The comprehensive complication index

Slankamenac, Ksenija ; Nederlof, Nina ; Pessaux, Patrick ; de Jonge, Jeroen ; Wijnhoven, Bas P L ; Breitenstein, Stefan ; Oberkofler, Christian E ; Graf, Rolf ; Puhan, Milo A ; Clavien, Pierre-Alain

DOI: https://doi.org/10.1097/sla.0000000000000948

The Comprehensive Complication Index

A Novel and More Sensitive Endpoint for Assessing Outcome and Reducing Sample Size in Randomized Controlled Trials

Ksenija Slankamenac, MD, PhD, Nina Nederlof, MD,† Patrick Pessaux, MD,‡ Jeroen de Jonge, MD, PhD,† Bas P. L. Wijnhoven, MD, PhD,† Stefan Breitenstein, MD,* Christian E. Oberkofler, MD,* Rolf Graf, PhD,* Milo A. Puhan, MD, PhD,§ and Pierre-Alain Clavien, MD, PhD*¶*

Objective: To test whether the newly developed comprehensive complication index (CCI) is more sensitive than traditional endpoints for detecting between-group differences in randomized controlled trials (RCTs).

Background: A major challenge in RCTs is the choice of optimal endpoints to detect treatment effects. Mortality is no longer a sufficient marker in studies, and morbidity is often poorly defined. The CCI, integrating all complications including their severity in a linear scale ranging from 0 (no complication) to 100 (death), is a new tool, which may be more sensitive than other traditional endpoints to detect treatment effects on postoperative morbidity.

Methods: The CCI was tested in 3 published RCTs from European centers evaluating pancreas, esophageal and colon resections. To compare the sensitivity of the CCI with traditional morbidity endpoints, for example, presence of any (yes/no) or only the most severe complications, all postoperative events were assessed, and the CCI calculated. Treatment effects and sample size calculations were compared using the CCI and traditional endpoints.

Results: Although RCTs failed to show between-group differences using any or most severe complications, the CCI revealed significant differences between treatment groups in 2 RCTs—after pancreas ($P = 0.009$) and esophageal surgery ($P = 0.014$). The CCI in the RCT on colon resections confirmed the absence of between-group differences ($P = 0.39$). The required sample sizes in trials are up to 9 times lower for the CCI than for traditional morbidity endpoints.

Conclusions: This study demonstrates superiority of the CCI to traditional endpoints. The CCI may serve as an appealing endpoint for future RCTs and may reduce the sample size.

Keywords: complication, comprehensive complication index, morbidity, outcome research, randomized controlled trial, sample size, valid endpoint

(*Ann Surg* 2014;260:757–763)

From the *Department of Surgery and Transplantation, University Hospital of Zurich, Zurich, Switzerland; †Department of Surgery, Division of Gastrointestinal and Transplantation Surgery, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands; ‡Pôle des Pathologies Digestives, Hépatiques et de la Transplantation, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Université de Strasbourg, Strasbourg, France; §Institute of Social and Preventive Medicine, University of Zurich, Zurich, Switzerland.; and ¶Hospital Paul Brousse, Université Paris Sud, Paris, France

S.K. and N.N. contributed equally to this work.

Disclosure: K.S. was supported by the PROSPER fellowship of the Swiss National Foundation (SNF 131633 and SNF 151049). P.A.C. was supported by a grant from the Liver and Gastrointestinal Disease (LGID) Foundation. P.A.C., M.P., K.S., and R.G. are members of the advisory board of the AssesSurgery GmbH. The authors declare no conflicts of interests.

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, University Hospital Zurich, Raemistrasse 100, CH-8091 Zurich, Switzerland. E-mail: clavien@access.uzh.ch.

ISSN: 0003-4932/14/26005-0757

DOI: 10.1097/SLA.0000000000000948

A major challenge in designing randomized controlled trials (RCTs) is the choice for objective, concise, and clinically relevant endpoints. Mortality is no longer an acceptable primary endpoint in surgical studies given the sharp decline in mortality for most procedures in the past decades. Morbidity is often poorly defined, which has led to inconsistent reporting and confusion in the literature.[1–7] Furthermore, most authors have reported only the most severe complications or only events judged to be relevant but ignored complications of lesser magnitude as well as the total number of complications.[5] To address this issue, the Comprehensive Complication Index (CCI) was recently introduced. It integrates all postoperative complications with their respective severities, on a scale ranging from 0 (no burden from complications) to 100 (death).[8]

The CCI, summarizing the entire postoperative experience of the patient with respect to complications, is based on the widely established Clavien-Dindo classification.[3,8] Validations from 4 different perspectives showed that the CCI is a valid endpoint for postoperative overall morbidity. Although the CCI is an attractive novel tool, which may serve as a primary or secondary endpoint in many types of studies, the external validity has not been tested in RCTs. Therefore, the aim of this study was to externally evaluate whether the CCI is more sensitive than traditional primary endpoints in detecting between-group differences.

