https://doi.org/10.1016/j.hpb.2022.02.008

ORIGINAL ARTICLE

Benchmarking postoperative outcomes after open liver surgery for cirrhotic patients with hepatocellular carcinoma in a national cohort

Simone Famularo[1,2,3], Nadia Russolillo[4], Matteo Donadon[1,2], Federica Cipriani[5], Francesco Ardito[6], Pasquale Perri[7], Alessandro Giani[3], Francesca De Stefano[8], Quirino Lai[9], Sarah Molfino[10], Matteo Zanello[11], Maurizio Iaria[12], Giuliano La Barba[13], Enrico Pinotti[14], Paola Germani[15], Simone Conci[16], Cecilia Ferrari[17], Luca Fumagalli[18], Maurizio Romano[19], Adelmo Antonucci[20], Giuseppe Zimmitti[21], Albert Troci[22], Antonio Floridi[23], Valentina Ferraro[24], Stefan Patauner[25], Antonio Frena[25], Riccardo Memeo[24], Michele Crespi[22], Mohammed A. Hilal[21], Giacomo Zanus[19], Marco Chiarelli[18], Andrea Percivale[17], Andrea Ruzzenente[16], Paola Tarchi[15], Mauro Zago[14,18], Giorgio Ercolani[13], Raffaele Dalla Valle[12], Elio Jovine[11], Gian Luca Baiocchi[10], Massimo Rossi[9], Marcello Maestri[8], Fabrizio Romano[3], Gian Luca Grazi[7], Felice Giuliante[6], Luca Aldrighetti[5], Alessandro Ferrero[4], Guido Torzilli[1,2] HE.RC.O.LE.S. Group, HERCOLES GROUP Contributors

[1]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, [2]Department of Hepatobiliary and General Surgery, IRCCS Humanitas Research Hospital, Rozzano, Milan, [3]School of Medicine and Surgery, University of Milano-Bicocca, San Gerardo Hospital, Monza, [4]Department of General and Oncological Surgery, Mauriziano Hospital "Umberto I", Turin, [5]Hepatobiliary Surgery Division, IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, [6]Hepatobiliary Surgery Unit, Fondazione Policlinico Universitario A. Gemelli, IRCCS, Catholic University of the Sacred Heart, [7]Division of Hepatobiliarypancreatic Unit, IRCCS - Regina Elena National Cancer Institute, Rome, [8]Unit of General Surgery 1, University of Pavia and Foundation IRCCS Policlinico San Matteo, Pavia, [9]Hepato-Biliary Surgery and Organ Transplantation Unit, Sapienza University of Rome, Umberto I Polyclinic of Rome, Rome, [10]Department of Clinical and Experimental Sciences, University of Brescia, Brescia, [11]Alma Mater Studiorum, University of Bologna, AOU Sant'Orsola Malpighi, IRCCS at Maggiore Hospital, Bologna, [12]Department of Medicine and Surgery, University of Parma, Parma, [13]General and Oncologic Surgery, Morgagni-Pierantoni Hospital, Forlì, [14]Department of Surgery, Ponte San Pietro Hospital, Bergamo, [15]Surgical Clinics, University Hospital of Trieste, Trieste, [16]Division of General and Hepatobiliary Surgery, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Verona, [17]HPB Surgical Unit, San Paolo Hospital, Savona, [18]Department of Emergency and Robotic Surgery, ASST Lecco, Lecco, [19]Department of Surgical, Oncological and Gastroenterological Science (DISCOG), University of Padua, Hepatobiliary and Pancreatic Surgery Unit - Treviso Hospital, [20]Department of Surgery, Monza Policlinic, Monza, [21]Department of General Surgery, Poliambulanza Foundation Hospital, Brescia, [22]Department of Surgery, L. Sacco Hospital, Milan, [23]Department of General Surgery, ASST Crema, Crema, [24]Department of Hepato-Pancreatic-Biliary Surgery, Miulli Hospital, Bari, and [25]Department of General and Pediatric Surgery, Bolzano Central Hospital, Bolzano, Italy

Abstract

Background: Benchmark analysis for open liver surgery for cirrhotic patients with hepatocellular carcinoma (HCC) is still undefined.

Methods: Patients were identified from the Italian national registry HE.RC.O.LE.S. The Achievable Benchmark of Care (ABC) method was employed to identify the benchmarks. The outcomes assessed were the rate of complications, major comorbidities, post-operative ascites (POA), post-hepatectomy liver failure (PHLF), 90-day mortality. Benchmarking was stratified for surgical complexity (CP1, CP2 and CP3).

Results: A total of 978 of 2698 patients fulfilled the inclusion criteria. 431 (44.1%) patients were treated with CP1 procedures, 239 (24.4%) with CP2 and 308 (31.5%) with CP3 procedures. Patients submitted to CP1 had a worse underlying liver function, while the tumor burden was more severe in CP3 cases. The ABC for complications (13.1%, 19.2% and 28.1% for CP1, CP2 and CP3 respectively), major complications (7.6%, 11.1%, 12.5%) and 90-day mortality (0%, 3.3%, 3.6%) increased with the surgical difficulty, but not POA (4.4%, 3.3% and 2.6% respectively) and PHLF (0% for all groups).

Conclusion: We propose benchmarks for open liver resections in HCC cirrhotic patients, stratified for surgical complexity. The difference between the benchmark values and the results obtained during everyday practice reflects the room for potential growth, with the aim to encourage constant improvement among liver surgeons.

Received 19 November 2021; accepted 21 February 2022

Correspondence Matteo Donadon, Department of Biomedical Sciences, Humanitas University, Via Montalcini 4, 20090 Pieve Emanuele, Milan, Italy. E-mails: matteo.donadon@hunimed.eu, , matteo.donadon@gmail.com

Introduction

In the last years, with surgical and perioperative care becoming safer, the focus on outcomes' improvement has shifted from mortality to complications' reduction and regaining a good quality of life.[1,2] Healthcare facilities and systems are increasingly using quality measures to assess and improve the quality of care. In this scenery, the surgical literature has recently introduced the concept of benchmarking as a quality assessment and improvement tool. This novel methodology has been applied to pancreatic[3,4] and oesophageal resections,[5] and also to liver surgery.[6–10] Benchmarking defines a process that defines the best achievable results by remarkable performers. Specifically in surgery, benchmarks represent the best attainable outcomes after a procedure and can be used as a point of reference for assessing and comparing performances for hospitals, departments, and individuals[3,5,11,12]. In this particular field, the ultimate goal of establishing benchmarks remains a systematic approach for determining referral standards to empower self-assessment and ease detection of areas for improvement.[12] Benchmarking should not be considered as a "tournament" among physicians, but a quantitative stimulus for health-systems in the identification of the critical issues where to intervene, and, ultimately, to accomplish excellence.

