Beyond *P* Values

Novel Minimal Important Difference of the Comprehensive Complication Index That Reflects a Meaningful Outcome for Patients Undergoing Major Abdominal Surgery

Fariba Abbassi, MD,†‡ Matthias Pfister, MD,‡§ Julia Braun, PhD,†*
Eva Angenete, MD, PhD,‖¶ Eva Haglind, MD, PhD,‖¶
*Aron Onerup, MD, PhD,#** Margot H.M. Heijmans, PhD,††*
Gerrit D. Slooter, MD, PhD,†† Charlotte J.L. Molenaar, MD, PhD,††
Julio F. Fiore Jr, PhD,‡‡ Liane S. Feldman, MD,‡‡
Nicolò Pecorelli, MD, Msc,§§ Giovanni Guarneri, MD,§§
Massimo Falconi, MD,§§ Sarkis K. Drejian, MD,‖‖¶¶
Åsmund A. Fretland, MD, PhD,‖‖¶¶## Bjorn Edwin, MD, PhD,‖‖¶¶##
*Sander Ubels, MD, PhD,*** Gerjon Hannink, PhD,†††*
Milo A. Puhan, MD, PhD,†‡ and Pierre-Alain Clavien, MD, PhD‡§‡‡‡✉*

Objective: To estimate the minimal important difference (MID) of the Comprehensive Complication Index (CCI®) in patients undergoing abdominal surgery.

Background: The CCI® is a validated metric that quantifies cumulative surgical morbidity. While the CCI® is a sensitive endpoint to detect treatment effects, a statistically significant effect does not necessarily translate into clinical relevance. Relevant differences from the patients' perspective are best captured by the MID.

Methods: Individual patient data were extracted from surgical studies reporting CCI® at 30 days and using patient-reported outcome measures with established MIDs at baseline and 30 days. To determine the MID for the CCI®, we used an anchor-based approach as recommended by methods guidelines. A patient-reported outcome measure was selected as an anchor only if the Spearman correlation coefficient between its change in score (baseline to 30 days postoperative) and the CCI® was $\geq |0.30|$. We used linear regression to estimate the MID of the CCI® across different anchors, and triangulation to determine a single MID.

Results: Data were extracted from 3 published randomized controlled trials and 1 prospective observational study (n = 1583 patients) in major abdominal surgery. In colorectal surgery cohorts, 2 subscores of the Short Form-36, 2 subscores of the Multidimensional Fatigue Inventory-20, the EuroQol-5-Dimension Index Score, and the EuroQol Visual Analog Scale showed a correlation with the CCI® of $\geq |0.30|$. This resulted in MID estimates for the CCI® ranging from 6.1 to 22.2. In hepato-pancreato-biliary surgery, 1 subscore of the Short Form-36, and 2 subscores of the Patient Reported Outcome Measure Information System-29 questionnaire qualified as anchors providing MID estimates ranging from 6.2 to 13.8.

Conclusions: We propose a mean difference of 12 points in the CCI® between treatment groups as a relevant difference in patients undergoing abdominal surgery. This MID provides an important foundation for sample size calculations and interpretation of randomized controlled trials and large real-world observational studies.

Key Words: anchor-based approach, CCI®, clinical significance, comprehensive complication index, MID, minimal important difference, outcome reporting, patient reported outcome measures

(*Ann Surg* 2025;282:670–677)

Received for publication February 24, 2025; accepted May 9, 2025.

From the *Department of Surgery and Transplantation, University of Zurich, Zurich, Switzerland; †Department of Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland; ‡Hub for Translational Research and Liver/GI Health, Zurich, Switzerland; §Wyss Zurich Translational Center, ETH Zurich and University of Zurich, Zurich, Switzerland; ‖Department of Surgery, SSORG-Scandinavian Surgical Outcomes Research Group, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; ¶Department of Surgery, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden; #Department of Pediatrics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; **Department of Pediatric Oncology, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden; ††Department of Surgery, Máxima Medical Center, Veldhoven, The Netherlands; ‡‡Department of Surgery, McGill University, Montreal, Canada; §§Division of Pancreatic Surgery, Pancreas Translational and Clinical Research Center, San Raffaele Scientific Institute, Milan, Italy; ‖‖The Intervention Centre, Oslo University Hospital, Oslo, Norway; ¶¶Institute of Clinical Medicine, University of Oslo, Oslo, Norway; ##Department of Hepato-Pancreato-Biliary Surgery, Oslo University Hospital, Oslo, Norway; ***Department of Surgery, Radboud University Medical Center, Nijmegen, The Netherlands; †††Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands; and ‡‡‡Swiss Medical Network, Zurich, Switzerland.

✉clavien@access.uzh.ch.

This study is supported by the Liver and Gastrointestinal Disease (LGID) Foundation, the Yvonne Jakob Foundation, and the FZ-Hhealthcare Foundation.

M.A.P. and P.A.C. contributed equally to this work.