METHODS

We externally validated the CCI[8] on recently published RCTs[9-11] that reported specific complications after different surgical procedures. The first step was to identify and contact a number of centers, which have conducted RCTs investigating specific and nonspecific morbidity endpoints. We considered RCTs regardless of their conclusions in the original analysis and focused on different types and complexity of general surgical procedures, different diseases, and countries. All RCTs with a proper study design according to the CONSORT guidelines were identified by a systematic literature search in peer-reviewed high impact journals in the last 3 years (2011-2013)[12-15] (Supplemental Digital Content 1: flow diagram, available at http://links.lww.com/SLA/A658). After contacting several centers in general, cardiac, and plastic surgery, we were granted full access to primary data and each patient record of 3 European trials[9-11] addressing different surgical interventions and diseases in compliance with our study design. The first trial focused on the rate of pancreatic fistulas after pancreaticoduodenectomy.[11] The second trial was evaluating the rate of anastomotic strictures after 2 different types of anastomosis after esophagectomy[9] and, finally, the third trial focused on the rate of overall complications after colon resection for perforated diverticulitis.[10]

Primary data[9–11] were reanalyzed including calculation of the CCI in each patient. The CCI of the comparative groups in each trial was tested along traditional reported morbidity endpoints in the literature including the presence of any complication (yes/no)

and the most severe complications (≥grade IIIb according to the Clavien-Dindo classification[3]). The first author (K.S.) visited each center to secure consistent and exhaustive reevaluation of the data.

The multicenter RCT originating from France, published in 2011, was designed to test whether an external pancreatic duct stent might reduce the rate of pancreatic fistulas after pancreaticoduodenectomy.[11] A sample size calculation postulated a 10% reduction in the incidence of pancreatic fistulas in patients with pancreatic stents compared to those without drainage[11]. Assuming a power of 80% and an α-error of 0.05, the investigators enrolled 158 patients comparing 77 patients with, versus 81 without insertion of an external stent drainage (Fig. 1).[11] The results indicated that external stent drainage of the pancreatic duct significantly reduces the risk of pancreatic fistulas, as well as overall morbidity rates after pancreaticoduodenectomy.[11]

The second RCT, performed in the Netherlands from 2005 to 2007, compared an end-to-end (E-E) with end-to-side (E-S) esophagogastrostomy after esophageal cancer resection (Fig. 1).[9] Their primary endpoint was the development of anastomotic stricture and need for dilatation within 1 year after surgery. They, therefore, performed a noninferiority trial with one-sided testing assuming a 50% reduction in the rate of stenosis in patients with E-S comparing to E-E anastomosis. Assuming a power of 80% and an α-error of 0.05, 64 patients per group were required. The authors observed a lower incidence of anastomotic stricture in patients with E-S anastomosis. They also concluded that E-S anastomosis was associated with significantly more anastomotic leaks than E-E anastomosis.[9] The authors, however, did not investigate the overall morbidity expressed as the presence of any complication, nor the most severe complication (≥grade IIIb).

Finally, the third multicenter RCT focused on patients with perforated left-sided diverticulitis enrolled from 4 surgical centers in Switzerland (including the Department of Surgery at the University Hospital Zurich).[10] The study was designed to test whether the conventional Hartmann's procedure (HP: colonic resection with closure of the rectal stump and an end-colostomy) is comparable to a primary anastomosis (PA) with diverting ileostomy (Fig. 1) using the rate of overall complications, regardless of the severity, as the main endpoint. Both strategies require a second operation, that is, stoma reversal and in HP a reestablishment of the continuity of the colon. For the sample size calculation, 25% reduction in the rate of overall complications was assumed with a power of 80% and an α-error of 0.05. This yielded an estimated group size of 68 patients. The planned interim analysis, enrolling 62 patients, lead to a discontinuation of the trial, as recommended by the data-monitoring board, due to significant differences in adequately powered relevant secondary endpoints. Thirty patients had been randomized for the HP procedure and 32 for a primary anastomosis. Although the overall complication rates for both resection and stoma reversal operations were comparable, severe complications (grades ≥IIIb) were significantly reduced after reversal operations in the PA group.[10]

Data Collection and Primary Endpoint

We used the existing databases of these 3 RCTs and calculated the CCI for each patient.[9–11] Each postoperative event in each patient was assessed on site and graded according to the Clavien-Dindo classification, which is based on the treatment used to correct the postoperative complication.[3] For the development of the CCI, we used the established Clavien-Dindo classification system[3] for complications, adopting methods from operation-risk-index analysis in marketing research and developed a formula that considers any combination of complications.[16–18] The CCI was finally calculated as the sum of all complications that are weighted for their severity by patients and physicians.[8,19] The final formula yields a score from 0 (*no complication*) to 100 (*death*).[8] The CCI can easily be calculated online by free access at www.assessurgery.com.