The present study was thought among the Italian hepatocellular carcinoma surgical study group (He.Rc.O.Le.S. Group), as the first attempt to measure the obtained surgical results and to create shared reference points to define the potential and points for improvement in the surgical treatment of a particular surgical setting: cirrhotic patients undergoing open liver resection (OLR) for hepatocellular carcinoma (HCC). Although the spread of laparoscopy in liver surgery, the open approach remains the most employed, at least in Italy.[13] Benchmarks were obtained using the Achievable Benchmark of Care (ABC) method.[14]

Methods

Study overview, patient selection, and study design

This retrospective study evaluated prospectively-collected data from patients enrolled between 2008 and 2019 in the Italian Register of HCC, promoted by the Hepatocarcinoma Recurrence on the Liver Study Group (He.Rc.O.Le.S. Group, clinicaltrial.gov registration number: NCT04053231, www.hercolesgroup.eu).[13]

Results are reported according to principles of Strengthening the Reporting of Observational Studies in Epidemiology (STROBE).[15] All consecutive adult patients (age ≥18 years) with histologically proven HCC who underwent surgery from January 2008 to December 2019 were evaluated. Inclusion criteria were: 1) patients who underwent surgery for the first diagnosis of HCC, 2) HCC confirmed with the histological specimen, 3) cirrhotic disease of the liver in the previous past, 4) being treated by OLR. Exclusion criteria were: 1) histological diagnosis of the combined primary liver tumor (e.g. hepato-cholangiocellular carcinoma), 2) missing data on the type of surgical procedure executed, on postoperative course or on ninety-day mortality. All the included patients were followed-up for 90 days after liver resection by in-office visits.

Aim and study endpoints

The aim of this study was to assess the best achievable outcomes in OLR for cirrhotic patients with HCC after risk adjustment based on technical difficulty, using a national registry. The endpoints were the rate of complications, major comorbidities, post-operative ascites (POA) occurrence, post-hepatectomy liver failure (PHLF) and 90-day mortality. Secondary endpoints were the rate of R1 resections and the length of stay.

Definitions

The presence of cirrhosis was evaluated by hepatologists through clinical, biochemical, radiological and histological information. Comorbidities were recorded by using the Charlson Comorbidity index.[16] The type of liver resection was defined based on the Brisbane nomenclature,[17] while the complexity of the procedures was staged according to the proposal of Kawaguchi et al.[18] as follows: grade 1 complexity (CP1) includes wedge resections and left lateral sectionectomy; grade 2 complexity (CP2) includes anterolateral segmentectomy and left hepatectomy; grade 3 complexity (CP3) includes posterosuperior segmentectomy, right hepatectomy, extended right hepatectomy, right posterior sectionectomy, central hepatectomy, extended left hepatectomy and right or left trisectionectomy. The volume of liver surgical procedures executed per year in each participating center has been recorded, and centers were divided between low (<50 procedures/year), medium (50–100 procedures/year) and high volume (>100 procedures/year) centres.[19] Of note, the centres' volume was estimated on the total number of liver procedures regardless of the specific pathology indication.

PHLF was defined according to the 50-50 criteria.[20] Postoperative complications were recorded using the Clavien-Dindo classification,[21] and major complications were established in the case of Clavien-Dindo grades 3 or 4. Comprehensive Complication Index (CCI)[22,23] has been measured locally by each center. POA was defined as the presence of 500 ml/24 h of ascites in the surgical drainage(s) (or the evidence in the post-operative course of intra-abdominal liquid by ultrasound in case of no drainage) for at least three consecutive days.[24–26]

Benchmark analysis

The Achievable Benchmark of Care (ABC™, University of Alabama, Birmingham, Alabama, USA) method[14] was employed to identify the benchmarks. Briefly, this method relies on the identification of the benchmark as the performance achieved by the top 10 percent of providers adjusted for the number of patients per provider. This approach reduces the disproportionate impact of procedures with a small denominator without eliminating them from the sample, and it has been already applied in different clinical, surgical and health-care settings.[7,27–29] In brief, these are the steps for this approach: 1. Rank order providers (e.g. hospitals) in descending order of performance relative to process indicator. 2. Beginning with the best-performing provider, add providers sequentially in descending order until this subset of providers represents at least 10% of all patients or subjects in the entire dataset. 3. Calculate benchmark based on subset as follows: total number of patients in subset experiencing the event of interest/total number of patients in subset.

The procedure was applied for each center within each group of surgical complexity. An adjusted performance fraction (APF) was calculated by adding 1 to the number of events (numerator), and adding 2 to the number of available procedures (denominator), then dividing the two obtained adjusted values[14]: this was done to adjust the impact of low cases centers. The APF was then used to sort the centers from lowest to highest and to create a ranking among centres. Using this ranking, the benchmark setters were identified by summing the number of procedures for each center, starting with the highest performer (with the lowest APF), and successively including the next highest performer, until the sum of the procedures among these highest performers comprised at least 10 percent of the number of procedures for the cluster. The centers that were included with their number in this 10 percent list were considered the benchmark setters. The ABC was the sum of all the adverse outcomes (e.g. complications, severe complications, etc.) of the benchmark setters (numerator) divided by the sum of the number of procedures of the benchmark setters (denominator). For this purpose, centers with zero complications or zero major comorbidities were excluded by the benchmarks' assessment for those cluster-analyses, since the absence of events was considered unrealistic, as suggested by Russolillo et al.[7] The benchmark value for the length of stay was defined as the 10th percentile of the median value across centres. After benchmarks identification, all the centers included in the register, regardless of their contribution to the benchmark selection, were compared.

Statistical analysis

The sample description was done using median and interquartile range (IQR) for numeric variables and number and proportion for categorical variables. Mann–Whitney, Kruskall-Wallis and Fisher tests were used to compare baseline patients' characteristics when appropriate. Pairwise post-hoc analyses were performed after Holm-Bonferroni correction. For the purpose of the study and for the methodology applied, patients with missing data were excluded. All statistical tests were two tails and a 5% significance level. All the analyses and graphics were computed by using the open-source R software (v4.0.3) and Office-Excel (Microsoft Office 2016, Microsoft Corp.).

Results

Between 2008 and 2019, 2698 patients were enrolled in the HE.RC.O.LE.S. register. At the time of analysis, 23 centers had data fully enrolled in the dataset. Of them, 186 (6.8%) were excluded because of missing data, while 1534 were excluded because they were operated by the laparoscopic approach or because of absence of cirrhosis at the final histology. The final sample of the study included 978 HCC patients. The number of patients per center ranged from 1 to 289 (Fig. 1): 5 centers had more than 100 procedures related to HCC in cirrhotic patients, while 2 of them enrolled more than 250 cases. Of 23 centers participating in the study, 5 (20.9%) were classified as high volume centers, 11 (45.9%) as medium volume and 7 (29.2%) as low volume centers. The considered cohort was composed of 236 (24.1%) female patients, with a median age of 70 years (IQR 63.0–75.0), and a median Charlson Comorbidity Index of 6 (IQR 5–8). Eight-hundred eighty-one (90.1%) cases were Child-Pugh A; median MELD score was 7 (IQR 7–9). Median number of nodules was 1 (IQR 1–2), and the median size was 3.45 cm (IQR 2.4–5.0). According to the BCLC staging system, 539 (55.1%) patients were BCLC 0-A, while 199 (20.3%) were BCLC B, and 148 (15.1%) were BCLC C. Baseline characteristics are summarized in Table 1.

