Cucchetti Alessandro (Orcid ID: 0000-0001-5269-1964)
Ratti Francesca (Orcid ID: 0000-0002-8108-9064)
Cillo Umberto (Orcid ID: 0000-0002-2310-0245)

Variations in risk-adjusted outcomes following 4318 laparoscopic liver resections

Alessandro Cucchetti[1,2], Luca Adrighetti[3], Francesca Ratti[3], Alessandro Ferrero[4], Alfredo Guglielmi[5], Felice Giuliante[6], Umberto Cillo[7], Vincenzo Mazzaferro[8] Luciano De Carlis[9] and Giorgio Ercolani[1,2], on behalf of the I Go MILS Group*

A.Cucchetti and L.Aldrighetti shared first co-authorship.

[1] Department of Medical and Surgical sciences – DIMEC; Alma Mater Studiorum – University of Bologna, Bologna, Italy. alessandro.cucchett2@unibo.it, giorgio.ercolani2@unibo.it

[2] Morgagni – Pierantoni Hospital, Forlì, Italy

[3] Hepatobiliary Surgery Division, Department of Surgery, IRCCS San Raffaele Hospital, School of Medicine, Milan, Italy. luca.aldrighetti@hsr.it, ratti.francesca@hsr.it

[4] Department of General and Oncological Surgery, Mauriziano Hospital, Turin, Italy. aferrero@mauriziano.it

[5] Department of Hepatobiliary Surgery, G. B. Rossi Hospital, University of Verona, Verona, Italy. alfredo.guglielmi@univr.it

[6] Unit of Hepato-Biliary Surgery, Foundation 'Policlinico Universitario A. Gemelli', Università Cattolica del Sacro Cuore, Rome, Italy. felice.giuliante@policlinicogemelli.it

[7] Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy. cillo@unipd.it

[8] Hepatopancreatobiliary Surgery and Liver Transplantation, Department of Surgery, University of Milan, Milan, Italy. vincenzo.mazzaferro@istitutotumori.mi.it

[9] Department of General Surgery and Transplantation, Niguarda Ca' Granda Hospital, Milan, Italy. luciano.decarlis@ospedaleniguarda.it

* Other collaborators of the I Go MILS Group: Giuseppe Maria Ettorre, Fabrizio di Benedetto, Raffaele Dalla Valle, Salvatore Gruttadauria, Elio Jovine, Ugo Boggi, Leonardo Vincenti, Roberto Santambrogio, Antonio Giuliani, Guido Torzilli, Giuliano Zimmiti, Alberto Brolese, Andrea Belli, Matteo Ravaioli, Antonio Frena, Giorgio Ettore Rossi, Gian Luca Grazi, Fausto Zamboni, Stefano Berti, Fulvio Calise, Marco Massani, Luca Morelli, Marco Filauro, Giuseppe Tisone, Andrea Coratti, Giuseppe Navarra, Raffaele Romito, Graziano Ceccarelli, Giulio Belli, Guido Griseri, Adelmo Antonucci, Pietro Mezzatesta, Luigi Veneroni, Marcello Schiavo, Michele Colledan, Amilcare Parisi, Silvio Guerriero, Marco Spada, Giacomo Batignani, Giovanni Sgroi, Piero Floridi, Luigi Boni, Pietro Maida, Dario Ribero, Giuliano La Barba.

Corresponding Author: Alessandro Cucchetti, MD. Department of Medical and Surgical sciences – DIMEC; Alma Mater Studiorum – University of Bologna, Via Massarenti, 9. 40138 Bologna, Italy. alessandro.cucchett2@unibo.it

Abstract

Background/Purpose: Quality measures in surgery are important to establish appropriate levels of care and to develop improvement strategies. Purpose of this study was to provide risk-adjusted outcome measures after laparoscopic liver resection (LLR).

Methods: data from a prospective, multicenter database involving 4318 patients submitted to LLRs in 41 hospitals from an intention-to-treat approach (2014-2020) were used to analyze heterogeneity (I^2) among centers and to develop a risk-adjustment model on outcome measures through multivariable mixed-effect models to accounts for confounding due to case-mix.

