

From Survival to Success

Evolving Metrics in Surgical Quality Assessment



Dr. med. Fariba Abbassi

This article reviews the evolution and current state of surgical quality assessment, emphasizing standardized outcome measures like the Clavien-Dindo Classification and Comprehensive Complication Index (CCI®), and introducing benchmarking for fair quality comparisons. It promotes for improved reporting, patient-centered metrics, and standardized guidelines for outcome evaluation.

The interest in measuring, comparing, and improving the quality of health care is enormous. However, the ideal mechanism of measuring quality in healthcare is under critical worldwide debate in many fields. In surgery, survival rates of surgical procedures were the main quality indicator for a long time. But with the dramatic decrease in perioperative mortality rates following most procedures, the focus has turned toward postoperative morbidity, and the incidence of postoperative complications has become one of the most used parameters to assess the quality of surgery in the literature.

However, the absence of standardized and universally accepted surgical definitions and endpoints has resulted in inconsistent, arbitrary, and frequently clinically inconsequential outcome assessments. This situation has paved the way for biased interpretations and has also impeded the improvement of healthcare quality (1, 2).

Definitions

Meaningful quality assessment and accurate outcome reporting require relevant, objective, standardized and well-defined outcome measures. These measures enable proper evaluation of postoperative results and comparative analyses between different therapeutic approaches or between different centers.

The definition of surgical complications was lacking until the early 1990's. In 1992, Clavien and his colleagues introduced the term "postoperative adverse outcome", which is based on complications, failure to cure, and sequelae (3):

Failure to cure: Signifies that the intended outcome of an intervention was not achieved (such as the absence of curative resection for a malignant tumor)

Sequelae: Refers to unfavorable events inherent to the procedure (for instance, amputation inevitably leading to disability)

Complications: Encompasses all other incidents.

Assessment tools

To prevent complications after an intervention and enable credible comparisons of competing therapies or care providers, standardized assessment tools are essential. These tools must be relevant to patients and healthcare providers, as well as all other stakeholders within society, and must be widely accepted across various healthcare systems and cultures. Several efforts were made to categorize surgical complications prior to 1990, yet none gained widespread acceptance or popularity. Therefore, when Clavien et al. compared laparoscopic cholecystectomy with the open technique in 1992, they not only had to redefine the term “complication” but also develop a standardized method for classifying them. This so-called Toronto classification determined the severity of morbidity based on the therapeutic treatment used for complication management (3). However, the dissemination of the classification remained very modest, with only a few citations in the literature. In 2004, Clavien and Dindo revised the fundamental Toronto model, leading to the “Clavien–Dindo Classification” (CDC), which consists of five grades, two of which are further subdivided into two subgroups (Table 1) (4). This new classification was validated in a cohort of 6336 patients who underwent elective surgery at the University Hospital of Zurich between 1988 and 1997. In addition, the acceptability and reproducibility of the classification was demonstrated in an international survey of surgeons at different levels of training. In conclusion, the CDC provides a simple and reliable method for standardized classification of the severity of complications and is the best established and most widely used classification in surgery today.

Table 1. The Clavien-Dindo Classification

Grades	Definition
Grade I	Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic and radiological interventions. Allowed therapeutic regimens are drugs as antiemetics, antipyretics, analgesics, diuretics and electrolytes and physiotherapy. This grade also includes wound infections opened at the bedside.
Grade II	Requiring pharmacological treatment with drugs other than such allowed for grade I complications. Blood transfusions and total parenteral nutrition are also included.
Grade III	Requiring surgical, endoscopic or radiological intervention
- <i>IIIa</i>	<i>Intervention not under general anesthesia</i>
- <i>IIIb</i>	<i>Intervention under general anesthesia</i>
Grade IV	Life-threatening complication (including CNS complications)* requiring IMC/ICU-management.
- <i>IVa</i>	<i>Single organ dysfunction (including dialysis)</i>
- <i>IVb</i>	<i>Multiorgan dysfunction</i>
Grade V	Death of a patient

*Brain hemorrhage, ischemic stroke, subarachnoidal bleeding, but excluding transient ischemic attacks (TIA); CNS: central nerve system; IMC: Intermediate care; ICU: Intensive care unit.

Table 1. Clavien–Dindo Classification versus CCI®.

Comparing outcomes of patients with multiple complications according to the CDC can be challenging. The CCI® simplifies this by reflecting a patient’s overall morbidity. It combines all complications and their relative severity into a single number, making it easier to compare outcomes of patients with multiple complications. With the CCI® you can easily say, that patient B had the worse postoperative course.