Regarding the complexity of the surgical procedure executed, 431 (44.1%) patients were treated with grade 1 procedures, 239 (24.4%) with grade 2 and 308 (31.5%) with grade 3 procedures. Overall, wedge resections were the most common procedure (393 cases, 40.2%), followed by segmentectomy (319 cases, 32.6%) and left hepatectomy (76, 7.8%). The median length of surgery was 235 min (IQR 180–295), and median blood loss was 350 ml (IQR 200–550), but blood transfusion was required only in 118 (12.1%) patients. The median duration of hospital recovery was 9 days (IQR 7–13).

Low-volume centers performed 142 (14.5%) liver resections: 54 (38.1%) CP1, 52 (36.6%) CP2 and 36 (25.3%) CP3 procedures. Medium-volume centers executed 424 (43.3%) resections:

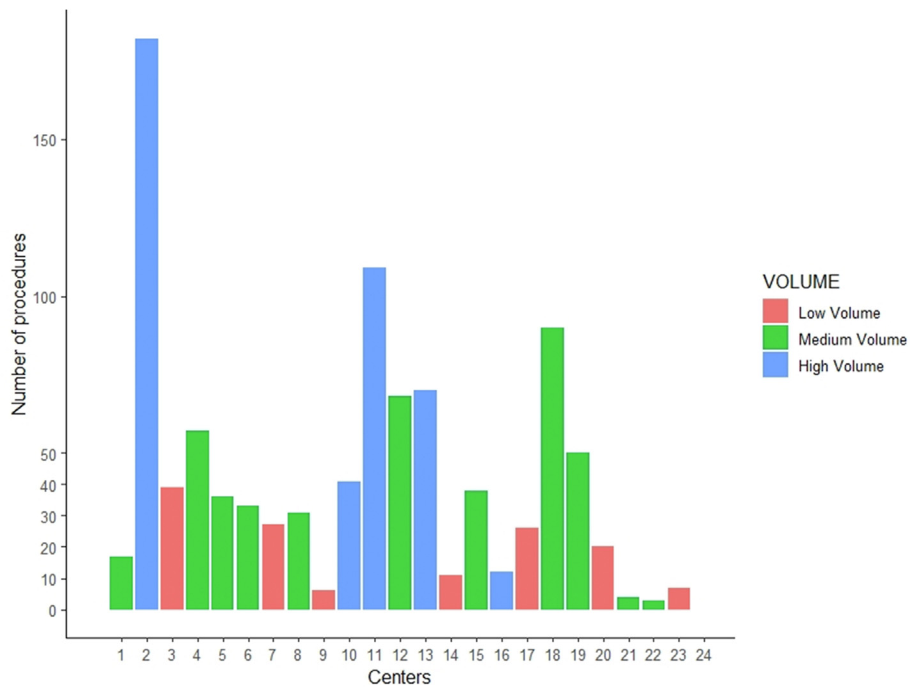

Figure 1 The distribution of the surgical procedures among participating centers. Those are subgrouped by the volume classification in low, medium and high volume centers (centres volume was defined as the number of liver resections executed per year for any medical reason; in the y-axis is represented the number of patients included in the present study per each centre previously classified)

204 (48.1%) CP1, 91 (21.5%) CP2, 129 (30.4%) CP3 procedures. High volume centers, indeed, performed 412 (42.2%) liver resections: 173 CP1, 96 CP2 and 143 CP3 procedures.

Complexity grades' comparison

Patients in CP3 had a lower burden of comorbidities when compared with CP1, as represented by the median Charlson Index which was 6 (IQR 5–7) versus 7 (IQR 5–8) respectively (p: 0.033). Seventy-six (17.6%) patients had steatosis in CP1, versus 34 (14.2%) in CP2 (p:0.006). CP1 cases had a higher median MELD (8, IQR 7–9) when compared with CP2 (7, IQR 6–8, p:0.001) and CP3 (7, IQR 7–9, p:0.048). Patients with BCLC B (76, 24.7%) and BCLC C (70, 22.7%) underwent more frequently CP3 procedures ($p < 0.001$). Median blood loss was higher in CP3 (450 ml, IQR 200–652.5) than in CP1 (265 ml, IQR 100–450, $p < 0.001$) and in CP2 (400 ml, IQR 200–600) versus CP1 ($p < 0.001$). Further data are available in Table 1.

Post-operative complications and mortality

Complications occurred more frequently among CP3 procedures (154, 50%) than in CP2 (87, 36.4%, p:0.002), and in CP2 than in CP1 (191, 44.3%, p:0.056). Median CCI was higher in CP3 (22.60, IQR 20.9–33.5) than in CP1 (20.9, IQR 9.0–26.2, p:0.012) and CP2 (20.9, IQR 9.0–33.5, p:0.001). Major comorbidities were more frequent in CP3 (41, 13.3%) than in CP1(34, 7.9%, p: 0.048) and CP2 (18, 7.5%, p:0.006). POA was significantly more frequent in CP1 (78, 18.1%) than in CP2 (24, 10%, p:0.026), and in CP3 (53, 17.2%) than in CP2 (p: 0.021). Therefore, PHLF occurred in 25 (5.8%) cases in CP1 group, while it was observed in 8 (3.3%) and 26 (8.4%) patients in CP2 and CP3 groups (CP2 vs CP3 p: 0.027). Ninety-day mortality occurred in 10 (2.3%), 9 (3.8%) and 15 (4.9%) patients respectively in CP1, CP2 and CP3 (p:0.169). Results are reported in Table 2.

Benchmarks estimation

The ABC method[14] was applied according to the surgical complexity, and 5 outcomes were evaluated. In the CP1 group, the benchmark for overall complications was 13.16% (25th percentile: 41.9, 75th percentile: 64.3), while for major morbidities was 7.69% (25th percentile: 15.9, 75th percentile: 34.0). The best achievable rate of POA was 4.44% (25th percentile: 10.5, 75th percentile: 41.4), while for PHLF was 0.0% (25th percentile: 6.7, 75th percentile: 19.8%) and for 90-days mortality was 0.0% (25th percentile: 3.1, 75th percentile: 9.1). In CP2, the complications benchmark was 19.28% (25th percentile: 31.7, 75th percentile: 69.1); the major complications one was 11.11% (25th percentile: 15.6, 75th percentile: 41.7), while for POA was 3.33% (25th percentile: 7.7, 75th percentile: 33.3). Benchmarks for PHLF and 90-days mortality were 0.0% (25th percentile: 5.3, 75th percentile: 11.1) and 3.33% (25th percentile: 8.0, 75th

Table 1 Baseline characteristics of the whole cohort and after subgrouping for the three grades of surgical complexity