In addition, for each patient, we assessed traditional morbidity endpoints: the total number of complications, the presence of any (yes/no), and the most severe (≥grade IIIb[3]) complications. In the esophageal stricture trial, we also calculated the CCI at discharge and a long-term CCI after 1 year after initial surgery in all patients. The longitudinal assessment of the overall morbidity using the CCI is novel allowing to present the cumulative effects of complication over time.

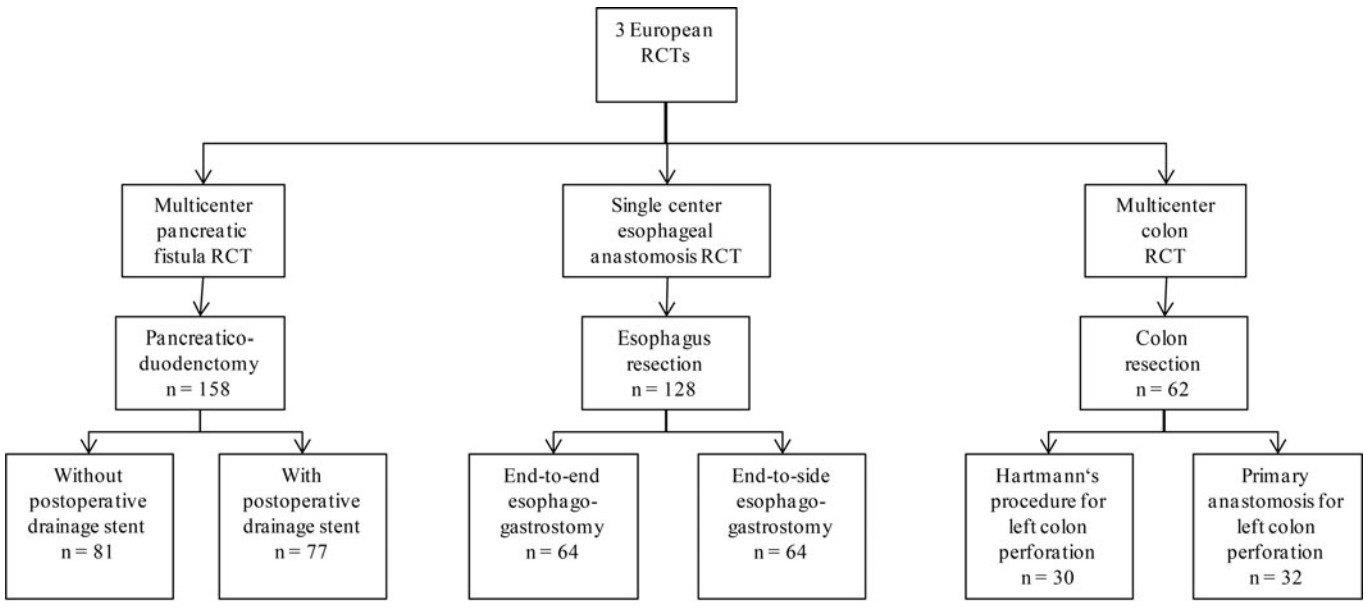

FIGURE 1. The 3 European randomized controlled trials.[9–11]

In-hospital costs, the length of hospital stay (LOS) and intensive care unit (ICU) stay were also extracted from the respective original database. LOS and ICU stay were available in all 3 RCTs, whereas the in-hospital costs were only available in the multicenter colon RCT.

We additionally performed a sample size calculation to properly evaluate the sensitivity of the CCI as primary endpoint in trials. We assumed a difference of 10 points for the CCI, a relative risk reduction of 40% for the specific and traditional morbidity endpoints, a power of 80%, and an α-error of 0.05. The difference of 10 points on the CCI scale is chosen because it reflects 1-grade difference in the established Clavien-Dindo classification. For the colon trial, we were not able to calculate a sample size because of premature termination of the initial trial, a 2-step procedure and therefore unequal group sizes (n = 15 vs n = 22) in the second surgery.

Eligibility criteria (Supplementary Digital Content 2, available at http://links.lww.com/SLA/A658) and the paragraph about the statistical analysis (Supplementary Digital Content 3, available at http://links.lww.com/SLA/A658) are reported in detail in the supplemental materials.[20–22] For all results, we reported point estimates, 95% confidence intervals (CIs), and *P* values (≤0.05 considered significant). We performed statistical analyses using STATA (version 11, Stata Corp, College Station, Texas).