Results: Involved hospitals operated very different patients: the largest heterogeneity was observed for operating in presence of previous abdominal surgery (I^2:79.1%), in cirrhotic patients (I^2:89.3%) suffering from hepatocellular carcinoma (I^2:88.6%), or requiring associated intestinal resections (I^2:82.8%) and in regards to technical complexity (I^2 for the most complex LLRs: 84.1%). These aspects determined substantial or large heterogeneity in overall morbidity (I^2:84.9%), in prolonged in-hospital stay (I^2:86.9%) and in conversion rate (I^2:73.4%). Major complication had medium heterogeneity (I^2:46.5%). The heterogeneity of mortality was null. Risk-adjustment accounted for all this variability and the final risk-standardized conversion rate was 8.9%, overall morbidity was 22.1%, major morbidity was 5.1% and prolonged in-hospital stay was 26.0%. There were no outliers among the 41 participating centres. An online tool was provided.

Conclusions: A benchmark for LLRs including all eligible patients was provided, suggesting that surgeons can act accordingly in the interest of the patient, modifying their approach in relation to different indications and different experience, but finally providing the same quality of cares.

Introduction

Measurements of surgical care outcomes serve many purposes. Performance appraisal is used by providers to establish appropriate levels of care, as well as for the development and monitoring of improvement strategies. Additionally, providing potential consumers with awareness of hospital performance serves to gain confidence in the healthcare system and inform patients about expected outcomes after surgery. However, direct comparison of outcome measures between different centres may discourage some healthcare professionals from treating patients considered to be at high risk, particularly in the case of low-volume activities, due to the need to meet precisely these predetermined outcome levels, ensuing a risk-avoidance perspective *[1]*.

In this context, laparoscopic liver resection (LLR) has evolved considerably in the last decade, expanding on the territory and facing increasingly complex cases *[2-5]*. With the recognized benefits of the minimally invasive approach and widespread knowledge in the population, more and more surgical centres modified their clinical practice to offer it to patients in need. As a result, several hospitals now offer and perform LLR, with different patient selection and technical approaches.

Risk adjustment is an analytical process that starts from the study of the association between potential confounders and desired outcomes, in order to compare homogeneous groups with respect to their "a priori" risk of encountering the studied results *[6,7]*. In other words, it allows to isolate the hospital effect by analysing the case-mix of patients treated in each hospital. For LLRs there are some data on the expected outcomes of operating in the best patient in the best clinical circumstances *[8]*, but this does not represent routine clinical practice *[9-11]*. Previously, some risk adjustment was attempted to extend the estimates of expected outcomes to a larger surgical population, but without solving the problem of comparing various centres with different volumes, attitudes and experience *[10]*.

In the present study we aimed to evaluate the heterogeneity of clinical conditions influencing outcomes after LLR in different surgical centres. This heterogeneity was then used, using a risk-adjustment approach, to isolate and quantify the variance in desired outcomes due to the hospital effect. Finally, a standardized direct comparison between participating centres was provided with the aim of establishing desired expected outcomes and identifying potential health care provider outliers.

Methods

The study population derived from the Italian "I Go MILS" registry *[5,10]*. This is a prospective registry implemented at the end of 2014 collecting data from LLRs from 54 surgical centres in an intention-to-treat approach. Participation in this register does not depend on the number of procedures performed or registered. Its completeness is periodically monitored by an external provider and cases with missing data are reported to the participating centre asking for it to be completed. This register is approved by the Ethics Committee of the promoting centre and shared among the participants (IGOMILS - OSR of 06 March 2014 – available at: https://www.cr-technology.com/igomils/eclinical/website/documents.aspx). All entries had informed consent signed for the use of personal data. the present study fulfilled the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data.

At the time of data extraction (January 31, 2021), the database had 5015 unique entries since November 2014. Cases still in the completion phase were excluded from the analysis. This led to a first selection of 4463 patients with complete data entries. Risk adjustment requires that the outcomes of interest must be present in the study cohort. For this reason, we were forced to exclude centres with a low number of cases where the possible occurrence of the desired outcomes might not be observable. This was achieved by limiting the study population to centres that provided cases $\geq$ 25th percentile (24 cases per centre), with the consequent exclusion of 13 centres for a total of 145 cases, which however represented only 3.2% of the overall cohort. The final study population consisted of 4318 patients undergoing LLR in 41 hospitals with an intent-to-treat approach.

Definitions and outcome measures

The outcome measures considered were conversion rate, morbidity, major morbidity, mortality and prolonged hospital stay. All these outcomes were recorded within 90 days of surgery. Morbidity included all postoperative complications and has been classified according to the Clavien-Dindo classification, with grade III or higher being defined as major complications *[12]*. Prolonged hospital stay was defined based on the 75th percentile of the median values for post-operative hospital stays in each centre *[8]*. The technical complexity of LLR was defined according to the Kawaguchi classification *[13]*: grade I included wedge resection and left lateral sectionectomy; grade II included anterolateral segmentectomy and left hepatectomy; and grade III included posterosuperior segmentectomy, right posterior sectionectomy, right hepatectomy, central hepatectomy and extended left/right hepatectomy.