P.A.C. developed the Clavien-Dindo classification. P.A.C. and M.A.P. developed the Comprehensive Complication Index (CCI®). The remaining authors report no conflicts of interest.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.annalsofsurgery.com.

DOI: 10.1097/SLA.0000000000006877

The recent Outcome4Medicine consensus conference[1] on how to assess quality of surgical interventions recommended the Comprehensive Complication Index (CCI®)[2,3] as the only validated metric quantifying cumulative surgical morbidity. Unlike traditional classification systems, such as the Clavien-Dindo (CD) classification,[4,5] which categorize complications into discrete grades based solely on severity, the CCI® reflects the cumulative effect of all complications experienced by an individual patient, ranging from 0 (indicating an uneventful postoperative course) to 100 (indicating the death of the patient). The formula for calculating the CCI® has been developed by weighting each CD grade through an evaluation of clinical scenarios, incorporating perspectives from both patients and physicians. This approach ensures that the CCI® takes into account the varying significance of complications from both clinical and patient-centered viewpoints. The CCI® is significantly more sensitive than the CD classification and other metrics in detecting treatment effects allowing for smaller sample sizes in randomized controlled trials (RCTs).[6] This makes the CCI® particularly well-suited as primary endpoint in surgical research.[6,7] To date, more than 120 RCTs have employed the CCI® as an endpoint.[8]

The interpretation of CCI® scores and the determination of their clinical importance remain, however, challenging. As with any metric, statistically significant results may not always be clinically meaningful. A given between-group difference in CCI® may reach statistical significance due to a large sample size, while having little to no impact on patient well-being.[9,10] This discrepancy raises concerns about the real-world relevance of such findings, particularly when changes are statistically significant but fail to make a tangible difference in patients' experiences and outcomes.

To address this, the concept of the minimal important difference (MID) has been introduced more than 30 years ago. The MID is defined as the smallest mean difference between treatment groups in any given metric that is perceived by patients as meaningful, and that would, therefore, in the absence of troublesome side effects and excessive costs, justify a change in clinical management.[11] The MID is crucial for interpreting the clinical significance of treatment effects, as it helps to differentiate between changes that are statistically significant but clinically negligible and those that truly improve patient outcomes. Consequently, it serves as an essential tool for assessing the effectiveness of interventions and guiding clinical decision-making.

Estimates of the MID for a specific outcome are typically derived using specific methods, including anchor-based approaches.[12] These approaches rely on external references, known as "anchors," which are variables or measures that reflect a meaningful change from the patient's or clinician's perspective. The anchors provide a reference point against which changes in the outcome of interest – such as the CCI® – can be assessed. These anchors should reflect the patient perspective, such as patient-reported improvements in symptoms or quality of life. The anchor-based approach requires a relationship between the outcome measure and the anchor. This relationship is important because it allows researchers to determine what degree of change in the outcome measure is considered meaningful from the patient's perspective. The strength of this relationship can vary depending on the type of anchor used. For example, an anchor that directly reflects a critical aspect of a patient's health, such as functional recovery after surgery, will likely have a stronger and more direct relationship with the CCI® than an anchor measuring an aspect of health that is not related to surgery.

To determine the MID, researchers analyze how changes in the outcome measure of interest correspond to changes in the selected anchors. The goal is to identify the smallest difference in the outcome measure that corresponds to a meaningful change in the anchor. For example, a change in the CCI® that correlates with a deterioration in self-reported physical function can be interpreted as clinically meaningful if it represents a shift that patients would perceive as a deterioration in their health.

By using different anchors, each reflecting a meaningful change, researchers can ensure that the resulting MID estimates are robust and reliable. Therefore, it is recommended to use multiple anchors with strong enough correlations and to triangulate the results, ultimately arriving at a single value or a narrow range of values that best represent the MID.[13]

Predefined differences in the CCI® have been used to calculate target sample sizes in some trials, but in contrast to other metrics, for example, used in patient-centered outcome reporting [ie, patient-reported outcome measures (PROMs), patient-reported experience measures], an MID for interpreting the clinical significance of changes in complication burden is completely lacking. To address this gap, we conducted an analysis to determine the MID of the CCI® in patients undergoing abdominal surgery.

METHODS

Study Design

We conducted a literature review identifying all RCTs in abdominal surgery using the CCI® as an endpoint.[8] All studies reporting the CCI® at 30 days and, in addition, PROMs at baseline and after 30 days were eligible for inclusion. Feasibility trials were excluded. There was no study on pancreatic surgery among the eligible RCTs. To depict a comprehensive spectrum of abdominal surgery, we, in addition, included data from a high-quality prospective observational study investigating pancreatic resection.

Approach to Determine the MID of the CCI®

To determine the MID for the CCI®, we used an anchor-based approach, in which PROMs with known MIDs served as anchor to predict the MID of the CCI®. Since the MID must reflect the patient's perspective, it is essential to select anchors that capture this aspect. PROMs, which provide direct patient-reported information, are, therefore, well-suited for this purpose. Furthermore, anchors must have a proven association with the outcome measure of interest. To select appropriate PROMs, Spearman correlation analyses between the change of the PROMs and the "change" in CCI® were conducted, with a threshold of $r \geq |0.30|$.[13] As the baseline of the CCI® is always 0 (no preoperative complications), we used the CCI® score at 30 days as the "change" in CCI®.

