Open access • Journal Article • DOI:10.1097/SLA.0000000000002477

Defining Benchmarks in Liver Transplantation: A Multicenter Outcome Analysis Determining Best Achievable Results. — Source link

Xavier Muller, F. Marcon, Gonzalo Sapisochin, Max Marquez ...+45 more authors

Institutions: University of Zurich, Children's of Alabama, Toronto General Hospital, Washington University in St. Louis ...+10 more institutions

Published on: 01 Sep 2017 - Annals of Surgery (Ann Surg)

Topics: Liver transplantation

Defining Benchmarks in Liver Transplantation: A Multicenter Outcome Analysis Determining Best Achievable Results

Muller, Xavier ; Marcon, Francesca ; Sapisochin, Gonzalo ; Marquez, Max ; Dondero, Federica ; Rayar, Michel ; Doyle, Majella M B ; Callans, Lauren ; Li, Jun ; Clavien, Pierre-Alain ; Puhan, Milo ; et al

DOI: https://doi.org/10.1097/SLA.0000000000002477

Defining Benchmarks in Liver Transplantation

A Multicenter Outcome Analysis Determining Best Achievable Results

Xavier Muller, MD,[*] Francesca Marcon, MD,[†] Gonzalo Sapisochin, MD,[‡] Max Marquez, MD,[‡] Federica Dondero, MD,[§] Michel Rayar, MD,[¶] Majella M. B. Doyle, MD,[||] Lauren Callans,[**] Jun Li, MD,[††] Greg Nowak, MD,[‡‡] Marc-Antoine Allard, MD,[§§] Ina Jochmans, MD, PhD,[¶¶] Kyle Jacskon, MD,[||||] Magali Chahdi Beltrame, MD,[***] Marjolein van Reeven, MD,[†††] Samuele Iesari, MD,[‡‡‡] Alessandro Cucchetti, MD,[§§§] Hemant Sharma, MD,[¶¶¶|||||||] Roxane D. Staiger, MD,[*] Dimitri A. Raptis, MD, PhD,[*] Henrik Petrowsky, MD,[*] Michelle de Oliveira, MD,[*] Roberto Hernandez-Alejandro, MD,[¶¶¶|||||||] Antonio D. Pinna, MD,[§§§] Jan Lerut, MD, PhD,[‡‡‡] Wojciech G. Polak, MD, PhD,[†††] Eduardo de Santibañes, MD,[***] Martín de Santibañes, MD,[***] Andrew M. Cameron, MD, PhD,[||||] Jacques Pirenne, MD, PhD,[¶¶] Daniel Cherqui, MD,[§§] René A. Adam, MD, PhD,[§§] Bö-Göran Ericzon, MD, PhD,[‡‡] Bjoern Nashan, MD, PhD,[††] Kim Olthoff, MD,[**] Avi Shaked, MD,[**] William C. Chapman, MD,[||] Karim Boudjema, MD,[¶] Olivier Soubrane, MD, PhD,[§] Catherine Paugam-Burtz, MD, PhD,[§] Paul D. Greig, MD,[‡] David R. Grant, MD,[‡] Amanda Carvalheiro, MD,[†] Paolo Muiesan, MD,[†] Philipp Dutkowski, MD,[*] Milo Puhan, MD, PhD,[****] and Pierre-Alain Clavien, MD, PhD[*]

Objective: To propose benchmark outcome values in liver transplantation, serving as reference for assessing individual patients or any other patient groups.

Background: Best achievable results in liver transplantation, that is, benchmarks, are unknown. Consequently, outcome comparisons within or across centers over time remain speculative.

Methods: Out of 7492 liver transplantation performed in 17 international centers from 3 continents, we identified 2024 low risk adult cases with a laboratory model for end-stage liver disease score ≤20 points, a balance of risk score ≤9, and receiving a primary graft by donation after brain death. We chose clinically relevant endpoints covering intra- and postoperative course, with a focus on complications graded by severity including the complication comprehensive index (CCI®). Respective benchmarks were derived from the median value in each center, and the 75 percentile was considered the benchmark cutoff.

Results: Benchmark cases represented 8% to 49% of cases per center. One-year patient-survival was 91.6% with 3.5% retransplantations. Eighty-two percent of patients developed at least 1 complication during 1-year follow-up. Biliary complications occurred in one-fifth of the patients up to 6 months after surgery. Benchmark cutoffs were ≤4 days for ICU stay, ≤18 days for hospital stay, ≤59% for patients with severe complications (≥ Grade III) and ≤42.1 for 1-year CCI®. Comparisons with the next higher risk group (model for end stage liver disease 21–30) disclosed an increase in morbidity but within benchmark cutoffs for most, but not all indicators, while in patients receiving a second graft from 1 center (n = 50) outcome values were all outside of benchmark values.

Conclusions: Despite excellent 1-year survival, morbidity in benchmark cases remains high with half of patients developing severe complications during 1-year follow-up. Benchmark cutoffs targeting morbidity parameters offer a valid tool to assess higher risk groups.