Statistical analysis

As the first aim of the study was to assess heterogeneity on clinical characteristics, these were reported after pooling across centre through a random-effect model *[14]*. This approach returned appropriate weighted values and I^2 values, a measure of heterogeneity which was interpreted according to Higgins as follows: <25% = low heterogeneity; 25%-50% = medium, 51%-75% = substantial and >75% = considerable *[15]*. The same was applied for weighted estimation of outcome measures considered.

The potential impact of each clinical variable on the desired outcomes was then investigated through multilevel mixed-effects logistic regression. This analysis is used when some sort of clustering in the data exists (as in the present study where clusters were represented by different hospitals) considering that observations from the same cluster are usually more similar to each other than observations from different clusters. It contains both fixed and random effects. The fixed part models variables as a common logistic regression, whereas the random part introduces intercepts that are different for each cluster considered. The analysis of the variance of the random intercept finally provides how much of it is due to clustering, providing the intraclass correlation coefficient (ICC), that is the heterogeneity of the outcome due to the hospital' effect. This analysis produced expected outcomes for each center accounting for the case-mix. Subsequently, expected and observed outcomes were compared returning the risk-standardized outcome for each center involved. No a-priori level of significance was set *[16]* and single variables were considered for multivariable regressions when their confidence intervals (CI) did not include the 1. Collinearity was checked through Variance Inflation Factor (VIF) evaluation. Analyses were performed using R-Project 3.2.5 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) and STATA (StataCorp. 2017. Release 15. College Station, TX: StataCorp LLC.).

Results

The median number of cases per centre was 62 (range: 24 - 548). The characteristics of the study population are described in *Table 1* where the distributions are weighted among the participating centres. Almost all the parameters considered have substantial (I^2 = 51% -75%) or considerable (I^2> 75%) heterogeneity, arguing that different centres operate on very different patients. The parameter with the lowest heterogeneity is the presence of a previous liver resection (I^2 = 31.4%). In this context, also the outcome measures (*Table 2*), with the exception of mortality (I^2 = 0.0%), are influenced by a great heterogeneity, being the weighted conversion rate of 8.1%, the morbidity of the

21.8%, the greater morbidity of 5.9%, the mortality of 0.7% and the prolonged postoperative hospital stay of 27.1%. Since the heterogeneity of mortality is nil, it was excluded in subsequent regression and risk adjustment models.

Predictors of conversion

Results from multivariable mixed-effects model are shown in *Table 3*. Conversion rate increased in presence of previous liver resection (p<0.001), of being operated for metastases (p<0.001 compared to benign disease), for hepatocellular carcinoma (HCC; p<0.001) or for cholangiocarcinoma (CCC; p<0.001). The need for multiple liver resections also increased the odds of conversion, along with the increase in diameter (p<0.001 for both). Technical difficulty increased conversion rate (p<0.001 for each Kawaguchi grade compared to easier resections). Finally, the need for concomitant bowel resection also increased conversion rate (p<0.001). The model estimated that the variance in the conversion rate was attributable to the operating hospital for the 15.3%.

Predictors of morbidity

As reported in *Table 3*, overall morbidity increased in presence of cirrhosis (p=0.006). Patients operated for benign disease and metastasis shared similar risk, with respect to which HCC and CCC increased the risk (p=0.001 for both). The need for multiple resections also increased morbidity (p<0.001) as well as technical difficulty (p<0.001 for each Kawaguchi grade). Finally, the need for concomitant gastrointestinal resection also increased morbidity (p<0.001). The 10.9% of the residual variance in overall morbidity was attributable to the operating hospital.

Predictors of major complications

Being operated for metastases increased major complications (p=0.017) as well as being operated for HCC (p=0.007) or for CCC (p<0.001) when compared to patients with benign disease (*Table 3*). Major complications also increased for more complex liver resections (p<0.001 for each Kawaguchi grade), and in presence of concomitant gastrointestinal surgery (p<0.001). The residual variance observed in major morbidity was attributable to hospitals for 5.1%.

