteristics                         | X                         | X                    | X                                | X                 |                           |                           |                           |                           |
| Laboratory values                               | –                         | X                    | –                                | –                 |                           |                           |                           |                           |
| Radiology characteristics                       | X                         | X                    | –                                | X                 |                           |                           |                           |                           |
| <i>Included variables</i>                       |                           |                      |                                  |                   |                           |                           |                           |                           |
| Tumour location                                 | X                         | X                    | –                                | –                 |                           |                           |                           |                           |
| Extent of liver resection                       | X                         | X                    | –                                | X                 |                           |                           |                           |                           |
| Tumour size                                     | X                         | –                    | –                                | X                 |                           |                           |                           |                           |
| Proximity to major vessel                       | X                         | –                    | –                                | –                 |                           |                           |                           |                           |
| Liver function                                  | X                         | –                    | –                                | –                 |                           |                           |                           |                           |
| Obesity (body mass index)                       | –                         | X                    | –                                | –                 |                           |                           |                           |                           |
| Thrombocytopenia                                | –                         | X                    | –                                | –                 |                           |                           |                           |                           |
| Operative time                                  | –                         | –                    | X                                | –                 |                           |                           |                           |                           |
| Estimated blood loss                            | –                         | –                    | X                                | –                 |                           |                           |                           |                           |
| Conversion                                      | –                         | –                    | X                                | –                 |                           |                           |                           |                           |
| <i>Outcomes tested against the tool's score</i> |                           |                      |                                  |                   |                           |                           |                           |                           |
| Difficulty index (assessed by experts)          | X                         | –                    | –                                | –                 | –                         | –                         | –                         | –                         |
| Blood loss                                      | X                         | X                    | –                                | X                 | X                         | X                         | X                         | X                         |
| Operative time                                  | X                         | –                    | –                                | –                 | X                         | X                         | X                         | X                         |
| Morbidity                                       | X                         | X                    | X                                | –                 | –                         | X                         | X                         | –                         |
| Mortality                                       | –                         | –                    | X                                | –                 | –                         | –                         | –                         | –                         |
| Conversion                                      | –                         | X                    | –                                | X                 | –                         | –                         | X                         | X                         |
| Length of stay                                  | –                         | X                    | –                                | –                 | –                         | X                         | X                         | X                         |
| Pringle   | –                         | –                    | –                                | –                 | –                         | –                         | –                         | X                         |
| Transfusion                                     | –                         | –                    | –                                | –                 | –                         | X                         | –                         | X                         |
| Intra-operative complications                   | –                         | –                    | –                                | X                 | –                         | –                         | –                         | –                         |

LLR laparoscopic liver resection, NA not applicable

risk score's association with operative time, blood loss and post-operative morbidity and mortality. The Hasegawa tool examined the association between the risk score and blood loss, conversion, length of stay and post-operative morbidity.

### Internal and external validity

Each tool's predictive performance is summarised in Table 4. The Halls score included internal validation, using 1/3 of the cohort not used for score development (split-sample) [22], and reported calibration (visual inspection) and discrimination (area under the curve). The Difficulty Scoring System's performance was externally validated in four studies, of which three were independent validations.

The outcomes tested in the validation studies varied and did not match the outcomes that the tool was developed to predict (Table 2). The studies examined the association between risk scores from the tool and blood loss, operative time, conversion, blood transfusion, post-operative morbidity, intra-operative hepatic pedicle clamping and length of stay, using univariate and multivariate regression models. One study assessed the model's calibration with Spearman's rank correlation statistic and visual inspection of scatter plots. One study examined the model's calibration by computing

the Area Under the Receiving Operating Curve (AUROC), which was 0.53.

### Discussion

This systematic review identified 4 tools to estimate the difficulty of LLR, including one subjected to 4 external validations. These scores predicting the difficulty of LLR are limited by methodological flaws and lack of proper validations. This represents a gap in LLR practice.