CCI®: Comprehensive Complication Index

Despite the widespread acceptance of the CDC, the classification is not used uniformly in the literature, leading to high interobserver variability. Numerous studies primarily report only on “major complications,” which in some cases is defined as CD grade ≥ 3 , while in others as CD grade $\geq 3b$, omitting complications of lesser severity (5). Furthermore, the management of complications can vary significantly among centers. Thanks to recent technical advances, treatments can now be performed endoscopically or via interventional radiology, no longer requiring immediate reoperation.

Consequently, many complications have been reclassified from Clavien-Dindo grade 3b to 3a, as they are now managed with local or regional anesthesia. These discrepancies are exacerbated by local circumstances and the availability of medical resources, and must be taken into account when interpreting study results.


A disadvantage of the CDC is that it requires extensive tabulation of complication details, making it difficult to compare outcomes, especially of patients with multiple complications (1). In response to this limitation, the Comprehensive Complication Index (CCI®) was developed in 2013. The CCI® is a metric that reflects the overall morbidity of an individual patient by summarizing all experienced complications and their relative severity into a single number, normalized on a scale from 0 (no complication) to 100 (death) (6, 7). Figure 1 illustrates the advantage of the CCI® in patients with multiple complications. The development of the CCI® formula explicitly considered the patients' perspective by allowing the patient to assign weights to the respective CD grades.

Which patient had the "worse" postoperative course?



Patient A

Patient A	
Wound infect	1
Abscess	3a
Gastric ulcer	3a



Patient B

Patient B	
Urinary infect	2
Severe pain	2
Bleeding	3b

Clavien-Dindo Classification		
	38	45
CCI®	38	45

To calculate the CCI®, a web application is available (<https://www.cci-calculator.com>), which requires only knowledge of the CD grades of all complications. The validity of the CCI® has been confirmed by evaluations in several independent patient groups (8, 9), showing a strong correlation with costs (10, 11) and serving as a highly sensitive endpoint for randomized trials (12, 13). In Table 2 some clinical scenarios are summarized to illustrate, how the CDC and CCI® can be used.

Table 2. Application of the Clavien-Dindo classification and the Comprehensive Complication Index (CCI®) in clinical scenarios.

Scenario	CD Classification	CCI*
A 45-year-old woman presented with abdominal pain 24 hours after a cholecystectomy. The liver function tests were found to be abnormal with elevated levels of pancreatic enzymes. An ERCP was performed, and a gallstone was removed from the common bile duct.	Grade IIIa for choledocholithiasis	26.2
A 25-year-old woman underwent appendectomy for perforated appendicitis. Postoperatively a deep vein thrombosis in the right leg was proven by phlebography, and full anticoagulation therapy was initiated on the ward. In addition, there was a severe paralytic ileus, which required prokinetic medication and parenteral nutrition.	Grade II for deep vein thrombosis, and grade II for paralytic ileus	29.6
A sigmoid resection was performed in a 56-year-old patient for acute diverticulitis. The patient developed fever and abdominal pain four days after surgery. A CT scan revealed an intra-abdominal abscess, which was treated by relaparotomy. The patient was discharged 2 weeks after surgery without further complications.	Grade IIIb for abscess	33.7
A 63-year-old man underwent a right hemi-liver resection for a hepatocellular carcinoma. The patient complained about right upper quadrant pain and developed a fever on postoperative day 4. A CT scan revealed an infected biloma in the right upper quadrant, which was drained percutaneously. After successful drainage of the biloma, the patient became acutely dyspneic due to a significant pneumothorax. A thorax drain was inserted to treat the pneumothorax. Finally, a subcutaneous seroma in the area of the abdominal wound had to be relieved bedside.	Grade IIIa for biloma, grade IIIa for pneumo-thorax, and grade I for wound seroma	38.1
A 62-year-old patient developed severe liver failure following a right hemihepatectomy. The patient became encephalopathic and hemodynamically unstable and was transferred to the ICU. He developed transient kidney failure requiring dialysis. A urinary tract infection was treated by antibiotics. The patient fully recovered after 3 weeks in the ICU.	Grade IVb for multi-organ failure, and grade II for urinary tract infection	50.7
A 47-year-old patient underwent a Whipple's procedure for a pancreatic cancer. The postoperative observation at the ICU was uneventful and the patient was shifted to the ward on postoperative day 1. After 4 days the fluid from a drain changed its color from serous to yellow. The fluid chemistry revealed a bilirubin of 500 µmol/l. The patient underwent a re-laparotomy confirming the diagnosis of an anastomotic leak of the hepaticojejunostomy. The bile duct was reconstructed. The next day, the patient developed significant ascites and developed anuric kidney failure. For this reason, the patient was readmitted to ICU for hemofiltration. Due to increasing amount of ascites, a bedside ultrasound was performed, demonstrating absence of flow in the portal vein. A CT-scan confirmed a portal vein thrombosis. The patient treated with therapeutic anticoagulation, and the amount of ascites subsequently decreased and stopped. Finally, the drains could be removed. After 3 weeks stay on the ICU the patient could be shifted to the floor with restored kidney function.	Grade IIIb for re-laparotomy, grade IVa for anuric kidney failure, and grade II for portal vein thrombosis	58.1

Table 2. Concept of Benchmarking.