Anchors which were suitable based on a correlation $\geq |0.30|$ were then used to calculate MID values for the CCI®. We used linear regression with the CCI® as dependent and the change in the respective PROMs as independent variables. Based on the results of the linear regression

model and the MID of the PROMs, we estimated the MID of the CCI® across different PROMs.

Assessment of the Certainty of MID Estimates

We used a modified Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach[14] to assess the level of certainty for each MID estimate. In line with this widely used methodology, we categorized certainty as high, moderate, low or very low. The decision to downgrade certainty for the MID estimate was based on 5 key criteria:

- Eligibility: We assessed whether changes in PROMs and the CCI® were sufficiently large to allow for a meaningful MID calculation. If variability in change from baseline to follow-up is low, the regression analyses may not yield valid estimates. Downgrading was applied if the median CCI® was 0 or if the median change in the PROM was smaller than the established MID for that PROM.
- Complication reporting: We evaluated whether procedure-specific complications were systematically recorded. The reporting of these complications served as an indicator of precise and accurate reporting, reducing the likelihood of relevant underreporting. Incomplete or inconsistent reporting could lead to an underestimation of postoperative morbidity and consequently to unreliable MID estimates. Studies that did not report procedure-specific complications or did not include all CD grades in the CCI® calculation were downgraded.
- Missing values: We examined how missing PROM data were addressed and whether imputation methods were applied. A high proportion of missing values or inadequate handling strategies could compromise the validity of the MID estimate. Downgrading was applied if either baseline or 30-day follow-up of a PROM was missing in $\geq 25\%$ of the study participants.
- Applicability: We considered whether the derived MID was generalizable to other studies in this field, where a MID would be typically used to determine sample size or interpret effects. Downgrading was applied in cases where the study population was highly specific and not representative of broader clinical contexts where the MID would be used.
- Imprecision: We assessed imprecision based on the CI of the MID estimate. We downgraded for imprecision if the lower boundary of the 95% CI was implausibly low (ie, below 6) and/or the upper boundary implausibly high (ie, above 15), based on our a priori expectation that the MID of the CCI® would lie between 6 and 15.

Each decision regarding the downgrading of certainty was justified in footnotes.

TABLE 2. Patient Reported Outcome Measures Used as Anchors

PROM	Used Score/Subscore	Score Range	MID
SF-36	Physical component summary*	5 to 80	5 [26–29]
SF-36	Physical functioning*	19 to 58	5 [26–29]
SF-36	RP*	21 to 57	5 [26–29]
MFI-20	General fatigue	4 to 20	2[30]
MFI-20	PF	4 to 20	2[30]
EQ-5D-3L	Index score	−0.594 to 1	0.02[31]
EQ VAS	EQ VAS	0 to 100	7.1[31]
PROMIS-29	PSS*	0 to 100	2.3[32]
PROMIS-29	Physical functioning	0 to 100	2.7[32]

*Norm-based Score.

EQ VAS indicates European Quality of Life Visual Analog Scale; EQ-5D-3L, European Quality of Life 5 Dimensions 3 Level; MFI-20, Multidimensional Fatigue Inventory-20; PF, physical fatigue; PSS, Physical Health Summary Score; RP, role physical; SF-36, Short Form-36.

Triangulation of MID Estimates

To determine a single MID estimate for the CCI®, we applied triangulation, a process of synthesizing multiple estimates from different methods or data sources to derive a single, more valid estimate, while accounting for the certainty of the available evidence. Specifically, we considered the MID estimates derived from different PROMs based on their GRADE certainty level, ensuring that estimates with higher certainty contributed more to the final MID determination, while those with lower certainty had less influence.

RESULTS

We obtained data from 5 RCTs and 1 prospective observational study. Three of these trials reported on colorectal surgery (n = 1022 patients),[15–17] 1 study on esophagectomy (n = 245 patients),[18] 1 study on resection of colorectal liver metastases (n = 250 patients),[19] and 1 study on pancreatic resection (n = 566 patients[20]; Table 1).

TABLE 1. Summary of Analyzed Studies

	Study 1[15]; N = 99*	Study 2[16]; N = 255*	Study 3[17]; N = 668*	Study 4[20]; N = 566*	Study 5[19]; N = 250*	Study 6[18]; N = 245*
Study design	RCT	RCT	RCT	Prospective observational study	RCT	RCT
Domain	Colorectal	Colorectal	Colorectal	HPB	HPB	Upper GIT
Disease						
Malignant	58 (59)	255 (100)	668 (100)	433 (77)	250 (100)	245 (100)
Benign	41 (41)	0	0	133 (23)	0	0
CCI®	0 (0, 12)	0 (0, 12)	21 (0, 38)	21 (0, 35)	0 (0, 0)	21 (0, 30)
Analyzed PROMs	SF-36 MFI-20	SF-36 EORTC C30	SF-36 EQ-5D-3L EQ VAS	PROMIS-29	SF-36	EORTC C30 EQ-5D-3L EQ VAS

*n (%); median (IQR).