Predictors of prolonged in-hospital stay

Aging increased the odds of prolonged hospitalization (p<0.001; *Table 3*). Being operated for metastases also increased hospitalization (p=0.016) as well as being operated for HCC (p=0.002) or for CCC (p<0.001). The need for multiple resections also prolonged subsequent hospitalization (p=0.002) as well as the increase of technical difficulty (p=0.006 for Kawaguchi grade II and p<0.001 for the grade III) and in presence of concomitant gastrointestinal surgery (p<0.001). The residual variance in major morbidity was attributable to hospitals for 11.6%

Risk-adjusted analyses

As detailed in *Figure 1*, after risk-adjustment, the standardized conversion rate was 8.9%, the overall morbidity was 22.1%, major morbidity was 5.1% and the prolonged hospital stay was 26.0%, being their 95% CIs related to the size of samples considered. As can be seen in each panel, all standardized risks fall within the 95% CI, supporting the fact that all participating centres can provide similar results when adjusted for the patient's case mix. Details on how data were calculated and a tool for eventual comparison of an additional centre were provided in *Supplementary materials*.

Discussion

The increase in the demand for minimally invasive surgery led laparoscopic liver resection outside large tertiary university hospitals towards smaller volumes distributed throughout the territory. This forms the basis of the considerable heterogeneity of the indications and results observed in the present study. While this may seem like a potential flaw, it represents the opportunity to examine these differences and to correlate them with the outcome of LLRs between different providers.

First, if all participating centres shared similar indications and expertise, the heterogeneities would all be <25%, but this was not observed for any of the parameters considered. The feature with the lowest heterogeneity was the presence of previous hepatic resection ($I^2 = 31.4\%$), suggesting that previous liver surgery is the most resistant concern in electing patients to the laparoscopic approach, due to the higher conversion risk expected *[17]*. On the contrary, all the other characteristics are affected by a substantial and considerable variance. This is likely the consequence of the spread of LLR outside of teaching/high-volume hospitals that kicked off the learning curve in neighbouring centres. The learning curve of LLR has been extensively studied, with several procedures gradually approached from the simplest to the most difficult, with a higher percentage of patients undergoing anatomic resections, and with the inclusion of more patients with cirrhosis in the subsequent experience *[18]*. The remarkable heterogeneity observed for technical complexity and cirrhosis (I^2 > 75.0%), argues that the case-mix presented here is the sum of several learning curves. The next necessary step is to evaluate the impact of this heterogeneity on the desired outcomes.

Predictors of conversion *[17,19]*, morbidity and prolonged hospital stay *[10,20]* were expected. However, the present analysis adds important information with the intra-class correlation coefficient (ICC), which is useful to understand how much of the overall variance of results is the simple consequence of clustering. For conversion, overall morbidity and prolonged hospital stay the

ICC was> 10%, corresponding to the percentage of variance related to the hospital effect, assuming a similar covariates distribution. Thus, along with heterogeneity on indications and learning curve, there is an effect of individual surgeons in obtaining the desired outcomes. This creates additional bias when comparing results from different hospitals if no adjustment is applied. It is also likely that for complications some variance can be related to their different recording, however conversion and prolonged hospital stay are objective measures, and for these two outcomes the estimated variance has to be considered robust.

The graphs in *Figure 1* show two very important results. First, all desired outcomes had values comparable or even lower than those reported from other national databases. The present standardized conversion rate of 8.9% was slightly lower than conversion rate of 13% reported from 20 Dutch centres *[20]*. The overall morbidity was 22.1%, very similar to the Dutch experience of 25%; however, major complication was 5.1%, lower than the 11% reported from the National Clinical Database of Japan *[21]* and of the 10% reported from the Netherlands database *[20]*. Additionally, in all cases mortality was lower than 1% since 2014 onwards (start date of I Go MILS database). These observations provide robust evidences of the goodness of our clinical and surgical approach. Second, all centres are within confidence bands. This means that different indications, different clinical approaches and different skills produced homogeneous end-results. All participating centres, with different volumes of cases provided, are around the average, depending their distance by the size of the volume itself. Consequently, two conclusions can be drawn. First, surgeons in different realities worked conscientiously, recognizing their own limitations in accordance with the relevant guidelines *[22,23]*. Secondly, risk avoidance was present but not aimed at rigid compliance with predefined standards, but in the interest of the patient. The I Go MILS promoted, since its foundation, hands-on training courses, inter-hospitals partnerships with itinerant surgeons, as well as webinars and labs, with the ultimate goal of developing LLR while minimizing the risks for the patients.