RESULTS

As summarized in Figure 1, we reanalyzed 348 patients for which detailed characteristics were reported previously.[9–11]

Conventional Morbidity Endpoints Versus CCI

In the *pancreatic fistula trial*, the overall burden of the postoperative morbidity represented by the median CCI after pancreaticoduodenectomy was significantly lower in patients with an external pancreatic stent than in those without any drainage [0 (IQR [interquartile range]: 0–26.2) vs 20.9 (IQR: 0–29.6), $P = 0.009$] (Table 1). In contrast, there was no between-group difference considering the "most severe complication" in patients with the external stent compared with those ones without stent (13% vs 11%, 95% CI: 0.5–3.1, $P = 0.72$). The authors of this trial published a significant between-group difference in the rate of pancreatic fistulas (42% without stent vs 26% with stent, $P = 0.035$).[11] These findings support the higher sensitivity of the CCI because the between-group difference for the CCI presented lower *P* values ($P = 0.009$) as indirect comparison of effect sizes, than the differences for the specific complication "pancreatic fistula" ($P = 0.035$), as well as for the "most severe complication" ($P = 0.72$) in the same patient population.[20,21] The *P* values for the CCI ($P = 0.009$) compared to those ones for the presence of any complications ($P = 0.008$) were similar in the same patient population (Table 1).

In the esophageal anastomosis trial, patients with an E-S anastomosis disclosed significantly higher CCI at discharge than patients with an E-E anastomosis [22.6 (IQR: 0–41.2) vs 10.5 (IQR: 0–24.4), $P = 0.014$]. In contrast, this trial failed to show any statistically significant between-group differences using traditional endpoints such as the presence of any or the most severe complications (≥grade IIIb) (Table 2).

TABLE 1. Effects of Stent Versus No Stent on Postsurgical Morbidity Using Different Measurements in Patients With Pancreatic Fistula Following Pancreaticoduodenectomy[11]

	Without Stent (n = 81)	With Stent (n = 77)	Unadjusted Difference (95% CI, *P*)
CCI	20.9 (0–29.6)	0 (0–26.2)	− 12.2 (−4.5 to −16.7, $P = 0.009$)
CCI of the pancreatic fistula	0 (0–20.9)	0 (0–8.7)	− 3.2 (−6.8 to 0.5, $P = 0.091$)
Number of complications	1 (0–2)	0 (0–1)	− 0.5 (−0.9 to −0.1, $P = 0.021$)
	Without Stent (n = 81)	**With Stent (n = 77)**	**Unadjusted Odds Ratio (95% CI, *P*)**
Presence of any complication (%)	50 (61.7%)	32 (41.5%)	0.4 (0.2–0.8, $P = 0.008$)
Severe complications ≥ IIIb* (%)	9 (11.1%)	10 (13.0%)	1.2 (0.5–3.1, $P = 0.72$)
Pancreatic fistula (%)	34 (42%)	20 (26%)	0.5 (0.2–0.95, $P = 0.035$)

CI = confidence interval, CCI = Comprehensive Complication Index; all results reported as median and interquartile range; *grading of complications according to the Clavien-Dindo classification system3.

TABLE 2. Effects of End-to-End Versus End-to-Side Anastomosis on Postsurgical Morbidity Using Different Measurements in Patients With Anastomotic Stricture After Esophagectomy[9]

	E-E (n = 64)	E-S (n = 64)	Unadjusted Difference (95% CI, *P*)
CCI at discharge	10.5 (0–24.4)	22.6 (0–41.2)	11.6 (2.4–20.8, $P = 0.014$)
CCI after 1 year	26.2 (20.9–40.6)	26.2 (8.7–38.2)	− 1.4 (−10.3 to 7.4, $P = 0.75$)
No. complications	1 (0–2)	1 (0–2.5)	0.5 (−0.1 to 1.1, $P = 0.08$)
	E-E (n = 64)	**E-S (n = 64)**	**Unadjusted Odds Ratio (95% CI, *P*)**
Presence of any complication (%)	42 (65.6%)	47 (73.4%)	1.5 (0.7–3.1, $P = 0.34$)
Severe complications ≥ IIIb* (%)	11 (17.2%)	20 (31.3%)	2.2 (0.95–5.1, $P = 0.07$)
Anastomotic stricture (%) after 1 year	20 (40%)	10 (18%)	0.3 (0.1–0.7, $P = 0.004$)
Anastomotic leakage (%)	14 (22%)	26 (41%)	3.4 (1.4–8.2, $P = 0.04$)

All results were presented in median and IQR.
*Grading of complications according to the Clavien-Dindo classification system.[3]

The original analysis of this trial reported a significantly higher rate of stricture in the anastomosis 1 year after the initial surgery (40% in E-E vs 18% in E-S, $P < 0.01$).[9] The CCI directly after discharge was significantly different between both groups, mostly related to a significantly higher rate of anastomotic leaks in the E-S anastomosis group [odds ratio: 3.4 (95% CI: 1.4–8.2, $P = 0.04$]. After calculating the CCI at 1-year follow-up (including stricture), however, the median CCI was similar between patients with E-E and E-S anastomosis [26.2 (20.9–40.6) vs 26.2 (8.7–38.2), $P = 0.75$]. This new finding suggests that there is no long-term difference (after 1 year) in the morbidity between both types of anastomoses. The CCI allows reporting of complications that occur during different time periods. After 1-year follow-up, the CCI ($P = 0.75$) balances the initially higher rate of anastomotic leaks in patients with E-S anastomosis ($P = 0.04$) with the increasing rate of anastomotic strictures in patients with E-E anastomosis (Table 2). Thus, although initially 1 patient group seems to have an advantage, long-term observations reverse this observation by considering late complications experienced by the other group of patients.