None of the studies adhered to recommendations for the development and validation of predictive tools (TRIPOD guidelines) [22]. Discrimination and calibration of the tools were not assessed in 3 out of 4 development studies. Development and validation studies used regression analyses to examine the association between score risk categories and outcomes, which is not an appropriate method in predictive score development and validation [22]. When assessed, discrimination was not better than chance along in predicting LLR difficulty (AUROC = 0.53). The lack of information on missing data on predictor variables, number of events and follow-up time were other significant issues. Finally, small sample sizes of some development studies combined with a high number of predictor variables (5 factors for 90 patients

**Table 4** Details of evaluations of tools internal and external validity

| Study               | Ban et al. [16]           | Hasegawa et al. [27] | Kawaguchi et al. [28]            | Halls et al. [18] | Im et al. [30]            | Uchida et al. [29]                           | Tanaka et al. [31]        | Pariyasamy et al. [32]    |
|---------------------|---------------------------|----------------------|----------------------------------|-------------------|---------------------------|--|---------------------------|---------------------------|
| Tool studied        | Difficulty Scoring System | Hasegawa model       | Difficulty of LLR Classification | Halls model       | Difficulty Scoring System | Difficulty Scoring System                    | Difficulty Scoring System | Difficulty Scoring System |
| Internal validation |                           |                      |                                  |                   |                           |  |                           |                           |
| Apparent            | NR                        | NR                   | NR                               | NR                | NA (validations)          |  |                           |                           |
| Bootstrapping       | NR                        | NR                   | NR                               | NR                |                           |  |                           |                           |
| Calibration         | NR                        | NR                   | NR                               | Visual            |                           |  |                           |                           |
| Discrimination      | NR                        | NR                   | NR                               | AUC               |                           |  |                           |                           |
| External validation |                           |                      |                                  |                   |                           |  |                           |                           |
| Independent         | NR                        | NR                   | NR                               | NR                | Yes                       | Yes  | No                        | Yes                       |
| Geographic          | NR                        | NR                   | NR                               | NR                | Yes                       | No   | No                        | Yes                       |
| Temporal            | NR                        | NR                   | NR                               | NR                | Yes                       | No   | Yes                       | Yes                       |
| Calibration         | NR                        | NR                   | NR                               | NR                | NR                        | Spearman's rank correlation<br>Scatter plots | NR                        | NR                        |
| Discrimination      | NR                        | NR                   | NR                               | NR                | NR                        | No   | NR                        | AUROC                     |

LLR laparoscopic liver resection, NR not reported, NA not applicable, AUC area under the curve, AUROC area under the received operating characteristics curve

in development of the DSS) could lead to overfitting of the models and inaccurate risk prediction.

Similar to other reviews of prognostication and prediction tools, significant variation was documented in the number and types of predictive factors included in the identified tools [23, 24]. Variation in variable inclusion may be driven by the use of statistical rules, rather than clinical knowledge, to select covariates in the model, or the use of data from sources not designed to develop a predictive tool. Indeed, the processes to select covariates included in the scores were not standardised. Some factors previously recognised by international LLR experts were neither considered nor included, especially patient-level factors [18]. Intra-operative variables were also used, which limits usability in the preoperative setting for surgical planning. Selecting prognostic factors using a data-driven strategy risks over-estimating the predictive ability, since this strategy optimises case mix and statistical relationships specific to the development dataset (overfitting). Models and tools developed that way are less likely to perform well in other populations. The ideal predictive tool would incorporate clinically relevant factors a priori in the model to improve face validity and facilitate subsequent tool adoption in the community.

Furthermore, LLR difficulty is complicated to predict; there is no consensus regarding the definition of LLR, as evident by the variety and inconsistency of anchor outcomes predicted by the existing scores. Some scores attempted to predict more than one outcome. Validation studies did not validate against the outcomes for which the score was developed, thereby introducing bias and confusion about validity and applicability. An ideal tool would use an universal, consistent and objective definition for LLR difficulty as the anchor outcome, which is not currently available.

A predictive tool for the difficulty of LLR that can be used reliably during preoperative surgical planning could help in improving the uptake of LLR. The Morioka International Consensus on LLR outlined the steep learning curve for LLR [13]. Therefore, preoperative assessment of the difficulty of LLR is important in selecting appropriate patients according to a surgeon's skills and experience at each stage of the learning curve. Mentoring from and discussions with more experienced minimally invasive surgeons can improve case selection and incorporation of minimally invasive techniques in one's practice [35]. However, these opportunities may not be readily available depending on practice settings. The availability and acceptability of a well-developed and validated tool may help to more individually and accurately estimate LLR difficulty, thereby assisting a broader group of surgeons in safely transitioning towards LLR. To be useful in practice, a score should be practical, easy to use and present generalisability proved via robust validation studies [22, 26, 36]. There is a significant opportunity to misclassify LLR difficulty with the existing tools, hence potentially not

selecting proper cases for LLR which can expose patients to harm. As many tools were developed and published prior to the circulation and widespread dissemination of the TRIPOD guidelines across the medical community, these limitations may be easily addressed in future research at the time of study design. Finally, a reliable and accurate predictive tool would also improve risk stratification to support the design and conduct of clinical trials focusing on LLR.