Keywords: benchmark, complication, liver transplantation, morbidity, outcome

(*Ann Surg* 2017;xx:xxx–xxx)

More than 300.000 liver transplantations (LT) have been performed worldwide, since its introduction in clinical practice in the early 1980s.[1–3] With major improvements in surgical techniques, immunosuppression therapy, and patient selection 1-year patient survival rates have continuously improved to over 85%.[2,4–6] LT is

From the [*]Department of Surgery and Transplantation, Swiss HPB Center Zurich, University Hospital Zurich, Zurich, Switzerland; [†]Department of HBP Surgery and Transplantation, Queen Elizabeth Hospital and Birmingham Children's Hospital, Birmingham, UK; [‡]Department of Surgery, Abdominal Transplant and HPB Surgical Oncology, Toronto General Hospital, University of Toronto, Toronto, ON, Canada; [§]Department of Surgery and Transplantation, Hôpital Beaujon, Clichy, France; [¶]Department of HBP Surgery and Transplantation, University Hospital Rennes, Rennes, France; [||]Division of General Surgery and Section of Abdominal Transplantation, Washington University School of Medicine, St. Louis, MO; [**]Division of Transplantation, University of Pennsylvania, Philadelphia, PA; [††]Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [‡‡]Division of Transplantation Surgery, CLINTEC Karolinska Institute, Karolinska University Hospital Huddinge, Stockholm, Sweden; [§§]Department of Surgery, The Hepatobilairy Center, Paul Brousse Hospital, University Paris Sud, Villejuif, France; [¶¶]Abdominal Transplant Surgery, KU Leuven, University Hospitals Leuven, Leuven, Belgium; [||||]Department of Surgery, The Johns Hopkins Hospital, Baltimore, MD; [***]Division of HBP Surgery, Department of Surgery, Liver Transplant Unit, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina; [†††]Department of Surgery and Transplantation, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands; [‡‡‡]Department of Abdominal and Transplant Surgery, University Hospital St. Luc, Brussels, Belgium; [§§§]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy; [¶¶¶]Division of Transplantation, University of Rochester Medical Center, New York, NY; [|||||||]Department of Surgery, Western University, London, ON, Canada; and [****]Department of Epidemiology, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.

CCI® refers to registered trademark from the University of Zurich, Zurich, Switzerland.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Pierre-Alain Clavien, MD, PhD, Department of Surgery and Transplantation, University Hospital Zurich, Raemistrasse 100, CH-8091 Zurich, Switzerland. E-mail: clavien@access.uzh.ch.

ISSN: 0003-4932/16/XXXX-0001
DOI: 10.1097/SLA.0000000000002477

currently the standard of care for many patients with end-stage liver disease and liver tumors; a success that has created a dramatic imbalance between organ need and availability.[7,8] To face the resultant organ shortage, higher-risk grafts are implanted in sick recipients, potentially jeopardizing the postoperative course. One of the main issues in evaluating the results of LT is the lack of reliable comparison to reference groups.

With this in mind, we embarked on the concept of benchmarking, which is widely known as quality assessment and improvement in the field of banking or industrial manufacturing. Benchmarking has been used in medicine as a tool for evaluating single-center outcomes by risk-adjusted comparisons to national data.[9–11] We introduced the idea of identifying best achievable results by selecting low risk cases in experienced centers and presented a large multicenter study at the American Surgical Association annual meeting in 2016, proposing benchmark values for major hepatectomies. The ideal low risk cases were those performed in healthy living-donor using a comprehensive analysis of relevant endpoints.[12]

To define benchmark values in LT, we targeted high-volume programs, holding an audited prospective database from areas, where cadaveric LT is common practice. Low risk cases were defined by pretransplant characteristics, using the model for end-stage liver disease (MELD) score, which predicts mortality on the waiting list, and the balance of risk (BAR) score, which takes into account recipient and donor factors. Other well-established risk factors were considered exclusion criteria to define the benchmark cohort.[13–16] We selected 13 endpoints – to compute benchmark cutoffs based on their clinical relevance and reproducibility. These benchmark values may serve as reference for future studies and help assess worldwide results and trends in the field of LT.

METHODS

Center Selection and Data Sources

We screened countries on every continent for centers performing at least 50 cadaveric LT per year during the 5-year study period (2010–2015) and holding a prospectively collected and audited database. The highest volume centers in the respective countries were selected and the final collaborative consortium included 17 centers: 11 from Europe, 5 from North America, and 1 from South America. No Asian center could be included due to the small number of available cadaveric grafts.

Study Population and Variables of Interest

Benchmark cases were identified in every center database. To define ''ideal'' LT cases, we considered the mortality risk of the patients on the waiting list as well as the donor/recipient match, and thus selected patients based on the MELD score and the BAR score.[17,18]

The MELD score is a scoring system widely used for prioritization of patients awaiting LT – based on serum levels of bilirubin and creatinine, and the international normalized ratio for prothrombin time – predicting the risk of mortality on the waiting list in patients with chronic liver diseases.[18–20] We selected only patients with a MELD score $\leq$20 at the time of transplantation, as patients with higher figures have been shown to display a increasing mortality risk.[20]

The BAR score predicts post-LT mortality by including donor age, recipient MELD score, recipient age, retransplant status, the need for mechanical ventilation, and cold ischemia time.[17] Patients with a BAR score $\leq$9 present an ideal donor–recipient match with very low mortality risk, and this cutoff was thus included as a selection criteria for benchmark cases.[7,17,21]

To further narrow the selection of the ''best cases,'' we excluded patients with acute liver failure, patients on mechanical ventilation at the time of surgery, and patients receiving a graft from donors after circulatory death (DCD).[13,16,22–24] In addition, special situations and technical difficulties impacting outcomes, such as recipient portal vein thrombosis, previous major abdominal surgery (hepato-biliary surgery and extensive colorectal surgery), partial graft implantation or retransplantation were excluded.[14,15,25]

In summary, patients with a MELD score $\leq$20, a BAR score $\leq$9 and receiving a standard LT from a donor after brain death were selected as benchmark cases.

Comparative Cohorts

To compare outcomes from the benchmark cohort, we identified 2 higher risk groups. First, we considered a slightly higher risk group from the same centers and study period using identical selection criteria except for the next higher MELD score ranging from 21 to 30 points. Second, to test the applicability of the established benchmark criteria for the assessment of a single-center cohort, we selected one of the highest risk groups from one of the benchmark centers – that is, those receiving a second graft due to failure of the original graft – over the same study period.