In the *multicenter colon trial*, there was no significant difference in the overall CCI, neither for the overall procedures nor in the respective steps (first and second operation). Nevertheless, the between-group difference of the CCI demonstrated a lower *P* value than the differences in traditional morbidity endpoints that emphasize the higher sensitivity of the CCI over the traditional endpoints (Table 3).[20,21] This trial showed comparable between-group complication rates for both surgical steps. Comparing the outcome of the first operation between the groups, there was also no significant difference in the rate of severe complications (44% vs 37%, $P = 0.57$; Table 3).

The CCI and the Sample Size

For 2 trials, a sample size calculation was performed for their specific endpoint, the CCI and traditional morbidity endpoints (any or most severe complications). The sample sizes are clearly lower for the CCI compared to the original and traditional endpoints as shown in Table 4. This illustrates the putative benefits of the CCI compared to complication endpoints such as any and more severe in minimizing the need for large sample sizes in the future. For example, in the *pancreatic fistula trial*, the required sample size would decrease from 695 patients/group to 76 patients per group when using the primary endpoint CCI *vs.* "most severe complication". Similar results were seen in the *esophageal anastomosis* trial (Table 4).

CCI Associated to LOS, ICU Stay, and In-Hospital Costs

The CCI was significantly associated to LOS and ICU stay in all 3 trials (Supplementary Digital Content 4, available at http://links.lww.com/SLA/A658). We also evaluated the costs of

TABLE 3. Effects of the Hartmann's Procedure Versus Primary Anastomosis on Postsurgical Morbidity Using Different Measurements in Patients With Perforated Diverticulitis[10]

	Hartmann's Procedure n = 30	Primary Anastomosis n = 32	Unadjusted Difference (95% CI, *P*)
CCI for both surgeries	40.3 ± 32.6	33.5 ± 28.3	6.8 (−8.7 to 22.3, $P = 0.39$)
CCI after first surgery	37.3 ± 33.1	32.2 ± 28.4	5.1 (−10.5 to 20.7, $P = 0.52$)
CCI after second surgery	n = 15, 12.4 ± 16.7	n = 22, 5.2 ± 9.7	7.2 (−1.7 to 16.0, $P = 0.11$)
	Hartmann's Procedure	**Primary Anastomosis**	**Unadjusted Odds Ratio (95% CI, *P*)**
First surgery	**n = 30**	**n = 32**	
Any morbidity (%)	24 (80%)	27 (84.4%)	0.74 (0.20–2.74; $P = 0.65$)
Severe complications ≥ IIIb* (%)	11 (36.7%)	14 (43.8%)	0.74 (0.27–2.1, $P = 0.57$)
Second surgery	**n = 15**	**n = 22**	**Unadjusted Odds Ratio (95% CI, *P*)**
Any morbidity (%)	6 (40%)	6 (27.3%)	1.78 (0.44–7.18, $P = 0.42$)
Severe complications ≥ IIIb* (%)	3 (20%)	0%	—

All results reported as mean ± standard deviation. No statistical analysis was performed if ≤5 events in a group.
*Grading of complications according to the Clavien-Dindo classification system.[3]

TABLE 4. Sample Size Calculation for Surgical RCTs Using Different Measurements for Postsurgical Morbidity

	Assumptions	Sample Size
Pancreatic fistula trial[11]		
Pancreatic fistula[11]	40% relative risk reduction	149 patients/group
Presence of any complication (yes/no)	40% relative risk reduction	75 patients/group
Most severe complication ≥ IIIb*	40% relative risk reduction	695 patients/group
CCI	Δ 10 points, SD = 22	76 patients/group
Esophageal anastomosis trial[9]		
Anastomotic stricture[9]	40% relative risk reduction	132 patients/group
Presence of any complication (yes/no)	40% relative risk reduction	76 patients/group
Most severe complication ≥ IIIb*	40% relative risk reduction	220 patients/group
CCI	Δ 10 points, SD = 20	63 patients/group

$\beta = 0.8$; $\alpha = 0.05$; Δ = difference in 10 points in the CCI scale.
*Grading of complications according to the Clavien-Dindo classification system.[3] Sample sizes were shown without considering loss of follow-up.

complications in the colon trial. An increase in one point on the CCI scale created additional costs of Swiss Franc (CHF) 883 (95% CI: 222–1543, $P = 0.010$) (about US$ 980, 95% CI: 247–1714). In other words, an increase of the CCI of 10 points on the scale increases the in-hospital costs to additional US$ 9800 (Supplementary Digital Content 4, available at http://links.lww.com/SLA/A658).