EORTC indicates European Organization for Research and Treatment of Cancer; EQ VAS, European Quality of Life Visual Analog Scale; EQ-5D-3L, European Quality of Life 5 Dimensions 3 Level; GIT, gastrointestinal tract; HPB, hepato-pancreato-biliary; MFI-20, Multidimensional Fatigue Inventory-20; SF-36, Short Form-36.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

The Short Form-36 questionnaire[21] was used in 4 studies, while the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30,[22] the European Quality of Life 5 Dimensions 3 Level version,[23] and the European Quality of Life Visual Analog Scale were used in 2 studies each. The Patient Reported Outcomes Measurement Information System 29 (PROMIS-29)[24] and the Multidimensional Fatigue Inventory-20[25] were only used in one study each (Table 1).

The PROMs for which we found a correlation with the CCI® of $\geq|0.30|$ are shown in Table 2. The highest correlation was found between PROMIS-29 and the CCI® ($r = 0.42$). Two studies[16,18] did not show a correlation $\geq|0.30|$ for any PROM (Supplemental Table, Supplemental Digital Content 1, http://links.lww.com/SLA/F576). Table 3 shows the linear regression and the resulting 9 MID estimates for the CCI® across the relevant PROMs. The MID estimates range between 6.1 and 22.2.

The assessment of the certainty of the 9 estimates according to the modified GRADE approach revealed moderate certainty for 3 estimates, low certainty for 4 estimates, and very low certainty for 2 estimates (Table 4).

Triangulation of the MID estimates, considering their GRADE certainty level, resulted in a proposed single MID of 12 CCI® points.

DISCUSSION

Based on our findings, we propose for the CCI® a 12-point mean difference between treatment groups as a relevant effect in patients undergoing abdominal surgery. This MID can help investigators and clinicians in designing future studies assessing morbidity in surgical populations and to determine whether treatment effects are meaningful to patients and indicate a true beneficial outcome.

This study evaluated multiple PROMs, including Short Form-36, European Organization for Research and Treatment of Cancer QLQ-C30, European Quality of Life 5 Dimensions 3 Level, European Quality of Life Visual Analog Scale, PROMIS-29, and Multidimensional Fatigue Inventory-20. By incorporating various instruments that capture different dimensions of patient well-being, we ensured a comprehensive assessment of the relationship between patient-perceived health status and postoperative morbidity.

Some previous studies have used an MID of 10 points for the CCI®.[6,33] Our results suggest that this threshold was slightly too low. It is possible that some studies may have somewhat overestimated the clinical significance of observed changes in the CCI®, while others may have dismissed meaningful differences as statistically but not clinically relevant. By establishing a data-driven MID of 12 points, our study contributes to a more accurate and standardized approach to the assessment of postoperative outcomes.

A crucial application of the MID lies in sample size calculation, as it ensures that clinical trials are designed to detect differences that are not only statistically significant but also clinically meaningful. In superiority trials, the MID defines the smallest effect size that would justify adopting a novel intervention over the current standard. To illustrate this, we consider a study comparing a new surgical technique to the established gold-standard procedure, which has a reported mean CCI® of, for example, 21 points. The trial aims to demonstrate a change in overall morbidity by 12 CCI® points, reflecting a meaningful difference in postoperative outcomes. Assuming a SD of 20,[6] the calculations yield a required total sample size of 88 patients (44 per group) for 80% power and 117 patients (59 per group) for 90% power at a significance level of 0.05.

Beyond superiority trials, the MID also plays a key role in noninferiority trial design. In these studies, it helps define the noninferiority margin (Δ) – the maximum difference at which a new intervention can still be considered clinically acceptable, that is, noninferior, compared with the standard treatment. To ensure this margin is meaningful rather than arbitrarily chosen, Δ is typically set as the MID or a fraction of the MID (eg, 50% or 75%), ensuring that any tolerated difference remains below the level of patient-perceived impact. For instance, in a surgical trial using the CCI®, if the standard procedure has a mean CCI® of 21 and the MID is 12 points, a reasonable noninferiority margin might be set at 9 points (75% of the MID). In this case, if the new technique results in a CCI® of $\leq$ 30 (21 + 9), it could be considered noninferior, as the difference remains below the threshold of clinical relevance. By connecting the noninferiority margin to the MID, these trials ensure that conclusions regarding noninferiority are based on patient-centered criteria, thereby preventing the adoption of interventions that introduce clinically meaningful harm.