Collected Data, Outcomes, and Follow-up

Local investigators collected center-specific data, using a secure online data entry management system, including the following recipient and donor characteristics: age, MELD before LT, underlying liver disease and comorbidities, cold ischemia time, and a variety of postoperative events at 4 specific time points (discharge, 3, 6, and 12 mo).

The primary end-point was morbidity and mortality. To assess LT morbidity, we selected peri- and postoperative parameters with proven impact on outcomes and costs, namely duration of surgery, intraoperative blood transfusion, hospital and intensive care unit (ICU) stay, as well as post-LT complications.[26–28]

A systematic and comprehensive classification of all postoperative complications was performed using the Clavien–Dindo grading system. This system ranks complications by severity according to their therapeutic consequences.[29,30] Briefly, grade I and II complications are events requiring only bedside procedures or a need for pharmacologic treatments. Grade III complications require surgical, radiological, or endoscopic treatment. Grade IV complications are life-threatening complications requiring ICU care, while grade V complications correspond to death. Patients requiring a retransplantation were graded IVa. To get additional information on morbidity, we then measured the cumulative morbidity using the comprehensive complication index (CCI®).[31,32] The CCI® expresses morbidity on a continuous numeric scale from 0 (no complications) to 100 (death) by weighing all postoperative complications according to the Clavien–Dindo classification for their respective severity.[31] Since in LT, complications related to the biliary tract are the most common complications and responsible for major costs and morbidity, they were additionally analyzed separately.[33,34]

Statistical Analysis

Benchmark Metrics

Benchmark cutoffs were determined for 1-year mortality, retransplantations, and peri- and postoperative morbidity indicators including length of surgery (skin-to-skin time), intraoperative blood transfusions, renal replacement therapy after transplantation until discharge, length of ICU, and hospital stay (from day of transplantation until hospital discharge), patients with mild ($\leq$grade II) and severe (grade $\geq$ III) complications, biliary complications, hepatic artery thrombosis (HAT) and the CCI® at discharge, 3 months, 6 months, and at 1 year. Every outcome indicator listed above was defined in the study protocol to secure standardization in data collection among all centers. Of note, biliary complications were

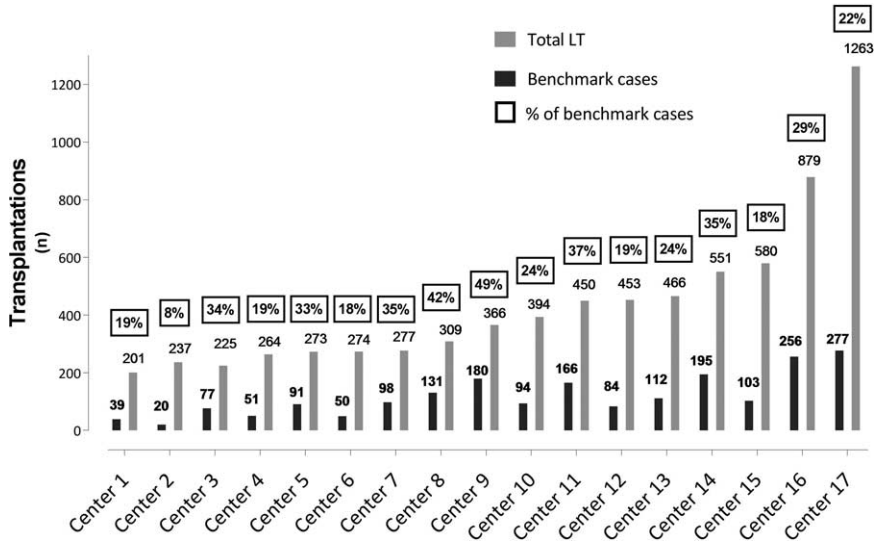

FIGURE 1. Distribution of LT benchmark cases among centers.

defined as every post-transplant event related to the biliary system such as leaks, bilioma, strictures or infection, and graded according to the Clavien–Dindo classification, and subsequently integrated in the CCI® formula. The median value for every indicator in each center was determined and the 75th percentile of all center-specific median values for a given indicator was considered the benchmark cutoff, as previously reported.[12] Thus, we opted for an equal count of every center in this study, regardless of differences in caseloads.

We conducted descriptive statistics for various intra- and postoperative parameters to characterize the patient population. Survival curves were compared by using the log rank test and linear regression analysis was performed for correlations. A value of $P < 0.05$ was considered statistically significant. Statistical analysis was performed using Stat 10.1 and SPSS, version 19.

RESULTS

Benchmark Cohort

Seventeen centers provided 7492 cases of LT over the 5-year study period. Approximately one quarter (27%, n = 2024) were benchmark cases with a large variation among centers (range: 8% to 49%) (Fig. 1).