Considering all these aspects, risk-adjustment' results seem more suited to provide reference values than the benchmark approach based on the *"best patient in the most expert centre"* [10,11]. For laparoscopic right hepatectomy and left lateral sectionectomy, an attempt was made on behalf of the French surgical association *[8]*. Without attempting to standardize risks among eight experienced French centres, the proposed references based on the *"best patient in the most expert centre"* seem hardly applicable in external validation cohorts *[11]*. In other words, any eventual external validation must strictly meet inclusion criteria (i.e. it should be an expert centre) and distributions of covariates (i.e. age, tumour characteristics and liver characteristics) for a reliable comparison. Consequently, any attempt at comparison outside of this approach would be unreliable. Rossler et al. analysed major hepatectomies collected from living donors for liver transplantation to measure and define the best

achievable results *[25]*. Liver donors are ideal candidates because they are young, without comorbidities or liver disease, so that outside of living donation programs, it would be virtually impossible to meet these prerequisites and produce reliable external comparisons *[11]*. Any attempt without criticism would lead to the avoidance of the risk in the sole interest of complying with this type of benchmark. The final step of the present analysis tried to give an answer to this latter aspect, through provision of a tool for external risk-standardized calculation of centres' outcomes.

An example is introduced in the attached spreadsheet (*Supplementary Materials*). Suppose a new centre wants to compare its conversion rate in its initial 100 LLR experience. This new centre operated only benign lesions located in easily accessible liver segments, such as left lateral segmentectomy. This centre converted 7 patients, with an observed conversion rate of 7.0%. This may seem like a good result since the conversion rate was below the weighted present value of 8.1% of *Table 2*, and in line with the French surgical association which suggested a benchmark of 7.2% *[8]*. However, this centre operated the best clinical situation, and assuming that the other covariates were as in *Figure 2*, the number of expected conversions would be 2, returning a standardized conversion risk of 29.3% with a lower confidence band of 11.8%. As the latter is higher than the standardized conversion rate of 8.9%, it can be concluded that this new centre did not perform optimally.

The present study suffers from at least two technical limitation. First, at the time of writing operative difficulty was assessed in the I Go MILS database through the Kawaguchi classification, whereas more comprehensive scores, such as Iwate criteria *[26,27]*, would have provided more granular data. This availability could have increased the reliability of the risk-adjusted model. Unfortunately, while some centres already have included this detail in the database, the exclusion of missing cases from the analysis would have returned a considerable decrease of the sample size, thus, reducing the robustness of the analysis itself. It is advisable for future studies to complete this classification in our database on which we have been working for years. However, it should be noted that available evidence did not prove superior ability of Iwate criteria in the prediction of post-operative outcomes *[28,29]*. Second, the model included cases from hospitals located in a single country, thus, present risk-adjusted outcomes may not be applicable to foreign centres. However, the present model served to compare performances between different providers, rather than produce a predictive model, but it would be desirable for future risk-adjusted analyses to include different countries as potential modifiers of the outcome. Another limitation is that the evaluation of the outcome is restricted to safety measures only, without efficacy analysis, and this is of importance for malignant lesions. However, the introduction of efficacy measures needs to be focused on the indication to the LLRs. For benign lesions, the quality of life should be a valuable measure of efficacy,

for HCC this should be the margin status, for CCC both the margin status and the number of lymph nodes retrieved and, finally, for metastases the efficacy would be further complex because it would require a balance between parenchymal and vascular margins and response to chemotherapy. As a result, it was virtually impossible to summarize all these different aspects in one measure of effectiveness. Further stratified analyses are therefore needed to address this problem and despite an attempt has been provided in the literature *[30]* this unfortunately lacks in-hospital analyses and modelling.

In conclusion, we have shown here in a national cohort, that surgeons can act according to guidelines in the interest of the patient, modifying their approach in relation to different indications and different experiences, but ultimately providing the same quality of care. The calculator provided for risk-standardization outcomes can be a valuable tool for monitoring LLR performance and for taking timely health measures to correct any eventual underperformance.

References:

1. Ravaioli M, Grande G, Di Gioia P, et al. Risk Avoidance and Liver Transplantation: A Single-center Experience in a National Network. Ann Surg. 2016;264(5):778-786.

2. Soubrane O, Eguchi S, Uemoto S, et al. Minimally Invasive Donor Hepatectomy for Adult Living Donor Liver Transplantation: An International, Multi-Institutional Evaluation of Safety, Efficacy and Early Outcomes. Ann Surg. 2020 Mar 27. doi: 10.1097/SLA.0000000000003852. Online ahead of print.

3. Nguyen KT, Gamblin TC, Geller DA. World review of laparoscopic liver resection-2,804 patients. Ann Surg. 2009;250(5):831-41.