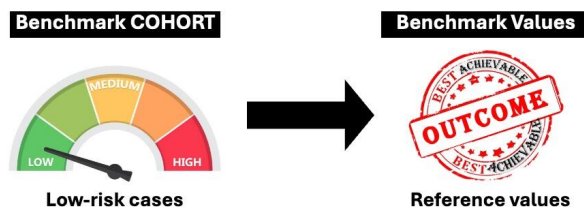
Data from international reference centers are collected and based on an ideal low-risk population, the so-called benchmark cohort, Benchmark cutoff values for different outcome parameters are determined, which can serve as reference values.

Meanwhile the CDC and CCI® were used in many centers and surgical disciplines worldwide for 20 and 10 years, respectively. In 2024, the pioneers developed recommendations to improve guidance on how to count and rate complications using the CDC and CCI® in scenarios that proved challenging over the years (14). In view of the steadily increasing use of CDC and CCI® in practice and research, these guidelines were crucial to further strengthen their consistent and standardized use.

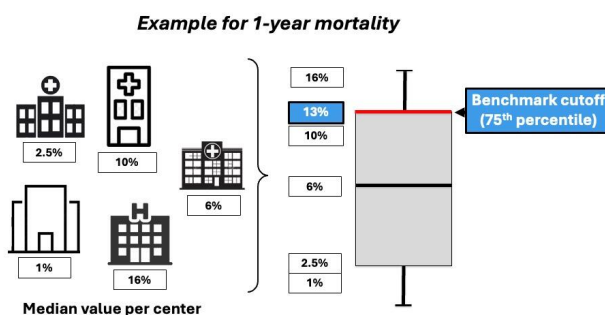
Benchmarking for fair quality comparisons

Fair comparisons among institutions or countries are difficult in heterogeneous groups of patients. Consequently, comparisons of quality indicators between centers should always be adjusted for the case-mix of patients (i.e. the proportion of low-risk vs. high-risk cases). Otherwise there is a danger of risk aversion, as providers may adopt avoidance strategies toward high-risk or complex cases (15, 16). To mitigate this risk, recently a novel approach called "Benchmarking" was proposed to estimate the best achievable outcomes for a given surgical procedure (17–19).

Originally, benchmarking comes from the economic practice, and its concept appeals to most surgeons: striving for the highest possible level of performance, not just the average. The aim of Benchmarking is to determine the best achievable outcome of a surgical procedure. Benchmark values for a specific surgical procedure are based on the outcomes of low-risk patients treated in international high-volume reference centers (Figure 2).



To create ambitious but achievable benchmarks, the benchmark cutoff is set at the 75th percentile of the centers' median (Figure 3).



To create a valid benchmark value, it is necessary to define the group of patients associated with the lowest risk for complications, such as young age, low body mass index, and the absence of comorbidities. Eligible centers for benchmark determination should be high-volume centers holding a prospective database, be involved in clinical research in the field of interest, and be from at least 2 continents (18, 19).

Standardized outcome reporting

Despite the availability of many tools to measure surgical outcomes in a standardized and objective manner, reporting has not improved over the last 20 years, as demonstrated by a recent critical appraisal of surgical literature (5). Major deficiencies in reporting surgical outcomes were observed in most journals, even at the highest levels, with a lack of information on key parameters such as readmission rates, outpatient events, overall morbidity, or patient-reported outcome measures (PROMs).

This situation has led to biased interpretations, especially when comparing competing treatment options or different care providers, and has hindered the improvement of healthcare quality (1, 2). In 2022, a major international effort brought together various stakeholders, including patients, payers, and policymakers, with the aim of developing guidelines for standardized and improved surgical outcome reporting, considering developments made over the past decades. Based on extensive literature reviews and expert input, an independent Jury offered a framework for outcome assessment and quality improvement following surgical procedures (20). The core document, published in Nature Medicine, presents the Jury's recommendations, integrating seven final statements:

1. Record outcome parameters at multiple, standardized time points perioperatively and postoperatively
2. Routinely use patient-reported outcomes in clinical care
3. Record morbidity using the CDC and the CCI®

4. Define benchmark values and compare results between centers and over time
5. Conduct routine interdisciplinary mortality and morbidity conferences
6. Appoint a 'data quality guarantor' at each institution
7. Follow the TRACK principle of transparency, respect, accountability, continuity and kindness in the event of unwarranted results.