Characteristics and outcomes in the benchmark cohort: Patients in the benchmark cohort displayed a median donor and recipient age of 55 (IQR: 43–67) and 57 (IQR: 50–62) years, respectively. Median MELD and BAR scores were 12 (IQR: 9–16) and 4 (IQR: 2–6), respectively. Underlying liver diseases included hepatocellular carcinoma (47.4%), hepatitis C (78.5%), alcoholic cirrhosis (26.5%), and hepatitis B (2%). The median graft preservation time was short with 7 hours cold storage (IQR: 6–9). The median duration of LT was 5 hours (IQR: 290–480) with low transfusion requirements [2U RBC (unit of red blood cell), IQR: 0–5]. Postoperative renal replacement therapy was performed in 6.6% of cases. Consistently, median length of ICU and hospital stay were short with 2 (1–5) and 13 (9–20) days, respectively. Overall 1- and 5-year patient survival-rates were 91.6% (actual survival) and 78.2% (actuarial survival), respectively. Post-LT complications, however, occurred frequently during the 1-year follow-up; the majority of patients (82%) presented at least 1 complication and 58% developed more than 1 complication. Overall, two-thirds (68.7%) of the patients developed mild complications ($\leq$ grade II), while almost half (48.9%) experienced a severe complication requiring intervention ($\geq$ grade III). Biliary complications occurred in 18.8% of the patients; the majority (57.7%) after hospital discharge. Accordingly, the overall median post-transplant morbidity index (CCI®) increased over time from 20.9 at discharge (IQR: 0–34.6) to 26.6 at 3 months (IQR: 8.7–42.4), 29.6 at 6 months (IQR: 12.2–44.9) and 33.5 at 12 months (IQR: 20.9–49.5). Omitting grade I complications for the CCI® calculation did not affect the overall median CCI® value at the 4 different time points.

Center-specific and Volume Effect on Outcomes

We recorded a significant variability in perioperative morbidity indicators among centers, namely median length of surgery (IQR: 4.9–8.1 h), median ICU (IQR: 1–8 d), and hospital stay (IQR: 6–24 d). The same applied to the CCI® from discharge (IQR: 0–51.7) to 1 year (IQR: 20.9–62.9). These results, however, did not correlate with the respective center volume. Interestingly, we noted that centers performing less benchmark cases; that is, higher proportion of more difficult cases, had less biliary complications in the benchmark cohort (Pearson R = 0.68, $P = 0.002$, Fig. 2).

Benchmark Values

The benchmark cutoffs, calculated as the 75th percentile of the medians of each center, served to indicate the best achievable results for each post-LT parameter (Tables 1 and 2). The benchmark for 1-year mortality and graft-loss are 9% and 11%, respectively.

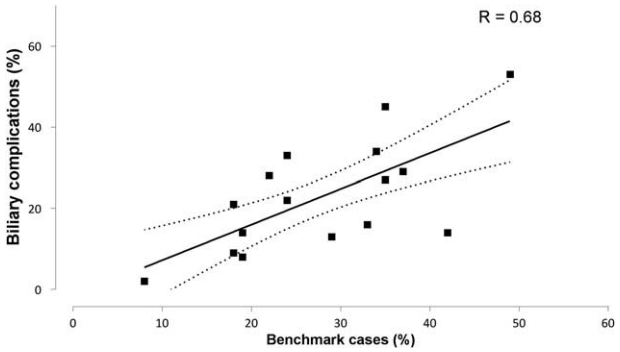

FIGURE 2. More biliary complication occur in centers with a higher proportion of benchmark cases.

TABLE 1. Benchmark Cutoffs in Liver Transplantation

Perioperative Course				
OP duration	≤6 hours			
Intraoperative blood transfusions	≤3U RBC			
Renal replacement therapy	≤8%			
ICU stay	≤4 days			
Hospital stay	≤18 days			
Morbidity and Mortality	**Discharge**	**3 months**	**6 months**	**1 year**
Any complication	≤80%	≤90%	≤90%	≤94%
≤Grade II complication	≤69%	≤81%	≤83%	≤83%
≥Grade III complication	≤42%	≤54%	≤58%	≤59%
Biliary complications	≤12%	≤18%	≤20%	≤28%
Hepatic artery thrombosis (HAT)*	≤4.1%	≤4.4%	≤4.4%	≤4.4%
CCI® points	≤29.6	≤34.5	≤37.2	≤42.1
Graft-loss	≤4%	≤6%	≤9%	≤11%
Mortality	≤2%	≤4%	≤7%	≤9%

*HAT are commonly divided into two distinct entities based on the time of occurrence after LT. The benchmark cutoffs for early HAT occurring within the first 30 days and for late HAT occurring thereafter are ≤4.1% and ≤1%, respectively.

Benchmark cutoffs – set for specific perioperative parameters – are ≤6 hours surgical time, need for ≤3 red blood units, and ≤8% of patients requiring postoperative renal replacement therapy. Benchmarks for ICU and hospital stays are ≤4 and ≤18 days, respectively. The cutoff for severe (grade ≥III) complications is 59%, for overall biliary complications 28% and for HAT 4.4%, while the benchmark values for cumulative morbidity, expressed by the CCI® at discharge, 3, 6, and 12 months, are 29.6, 34.5, 37.2, and 42.1, respectively.

Higher MELD Cohort

In a next step, we compared outcomes in the benchmark cohort with sicker transplant recipients, displaying a MELD score ranging from 21 to 30 (median MELD 24, IQR: 22–27), but otherwise same selection criteria (Table 2). In this cohort (n = 699), 1-year survival was similar to the benchmark cohort (88.7% vs 91.6%, $P = 0.154$). Although morbidity increased, most outcome endpoints remained inside the benchmark cutoffs. Only intraoperative transfusion rates (5 U RBC vs ≤3 U RBC), patients with renal replacement therapy after LT (13.7% vs ≤3%), grade IV complications (24% vs ≤20%) and retransplantations (7.1% vs ≤4%) were outside the benchmark cutoffs. Cumulative post-LT morbidity, expressed by the CCI® at discharge and after 1 year, remained within the benchmark threshold (29.6 vs ≤29.6, and 39.7 vs ≤42.1).