	Overall	CP1	CP2	CP3	p (CP1vsCP2)	p (CP1vsCP3)	P (CP2vsCP3)
n	**978**	**431**	**239**	**308**			
Age (years) (median [IQR])	70.00 [63.00, 75.00]	70.00 [63.00, 75.00]	72.00 [65.25, 75.75]	70.00 [62.00, 74.00]	0.264	0.264	0.018
Female (%)	236 (24.1)	95 (22.0)	55 (23.0)	86 (27.9)	0.848	0.243	0.458
Charlson Index (median [IQR])	6.00 [5.00, 8.00]	7.00 [5.00, 8.00]	6.00 [5.00, 8.00]	6.00 [5.00, 7.00]	0.13	0.033	0.722
Steathosis (%)	156 (16.0)	76 (17.6)	34 (14.2)	46 (14.9)	0.006	0.352	0.352
NA	82 (8.4)	24 (5.6)	32 (13.4)	26 (8.4)			
Child Pugh Grade (%)							
A	881 (90.1)	402 (93.3)	210 (87.9)	269 (87.3)	0.162	0.162	0.766
B	63 (6.4)	20 (4.6)	17 (7.1)	26 (8.4)			
NA	34 (3.5)	9 (2.1)	12 (5.0)	13 (4.2)			
MELD score (median [IQR])	7.00 [7.00, 9.00]	8.00 [7.00, 9.00]	7.00 [6.00, 8.00]	7.00 [7.00, 9.00]	0.001	0.048	0.122
HBV (%)	200 (20.4)	86 (20.0)	53 (22.2)	61 (19.8)	0.18	0.18	0.781
NA	11 (1.1)	1 (0.2)	4 (1.7)	6 (1.9)			
HCV (%)	553 (56.5)	242 (56.1)	130 (54.4)	181 (58.8)	0.224	0.054	0.472
NA	12 (1.2)	1 (0.2)	4 (1.7)	7 (2.3)			
Alcohol intake (%)	224 (22.9)	103 (23.9)	50 (20.9)	71 (23.1)	0.678	0.954	0.822
NA	21 (2.1)	9 (2.1)	5 (2.1)	7 (2.3)			
Varices (%)	179 (18.3)	94 (21.8)	31 (13.0)	54 (17.5)	0.039	0.24	0.306
NA	156 (16.0)	58 (13.5)	42 (17.6)	56 (18.2)			
Total Bilirubin (median [IQR])	0.82 [0.60, 1.19]	0.84 [0.60, 1.20]	0.80 [0.60, 1.18]	0.80 [0.60, 1.15]	1	1	1
Albumin (mg/dl) (median [IQR])	3.90 [3.50, 4.30]	0.88 [0.75, 1.04]	0.87 [0.73, 1.00]	0.86 [0.74, 1.00]	0.342	0.144	0.696
Platelet (median [IQR])	151.00 [111.00, 205.00]	138.50 [99.00, 188.00]	169.00 [129.00, 224.00]	157.50 [118.75, 220.75]	<0.001	<0.001	0.087
INR (median [IQR])	1.11 [1.04, 1.21]	1.11 [1.05, 1.20]	1.11 [1.02, 1.29]	1.11 [1.05, 1.21]	0.628	0.544	0.999
Alpha-Feto-Protein (median [IQR])	13.00 [4.80, 93.20]	9.31 [4.53, 61.00]	15.80 [5.00, 142.50]	18.00 [5.00, 129.00]	0.116	0.039	0.679
N. of Nodules (median [IQR])	1.00 [1.00, 2.00]	1.00 [1.00, 2.00]	1.00 [1.00, 1.00]	1.00 [1.00, 2.00]	<0.001	0.049	0.098
Size (cm) (median [IQR])	3.45 [2.40, 5.00]	3.00 [2.00, 4.30]	3.60 [2.30, 6.00]	4.00 [2.90, 6.77]	0.002	<0.001	0.005
BCLC (%)					NaN	<0.001	0.016
0	107 (10.9)	66 (15.3)	26 (10.9)	15 (4.9)			
A	432 (44.2)	200 (46.4)	99 (41.4)	133 (43.2)			
B	199 (20.3)	76 (17.6)	47 (19.7)	76 (24.7)			
C	148 (15.1)	32 (7.4)	46 (19.2)	70 (22.7)			
NA	92 (9.4)	57 (13.2)	21 (8.8)	14 (4.5)			
Type of procedure (%)					–	–	–
Wedge Resection	393 (40.2)	393 (91.2)	0 (0.0)	0 (0.0)			
segmentectomy	319 (32.6)	0 (0.0)	163 (68.2)	156 (50.6)			
right hepatectomy	71 (7.3)	0 (0.0)	0 (0.0)	71 (23.1)			
left hepatectomy	76 (7.8)	0 (0.0)	76 (31.8)	0 (0.0)			
right anterior sectionectomy	31 (3.2)	0 (0.0)	0 (0.0)	31 (10.1)			

(*continued on next page*)

Table 1 (*continued*)

	Overall	CP1	CP2	CP3	p (CP1vsCP2)	p (CP1vsCP3)	P (CP2vsCP3)
right posterior sectionectomy	34 (3.5)	0 (0.0)	0 (0.0)	34 (11.0)			
left lateral sectionectomy	38 (3.9)	38 (8.8)	0 (0.0)	0 (0.0)			
right trisectionectomy	15 (1.5)	0 (0.0)	0 (0.0)	15 (4.9)			
left trisectionectomy	1 (0.1)	0 (0.0)	0 (0.0)	1 (0.3)			
Pringle (%)	569 (58.2)	224 (52.0)	141 (59.0)	204 (66.2)	0.006	0.003	0.018
NA	23 (2.4)	7 (1.6)	12 (5.0)	4 (1.3)			
Pringle Time (min) (median [IQR])	40.00 [25.00, 60.00]	35.00 [21.75, 62.25]	40.00 [25.00, 55.00]	45.00 [30.00, 60.00]	0.964	0.243	0.243
Length of Surgery (min) (median [IQR])	235.00 [180.00, 295.00]	205.00 [170.00, 270.00]	232.50 [180.00, 285.00]	265.00 [210.00, 320.00]	0.023	<0.001	<0.001
Blood los (ml) (median [IQR])	350.00 [200.00, 550.00]	265.00 [100.00, 450.00]	400.00 [200.00, 600.00]	450.00 [200.00, 652.50]	<0.001	<0.001	0.110
Intraop Blood transfusion (%)	118 (12.1)	31 (7.2)	28 (11.7)	59 (19.2)	0.070	NaN	0.010
NA	89 (9.1)	40 (9.3)	29 (12.1)	20 (6.5)			
Length of Stay (days) (median [IQR])	9.00 [7.00, 13.00]	9.00 [7.00, 12.00]	9.00 [7.00, 12.00]	11.00 [8.00, 17.00]	0.211	<0.001	<0.001
R (%)					0.078	<0.001	0.312
R0	757 (77.4)	327 (75.9)	185 (77.4)	245 (79.5)			
R1	100 (10.2)	30 (7.0)	28 (11.7)	42 (13.6)			
R2	13 (1.3)	6 (1.4)	3 (1.3)	4 (1.3)			
NA	108 (11.0)	68 (15.8)	23 (9.6)	17 (5.5)			
Procedures in Low Volume Centers (%)	142 (14.5)	54 (12.5)	52 (21.7)	36 (11.7)	–	–	–
Procedures in Medium Volume C.(%)	424 (43.3)	204 (47.3)	91 (38.1)	129 (41.9)	–	–	–
Procedures in High Volume C.(%)	412 (42.2)	173 (40.2)	96 (40.2)	143 (46.4)	–	–	–

MELD model for end stage liver disease; HBV hepatitis B virus; HCV hepatitis C virus; INR International normalized ratio; BCLC Barcelona Clinic Liver Cancer staging system.