4. Cherqui D. Evolution of laparoscopic liver resection. Br J Surg. 2016;103(11):1405-7.

5. Aldrighetti L, Ratti F, Cillo U, et al. Diffusion, outcomes and implementation of minimally invasive liver surgery: a snapshot from the I Go MILS (Italian Group of Minimally Invasive Liver Surgery) Registry. Updates Surg. 2017;69(3):271-283.

6. Iezzoni LI. The risks of risk adjustment. JAMA. 1997;278(19):1600-7.

7. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. J Am Coll Surg. 2013;217(2):336-46.e1.

8. Hobeika C, Fuks D, Cauchy F, et al. Benchmark performance of laparoscopic left lateral sectionectomy and right hepatectomy in expert centers. J Hepatol. 2020;73(5):1100-1108.

9. Russolillo N, Aldrighetti L, Guglielmi A, Giuliante F, Ferrero A. Correspondence on "Benchmark performance of laparoscopic left lateral sectionectomy and right hepatectomy in expert centers". J Hepatol. 2021;74(4):985-986.

10. Russolillo N, Aldrighetti L, Cillo U,et al. Risk-adjusted benchmarks in laparoscopic liver surgery in a national cohort. Br J Surg. 2020;107(7):845-853.

11. Ercolani G, D'Acapito F, Solaini L, et al. Benchmarking a new tertiary referral center for hepato-biliary surgery through a critical systematic review of available literature. Int J Surg. 2020;84:78-84.

12. Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250(2):187-96.

13. Kawaguchi Y, Fuks D, Kokudo N, Gayet B. Difficulty of Laparoscopic Liver Resection: Proposal for a New Classification. Ann Surg. 2018 Jan;267(1):13-17.

14. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.

15. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.

16. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305-307.

17. Cipriani F, Ratti F, Fiorentini G, Catena M, Paganelli M, Aldrighetti L. Effect of Previous Abdominal Surgery on Laparoscopic Liver Resection: Analysis of Feasibility and Risk Factors for Conversion. J Laparoendosc Adv Surg Tech A. 2018;28(7):785-791.

18. Berardi G, Aghayan D, Fretland ÅA, et al. Multicentre analysis of the learning curve for laparoscopic liver resection of the posterosuperior segments. Br J Surg. 2019;106(11):1512-1522.

19. Troisi RI, Montalti R, Van Limmen JG, et al. Risk factors and management of conversions to an open approach in laparoscopic liver resection: analysis of 265 consecutive cases. HPB (Oxford). 2014;16(1):75-82.

20. Halls MC, Cipriani F, Berardi G, et al. Conversion for Unfavorable Intraoperative Events Results in Significantly Worse Outcomes During Laparoscopic Liver Resection: Lessons Learned From a Multicenter Review of 2861 Cases. Ann Surg. 2018;268(6):1051-1057.

21. van der Poel MJ, Fichtinger RS, Bemelmans M, et al. Implementation and outcome of minor and major minimally invasive liver surgery in the Netherlands. HPB (Oxford). 2019 Dec;21(12):1734-1743.

22. Ban D, Tanabe M, Kumamaru H, et al. Safe Dissemination of Laparoscopic Liver Resection in 27,146 Cases Between 2011 and 2017 From the National Clinical Database of Japan. Ann Surg. 2021 Dec 1;274(6):1043-1050.

23. Wakabayashi G, Cherqui D, Geller DA, et al. Recommendations for laparoscopic liver resection: a report from the second international consensus conference held in Morioka. Ann Surg. 2015;261(4):619-29.

24. Abu Hilal M, Aldrighetti L, Dagher I, et al. The Southampton Consensus Guidelines for Laparoscopic Liver Surgery: From Indication to Implementation. Ann Surg. 2018;268(1):11-18.

25. Rössler F, Sapisochin G, Song G, et al. Defining Benchmarks for Major Liver Surgery: A multicenter Analysis of 5202 Living Liver Donors. Ann Surg. 2016;264(3):492-500.

26. Wakabayashi G. What has changed after the Morioka consensus conference 2014 on laparoscopic liver resection? Hepatobiliary Surg Nutr. 2016 Aug;5(4):281-9.

27. Ban D, Tanabe M, Ito H, et al. A novel difficulty scoring system for laparoscopic liver resection. J Hepatobiliary Pancreat Sci. 2014 Oct;21(10):745-53.

28. Goh BKP, Prieto M, Syn N, et al. Validation and comparison of the Iwate, IMM, Southampton and Hasegawa difficulty scoring systems for primary laparoscopic hepatectomies. HPB (Oxford). 2021 May;23(5):770-776.