Second Graft (Retransplantation) Cohort

To test the applicability of the benchmark thresholds in a high-risk group, we looked at 50 patients receiving a second transplantation due to graft failure at 1 center. The median MELD score in that group was 30 (IQR: 24–36) and the median BAR was 17 (IQR: 11–20). Overall, 28% of retransplantations were performed in the first 2 weeks following primary LT, 34% between day 15 and the first year, and 38% thereafter. The main etiologies for a retransplantation were primary nonfunction in 26%, nonanastomotic biliary stenosis in 20%, hepatitis C recurrence in 16%, and hepatic artery thrombosis in 10% of cases. Except for the use of 1 DCD graft, all retransplantations were performed with whole grafts from donors after brain death (n = 49). We observed a 30% reduction in 1st-year patient survival, when compared with the benchmark cutoffs (58% vs ≥ 88%), and all outcome parameters in the re-LT groups were unambiguously inferior to the benchmark cutoffs (Table 2). For example, ICU- and hospital stays were twice as high (8 d vs ≤ 4 d and 28 d vs ≤18 d, respectively), and severe complications (≥ Grade III) (96% vs ≤59%) as well as biliary complications (42% vs ≤28%), all largely exceeded benchmark values.

DISCUSSION

This is the first attempt to quantify best possible outcomes after LT using an international cohort of well-defined benchmark cases, representing approximately 1 quarter of overall LTs. This

TABLE 2. First-year Outcomes After Liver Transplantation in Two Higher Risk Groups Compared With First-year Benchmark Cutoffs

	MELD 21–30 (n = 699)	**Retransplantations (n = 50)†**	**Benchmark Cutoffs (at 12 mo)**
OP duration, h (IQR)	6 (5–7.5)	6.1 (5–7.6)	≤6
Intraoperative blood transfusions, number (IQR)	5 (2–8)*	6 (4–10)	≤3
Hospital stay (d), median (IQR)	16 (10–26)	28 (16–51)	≤18
ICU stay (d), median (IQR)	4 (2–9)	8 (4–22)	≤4
Renal replacement therapy, patient number (%)	96 (13.7)*	21 (42)	≤8%
Any complication, patient number (%)	616 (88.1)	50 (100)	≤94%
Grade II complications, patient number (%)	460 (65.8)	39 (78)	≤76%
Grade IIIa complications, patient number (%)	188 (26.8)	34 (68)	≤41%
Grade IIIb complications, patient number (%)	180 (25.8)	20 (40)	≤28%
Grade IV complications, patient number (%)	173 (24)*	35 (70)	≤20%
≤Grade II complication, patient number (%)	516 (73.8)	42 (84)	≤82%
≥Grade III complication, patient number (%)	406 (58.1)	48 (96)	≤59%
Biliary complications, patient number (%)	109 (15.5)	21 (42)	≤28%
Retransplantations, patient number (%)	50 (7.1)*	5 (10)	≤4%
CCI® discharge, median (IQR)	29.6 (8.7–43)	62.1 (42.2–98.7)	≤29.6
CCI® 3 mo, median (IQR)	33.5 (20.9–47.5)	72.7 (54.6–100)	≤34.5
CCI® 6 mo, median (IQR)	36.1 (20.9–51.2)	81 (60–100)	≤37.2
CCI® at 1 yr, median (IQR)	39.7 (22.6–55.8)	85.8 (64.2–100)	≤42.1
1-year mortality, patients (%)	62 (8.8)	18 (36)	≤9%

*Value is outside the benchmark cutoff.
†All values in the retransplantation cohort are outside the benchmark cutoffs.

analysis is crucial for conclusive comparisons among different groups of patients, centers, and over time. We chose a number of clinically relevant endpoints in this ''ideal'' LT cohort, disclosing an excellent 1-year survival, but with high postoperative morbidity, increasing significantly during the first year after LT. The outcome analysis of 2 higher LT cohorts illustrated the relevance of the benchmark values.

LT belongs to one of the most successful therapies developed over the past half century rescuing many patients with almost certain short-term death to a close-to-normal quality of life and full function in the society.[7] The success of this complex procedure has emerged – besides center and team experience – through proper patient and donor selection, at the cost, however, of significant perioperative morbidity.[35] With an ever-increasing gap between need and organ supply, many groups have developed novel strategies to augment the number of available grafts such as living donation, donation after cardiac arrest, or the use of injured grafts, for example, fatty livers. This raises the question of how to evaluate these strategies, not only in terms of patient survival, but also by using more subtle morbidity endpoints with eventually financial burden. This study provides – through a careful methodology – baseline values for various outcome indicators in ''ideal'' LT patients, a concept called ''benchmarking,'' and a new tool for comparisons with other LT patient groups or even competitive procedures.

The selection of benchmark patients was performed by a stepwise risk stratification. First, we selected patients with a BAR score $\leq$9 points since this threshold has been reported to have an increasing morbidity and length of hospital stay.[17,21] Of note, the combination of high donor age (>40 yrs), high recipient age (>60 yrs), and prolonged cold ischemia (> 12 h) is in theory within the BAR $\leq$9 threshold. Second, we chose recipients with a MELD score $\leq$20, as we have shown inferior results at this cutoff.[36] Third, we excluded established risk factors such as retransplantation, portal vein thrombosis, acute liver failure, or high-risk donors to narrow the selection toward the optimal scenarios.[13,14,16] We did not consider additional cardiac, pulmonary, or connective tissue diseases of the recipient since these characteristics are less frequently captured in a prospective manner, and may lead to missing values.

We solely selected centers with a sufficient caseload, besides a well-mixed distribution around the world. We, however, made no restrictions regarding local strategies for graft preservation, implantation, and reperfusion techniques to be representative of the ''real world.''