In the same year, the Consolidated Standards for Reporting of Trials (CONSORT) statement published a CONSORT–Outcomes 2022 Extension (21), which integrates outcome reporting standards for clinical trials. Such reporting guidelines can greatly improve reporting if adopted and enforced by the scientific community and journals (22–27).

Patient-centered medicine

Finally, it is important to highlight the following aspect. Most metrics, except the CCI®, were developed solely from the healthcare providers' perspective, often overlooking the most important stakeholder: the patients. Physicians aim to provide the best care to patients. However, patients define "the best" using different criteria, such as postoperative functional status and nonmedical services like food quality. Therefore, to accurately assess surgical quality in both research and clinical practice, standardized assessment of outcomes from the patient's perspective including patient-reported outcome and experience measures (PROMs and PREMs) are essential and should be integral to any surgical outcome assessment. The importance of PROMs and PREMs has been widely demonstrated, providing valuable information for surgeons, healthcare providers, and policymakers to evaluate the effectiveness of surgical interventions. Surgeons can use these measures to assess quality of life, understand postoperative symptoms, and identify areas for improvement in patient care (28).

Additionally, there is an urgent need for a shift in the interpretation of study findings. Many authors find contentment when their analysis yields a P-value <0.05, often without considering whether the results also have clinical relevance (29). However, such consideration is paramount for patients and their families, healthcare providers, payers, and the public to make informed decisions. Thus, it is time for researchers to not only assess and report statistical significance but also clinical significance (29, 30). This is crucial, because P values merely indicate the presence of a treatment effect, without revealing its magnitude. Consequently, statistically significant findings may or may not be linked with clinically significant differences in treatment effects. The predominant concept used to delineate clinical significance is the minimal important difference (MID) (31). MID defines the smallest treatment effect that patients still perceive as beneficial and, in the absence of severe side effects and excessive costs, justifies a change in patient management.

In conclusion, the landscape of healthcare quality assessment in surgery is evolving rapidly, with an increasing emphasis on standardized outcome reporting and patient-centered measures. The adoption of tools like the CDC and CCI® has facilitated the objective evaluation of postoperative outcomes. However, the persistent lack of standardized outcome reporting across studies and journals remains a significant challenge, leading to biased interpretations and hindering quality improvement efforts. It is imperative for researchers, healthcare providers, and policymakers to embrace and enforce standardized outcome reporting guidelines. Only through consistent and

transparent reporting can we ensure meaningful comparisons and advancements in surgical care that truly benefit patients worldwide.

AUTOREN



Dr. med. Fariba Abbassi

Research Fellow

fariba.abbassi@usz.ch

REFERENZEN

1. Clavien PA, Puhan MA. Biased reporting in surgery. *Br J Surg.* 2014;101(6):591-2.
2. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet.* 1996;347(9007):984-5.
3. Clavien PA, Sanabria JR, Strasberg SM. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery.* 1992;111(5):518-26.
4. 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(2):205-13.
5. Abbassi F, Pfister M, Domenghino A, Puhan MA, Clavien PA. Surgical Outcome Reporting. Moving from a Comic to a Tragic Opera? *Annals of Surgery.* 2024:10.1097/SLA.0000000000006226.
6. Clavien PA, Vetter D, Staiger RD, Slankamenac K, Mehra T, Graf R, et al. The Comprehensive Complication Index (CCI(R)): Added Value and Clinical Perspectives 3 Years "Down the Line". *Ann Surg.* 2017;265(6):1045-50.
7. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien PA. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg.* 2013;258(1):1-7.
8. de la Plaza Llamas R, Ramia Angel JM, Bellon JM, Arteaga Peralta V, Garcia Amador C, Lopez Marcano AJ, et al. Clinical Validation of the Comprehensive Complication Index as a Measure of Postoperative Morbidity at a Surgical Department: A Prospective Study. *Ann Surg.* 2018;268(5):838-44.
9. Kim TH, Suh YS, Huh YJ, Son YG, Park JH, Yang JY, et al. The comprehensive complication index (CCI) is a more sensitive complication index than the conventional Clavien-Dindo classification in radical gastric cancer surgery. *Gastric Cancer.* 2018;21(1):171-81.
10. Dell-Kuster S, Gomes NV, Gawria L, Aghlmandi S, Aduse-Poku M, Bissett I, et al. Prospective validation of classification of intraoperative adverse events (ClassIntra): international, multicentre cohort study. *BMJ.* 2020;370:m2917.
11. Staiger RD, Cimino M, Javed A, Biondo S, Fondevila C, Perinel J, et al. The Comprehensive Complication Index (CCI®) is a Novel Cost Assessment Tool for Surgical Procedures. *Ann Surg.* 2018;268(5):784-91.
12. Boxhoorn L, van Dijk SM, van Grinsven J, Verdonk RC, Boermeester MA, Bollen TL, et al. Immediate versus Postponed Intervention for Infected Necrotizing Pancreatitis. *N Engl J Med.* 2021;385(15):1372-81.
13. Slankamenac K, Nederlof N, Pessaux P, de Jonge J, Wijnhoven BP, Breitenstein S, 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(5):757-62; discussion 62-3.