Table 2 Post-operative outcomes in the whole cohort and after subgrouping by surgical complexity grades

	Overall	CP1	CP2	CP3	p (CP1vsCP2)	p (CP1vsCP3)	p (CP2vsCP3)
n	978	431	239	308			
Complications (%)	432 (44.2)	191 (44.3)	87 (36.4)	154 (50.0)	0.056	0.146	0.002
Major Complications (%)	93 (9.5)	34 (7.9)	18 (7.5)	41 (13.3)	0.118	0.048	0.006
CCI (median [IQR])	20.90 [20.45–29.60]	20.90 [9.00, 26.20]	20.90 [9.00–33.50]	22.60 [20.90–33.55]	0.186	0.012	0.001
POA (%)	151 (15.4)	78 (18.1)	24 (10.0)	53 (17.2)	0.026	0.636	0.021
PHLF (%)	59 (6.0)	26 (6.0)	9 (3.7)	26 (8.4)	0.343	0.267	0.027
90 Days mortality (%)	34 (3.5)	10 (2.3)	9 (3.8)	15 (4.9)	0.403	0.092	0.678

CCI- Comprehensive Complication Index; POA-post-operative ascites; PHLF- post-hepatectomy liver failure.

percentile: 20.8). In case of CP3 procedures, benchmarks were: overall complication 28.05% (25th percentile: 47.5, 75th percentile: 70.0); major complications 12.50% (25th percentile: 17.5, 75th percentile: 51.8); POA 2.60% (25th percentile: 16.7, 75th percentile: 36.7); PHLF 0.0% (25th percentile: 8.5, 75th percentile: 16.3); 90-days mortality 3.57% (25th percentile: 5.8, 75th percentile: 18.3). Benchmarks, percentiles, and graphical representations are available in Fig. 2a–e. Center-specific results are indeed depicted in Fig. 3a–e, which is also the official report image that is sent to all the participating centers on an annual basis.

Regarding R1 resections, the benchmark value was 0% (25th percentile: 0, 75th percentile: 6.50), 0.02% (25th percentile: 0.1, 75th percentile: 15) and 0.03% (25th percentile (25th percentile: 0.2, 75th percentile: 17.7%) for CP1, CP2 and CP3 procedures respectively.

Considering the length of stay, the benchmarks values were 6 days (25th percentile: 7, 75th percentile: 10) for CP1, 6 days (25th percentile: 7, 75th percentile: 12) for CP2 and 7 days (25th percentile: 8, 75th percentile: 17) for CP3 procedures.

Discussion

The ultimate goal of establishing benchmarks remains the improvement of patients' outcomes, through stimulating surgeons to accomplish excellence. A practical application of this study was to generate a virtuous routine, in which each year every center receives a summary (as the one depicted in Fig. 3) with their results compared with the other members.

Despite the recent spread of the laparoscopic approach, open liver resections are still the more frequent type of curative procedure for HCC,[13] even in case of cirrhosis which is the most common presentation. Albeit benchmarking of specific technical procedures (laparoscopy,[7] major hepatectomies,[6] liver transplantation[8]) were already established, a whole benchmark analysis for all the possible surgical resections performed in the setting of cirrhotic HCC is still lacking, and the best achievable outcomes are globally unknown, particularly regarding the specific complications of those procedures, as per POA and PHLF, which are pivotal to be avoided in the setting of a fragile hepatic functional balance. Our study addresses this need, proposing the

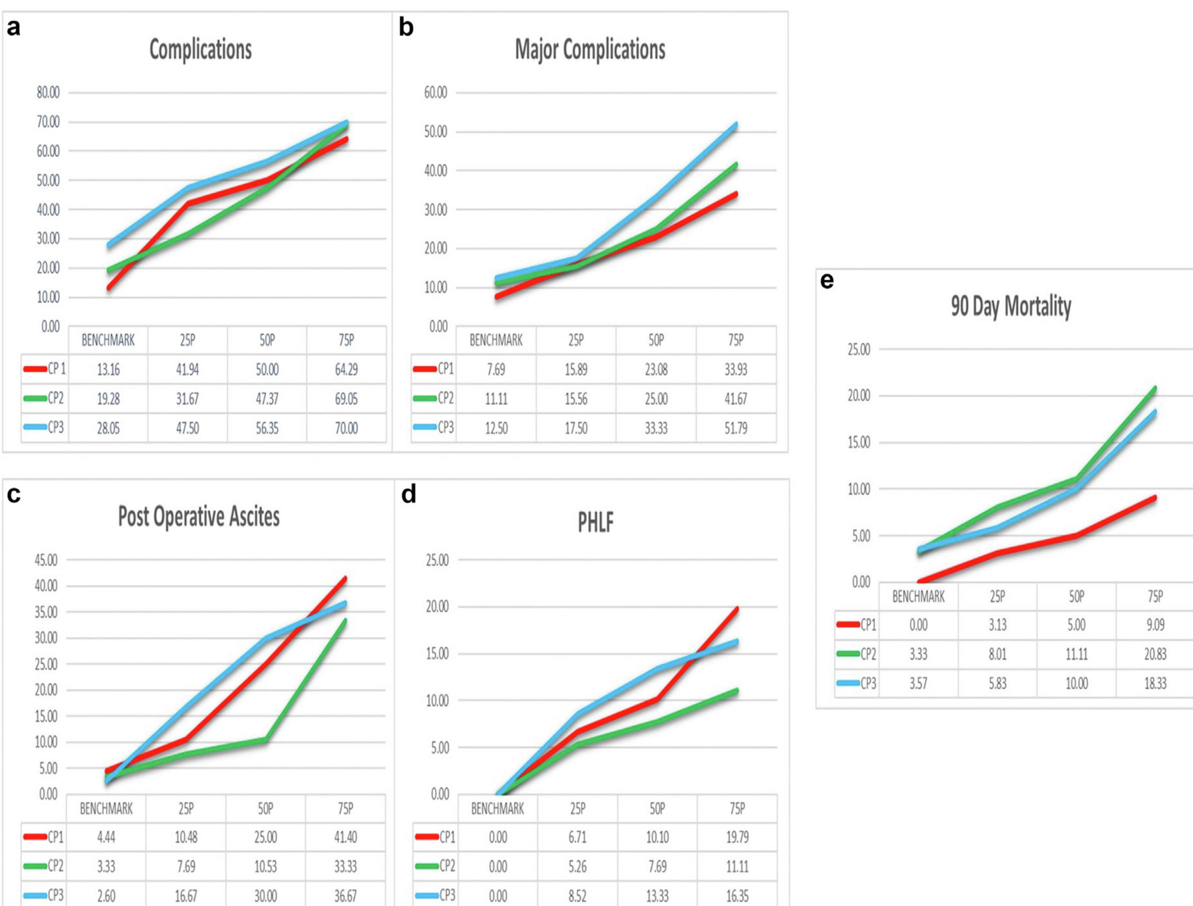

Figure 2 Benchmarks and trends are depicted for each considered outcome: a) overall complications; b) major complications (presented as the rate on the overall complications); c) post-operative ascites; d) PHLF: post-hepatectomy liver failure; e) 90-days mortality