29. Kawaguchi Y, Tanaka S, Fuks D, et al. Validation and performance of three-level procedure-based classification for laparoscopic liver resection. Surg Endosc. 2020 May;34(5):2056-2066.

30. Bagante F, Ruzzenente A, Beal EW, et al. Complications after liver surgery: a benchmark analysis. HPB (Oxford). 2019;21(9):1139-1149.

LEGEND TO FIGURE

Figure 1. Risk-standardized conversion rate, overall morbidity, major morbidity and prolonged in-hospital stay prevalence. As noticeable, all participating centers fall within 95% confidence intervals, supporting that despite different patients and interventions, all surgeons can accordingly act to produce similar safety outcomes.

Figure 2. Exemplification of how the risk standardization model works, and how an additional center can compare its performance against that of the I Go MILS. An excel spreadsheet for calculation is provided as *Supplementary Materials.*

Table 1. Baseline weighted characteristics of the 4318 patients submitted to laparoscopic liver resection at 41 Italian surgical centres between November 2014 and January 2021.

Variable	Weighted values (95%C.I.)	Heterogeneity (I^2)*
Age (years)	64.6 (64.0 – 65.4)	59.1%
Male	58.2% (55.6 – 60.7)	52.3%
BMI (Kg/m^2)	25.6 (25.4 – 25.8)	58.0%
Previous abdominal surgery	50.7% (46.5 – 54.9)	79.1%
Hepatic resection	14.8% (13.3 – 16.4)	31.4%
Gastrointestinal	28.0% (24.0 – 32.3)	78.3%
Cirrhosis	25.8% (21.0 – 31.3)	89.3%
Diagnosis		
Hepatocellular carcinoma	37.0% (31.8 – 42.6)	88.6%
Metastases	30.9% (26.0 – 36.2)	89.9%
Benign	20.5% (18.0 – 23.3)	76.9%
Cholangiocarcinoma	7.2% (5.9 – 8.7)	53.8%
Multiple lesions	12.5% (10.5 – 14.8)	73.5%
Diameter of the largest (cm)	4.1 (3.9 – 4.2)	70.9%
≥3cm	57.8% (55.2 – 60.4)	60.3%
Technical complexity †		
<u>Grade I</u>	70.9% (65.6 – 75.7)	89.6%
<u>Grade II</u>	16.7% (14.2 – 19.6)	74.7%
<u>Grade III</u>	12.0% (9.5 – 15.2)	84.1%
Associated intestinal resection	15.9% (13.0 – 19.3)	82.8%

Values are weighted proportions or weighted means estimated through random effect model. Continuity correction of 0.5 in studies with zero cell frequencies was adopted.

* I^2 statistic can be interpreted as follows: values of <25% =low heterogeneity; 25% - 50% = medium, 51% - 75% = substantial and >75% = considerable heterogeneity.

† Based on Kawaguchi classification as follows: <u>Grade I</u> = wedge resection or left lateral sectionectomy; <u>Grade II</u> = anterolateral segmentectomies or left hepatectomy; <u>Grade III</u> = posterosuperior segmentectomy, right posterior sectionectomy, right hepatectomy, central hepatectomy or extended left/right hepatectomy.

Table 2. Outcome measures of the 4318 patients submitted to laparoscopic liver resection at 41 Italian surgical centres between November 2014 and January 2021.

Outcome	Weighted values (95%C.I.)	Heterogeneity (I²)*
Conversion	8.1% (6.5 – 10.)	73.4%
Any complication	21.8% (18.5 – 25.5)	84.9%
Major complications	5.9% (4.9 – 7.1)	46.5%
Mortality	0.7% (0.5 – 1.1)	0.0%
LOS (days)	6.0 (5.6 – 6.3)	88.5%
>6 days†	27.1% (23.2 – 31.3)	86.9%

Values are weighted proportions estimated through random effect model. Continuity correction of 0.5 in studies with zero cell frequencies was adopted.

* I² statistic can be interpreted as follows: values of <25% =low heterogeneity; 25% - 50% = medium, 51% - 75% = substantial and >75% = considerable heterogeneity.

† Based on the 75th percentile of the median values of each center.

Table 3. Results from multivariable mixed-effect model on outcome measures.