This systematic review is limited by the literature available on the difficulty of LLR and the methodological weaknesses of the included studies, most of which are related to lack of adherence to guidelines for predictive tools development and validation [22, 26]. While there is a vast body of literature regarding the results of LLR, the difficulty of LLR has not been defined and few studies have systematically addressed how to estimate it [9]. Studies looking at factors associated with the results of LLR as part of descriptive multivariable analyses were not included as they did not aim to predict the difficulty of LLR, which requires a different methodology. The strengths of this systematic review include a comprehensive, systematic and sensitive search conducted without restriction for language or types of publication, which also considered the grey literature, and was performed in duplicate. The included studies were rigorously and critically appraised following recommendations from the TRIPOD and CHARMS guidelines, which limits the potential for bias in the information captured and conclusions drawn [22, 26]. The current results highlight the limitations of existing tools and the need to systematically develop and validate tools based on a well-defined anchor outcome for difficulty of LLR.

## Conclusion

This systematic review identified limitations to existing tools predicting the difficulty of LLR, mostly due to lack of adherence to best practices for predictive tools development and validation. Therefore, existing tools cannot be used confidently or safely to predict LLR difficulty. Consistent objective clinical outcomes to define LLR difficulty should be established, and better-quality tools developed and validated in a wide array of populations and clinical settings. This could contribute to improve uptake of LLR as well as risk stratification for future trials.

## Compliance with ethical standards

**Disclosures** Julie Hallet has received speaker honoraria from Ipsen Biopharmaceuticals, Canada, and Novartis Oncology, and is the recipient of the Society of Surgery of the Alimentary Tract (SSAT) Career Development Award for Clinical/Outcomes/Education Research. Shiva Jayaraman has received speaker honoraria from Ethicon, Olympus and Baxter Corporation, and advisory board honoraria from Ipsen Biop-

harmaceuticals Canada and Pendopharm. Natalie Coburn holds the Sherif and Mary-Lou Hanna Chair in Surgical Oncology Research. Patrick Pessaux, Kaitlyn Beyfuss, Pablo Serrano, Guillaume Martel, Tullio Piardi and Alyson Mahar have no conflict of interest or financial ties to disclose.