Our first observation was the variation in the proportion of benchmark cases among centers, ranging from less than 10% to more than 40% (Fig. 1). This difference likewise reflects how much risk centers are willing to take in terms of recipient and donor selection. One factor influencing this decision is the availability of donors, ranging from 8 to more than 45 donors per million population among the 17 participating centers.[37,38] Low donation rates force LT centers to take more risk to serve accumulating very sick patients awaiting a graft. In support of this argument, we noted that higher donation rates were associated with a higher amount of benchmark cases (data not shown). Additionally, some centers in the cohort must deal with more than two-thirds of so-called ''extended criteria donors,'' the most frequent being fatty liver grafts, which obviously are associated with more risk.[39,40] Thus, the methodology used for this study further validates the selection process identifying the ''benchmark cohort.''

We confirmed excellent survival rates close to 90% in this low risk population, as documented in most national and regional registries, but the study revealed impressively high morbidity rates with many severe complications occurring all over the 1st year post-LT follow-up. Comparing these data with the results for benchmark cases in major hepatectomy, we see a significant higher morbidity rate in the benchmark LT cases.[12] For example, benchmark values in LT cases are 3 times higher for overall and 5 times higher for severe complications, compared with benchmark values in major hepatectomy cases ($\leq$90% vs $\leq$31.2% and $\leq$58% vs $\leq$9.2%). These results further underline the need to select fine indicators of morbidity as endpoints for comparative studies, while mortality as endpoint appears too crude and with poor correlation with costs.[12,32,41]

The main issue in reporting complications has been the absence of standardization in definition and severity grading of postoperative negative events, as well as the lack of a reproducible and ''easy-to-understand'' metric system.[42,43] The Clavien–Dindo classification of complications – currently adopted in most areas of the world – ranks complications by severity based on their respective treatment.[29,30] The subsequent development of the CCI®, integrating all recorded complications into a single formula providing an index ranging from 0 (no complication) to 100 (death), is a validated metric system for comparative studies.[32,44–48] Together with other perioperative patient and cost-related outcome indicators, these morbidity tools were logically selected for establishing benchmark cutoffs.[43]

Another controversial and inconsistently reported issue is the minimum length of postoperative follow-up to accurately assess morbidity. We previously identified a minimal follow-up of 3 months to adequately document morbidity after major liver resection.[12] Here looking at a procedure associated with dramatically higher morbidity, we could document that the minimal follow-up must cover 1 year, because major complications, mostly biliary, continued to rise between 6 and 12 months after transplantation. Also interestingly, mortality increased from 4% to 7% between 3 and 6 months, and further increased to 9% at 1 year. In contrast, almost all HAT occurred before hospital discharge; i.e. within 4 weeks; which appears to be a serious event, since all cases required a retransplantation (data not shown). Of note, we cannot exclude from our database that other clinically silent HAT may occur later after surgery.

We also looked at the impact of center-volume on morbidity and mortality rates in the benchmark cohort. We observed that centers with a higher proportion of benchmark cases, that is, performing mostly low risk cases, had more biliary complications. One explanation could be that centers performing higher proportion of complex cases may develop better skills in preventing or correcting biliary problems.

Validation of the benchmark cohort was performed using 2 comparative cohorts. First, we used a slightly higher risk group (MELD 21–30) to test the ability of the benchmark analysis to quantify a minor increase in morbidity. Second, we used a single-center retransplant cohort to test the applicability of the benchmark cutoffs to quantify the magnitude of a more consequent increase in mortality and morbidity. In the first comparative group with slightly higher MELD scores (21–30), we failed to show any impact on mortality, but overall morbidity increased in this higher MELD group, with significantly higher need for intraoperative transfusion, postoperative renal replacement, and more grade IV complications. This analysis highlights differences, which have escaped to all previous reports, supporting a high sensitivity for benchmark thresholds. When considering the second comparative group, one of the highest risk groups from a single center, patients requiring a 2nd graft, the increase in morbidity becomes obvious covering all measured endpoints, and consequently quality of life of the patients, at least for the first year following surgery. This high morbidity probably has a massive effect on cost, as postoperative complications were found to be the most significant factors affecting direct and indirect costs.[39]

Another important issue to secure quality in outcome research is the accuracy in capturing postoperative complications, which depends on prospective collection of data, optimally by an

TABLE 3. Proposed Applications of Benchmark Values in Surgery

Performance Metric	Benchmark Tool
Center-specific case mix	Use the benchmark cases/center ratio to evaluate the number of difficult surgical cases in a center
Risk comparison among different patient populations undergoing the same surgical intervention	Use benchmarks to assess the gap in morbidity and mortality with higher risk groups
Estimating and comparing costs	Correlate morbidity in benchmark cases to costs to quantify financial burden in more complex cases
Self-evaluation	Select patients outside the benchmark CCI® value to be discussed at morbidity and mortality conferences
Surgical training	Define optimal teaching operations by selecting patients within benchmark criteria and thus establish minimal required teaching cases for a given center
Patient quality-of-life	Use benchmarks to quantitatively assess the negative impact of increased morbidity on quality-of-life in complex surgical procedures and selectively optimize postoperative patient care
Implementation of innovative surgical strategies and techniques	Select appropriate study candidates using benchmark criteria Compare the results to the established benchmarks for validation
Interdisciplinary comparisons	Comprehensively compare outcomes for a surgical procedure to those in other surgical fields by using benchmark values

independent observer and regular auditing.[49] While our study targeted those objectives, we observed variability in grade I complications among centers calling for suspicion of missing data. We therefore tested the impact of omitting grade I events for the CCI®. As shown in the inaugurating report on the CCI®, the contribution of grade I complications decrease exponentially with the occurrence of more severe complication (figure in supplementary material online, http://links.lww.com/SLA/B317).[32] Thus, considering high morbidity in the LT cohort, collecting grade I complications can be omitted in future studies with minimal consequence on outcome data.