A potential criticism of our findings could be that the MID of 12 for the CCI® may not be universally applicable to all abdominal surgery, as procedures with inherently different morbidity profiles may require different thresholds

TABLE 3. Linear Regression Model to Determine the MID of the CCI®

	Regression	Estimated MID of the CCI® (95% CI)	Certainty in MID Estimate (GRADE)
Study 1[15]	4.24 + 0.38*MID SF-36 PCS	6.1 (-10.4 to 22.6)	+ + - -
	7.01 + 0.97*MID MFI-20 GF	8.9 (3.9 to 14.0)	+ + - -
	6.50 + 1.09*MID MFI-20 PF	8.7 (3.9 to 13.4)	+ + - -
Study 3[17]	16.00 + 0.75*MID SF-36 PF**	19.8 (10.3 to 29.3)	+ + + -
	19.90 + 28.37*MID EQ-5D-3L index score	20.5 (20.3 to 20.6)	+ + - -
	20.04 + 0.30*MID EQ VAS	22.2 (9.3 to 35.0)	+ - - -
Study 4[20]	12.13 + 0.72*MID PROMIS-29 PSS	13.8 (7.9 to 19.7)	+ + + -
	9.06 + 0.87*MID PROMIS-29 PF**	11.4 (4.6 to 18.2)	+ + + -
Study 5[19]	4.56 + 0.33*MID SF-36 RP	6.2 (-2.6 to 15.1)	+ - - -

EQ VAS indicates European Quality of Life Visual Analog Scale; EQ-5D-3L, European Quality of Life 5 Dimensions 3 Level; MFI-20, Multidimensional Fatigue Inventory-20; PF, physical fatigue; PSS, Physical Health Summary Score; RP, role physical; SF-36, Short Form-36; PCS, physical component summary; GF, general fatigue, PF**, physical functioning.

TABLE 4. Assessment of the Certainty in MID Estimates According to the GRADE Approach

Study and PROM	Eligibility	Complication Reporting	Missing Values	Applicability	Imprecision	Certainty in MID Estimate (GRADE)
Study 1,[15] SF-36 PCS	SF-36 PCS change: Median: 6.7 Mean: 9.2 Median CCI®: 0	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing SF-36 PCS change values: 18 of 99 (18%)	Colorectal resection for benign and malignant diseases in university hospital. Exclusion criteria: Metastasis, diseases that preclude mobilization, ICU immediately postoperative	MID estimate: 6.1 (-10.4 to 22.6)	+ + - - Low*
Study 1,[15] MFI-20 GH	MFI-20 GH change: Median : 0 Mean: 0.6 Median CCI®: 0	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing MFI-20 GH change values: 16 of 99 (16%)	Colorectal resection for benign and malignant diseases in university hospital. Exclusion criteria: Metastasis, diseases that preclude mobilization, ICU immediately postoperative	MID estimate: 8.9 (3.9 to 14.0)	+ + - - Low*
Study 1,[15] MFI-20 PF*	MFI-20 PF change: Median: 0 Mean: 1.0 Median CCI®: 0	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing MFI-20 PF change values: 16 of 99 (16%)	Colorectal resection for benign and malignant diseases in university hospital. Exclusion criteria: Metastasis, diseases that preclude mobilization, ICU immediately postoperative	MID estimate: 8.7 (3.9 to 13.4)	+ + - - Low*
Study 3,[17] SF-36 PF	SF-36 PF change: Median: 10 Mean: 14 Median CCI®: 21	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing SF-36 PF change values: 162 of 668 (24%)	Elective colorectal cancer resection in regional and university hospitals. Exclusion criteria: Local resection, cytoreductive, inability to perform exercise	MID estimate: 19.8 (10.3 to 29.3)	+ + + - Moderate†
Study 3,[17] EQ-5D-3L Index score	EQ-5D-3L Index score change: Median: 0 Mean: 0.05 Median CCI®: 21	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing EQ-5D-3L index score change values: 192 of 668 (29%)	Elective colorectal cancer resection in regional and university hospitals. Exclusion criteria: Local resection, cytoreductive, inability to perform exercise	MID estimate: 20.5 (20.3 to 20.6)	+ + - - Low‡
Study 3,[17] EQ VAS	EQ VAS Change: Median: 0 Mean: 3.5 Median CCI®: 21	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing EQ VAS change values: 193 of 668 (29%)	Elective colorectal cancer resection in regional and university hospitals. Exclusion criteria: Local resection, cytoreductive, inability to perform exercise	MID estimate: 22.2 (9.3 to 35.0)	+ - - - Very low§
Study 4,[20] PROMIS-29 PSS	PROMIS-29 PSS change: Median: 13.1 Mean: 13.0 Median CCI®: 21	Very detailed. Procedure-specific complications available. Complications for each Organ system separately.	Imputation of PROMIS-29 PSS	Elective pancreatic surgery, malignant and benign lesions.	MID estimate: 13.8 (7.9 to 19.7)	+ + + - Moderate†
Study 4,[20] PROMIS-29 PF	PROMIS-29 PF change: Median: 13.5 Mean: 13.3 Median CCI®: 21	Very detailed. Procedure-specific complications available. Complications for each Organ system separately	Missing PROMIS-29 PF change values: 95 of 566 (17%)	Elective pancreatic surgery, malignant and benign lesions	MID estimate: 11.4 (4.6 to 18.2)	+ + + - Moderate†
Study 5,[19] SF-36 RP	SF-36 RP change: Median: 4.5	Very detailed. Procedure-specific complications available.	Missing SF-36 RP change values: 4 of 196 (2%)	Patients with CRLM, radical resectable in parenchyma-sparing way (< 3 consecutive segments). Resectable	MID estimate: 6.3 (-2.6 to 15.1)	+ - - - Very low‖