## References

- Mullen JT, Ribero D, Reddy SK, Donadon M, Zorzi D, Gautam S, Abdalla EK, Curley SA, Capussotti L, Clary BM, Vauthey JN (2007) Hepatic insufficiency and mortality in 1,059 noncirrhotic patients undergoing major hepatectomy. *J Am College Surg* 204:854–862
- Jarnagin WR, Gonen M, Fong Y, DeMatteo RP, Ben-Porat L, Little S, Corvera C, Weber S, Blumgart LH (2002) Improvement in perioperative outcome after hepatic resection: analysis of 1,803 consecutive cases over the past decade. *Ann Surg* 236:397–406 (**Discussion 406–7**)
- Sauerland S, Lefering R, Neugebauer EA (2004) Laparoscopic versus open surgery for suspected appendicitis. *Cochrane Database Syst Rev* 4:CD001546
- Leung KL, Kwok SP, Lam SC, Lee JF, Yiu RY, Ng SS, Lai PB, Lau WY (2004) Laparoscopic resection of rectosigmoid carcinoma: prospective randomised trial. *Lancet* 363:1187–1192
- Lujan J, Valero G, Biondo S, Espin E, Parilla P (2013) Laparoscopic versus open surgery for rectal cancer: results of a prospective multicentre analysis of 4,970 patients. *Surg Endosc* 27:295–302
- Ogiso S, Nomi T, Araki K, Conrad C, Hatano E, Uemoto S, Fuks D, Gayet B (2015) Laparoscopy-specific surgical concepts for hepatectomy based on the laparoscopic caudal view: a key to reboot surgeons' minds. *Ann Surg Oncol* 22(Suppl 3):S327–S333
- Nomi T, Fuks D, Kawaguchi Y, Mal F, Nakajima Y, Gayet B (2015) Learning curve for laparoscopic major hepatectomy. *Br J Surg* 102:796–804
- Dagher I, Belli G, Fantini C, Laurent S, Tayar C, Lainas P, Tranchart H, Franco D, Cherqui D (2010) Laparoscopic hepatectomy for hepatocellular carcinoma: a European experience. *J Am Coll Surg* 211:16–23
- Nguyen KT, Gamblin TC, Geller D (2009) World review of laparoscopic liver resection—2,804 patients. *Ann Surg* 250:831–841
- Wakabayashi G (2015) From Louisville to Morioka: where is now MILS? *Updates Surg* 67:101–104
- Goutte N, Bendersky N, Barbier L, Falissard B, Farges O (2017) Laparoscopic left lateral sectionectomy: a population-based study. *HPB* 19:118–125
- Varley PR, Tohme ST, Chidi AP, Goswami J, van der Windt D, Geller DA, Tsung A (2018) Dissemination of minimally invasive liver resection for primary malignancy: reevaluating effectiveness. *Ann Surg Oncol* 25:808–817
- Wakabayashi G, Cherqui D, Geller DA, Buell JF, Kaneko H, Han HS, Asbun H, O'Rourke N, Tanabe M, Koffron AJ, Tsung A, Soubrane O, Machado MA, Gayet B, Troisi RI, Pessaux P, Van Dam RM, Scatton O, Abu Hilal M, Belli G, Kwon CHD, Edwin B, Choi GH, Aldrighetti LA, Cai X, Cleary S, Chen K-H, Schön MR, Sugioka A, Tang C-N, Herman P, Pekolj J, Chen X-P, Dagher I, Jarnagin W, Yamamoto M, Strong R, Jagannath P, Lo CM, Clavien P-A, Kokudo N, Barkun J, Strasberg SM (2015) Recommendations for laparoscopic liver resection: a report from the second international consensus conference held in Morioka. *Ann Surg* 261(4):619–629
- Ban D, Kudo A, Ito H, Mitsunori Y, Matsumura S, Aihara A, Ochiai T, Tanaka S, Tanabe M, Itano O, Kaneko H, Wakabayashi G (2015) The difficulty of laparoscopic liver resection. *Updates Surg* 67(2):123–128. <https://doi.org/10.1007/s13304-015-0302-7>
- Allard M-A, Cunha AS, Gayet B, Adam R, Goere D, Bachellier P, Azoulay D, Ayav A, Navarro F, Pessaux P (2015) Early and long-term oncological outcomes after laparoscopic resection for colorectal liver metastases. *Ann Surg* 262:794–802
- Ban D, Tanabe M, Ito H, Otsuka Y, Nitta H, Abe Y, Hasegawa Y, Katagiri T, Takagi C, Itano O, Kaneko H, Wakabayashi G (2014) A novel difficulty scoring system for laparoscopic liver resection. *J Hepato-Biliary-Pancreat Sci* 21:745–753
- Kawaguchi Y, Fuks D, Kokudo N, Gayet B (2017) Difficulty of laparoscopic liver resection: proposal for a new classification. *Ann Surg* 267(1):13–17
- Halls MC, Cherqui D, Taylor MA, Primrose JN, Hilal MA, Aldrighetti L, Saati Al H, Alseidi A, Aroori S, Belli G, Besselink M, Edwin B, D'Hondt M, Dagher I, Dejong C, Geller D, Hamady Z, Hamoui M, Isaksson B, Ivanecz A, Le Roux G, Lesurtel M, O'Rourke N, Prasad R, Calvo MP, Reddy S, Retollar F, Santoyo J, Soonawalla Z, Soubrane O, Stavrou G, Subar D, Sutcliffe R, Tanis P, Troisi R, Van Dam R, Wakabayashi G, White S (2017) Are the current difficulty scores for laparoscopic liver surgery telling the whole story? An international survey and recommendations for the future. *HPB* 20(3):231–236
- Hasegawa Y, Wakabayashi G, Nitta H, Takahara T, Katagiri H, Umemura A, Makabe K, Sasaki A (2015) A novel model for prediction of pure laparoscopic liver resection surgical difficulty. *Surg Endosc* 31(12):5356–5363
- Geersing G-J, Bouwmeester W, Zuihthoff P, Spijker R, Leeflang M, Moons K (2012) Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. *PLoS ONE* 7:e32844–e32846
- Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS (2014) Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 11:e1001744–e1001712
- Collins GS, Reitsma JB, Altman DG, Moons KGM (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 102:148–158
- Mahar AL, Compton C, Halabi S, Hess KR, Gershenwald JE, Scolyer RA, Groome PA (2016) Critical assessment of clinical prognostic tools in melanoma. *Ann Surg Oncol* 23:2753–2761
- Mahar AL, Compton C, Halabi S, Hess KR, Weiser MR, Groome PA (2017) Personalizing prognosis in colorectal cancer: a systematic review of the quality and nature of clinical prognostic tools for survival outcomes. *J Surg Oncol* 116:969–982
- Mahar AL, Compton C, McShane LM, Halabi S, Asamura H, Rami-Porta R, Groome PA, Molecular Modellers Working Group of American Joint Committee on Cancer (2015) Refining prognosis in lung cancer: a report on the quality and relevance of clinical prognostic tools. *J Thorac Oncol* 10:1576–1589
- Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 162:W1–W73
- Hasegawa Y, Wakabayashi G, Nitta H, Takahara T, Katagiri H, Umemura A, Makabe K, Sasaki A (2017) A novel model for prediction of pure laparoscopic liver resection surgical difficulty. *Surg Endosc* 31:5356–5363
- Kawaguchi Y, Fuks D, Kokudo N, Gayet B (2018) Difficulty of laparoscopic liver resection. *Ann Surg* 267:13–17
- Uchida H, Iwashita Y, Saga K, Takayama H, Watanabe K, Endo Y, Yada K, Ohta M, Inomata M (2016) Clinical utility