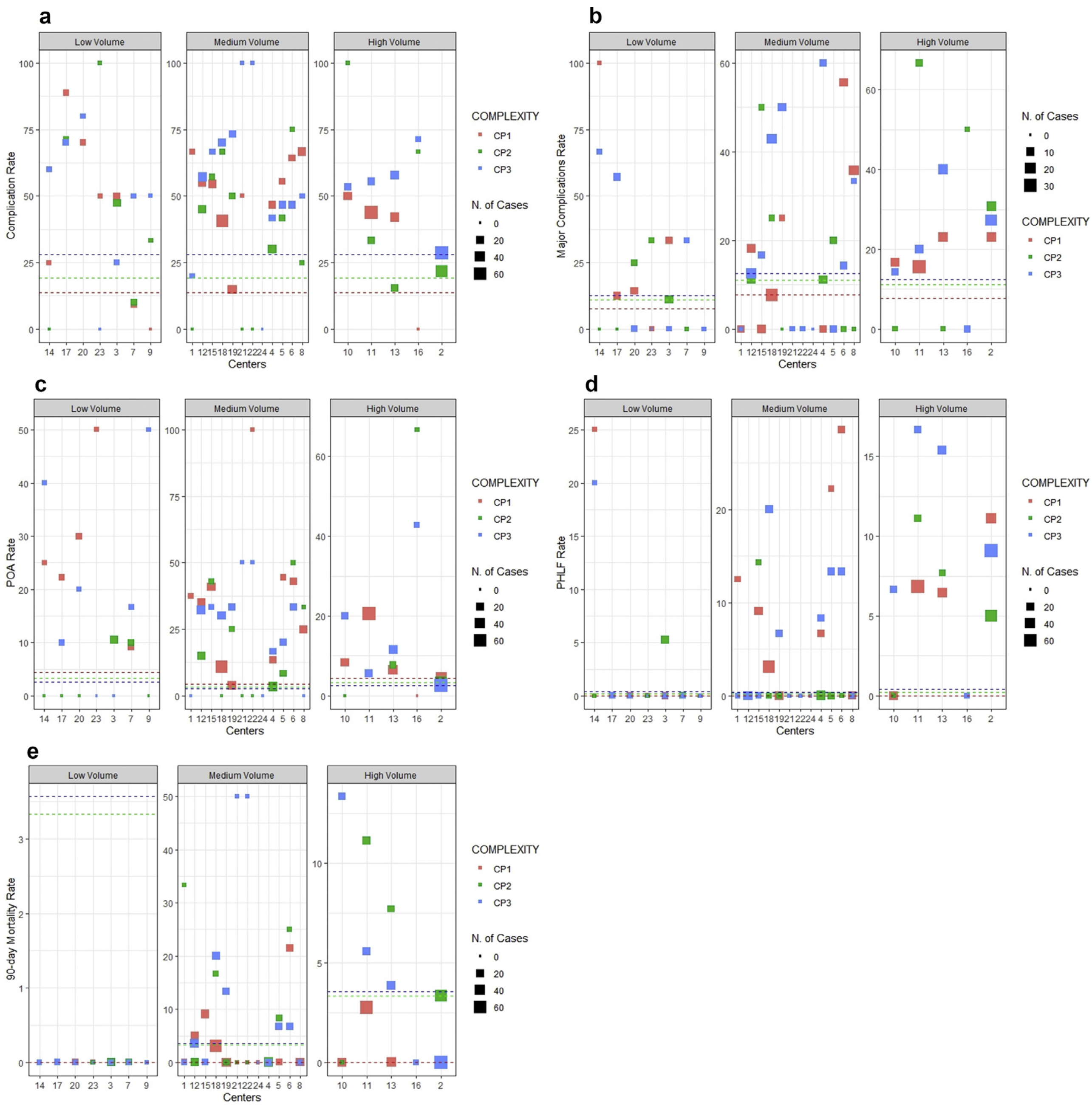

Figure 3 For each outcome, the reported results obtained by each center according to their volume classification are depicted. The colours summarized the volume classification, while the size of the squares is proportional to the number of cases treated by each center. a) overall complications; b) Major complications (presented as the rate on the overall complications); c) post-operative ascites; d) PHLF: post-hepatectomy liver failure; e) 90-days mortality. The dot-lines represent the benchmark values for each complexity grade: red: CP1, green: CP2 and blue: CP3

best achievable outcomes after stratification for surgical complexity,[18] based on a multicentric national database. As a consequence, the benchmark values for overall and major morbidity progressively increased with technical complexity. Also, 90-days mortality rates displayed an incremental trend related to surgical difficulty, even if CP2 and CP3 presented superimposable values. Interestingly, POA rate was not strictly related to the magnitude of resection but was higher in the CP1 group. Even if potentially misleading, this can be explained by the different grades of intrinsic hepatic impairment of the groups: less extensive surgery has been offered to patients with poorer residual liver function. In this case, this cohort would be more at risk for postoperative hepatic decompensation. Similarly, this occurrence may also be explained by the fact that high-volume centers more often deal with major resections, and usually they have the availability of high-quality in-hospital services, such as interventional radiology and intensive care unit: this has been proven to reduce not only the incidence of complications but also the impact of them on the failure-to-rescue syndrome.[30]

The benchmark value for PHLF of 0.0% for all three surgical complexity grades deserves a brief explanation. While PHLF is one of the main sources of morbidity and mortality after hepatectomy, its occurrence remains at low frequency. In a recent report assessing benchmarks in a population of living liver donors who underwent major hepatectomies, PHLF was set at 1%.[6] It is clear that these results are driven by the PHLF classification adopted: in fact, in the HERCOLES register PHLF was measured by using the 50-50 criteria,[20] which are designed to capture the most severe forms of PHLF, not considering the mild and intermediate presentations that are more common. Although other classifications could be applied to depict also the minor forms of PHLF, it is worthwhile to note that benchmarks represent the best possible outcome, but this does not mean that the best possible outcome is always achievable.[12]

R1 resection benchmark, which is a well-known risk factor for recurrence, was near 0 in each complexity grade, reflecting the inevitable surgeons' attention to achieve the best oncological outcome. Again, as for other parameters, this benchmarking value can be read as a reference point for potential improvement in those particular centers in which their benchmarking values may be discordant.