Future possible applications of benchmarks could help to identify relevant performance metrics, and thus contribute to improve surgical care, techniques, and training (Table 3). In this context, contrarily to suggestions that reporting on highly selected cases may cause risk-averse behavior or threats to innovation due to a fear of not being able to match results and consequently being blamed,[50,51] we would argue that benchmarking rather offers new horizons. Benchmarking, as proposed here and in our previous work on major hepatectomy,[12] that is, following a strict methodology, offers reference values applicable somewhat universally, which is different from traditional comparisons, even when a heterogeneous group is risk adjusted. In our view, those comparisons are associated with 3 significant limitations. First, since no reference value is available, such studies provide only information related to the examined groups. Second, the lack of standardization in reporting weakens generalizability of most studies, which typically fail to provide convincing information on morbidity, and third understanding of the many risk adjustment methodologies is cumbersome for many clinicians. The concept of benchmarking, as reference values is simple, informative and thus likely to be adopted for many procedures. Benchmarking would, however, not replace major comprehensive databases like the American College of Surgeons National Surgical Quality Improvement Program, which best provide detailed outcome information on specific events on individual patients, who undergo specific interventions.

In conclusion, this multicentric study proposes 13 benchmark values, which may serve as reference to evaluate LT around the globe. Any patient or group of patients, who disclose an outcome fitting within the benchmark values, should be considered ''optimal'' results. Deviation from those values must of course be discussed with caution, and interpreted in the context of additional risk factors. Important subgroups of patients, for example those receiving a DCD graft or a second graft, may benefit from new benchmark studies targeting those specific populations. The next logical application of the current study may be to test how far is the outcome of marginal grafts or sicker recipients from benchmark values. Several of the authors currently select their cases to be presented at morbidity/mortality conference based on the benchmark cutoffs. We believe that this new concept of benchmarking, now also available for major liver resection and oesophagectomy,[12,52] may find wide acceptance in daily clinical practice and for future studies.

REFERENCES

1. Starzl TE, Marchioro TL, Vonkaulla KN, et al. Homotransplantation of the liver in humans. *Surg Gynecol Obstet*. 1963;117:659–676.
2. Adam R, Karam V, Delvart V, et al. Evolution of indications and results of liver transplantation in Europe. A report from the European Liver Transplant Registry (ELTR). *J Hepatol*. 2012;57:675–688.
3. Sharing UNfO. Transplants By Organ Type January 1, 1988–January 31, 2017 February 16, 2017. Available at: https://www.unos.org/data/transplant-trends/#transplants_by_organ_type+year+All. Accessed February 17, 2017.
4. Bismuth H. Consensus conference on indications of liver transplantation. Paris, France, June 22-23, 1993. *Hepatology*. 1994;20(1 pt 2):1S–68S.
5. Agopian VG, Petrowsky H, Kaldas FM, et al. The evolution of liver transplantation during 3 decades: analysis of 5347 consecutive liver transplants at a single center. *Ann Surg*. 2013;258:409–421.
6. Burroughs AK, Sabin CA, Rolles K, et al. 3-month and 12-month mortality after first liver transplant in adults in Europe: predictive models for outcome. *Lancet*. 2006;367:225–232.
7. Dutkowski P, Linecker M, DeOliveira ML, et al. Challenges to liver transplantation and strategies to improve outcomes. *Gastroenterology*. 2015;148:307–323.
8. Daniel KE, Eickhoff J, Lucey MR. Why do patients die after a liver transplantation? *Clin Transplant*. 2017;31:e12906.
9. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: a method for continuous quality improvement in health. *Healthc Policy*. 2012;7:e101–e119.
10. Brennan MF. Postoperative complication reporting: more than mortality and morbidity. *Ann Surg*. 2013;258:8–9.
11. Oelschlager BK, Ma KC, Soares RV, et al. A broad assessment of clinical outcomes after laparoscopic antireflux surgery. *Ann Surg*. 2012;256:87–94.
12. Rossler F, Sapisochin G, Song G, et al. Defining benchmarks for major liver surgery: a multicenter analysis of 5202 living liver donors. *Ann Surg*. 2016;264:492–500.