Mean: 5.0	Complications for each Organ system	lung or adrenal metastases were
Median CCI®: 0	separately.	allowed
	CD grade 1 complications not assessed	

*Concerns about imprecision: The 95% CI is wide and exceeds plausible MID estimates. Events were rare. Downgraded by 2 levels.
†Concerns about imprecision: The 95% CI is wide and exceeds plausible MID estimates. Downgraded by one level.
‡Concerns about imprecision: The 95% CI exceeds plausible MID estimates. ≥ 25% of PROMs are missing. Downgraded by 2 levels.
§Concerns about imprecision: The 95% CI is wide and exceeds plausible MID estimates. Mean change of PROM is small. ≥ 25% of PROM change values are missing. Downgraded by 3 levels.
‖Concerns about imprecision: The 95% CI is wide and exceeds plausible MID estimates. Events are rare. No assessment of CD grade 1 complications. Downgraded by 3 levels.
CRLM indicates colorectal liver metastasis; EQ VAS, European Quality of Life Visual Analog Scale; EQ-5D-3L, European Quality of Life 5 Dimensions 3 Level; GH, general health, PF, physical fatigue; MFI-20, Multidimensional Fatigue Inventory-20; PCS, physical composite score; PF, physical functioning; PSS, Physical Health Summary Score; RP, role physical; SF-36, Short Form-36.

for clinical relevance. However, this argument does not hold as the CCI® is a nonlinear scale. In the low morbidity range, a change of 12 CCI® points corresponds approximately to 2 CD grade 1 complications. In contrast, at higher morbidity levels, the same 12-point increase reflects a more severe complication. For example, if a patient already has a CCI® of 33.7 (corresponding to a single CD grade 3b complication), an additional 12 CCI® points would correspond to an event of almost the same severity. This property of the CCI® ensures that an MID of 12 remains a meaningful threshold in different surgical contexts, as it reflects a proportionally relevant change in complication burden, regardless of baseline morbidity.

While our findings are based on data from abdominal surgery, the CCI® is widely used across surgical specialties.[1,33,34] The MID of 12 points may be applicable beyond abdominal procedures, but further validation is necessary. Different surgical fields, such as orthopedic, cardiac, or neurosurgical procedures, have distinct morbidity profiles and patient expectations, which may influence the perception of clinically meaningful changes in complication burden. Future research should explore whether an MID of 12 holds true in other surgical populations or whether specialty-specific thresholds need to be established.

The determination of an MID for the CCI® is an important step towards a more patient-centered approach in surgery, where decision-making ensures that each patient's values, expectations, and outcome priorities are carefully considered. Beyond its role in quantifying perioperative morbidity, its integration into a broader benefit-harm analysis could improve clinical decision-making.[35,36] Surgical interventions inherently involve a trade-off between benefits and risks, and the implementation of new treatment modalities should not be based solely on the demonstration of a meaningful improvement in a single outcome. Instead, decision making must consider a combination of perioperative morbidity, functional recovery, quality of life, and long-term patient well-being. The concept of MID should evolve beyond isolated measures to determine when the benefits of an intervention outweigh potential harms in a clinically meaningful way. Incorporating advanced benefit-harm analyses that integrate clinical data and patient preferences could significantly advance surgical outcomes research and provide a stronger evidence base for improving patient care.[10]

This study has some limitations. The analysis was restricted to 30-day postoperative outcomes. While this time frame is commonly used in surgical research, it does not capture longer-term complications or patient recovery trajectories. However, the analysis was deliberately limited to 30-day postoperative outcomes, as this represents the period in which complications are still correlated to PROMs whereas later during follow-up many effects other than the postsurgical course impact on PROMs, making it difficult to isolate the direct impact of surgical complications.[37,38] By limiting our analysis to this early phase, we ensure that the estimated MID accurately reflects morbidity directly attributable to surgery.

The determination of a data-driven MID for the CCI® represents a critical step towards improving the interpretation of postoperative morbidity in surgical research. By anchoring the assessment of treatment effects to patient-centered outcomes, this threshold enhances the clinical relevance of surgical trials and ensures that observed differences reflect meaningful changes in patient well-being.