- of the difficulty scoring system for predicting surgical time of laparoscopic liver resection. *J Laparoendosc Adv Surg Tech A* 26:702–706
30. Im C, Cho JY, Han HS, Yoon Y-S, Choi Y, Jang JY, Choi H, Jang JS, Kwon SU (2016) Validation of difficulty scoring system for laparoscopic liver resection in patients who underwent laparoscopic left lateral sectionectomy. *Surg Endosc* 31:430–436
  31. Tanaka S, Kubo S, Kanazawa A, Takeda Y, Hirokawa F, Nitta H, Nakajima T, Kaizu T, Kaneko H, Wakabayashi G (2017) Validation of a difficulty scoring system for laparoscopic liver resection: a multicenter analysis by the Endoscopic Liver Surgery Study Group in Japan. *J Am Coll Surg* 225:249–258
  32. Periyasamy M, Cho JY, Ahn S, Han HS, Yoon Y-S, Choi Y, Jang JS, Kwon SU, Kim S, Choi JK, Guro H (2017) Prediction of surgical outcomes of laparoscopic liver resections for hepatocellular carcinoma by defining surgical difficulty. *Surg Endosc* 31:5209–5218
  33. Dindo D, Demartines N, Clavien P-A (2004) Classification of surgical complications. *Ann Surg* 240:205–213
  34. Kazaryan AM, Røsok BI, Edwin B (2013) Morbidity assessment in surgery: refinement proposal based on a concept of perioperative adverse events. *ISRN Surg* 2013:625093–625097
  35. Schlachta CM, Lefebvre KL, Sorsdahl AK, Jayaraman S (2010) Mentoring and telementoring leads to effective incorporation of laparoscopic colon surgery. *Surg Endosc* 24:841–844
  36. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM (2015) A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 68:279–289

## Affiliations

Julie Hallet<sup>1,2,3</sup> · Patrick Pessaux<sup>4,5</sup> · Kaitlyn A. Beyfuss<sup>3</sup> · Shiva Jayaraman<sup>2,6</sup> · Pablo E. Serrano<sup>7</sup> · Guillaume Martel<sup>8</sup> · Natalie G. Coburn<sup>1,2,3</sup> · Tullio Piardi<sup>9</sup> · Alyson L. Mahar<sup>10</sup>

<sup>1</sup> Department of Surgery, Odette Cancer Centre – Sunnybrook Health Sciences Centre, 2075, Bayview Avenue, Toronto, ON M4N 3M5, Canada

<sup>2</sup> Department of Surgery, University of Toronto, Toronto, ON, Canada

<sup>3</sup> Sunnybrook Research Institute, Toronto, ON, Canada

<sup>4</sup> Department of Digestive and Endocrine Surgery, Nouvel Hôpital Civil, Strasbourg, France

<sup>5</sup> Institut Hospitalier Universitaire de Strasbourg, Strasbourg, France

<sup>6</sup> St-Joseph Health Centre, Toronto, ON, Canada

<sup>7</sup> Department of Surgery, McMaster University, Hamilton, ON, Canada

<sup>8</sup> Department of Surgery, The Ottawa Hospital, Ottawa, ON, Canada

<sup>9</sup> Department of Surgery, Centre Hospitalier Universitaire de Reims, Reims, France

<sup>10</sup> Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, USA