Considering the length of stay, the benchmarks values were in line with the recent literature regarding the open technique in liver surgery. However, we may note that, as highlighted recently by the literature regarding the fast-track programmes likes ERAS,[31] this parameter is highly susceptible of variance among centres, as the sum of different factors not always connected to the clinical outcome (patients will, hospital organization, surgeons attitude, etc): our results should be interpreted with caution, and in the future this benchmark should be evaluated considering the readiness for discharge rather than the effective discharge date.

Interestingly, beyond the 50th percentile for each outcome, there was a sharp increase of the adverse events' incidence: this result may help to address health systems towards the identification of the best potential centers which may centralize those types of surgical procedures, thanks to the optimal results achieved. From a health-system point of view, in order to guarantee the efficacy without affecting the local availability across the country of the health services, not exclusively benchmark setters should be considered as the natural providers of those procedures: however, one approach might be the identification of those centres that are near to the benchmark (as per the 25th - 50th percentiles) as regional referees, to avoid the dramatic increase of adverse events. It is also important to state that the best results obtained are not only the expression of the surgical skills: this is a simplistic interpretation. Dealing with HCC means managing not only the surgical fields but also the possible response of the damaged liver to the surgical stress, which needs efficient other facilities strictly connected with the surgical department. This is evident by the higher incidence of postoperative liver decompensation in the CP1 procedures: in HCC surgery, in fact, an apparent easy procedure may be set in a very difficult pathological environment which may significantly increase the risks. Another argument might be that CP1 procedures are very frequently performed in low-volume centers, which may have lower experience in patients selections, also as the results of the potential lack of multidisciplinary teams.

Although several methods[3,6,10] are available to estimate the benchmarks, we adopted the ABC method[14] since its ability to better capture the measure of interest in an unselected population, as recently discussed.[32] This way of measuring the performance helps to attribute the appropriate weight to all the possible factors which can modify the outcomes obtained, increasing the complexity - and consequently the reliability - of the health systems evaluation. In this sense, the correlation between volume and efficacy is not linear, and it depends on several other factors such as the local expertise, the availability of services, and the patient's selection criteria.[33] As our results showed, high-volume centers were not always the benchmarks setters for all the types of procedures: this may reflect exactly the tendency of the largest centers to treat the most difficult cases in terms of underlying comorbidities and functional reserve. In this sense, the research in quality benchmarking in liver surgery should continue with the aim to propose realistic cut-off values of outcome measures to avoid the risk of having top performers in low or middle-volume centers performing small liver resections in healthy patients and low performers in high-volume centers performing complex liver resection in diseased patients. Last but not least, in the current era of benchmarking in surgery, it is important to make distinctions between safety and quality in surgery as recently pointed out.[34] In particular, safety and quality in surgery may be, paradoxically, associated with some downsized to the strategy of aiming at zero harm such as the perfection in patients' selection, which limits the patient access to surgery, or the performance of innovative operations, which at the beginning might not be compatible with perfection, or the training of residents and fellows, which similarly to new and innovative procedures might not be associated with zero harm. Thus, when searching and interpreting benchmarking data in surgery wise and cautions are requested to any representative facing these features.

We acknowledge some inherent limitations of this study. Due to retrospective nature, the results may be affected by some confounders. Both the exclusion of 6.8% of patients because of missing data and the absence of other variables not reported in the registry might have impacted the measured results. Since data were analyzed from centers from a single country, the results herein described might not be easily generalized worldwide, and an external international validation appears to be necessary. While in the modern era liver resection may be delivered in some extremely selected cases of deteriorated liver function (as per Child-B), data of this study considered only cirrhotics with preoperative well-compensated liver function, consequently the results might be generalized only for that type of HCC patients. Moreover, due to its novelty, there is no agreement concerning the methodology used for benchmarking surgical procedures, and standardization of the statistical approach is crucial. Thus, CP1 and CP2 procedures may be often executed by laparoscopy (laparoscopic liver resection, LLR), leading to a sort of selection bias among the equivalent procedures here represented, which may be more complex and consequently excluded from a laparoscopic approach. This may cause overestimates of benchmarking grade I and II procedures (in Supplementary Table 1 a comparison among the number of procedures between OLR and LLR per each complexity grade is made available). More importantly, benchmarks may differ in case of varying the complexity classification employed. Finally, our results reflect best-achievable outcomes that are otherwise related to a certain period, thus benchmarks must be revised periodically concerning future advances.

In conclusion, since benchmarking defines the best achievable outcomes, the difference between the benchmarks values and the results obtained reflects the room for potential growth. This does not imply that the best possible outcome is always feasible, but since this novel method can be applied by departments and surgeons, it should be a tool for institutions and individuals to encourage improvement and to imagine a new health systems organization. In this scenario, we proposed benchmarks for open liver resection for HCC in cirrhotic patients, stratified for surgical complexity.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

None to declare.

References

1. Vonlanthen R, Clavien P-A. (2012) What factors affect mortality after surgery? *Lancet* 380:1034–1036. https://doi.org/10.1016/S0140-6736(12)61417-2.
2. Vonlanthen R, Slankamenac K, Breitenstein S, Puhan MA, Muller MK, Hahnloser D *et al*. (2011) The impact of complications on costs of major surgical procedures. *Ann Surg* 254:907–913. https://doi.org/10.1097/sla.0b013e31821d4a43.
3. Sánchez-Velázquez P, Muller X, Malleo G, Park J-S, Hwang H-K, Napoli N *et al*. (2019) Benchmarks in pancreatic surgery: a novel tool for unbiased outcome comparisons. *Ann Surg* 270:211–218. https://doi.org/10.1097/SLA.0000000000003223.
4. Raptis DA, Sánchez-Velázquez P, Machairas N, Sauvanet A, Rueda de Leon A, Oba A *et al*. (2020) Defining benchmark outcomes for pancreatoduodenectomy with portomesenteric venous resection. *Ann Surg* 272:731–737. https://doi.org/10.1097/SLA.0000000000004267.
5. Schmidt HM, Gisbertz SS, Moons J, Rouvelas I, Kauppi J, Brown A *et al*. (2017) Defining benchmarks for transthoracic esophagectomy. *Ann Surg* 266:814–821. https://doi.org/10.1097/sla.0000000000002445.
6. Rössler F, Sapisochin G, Song G, Lin Y-H, Simpson MA, Hasegawa K *et al*. (2016) Defining benchmarks for major liver surgery: a multicenter analysis of 5202 living liver donors. *Ann Surg* 264:492–500. https://doi.org/10.1097/SLA.0000000000001849.
7. Russolillo N, Aldrighetti L, Cillo U, Guglielmi A, Ettorre GM, Giuliante F *et al*. (2020) Risk-adjusted benchmarks in laparoscopic liver surgery in a national cohort. *Br J Surg* 107:845–853. https://doi.org/10.1002/bjs.11404.
8. Muller X, Marcon F, Sapisochin G, Marquez M, Dondero F, Rayar M *et al*. (2018) Defining benchmarks in liver transplantation. *Ann Surg* 267: 419–425. https://doi.org/10.1097/sla.0000000000002477.
9. Raptis DA, Linecker M, Kambakamba P, Tschuor C, Müller PC, Hadjittofi C *et al*. (2019) Defining benchmark outcomes for ALPPS. *Ann Surg* 270:835–841. https://doi.org/10.1097/SLA.0000000000003539.
10. Bagante F, Ruzzenente A, Beal E, Campagnaro T, Merath K, Conci S *et al*. (2018) Benchmarks value for incidence of post-hepatectomy liver failure after major liver surgery: a validation and integration analysis. *HPB* 20:S215–S216. https://doi.org/10.1016/j.hpb.2018.06.100.
11. von Eiff W. (2015) International benchmarking and best practice management: in search of health care and hospital excellence. *Adv Health Care Manag* 17:223–252. https://doi.org/10.1108/s1474-823120140000017014.
12. Staiger RD, Schwandt H, Puhan MA, Clavien P-A. (2019) Improving surgical outcomes through benchmarking. *Br J Surg* 106:59–64. https://doi.org/10.1002/bjs.10976.
13. Famularo S, Donadon M, Cipriani F, Ardito F, Carissimi F, Perri P *et al*. (2020) Hepatocellular carcinoma surgical and oncological trends in a national multicentric population: the HERCOLES experience. *Updates Surg*. https://doi.org/10.1007/s13304-020-00733-6.
14. Weissman NW, Allison JJ, Kiefe CI, Farmer RM, Weaver MT, Williams OD *et al*. (1999) Achievable benchmarks of care: the ABCs of benchmarking. *J Eval Clin Pract* 5:269–281. https://doi.org/10.1046/j.1365-2753.1999.00203.x.
15. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, & Vandenbroucke JP , for the STROBE Initiative. (2007) The strengthening the reporting of observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational Studies. *PLoS Med* 4e296. https://doi.org/10.1371/journal.pmed.0040296.
16. Charlson ME, Pompei P, Ales KL, MacKenzie CR. (1987) A new method of classifying prognostic comorbidity in longitudinal Studies: development and validation. *J Chron Dis* 40:373–383. https://doi.org/10.1016/0021-9681(87)90171-8.
17. Strasberg SM, Belghiti J, Clavien P-A, Gadzijev E, Garden JO, Lau W-Y *et al*. (2000) The Brisbane 2000 terminology of liver anatomy and resections. *HPB* 2:333–339. https://doi.org/10.1016/S1365-182X(17)30755-4.