13. Foley DP, Fernandez LA, Leverson G, et al. Donation after cardiac death: the University of Wisconsin experience with liver transplantation. *Ann Surg*. 2005;242:724–731.
14. Ghabril M, Agarwal S, Lacerda M, et al. Portal vein thrombosis is a risk factor for poor early outcomes after liver transplantation: analysis of risk factors and outcomes for portal vein thrombosis in waitlisted patients. *Transplantation*. 2016;100:126–133.
15. Hashimoto K, Fujiki M, Quintini C, et al. Split liver transplantation in adults. *World J Gastroenterol*. 2016;22:7500–7506.
16. Levesque E, Winter A, Noorah Z, et al. Impact of acute-on-chronic liver failure on 90-day mortality following a first liver transplantation. *Liver Int*. 2017;37:684–693.
17. Dutkowski P, Oberkofler CE, Slankamenac K, et al. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. *Ann Surg*. 2011;254:745–753.
18. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001;33:464–470.
19. MdCalc. MELD Score (Original, Pre-2016, Model for End-Stage Liver Disease). Available at: https://www.mdcalc.com/meld-score-original-pre-2016-model-end-stage-liver-disease. Accessed March 8, 2017.
20. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124:91–96.
21. Dutkowski P, Schlegel A, Slankamenac K, et al. The use of fatty liver grafts in modern allocation systems: risk assessment by the balance of risk (BAR) score. *Ann Surg*. 2012;256:861–868.
22. Jay CL, Lyuksemburg V, Ladner DP, et al. Ischemic cholangiopathy after controlled donation after cardiac death liver transplantation: a meta-analysis. *Ann Surg*. 2011;253:259–264.
23. McLean KA, Camilleri-Brennan J, Knight SR, et al. Decision modelling in donation after circulatory death (DCD) liver transplantation. *Liver Transpl*. 2016;23:594–603.
24. Moon DB, Lee SG, Kang WH, et al. Adult living donor liver transplantation for acute-on-chronic liver failure in high- model for end-stage liver disease score patients. *Am J Transplant*. 2017;17:1833–1842.
25. Azoulay D, Castaing D, Adam R, et al. Split-liver transplantation for two adult recipients: feasibility and long-term outcomes. *Ann Surg*. 2001;233:565–574.
26. Birkmeyer JD, Gust C, Dimick JB, et al. Hospital quality and the cost of inpatient surgery in the United States. *Ann Surg*. 2012;255:1–5.
27. Vonlanthen R, Clavien PA. What factors affect mortality after surgery? *Lancet*. 2012;380:1034–1036.
28. Dimick JB, Osborne NH, Hall BL, et al. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg*. 2010;210:503–508.
29. Clavien PA, Barkun J, de Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. 2009;250:187–196.
30. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240:205–213.
31. Slankamenac K, Graf R, Barkun J, et al. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg*. 2013;258:1–7.
32. Slankamenac K, Nederlof N, Pessaux P, et al. The comprehensive complication index: a novel and more sensitive endpoint for assessing outcome and reducing sample size in randomized controlled trials. *Ann Surg*. 2014;260:757–762.
33. Kochhar G, Parungao JM, Hanouneh IA, et al. Biliary complications following liver transplantation. *World J Gastroenterol*. 2013;19:2841–2846.
34. Axelrod DA, Lentine KL, Xiao H, et al. National assessment of early biliary complications following liver transplantation: incidence and outcomes. *Liver Transpl*. 2014;20:446–456.
35. Oberkofler CE, Stocker R, Raptis DA, et al. Same quality—higher price? The paradox of allocation: the first national single center analysis after the implementation of the new Swiss transplantation law: the ICU view. *Clin Transplant*. 2011;25:921–928.
36. Oberkofler CE, Dutkowski P, Stocker R, et al. Model of end stage liver disease (MELD) score greater than 23 predicts length of stay in the ICU but not mortality in liver transplant recipients. *Crit Care*. 2010;14:R117.
37. Jochmans I, van Rosmalen M, Pirenne J, et al. Adult liver allocation in eurotransplant. *Transplantation*. 2017;101:1542–1550.
38. Johnson RJ, Bradbury LL, Martin K, et al. Organ donation and transplantation in the UK-the last decade: a report from the UK national transplant registry. *Transplantation*. 2014;97(suppl 1):S1–S27.
39. Jadlowiec CC, Taner T. Liver transplantation: current status and challenges. *World J Gastroenterol*. 2016;22:4438–4445.
40. Cameron AM, Ghobrial RM, Yersiz H, et al. Optimal utilization of donor grafts with extended criteria: a single-center experience in over 1000 liver transplants. *Ann Surg*. 2006;243:748–753.
41. Vonlanthen R, Slankamenac K, Breitenstein S, et al. The impact of complications on costs of major surgical procedures: a cost analysis of 1200 patients. *Ann Surg*. 2011;254:907–913.
42. McCulloch P, Schuller F. Innovation or regulation: IDEAL opportunity for consensus. *Lancet*. 2010;376:1034–1036.
43. Clavien PA, Puhan MA. Measuring and achieving the best possible outcomes in surgery. *Br J Surg*. 2017;104:1121–1122.
44. Clavien PA, Vetter D, Staiger RD, et al. The comprehensive complication index (CCI®)—added value and clinical perspectives 3 years ''down the line''. *Ann Surg*. 2017;265:1045–1050.
45. Nathan H, Dimick JB. Quality accounting: understanding the impact of multiple surgical complications. *Ann Surg*. 2017;265:1051–1052.
46. Yamashita S, Sheth RA, Niekamp AS, et al. Comprehensive complication index predicts cancer-specific survival after resection of colorectal metastases independent of RAS mutational status. *Ann Surg*. 2016 [Epub ahead of print].
47. Nederlof N, Slaman AE, van Hagen P, et al. Using the comprehensive complication index to assess the impact of neoadjuvant chemoradiotherapy on complication severity after esophagectomy for cancer. *Ann Surg Oncol*. 2016;23:3964–3971.
48. Kalisvaart M, de Haan JE, Polak WG, et al. Comparison of postoperative outcomes between donation after circulatory death and donation after brain death liver transplantation using the comprehensive complication indes. *Ann Surg*. 2017. Epub ahead of print.
49. Dindo D, Hahnloser D, Clavien PA. Quality assessment in surgery: riding a lame horse. *Ann Surg*. 2010;251:766–771.
50. Roberts JP. Impact of outcomes monitoring on innovation and risk in liver transplantation. *Liver Transpl*. 2012;18(suppl 2):S59–S63.
51. Abecassis MM, Burke R, Klintmalm GB, et al. American Society of Transplant Surgeons transplant center outcomes requirements—a threat to innovation. *Am J Transplant*. 2009;9:1279–1286.
52. Schmidt HM, Gisbertz SS, Moons J, et al. Defining benchmarks for transthoracic esophagectomy: a multicenter analysis of total minimally invasive esophagectomy in low risk patients. *Ann Surg*. 2017. Epub ahead of print.