DISCUSSION

This study demonstrates the superiority of the CCI over traditionally reported morbidity endpoints "most severe complication" and specific complications by detecting between-group differences in 3 external trial populations. Another finding is the easy and new applicability to longitudinal assessment of complications over time, as illustrated in the analysis of the 1-year CCI follow-up in the *esophageal anastomosis* trial. Next, the CCI is associated to LOS, length of ICU stay, and in-hospital costs, which add clinical value to this morbidity index. Finally, the most relevant finding for the CCI is that the required sample sizes in trials are up to 9 times lower for the CCI than for other endpoints.

Reporting outcomes of surgical or other invasive procedures using morbidity, as the primary endpoint, has been associated with serious limitations because of various definitions and different interpretation of postoperative events.[1,2,4,5] Assessing the overall morbidity by the presence of any complication causes the problem of ignoring either the number of different complications occurring in a patient after surgery or, more importantly, the severity of complications. Recording only the most severe complication does not give weight to either any complication of lesser importance or the total number of complications, even though they affect the patient. As one of the first attempts in outcome standardization, in 1992, a classification system was proposed to grade the severity of complications according to the degree of treatment needed to correct the complications.[23] In 2004, this original classification was revised to generate the "Clavien-Dindo classification" on the basis of the same principles but eliminating criteria such as the length of stay and newly grading complications with readmission to ICU units because of organ dysfunction.[3,6] The Clavien-Dindo classification gained wide acceptance and became increasingly used in a variety of studies and registries.[24–28] However, with this system, each complication is graded separately. For ease of reporting, usually only the most severe complication was included, which does not represent the "true" overall morbidity burden of surgical procedures.[5] The recently developed CCI is based on a formula used in the economy, which incorporates multiple factors influencing the globalization of a corporation decision. With this formula, all complications, weighted by severity, are integrated in a linear scale. It facilitates reporting not only of the in-hospital morbidity but also at various postoperative follow-up, for example, the 90-day morbidity or other time-span. Furthermore, the CCI is a primary outcome measure, which is calculated separately for each patient regardless of the population studied. Individual grade of complications or the CCI represent endpoints, and thus a risk-adjustment is necessary for proper interpretation in specific groups of patients. For example, a higher median CCI in a hospital compared to others might not mean better quality, particularly if the population of this specific center is at higher risk for surgery with high incidence of comorbidities.

If future trials aim at focusing on the "overall morbidity" and not on a specific complication related to the procedure under investigation, our current data strongly support the use of the CCI as a primary endpoint. It seems that the required sample size for superiority trials is impressively lower for the CCI than traditional endpoints, particularly the "most severe complications" endpoint. Surgical trials often require large sample sizes to detect a between-group difference, which are only feasible in large and costly multicenter endeavor. Switching to the CCI may result in a dramatic reduction of the required sample size, so the feasibility of a trial increases, whereas costs decrease. In addition, the number of negative trials associated with a type II error may be reduced. It is obvious that there is a substantial number of false-negative results in the surgical trial literature that are solely the result of insensitive endpoints.[29]

The strength of the current study is that the CCI was externally tested and was apparently more responsive than existing outcome endpoints. It may, therefore, serve as a standardized and easily applicable primary endpoint in surgical trials and other medical specialties. Furthermore, the CCI was performed on 3 different European patient populations enrolled for different surgical procedures and diseases supporting its broad use. Further strength in this report is the 1-year CCI as a longitudinal measure of the overall morbidity over a certain time. The 3 trials all enrolled patients by randomization to control for confounding factors and bias.

There are also some limitations. Even though the CCI was externally tested on a broad spectrum of patients undergoing a variety of major abdominal surgical procedures, it might be still important to test the CCI in other medical fields such as interventional radiology, urology, and cardiology. A further limitation might be that the 3 trials were not powered for the CCI as primary endpoint. Nevertheless, this study already showed in rather small patient populations significant between-group differences in the CCI whereas the differences in the frequencies of any and/or most severe complications were similar.

CONCLUSIONS

The CCI offers a novel sensitive endpoint for clinical trials that can be readily calculated (available at *www.assessurgery.com*). The CCI may also allow in future better information of patients, standardized reporting in outcome research including readily available assessment of morbidity at various time points, and increased comparability of quality of surgery across centers worldwide. Perhaps, one of the most attractive aspects of the CCI is the possibility to conduct conclusive trials with smaller sample sizes when focusing on the "overall morbidity" as surgical outcome.

ACKNOWLEDGMENTS

The authors thank all other coauthors from the 3 trials[9–11] for offering their patient databases to test the CCI on different patient populations.