REFERENCES

1. Domenghino A, Walbert C, Birrer DL, et al. Consensus recommendations on how to assess the quality of surgical interventions. *Nat Med*. 2023;29:811–822.
2. 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.
3. Clavien PA, Vetter D, Staiger RD, et al. The Comprehensive Complication Index (CCI®(R)): added value and clinical perspectives 3 years "down the line". *Ann Surg*. 2017;265:1045–1050.
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;240:205–213.
5. 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.
6. 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;260:757–763.
7. Kim TH, Suh YS, Huh YJ, 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;21:171–181.
8. Abbassi F, Pfister M, Lucas KL, et al. Milestones in surgical complication reporting: Clavien-Dindo classification 20 years and Comprehensive Complication Index 10 years. *Ann Surg*. 2024;280:763–771.
9. Gikandi A, Hallet J, Koerkamp BG, et al. Distinguishing clinical from statistical significances in contemporary comparative effectiveness research. *Ann Surg*. 2024;279:907–912.
10. 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;279:913–914.
11. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.
12. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56:395–407.
13. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102–109.
14. Schünemann HJ. GRADE: from grading the evidence to developing recommendations. A description of the system and a proposal regarding the transferability of the results of clinical research to clinical practice. *Z Evid Fortbild Qual Gesundhwes*. 2009;103:391–400.
15. Fiore JF Jr., Castelino T, Pecorelli N, et al. Ensuring early mobilization within an enhanced recovery program for colorectal surgery: a randomized controlled trial. *Ann Surg*. 2017;266:223–231.
16. Molenaar CJL, Minnella EM, Coca-Martinez M, et al. Effect of multimodal prehabilitation on reducing postoperative complications and enhancing functional capacity following colorectal cancer surgery: the PREHAB randomized clinical trial. *JAMA Surg*. 2023;158:572–581.
17. Onerup A, Andersson J, Angenete E, et al. Effect of short-term homebased pre- and postoperative exercise on recovery after colorectal cancer surgery (PHYSSURG-C): a randomized clinical trial. *Ann Surg*. 2022;275:448–455.
18. van Workum F, Verstegen MHP, Klarenbeek BR, et al. Intrathoracic vs cervical anastomosis after totally or hybrid minimally invasive esophagectomy for esophageal cancer: a randomized clinical trial. *JAMA Surg*. 2021;156:601–610.
19. Fretland ÅA, Dagenborg VJ, Bjørnelv GMW, et al. Laparoscopic Versus Open Resection for Colorectal Liver Metastases: The OSLO-COMET Randomized Controlled Trial. *Ann Surg*. 2018;267:199–207.
20. Pecorelli N, Guarneri G, Di Salvo F, et al. The impact of postoperative complications on recovery of health-related quality of life and functional capacity after pancreatectomy: findings from a prospective observational study. *Ann Surg*. 2024;280:719–727.
21. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–483.
22. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85:365–376.
23. Rabin R, de, Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33:337–343.
24. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63:1179–1194.
25. Smets EMA, Garssen B, Bonke B, et al. The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosomatic Research*. 1995;39:315–325.
26. Frendl DM, Ware JE. Patient-reported Functional Health and Well-Being Outcomes With Drug Therapy: A Systematic Review of Randomized Trials Using the SF-36 Health Survey. *Med Care*. 2014;52:439–445.
27. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41:582–592.
28. Ware JE, Kosinski M SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1: QualityMetric2001.
29. Ware JE Jr., Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995;33(4 Suppl):As264–279.
30. Purcell A, Fleming J, Bennett S, et al. Determining the minimal clinically important difference criteria for the Multidimensiona l Fatigue Inventory in a radiotherapy population. *Support Care Cancer*. 2010;18:307–315.
31. Cheng LJ, Chen LA, Cheng JY, et al. Systematic review reveals that EQ-5D minimally important differences vary with treatment type and may decrease with increasing baseline score. *J Clin Epidemiol*. 2024;174:111487.
32. Pecorelli N, Guarneri G, Vallorani A, et al. Validation of the PROMIS-29 Questionnaire as a Measure of Recovery After Pancreatic Surgery. *Ann Surg*. 2023;278.
33. Kowalewski KF, Müller D, Mühlbauer J, et al. The comprehensive complication index (CCI®): proposal of a new reporting standard for complications in major urological surgery. *World J Urol*. 2021;39:1631–1639.
34. Rele S, Shadbolt C, Schilling C, et al. Validation of the Clavien-Dindo classification and Comprehensive Complication Index as measures of morbidity following total hip and knee arthroplasty. *Bone Joint J*. 2025;107-b:81–88.
35. Yebyo HG, Aschmann HE, Puhan MA. Finding the balance between benefits and harms when using statins for primary prevention of cardiovascular disease: a modeling study. *Ann Intern Med*. 2019;170:1–10.
36. Aschmann HE, Boyd CM, Robbins CW, et al. Informing patient-centered care through stakeholder engagement and highly stratified quantitative benefit–harm assessments. *Value Health*. 2020;23:616–624.
37. Driscoll B, Leonard LD, Kovar A, et al. Surgeon perceptions of the integration of patient-reported outcome measures into clinical practice. *J Surg Res*. 2022;280:486–494.
38. Heerkens HD, Tseng DS, Lips IM, et al. Health-related quality of life after pancreatic resection for malignancy. *Br J Surg*. 2016;103:257–266.