14. Abbassi F, Pfister M, Lucas KL, Domenghino A, Puhan MA, Clavien PA. Milestones in Surgical Complication Reporting. Clavien–Dindo Classification 20 years & Comprehensive Complication Index (CCI®) 10 years. *Annals of Surgery*. 2024;in Press.
15. Aloia TA. Should Zero Harm Be Our Goal? *Ann Surg*. 2020;271(1):33–6.
16. Vonlanthen R, Lodge P, Barkun JS, Farges O, Rogiers X, Soreide K, et al. Toward a Consensus on Centralization in Surgery. *Ann Surg*. 2018;268(5):712–24.
17. Rossler F, Sapisochin G, Song G, Lin YH, Simpson MA, Hasegawa K, et al. Defining Benchmarks for Major Liver Surgery: A multicenter Analysis of 5202 Living Liver Donors. *Ann Surg*. 2016;264(3):492–500.
18. Gero D, Muller X, Staiger RD, Gutschow CA, Vonlanthen R, Bueter M, et al. How to Establish Benchmarks for Surgical Outcomes?: A Checklist Based on an International Expert Delphi Consensus. *Ann Surg*. 2022;275(1):115–20.
19. Staiger RD, Schwandt H, Puhan MA, Clavien PA. Improving surgical outcomes through benchmarking. *Br J Surg*. 2019;106(1):59–64.
20. Domenghino A, Walbert C, Birrer DL, Puhan MA, Clavien P–A, Heuskel D, et al. Consensus recommendations on how to assess the quality of surgical interventions. *Nature Medicine*. 2023;29(4):811–22.
21. Butcher NJ, Monsour A, Mew EJ, Chan AW, Moher D, Mayo–Wilson E, et al. Guidelines for Reporting Outcomes in Trial Reports: The CONSORT–Outcomes 2022 Extension. *Jama*. 2022;328(22):2252–64.
22. Moher D, Jones A, Lepage L, Group ftC. Use of the CONSORT Statement and Quality of Reports of Randomized TrialsA Comparative Before–and–After Evaluation. *JAMA*. 2001;285(15):1992–5.
23. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews*. 2012(11).
24. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross–Sectional Study. *PLoS Med*. 2016;13(5):e1002028.
25. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta–analysis (PRISMA) statement on the quality of published systematic review and meta–analyses. *PLoS One*. 2013;8(12):e83138.
26. Agha RA, Fowler AJ, Limb C, Whitehurst K, Coe R, Sagoo H, et al. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. *Int J Surg*. 2016;30:169–72.
27. Leclercq V, Beaudart C, Ajamieh S, Rabenda V, Tirelli E, Bruyère O. Meta–analyses indexed in PsycINFO had a better completeness of reporting when they mention PRISMA. *J Clin Epidemiol*. 2019;115:46–54.
28. Chiche L, Yang HK, Abbassi F, Robles–Campos R, Stain SC, Ko CY, et al. Quality and Outcome Assessment for Surgery. *Ann Surg*. 2023;278(5):647–54.
29. Gikandi A, Hallet J, Koerkamp BG, Clark CJ, Lillemoe KD, Narayan RR, et al. Distinguishing Clinical From Statistical Significances in Contemporary Comparative Effectiveness Research. *Ann Surg*. 2024;279(6):907–12.
30. Puhan MA, Clavien PA. Is Statistical Significance Alone Obsolete?: Let’s Turn to Meaningful Interpretation of Scientific and Real–world Evidence on Surgical Care. *Ann Surg*. 2024;279(6):913–4.

31. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-15.