REFERENCES

1. Feldman L, Barkun J, Barkun A, et al. Measuring postoperative complications in general surgery patients using an outcomes-based strategy: comparison with complications presented at morbidity and mortality rounds. *Surgery*. 1997;122:711–719; discussion 719–720.
2. Pomposelli JJ, Gupta SK, Zacharoulis DC, et al. Surgical complication outcome (SCOUT) score: a new method to evaluate quality of care in vascular surgery. *J Vasc Surg*. 1997;25:1007–1014; discussion 1014–1015.
3. 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.
4. Strasberg SM, Linehan DC, Clavien PA, et al. Proposal for definition and severity grading of pancreatic anastomosis failure and pancreatic occlusion failure. *Surgery*. 2007;141:420–426.
5. Sugawara Y, Tamura S, Makuuchi M. Systematic grading of surgical complications in live liver donors. *Liver Transpl*. 2007;13:781–782.
6. 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.
7. Marsh JW, Gray E, Ness R, et al. Complications of right lobe living donor liver transplantation. *J Hepatol*. 2009;51:715–724.
8. 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.
9. Nederlof N, Tilanus HW, Tran TC, et al. End-to-end versus end-to-side esophagogastrostomy after esophageal cancer resection: a prospective randomized study. *Ann Surg*. 2011;254:226–233.
10. Oberkofler CE, Rickenbacher A, Raptis DA, et al. A multicenter randomized clinical trial of primary anastomosis or Hartmann's procedure for

perforated left colonic diverticulitis with purulent or fecal peritonitis. *Ann Surg*. 2012;256:819–826; discussion 826–827.
11. Pessaux P, Sauvanet A, Mariette C, et al. External pancreatic duct stent decreases pancreatic fistula rate after pancreaticoduodenectomy: prospective multicenter randomized trial. *Ann Surg*. 2011;253:879–885.
12. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637–639.
13. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42–46.
14. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005;365:1159–1162.
15. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
16. Meffert H, Bolz J. *Internationales Marketing-Management*. 3rd ed. Stuttgart, Germany: Kohlhammer; 1998:78.
17. Welge M, Holtbrügge D. *Internationales Managment: Theorien, Funktione, Fallstudien*. 5th ed. Stuttgart, Germany: Schaeffer-Poeschel; 2006:100–121.
18. Zentes J, Swoboda B, Schramm-Klein H. *Internationales Marketing*. 2nd ed. Vahlen, Germany: Munchen; 2006:180.
19. Slankamenac K, Graf R, Puhan MA, et al. Perception of surgical complications among patients, nurses and physicians: a prospective cross-sectional survey. *Patient Saf Surg*. 2011;5:30.
20. Altman DG, Gore SM, Gardner MJ, et al. Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)*. 1983;286:1489–1493.
21. Ellis PD. *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*. Cambridge, UK: Cambridge University Press; 2010.
22. 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.
23. Clavien PA, Camargo CA, Jr, Croxford R, et al. Definition and classification of negative outcomes in solid organ transplantation. Application in liver transplantation. *Ann Surg*. 1994;220:109–120.
24. Tefekli A, Ali Karadag M, Tepeler K, et al. Classification of percutaneous nephrolithotomy complications using the modified Clavien grading system: looking for a standard. *Eur Urol*. 2008;53:184–190.
25. Rabbani F, Yunis LH, Pinochet R, et al. Comprehensive standardized report of complications of retropubic and laparoscopic radical prostatectomy. *Eur Urol*. 2010;57:371–386.
26. Seely AJ, Ivanovic J, Threader J, et al. Systematic classification of morbidity and mortality after thoracic surgery. *Ann Thorac Surg*. 2010;90:936–942; discussion 942.
27. Mitropoulos D, Artibani W, Graefen M, et al. Reporting and grading of complications after urologic surgical procedures: an ad hoc EAU guidelines panel assessment and recommendations. *Eur Urol*. 2012;61:341–349.
28. Ubbink DT, Visser A, Gouma DJ, et al. Registration of surgical adverse outcomes: a reliability study in a university hospital. *BMJ Open*. 2012;2:e000891.
29. Clavien PA, Puhan MA. Biased reporting in surgery. *Br J Surg*. 2014;101:591–592.

DISCUSSANTS

A. Fingerhut (Paris, France):

First of all, I thank the organizing committee for the privilege of commenting on this paper, complementing the previous, recently published paper in *Annals of Surgery* from the same team on the CCI (K. Slankamenac et al, *Ann Surg*. 2013;258:1–7). The senior author, Professor Clavien, has devoted more than 20 years of his career to postoperative complications, in terms of how to grade their severity and evaluating their impact on postoperative outcomes. I had the privilege of seeing and correcting the very first paper published on the subject, in Europe, when I taught medical writing in Switzerland several years ago.

I have several comments and questions.

One endpoint that was used in France for many years, when the current French Association for Research "FRENCH" was called "ARC" (Association pour la Recherche en Chirurgie), was the number of patients, which had one complication or more. Although a possible drawback of this endpoint might be that, admittedly, it may have overestimated rates of severe complications, it also has to be said that it never underestimated morbidity, which may be the case with many of the other systems, including this one, when subjective judgment, albeit limited here (the so-called "operation risk index" and use of analogue scales), intervenes. My first question is as follows: is our "ARC" endpoint (number of patients with 1 or more complications) much different from your calculation in Table 4: "presence (or not) of at least 1 complication"?