DISCUSSANTS

José M. Ramia-Angel (Alicante, Spain)

Many thanks for the opportunity to act as the first discussant of this "changing practice" special lecture. I don't

think that the concept of meaningful outcomes is just an idea; it's a moral obligation for all surgeons. It is a mandatory tool to measure everything when it comes to postoperative outcomes. I believe this special lecture represents a giant step ahead. First, we had the Clavien Classification, then the CCI®, and now, we are finally integrating the patient's perspective into outcome measurement, which is a crucial advancement.

My 3 questions are as follows:

First, while the concept is very appealing, it's somewhat difficult to understand how you arrived at the 12-point difference as the MID for the CCI®. Could this pose a limitation when it comes to implementation?

Second, when you measured the quality of life, you used several scoring systems, which produced slightly different results. Could this introduce bias in estimating the significance of the 12-point MID?

Finally, do we need a prospective validation of these results in future studies?

Response From Fariba Abbassi (Zurich, Switzerland)

Professor Ramia, many thanks for your compliments and constructive discussion. Regarding your first comment about the complexity of this concept, I wouldn't necessarily say that the concept is more complex; it's simply a different perspective. The CCI® and CD classification are tools to objectively measure what has happened, while the MID helps us understand how much this matters to the patient. In other words, it's not about replacing existing systems, but about adding value to them.

Regarding your question about the use of multiple PROMs. The use of several measures to assess the quality of life is not a bias, but rather a strength of our study. By applying several anchors to determine the MID, we are able to capture different dimensions of a patient's well-being and correlate them with postoperative morbidity. This approach enhances both the reliability and the robustness of the results.

Finally, thanks for your third question regarding the need for prospective validation. We used raw data from published RCTs, which were conducted prospectively, although not originally designed to estimate the MID. A dedicated prospective study could allow for better selection of anchor measures, which would be valuable. Therefore, yes, a prospective study would certainly be welcome. However, planning a study solely for this purpose may be somewhat unrealistic. However, it could be effectively integrated prospectively within a cohort study or an RCT.

Christiane Bruns (Cologne, Germany)

While your study analysis established an MID of 12 for pooled gastrointestinal surgeries, I question its applicability to specific subgroups. A patient with rectal anastomotic insufficiency receiving EVAC therapy faces fundamentally different challenges than an upper GI patient with dysphagia undergoing the same treatment. These differences – in anatomy, functional impairment (eg, swallowing), and procedural context – likely invalidate a shared MID. Specifically, upper GI patients may require a substantially higher MID to reflect their greater clinical burden and outcome-related needs.

Response From Fariba Abbassi (Zurich, Switzerland)

Dear Professor Bruns, thank you very much for this important question. The strength of the CCI® is its nonlinearity. For procedures with low morbidity, a MID of 12 CCI® points corresponds to 2 grade I complications, whereas in high-morbidity surgeries with a CCI® of around 37, the same 12 points reflect a 3b complication. This allows the same threshold to be meaningful across procedures with varying morbidity levels. Regarding different patient populations, our aim was to provide a generalizable estimate as a foundation for future refinement. While procedure-specific MIDs would ideally capture distinct burdens more precisely, their systematic determination across all surgical populations may become clinically inaccurate. The presented MID reflects this necessary compromise.

Mickaël Lesurtel (Clichy, France)

I have a similar question. For me, I don't find this MID too complicated; rather, I wonder if it might be too simplistic. How can you be confident that this value is applicable across all types of operations, diseases, and patient populations?

Response From Fariba Abbassi (Zurich, Switzerland)

Thank you very much, Professor Lesurtel, for raising this criticism. In clinical practice, we constantly deal with heterogeneous patient populations, and naturally, our metrics must adapt to this reality. The concept of an MID is not fixed; it evolves with accumulating evidence. For example, the MID for the 6-minute walk was around 54 meters, but this figure has evolved over time. Similarly, our estimate provides a starting point, not a definitive number for all settings. Future studies in specific subgroups or procedures may yield more tailored values.

Andreas Schnitzbauer (Bochum, Germany)

Congratulations on this paper. I like it very much. I have just one question: One of the biggest challenges in collecting the CCI® reproducibly, especially for grade I and II complications, is the variability when different individuals perform the assessment, which may introduce bias. Have you considered creating an algorithm that can automatically extract the CCI® from electronic health records to generate a highly reliable and consistent score for every patient, regardless of who assesses it?

Response From Fariba Abbassi (Zurich, Switzerland)

Professor Schnitzbauer, thank you very much for your positive feedback and excellent question. You are absolutely right that reproducible and consistent calculation of the CCI®, especially for lower-grade complications, can be challenging and prone to inter-rater variability. Developing an automated algorithm to extract the CCI® directly from electronic health records (not just from discharge summaries) is a promising approach. I believe this aligns well with what you are implementing at your center. Your pioneering work may pave the way for many institutions to follow in this direction.

Meanwhile, our online CCI® calculator offers a simple manual calculation method, supports data collection across various patient groups, and aims to facilitate the future development of automated institutional CCI® extraction from full electronic health records.