Variable	Conversion OR (95%C.I.)	Any complication OR (95%C.I.)	Major complications OR (95%C.I.)	LOS>6 days OR (95%C.I.)
Age (years)	-	-	-	1.02 (1.01 – 1.03)
Previous hepatic resection	1.78 (1.34 – 2.36)	-	-	-
Cirrhosis	-	1.39 (1.10 – 1.75)	-	-
Diagnosis				
Benign	Ref.	Ref.	Ref.	Ref.
Metastases	2.27 (1.50 – 3.44)	Ref.	1.82 (1.11 – 2.98)	1.37 (1.06 – 1.76)
Hepatocellular carcinoma	2.59 (1.67 – 4.01)	1.47 (1.17 – 1.85)	1.94 (1.20 – 3.15)	1.49 (1.16 – 1.91)
Cholangiocarcinoma	3.32 (2.00 – 5.53)	1.62 (1.22 – 2.13)	3.35 (1.90 – 5.91)	1.83 (1.32 – 2.56)
Multiple resected lesions	2.79 (2.11 – 3.68)	1.51 (1.22 – 1.87)	-	1.41 (1.14 – 1.74)
Diameter of the largest (cm)	1.09 (1.06 – 1.14)	-	-	-
Technical complexity †				
<u>Grade I</u>	Ref.	Ref.	Ref.	Ref.
<u>Grade II</u>	2.04 (1.52 – 2.75)	1.58 (1.29 – 1.93)	Ref.	1.34 (1.09 – 1.64)
<u>Grade III</u>	3.13 (2.35 – 4.18)	2.29 (1.86 – 2.81)	1.80 (1.31 – 2.49)	2.83 (2.31 – 3.47)
Associated intestinal resection	1.83 (1.37 – 2.42)	2.06 (1.70 – 2.51)	2.15 (1.56 – 2.94)	3.46 (2.85 – 4.19)
Constant	0.01 (0.01 – 0.02)	0.13 (0.10 – 0.17)	0.17 (0.06 – 0.47)	0.04 (0.02 – 0.06)
Intraclass correlation coefficient	15.3% (8.8 – 25.2)	10.9% (6.4 – 17.8)	5.1% (1.9 – 12.7)	11.6% (7.2 – 18.2)

Mortality was not estimated due to small number of cases and the absence of heterogeneity among participating center. It was included in "any complications" and "major complications" count. Variables not reported were not related to the outcome measures.

Intraclass correlation coefficient (ICC) estimates the proportion of the total residual variance due to center' effect.

AUC/Slope values for conversion: 0.776/1.08, for any complication: 0.719/1.08, for major morbidity: 0.717/1.14; for prolonged in-hospital stay: 0.752/1.06

JHBP_1141_Figure 1.jpg

A new centre wants to compare its conversion rate in its initial 100 LLR experience. This new centre operated only benign lesions located in easily accessible liver segments, such as left lateral segmentectomy. This centre converted 7 patients, with an observed conversion rate of 7.0%.

Number of patients in the cohort	100
Number of pts. converted	7
Number of pts. with any complication	16
Number of pts. with major complications	3
Number of pts. with LOS>6 days	19

However, in this cohort none of the patients had previous hepatic resection or cirrhosis or concomitant intestinal resection. Of them, 35 patients had multiple resections.

Mean age of the whole cohort	64.6
Number of pts. with previous hepatic resection	0
Number of pts. with cirrhosis	0
Number of pts. operated for HCC	0
Number of pts. operated for CCC	0
Number of pts. operated for metastases	0
Number of pts. with multiple resections	35
Mean diameter of the largest lesion (cm)	4.1
Number of pts. in Kawaguchi Group II	0
Number of pts. in Kawaguchi Group III	0
Number of pts. with concomitant intestinal resection	0

In the I Go MILS, operating this best clinical situation, only 2 conversions would be expected. The new centre had a risk-standardized conversion rate of 29.3% with the reference rate of 8.9% outside 95% confidence limit. This is an under-performance.

CONVERSION	Reference rate: 8.9%
Observed conversion rate	7.0%
Number of expected conversions	2
Risk-standardized conversion rate	29.3%
95% Confidence limits	11.8% - 60.4%

ANY COMPLICATIONS	Reference rate: 22.1%
Observed morbidity	16.0%
Number of expected complications	11
Risk-standardized morbidity	32.9%
95% Confidence limits	18.8% - 53.5%

MAJOR COMPLICATIONS	Reference rate: 5.1%
Observed major morbidity	3.0%
Number of expected major complications	2
Risk-standardized morbidity	7.8%
95% Confidence limits	1.6% - 22.7%

PROLONGED LENGTH OF STAY	Reference rate: 26.0%
Observed prolonged LOS	19.0%
Number of expected prolonged LOS	16
Risk-standardized prolonged LOS	31.5%
Lower 95% Confidence limit	19.0% - 49.2%

JHBP_1141_Figure 2.jpg