The difference seems to be semantic, at the worst.

Of note, the difference in the number of patients necessary for an RCT differs very little, when this number is calculated, based on the amount of patients with "at least one complication" and the "CCI" (eg, 75 patients and 76 patients, respectively, for the pancreatic fistula study as indicated in Table 4).

Our definition of complications also included postoperative death, whatever the cause [again, because death may be due to a complication that has occurred but is still unknown (undetected) to the investigator]. If I understand correctly, death was excluded from the CCI index calculation. How then does death (from a complication whatever its Clavien-Dindo severity, certainly the worst outcome by any means) intervene in the CCI?

Now to my third question, how does the CCI take into account that a patient might have more than 1 complication, related to the same cause (eg, surgical site infection, organ/space deep infection, and pulmonary infection can all be due to a gastrointestinal or pancreatic-enteric anastomotic leak)?

You stated that the main advantage of the CCI is that the power calculation, based on the CCI delta, should reduce the number of patients necessary for a clinical trial. Finally, how did you calculate that a 10-point delta in the CCI would correspond to a clinically relevant change in outcome? (here, in the first example. In Table 4, grades A, B, and C pancreatic fistula or not); would the same delta be pertinent to distinguish between patients with or without only grades B and C? Or, once again, would fistulas lead to death?

Congratulations to the authors for yet another milestone in the evaluation of postoperative complications; certainly, we will be hearing more about this in the near future. I am, albeit, curious to know how to use the power calculation in practical terms.

Response From K. Slankamenac (Zurich, Switzerland):

Thank you, Professor Fingerhut, for these relevant comments and questions. With regard to your first question, the definition of any complication is, indeed, the same as that of an ARC complication, as used in France. You suggested that sample size calculation is similar using the ARC and CCI, as the main endpoint. Although the use of an ARC complication seems better than the other conventional endpoints, CCI is still superior. For example, in the esophageal trial, the sample with the definition for any complication requires 76 patients per group, whereas 63 patients would be sufficient, if CCI were selected. It is not surprising that the CCI is superior, because it captures all complications and their different severities. For example, this traditional definition of "any complication" is only defined as yes or no, so we don't know the severity of these complications and the number of all complications.

Regarding your second question of whether death is included in the CCI scale, the answer is yes. Death was arbitrarily given 100 points, the highest value on the scale. Your third question focused on a chain of complications, asking whether each one is included in the CCI or only the most severe. We have applied the same principles, as reported in our 5-year experience article, published in 2009, in this journal. When complications clearly depend on each other (chain events), we report only the most severe one. Similarly, when complications do not depend on each other, they should be listed.

To answer your last question, concerning the choice of 10 points as a relevant delta for the sample size calculation, we would

like to emphasize that this 10-point difference corresponds to a 40% relative risk reduction in the traditional endpoints. This is explained in the full manuscript.

DISCUSSANTS

C. Bassi (Verona, Italy):

First, I would like to thank you for this excellent study, which increased our understanding of the best way to classify complications in our patients. The main problem I see is that I'm still perplexed about the application of the same principles in a different kind of surgery and a specific operation. What do you think about the patients, who are going to be treated by local anesthesia, even though they are bleeding? How can we consider these patients in a lower grade than patients with general anesthesia? With this system, how can we stratify different situations in different kinds of surgery? These are my only questions. In the future, I think that we will have to continue to search for the best way to apply this to general surgery. Thank you.

Response From K. Slankamenac (Zurich, Switzerland):

Thank you, Professor Bassi, for these comments and your 2 questions. First, you are questioning the validity of some aspects of the grading system, for example, by giving a grade 3a for a bleeding treated with local anesthesia and a grade 3b complication for the same bleeding treated under general anesthesia. Part of the severity grading system is based on the burden of a complication on the patient and the associated risk. In this respect, there is no doubt that general anesthesia is more invasive than local anesthesia. This fact was well-reflected in a large study, which asked patients, nurses, and physicians to estimate the severity of different complication scenarios on a numerous analog scale. Of note, the Clavien-Dindo grading system allows for the combination of grade 3a and 3b complications, and thus, only reports grade 3 complications, so that there is no difference in the use of general or local anesthesia. Similarly, grade 4a (single organ failure) and 4b (multiple organ failure) can be combined and presented as grade 4 complications.

Your second question relates to the principles of the morbidity scale applied in different types of surgery. We believe that this is, in fact, a strength of both the Clavien-Dindo system and CCI, which enable conclusive comparisons to be drawn among different or same procedures over time, or even now, with the CCI, to longitudinally assess the postoperative course, for example, at discharge, after 3 months, 1 year, etc.