18. Kawaguchi Y, Hasegawa K, Tzeng C-WD, Mizuno T, Arita J, Sakamoto Y *et al*. (2020) Performance of a modified three-level classification in stratifying open liver resection procedures in terms of complexity and postoperative morbidity. *Br J Surg* 107:258–267. https://doi.org/10.1002/bjs.11351.
19. Torzilli G, Viganò L, Giuliante F, Pinna AD. (2016) Liver surgery in Italy. Criteria to identify the hospital units and the tertiary referral centers entitled to perform it. *Updates Surg* 68:135–142. https://doi.org/10.1007/s13304-016-0373-0.
20. Balzan S, Belghiti J, Farges O, Ogata S, Sauvanet A, Delefosse D *et al*. (2005) The '50-50 criteria' on postoperative day 5: an accurate predictor of liver failure and death after hepatectomy. *Ann Surg* 242:824–828. discussion 828–9. https://doi.org/10.1097/01.sla.0000189131.90876.9e
21. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD *et al*. (2009) The clavien-dindo classification of surgical complications. *Ann Surg* 250:187–196. https://doi.org/10.1097/sla.0b013e3181b13ca2.
22. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien P-A. (2013) The comprehensive complication Index. *Ann Surg* 258:1–7. https://doi.org/10.1097/sla.0b013e318296c732.
23. Giani A, Cipriani F, Famularo S, Donadon M, Bernasconi DP, Ardito F *et al*. (2020) Performance of comprehensive complication Index and clavien-dindo complication scoring system in liver surgery for hepatocellular carcinoma. *Cancers* 12. https://doi.org/10.3390/cancers12123868.
24. Azoulay D, Eshkenazy R, Andreani P, Castaing D, Adam R, Ichai P *et al*. (2005) In situ hypothermic perfusion of the liver versus standard total vascular exclusion for complex liver resection. *Ann Surg* 241:277–285. https://doi.org/10.1097/01.sla.0000152017.62778.2f.
25. Chan K-M, Lee C-F, Wu T-J, Chou H-S, Yu M-C, Lee W-C *et al*. (2012) Adverse outcomes in patients with postoperative ascites after liver resection for hepatocellular carcinoma. *World J Surg* 36:392–400. https://doi.org/10.1007/s00268-011-1367-1.
26. Famularo S, HERCOLES Group, Donadon M, Cipriani F, Ardito F, Iaria M, Carissimi F *et al*. (2021) The impact of postoperative ascites on survival after surgery for hepatocellular carcinoma: a national study. *J Gastrointest Surg*. https://doi.org/10.1007/s11605-021-04952-z.
27. Parikh K, Hall M, Mittal V, Montalbano A, Mussman GM, Morse RB *et al*. (2014) Establishing benchmarks for the hospitalized care of children with asthma, bronchiolitis, and pneumonia. *Pediatrics* 134:555–562. https://doi.org/10.1542/peds.2014-1052.
28. Hatfield MD, Ashton CM, Bass BL, Shirkey BA. (2016) Surgeon-specific reports in general surgery: establishing benchmarks for peer comparison within a single hospital. *J Am Coll Surg* 222:113–121. https://doi.org/10.1016/j.jamcollsurg.2015.10.017.
29. Landercasper J, Fayanju OM, Bailey L, Berry TS, Borgert AJ, Buras R *et al*. (2018) Benchmarking the American society of breast surgeon member performance for more than a million quality measure-patient encounters. *Ann Surg Oncol* 25:501–511. https://doi.org/10.1245/s10434-017-6257-9.
30. Ardito F, Famularo S, Aldrighetti L, Grazi GL, DallaValle R, Maestri M *et al*. (2020) The impact of hospital volume on failure to rescue after liver resection for hepatocellular carcinoma. *Ann Surg* 272:840–846. https://doi.org/10.1097/sla.0000000000004327.
31. Celio DA, Poggi R, Schmalzbauer M, Rosso R, Majno P, Christoforidis D. (2019) ERAS, length of stay and private insurance: a retrospective study. *Int J Colorectal Dis* 34:1865–1870. https://doi.org/10.1007/s00384-019-03391-2.
32. Russolillo N, Aldrighetti L, Guglielmi A, Giuliante F, Ferrero A. (2020) Letter regarding 'benchmark performance of laparoscopic left lateral sectionectomy and right hepatectomy IN expert centers. *J Hepatol*. https://doi.org/10.1016/j.jhep.2020.11.009.
33. Franchi E, Donadon M, Torzilli G. (2020) Effects of volume on outcome in hepatobiliary surgery: a review with guidelines proposal. *Glob Health Med* 2:292–297. https://doi.org/10.35772/ghm.2020.01013.
34. Aloia TA. (2020) Should zero harm Be our goal? *Ann Surg* 271:33–36. https://doi.org/10.1097/SLA.0000000000003316.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.hpb.2022.02.008.