

Consensus recommendations on how to assess the quality of surgical interventions

Received: 24 October 2022

Accepted: 26 January 2023

Published online: 17 April 2023

 Check for updates

Anja Domenghino^{1,2}, Carmen Walbert³, Dominique Lisa Birrer¹, Milo A. Puhan²✉, Pierre-Alain Clavien¹✉ & The Outcome4Medicine consensus group*

Postoperative complications represent a major public health burden worldwide. Without standardized, clinically relevant and universally applied endpoints, the evaluation of surgical interventions remains ill-defined and inconsistent, opening the door for biased interpretations and hampering patient-centered health care delivery. We conducted a Jury-based consensus conference incorporating the perspectives of different stakeholders, who based their recommendations on the work of nine panels of experts. The recommendations cover the selection of postoperative outcomes from the perspective of patients and other stakeholders, comparison and interpretation of outcomes, consideration of cultural and demographic factors, and strategies to deal with unwarranted outcomes. With the recommendations developed exclusively by the Jury, we provide a framework for surgical outcome assessment and quality improvement after medical interventions, that integrates the main stakeholders' perspectives.

A large proportion of the world's population undergo surgical interventions during their lifetime, sometimes repeatedly. Surgical interventions (also known as 'medical interventions') are defined as any procedure on the human body with a therapeutic purpose, which includes invasive (open) or minimally invasive (laparoscopic, robotic, endoscopic or percutaneous) procedures. The World Health Organization (WHO) recognized in 2008 that complications of surgical interventions are a major burden and a global public health issue¹. Ten years later, postoperative complications were described as a hidden pandemic with largely under-recognized causes, which are often avoidable². A major barrier to reducing the burden of surgical interventions is the scarcity of data to act upon, and even if available, data on outcomes after surgical interventions are often of poor quality. The lack of consistent reporting is well highlighted in the medical literature³, and even top surgical journals often fail to provide proper information on postoperative events, for example, in defining the severity of complications or providing sufficient follow-up for assessment. Postoperative complications not only cause suffering and dissatisfaction with a reduction in quality of life (QoL) for patients, but also have

serious implications on many levels of society and are associated with tremendous financial cost^{4,5}.

The first step to preventing harmful events after an intervention and allowing for credible comparisons of competing therapies or care providers is to develop standardized tools assessing both the positive and negative outcomes of a procedure. Such tools must be relevant for patients and health care providers, as well as all other stakeholders within society, and must be widely accepted among various health care systems and cultures. Considering that the subject area remains complex and tools for outcome measurements covering perspectives of a broad range of stakeholders are not available, we opted for the format of a consensus approach to develop guidelines on how to assess the outcomes of a surgical intervention. The assumption was that the best available evidence, together with a consensus developed among diverse representatives of society, would yield the most convincing approach for broad adoption.

For this purpose, we relied on the Zurich–Danish model, where an independent Jury frames recommendations based on evidence reports prepared by a multidisciplinary panel of experts and on its

¹Department of Surgery and Transplantation, University Hospital Zurich, Zurich, Switzerland. ²Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich (UZH), Zurich, Switzerland. ³Biogen International GmbH, Baar, Switzerland. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: miloalan.puhan@uzh.ch; clavien@access.uzh.ch

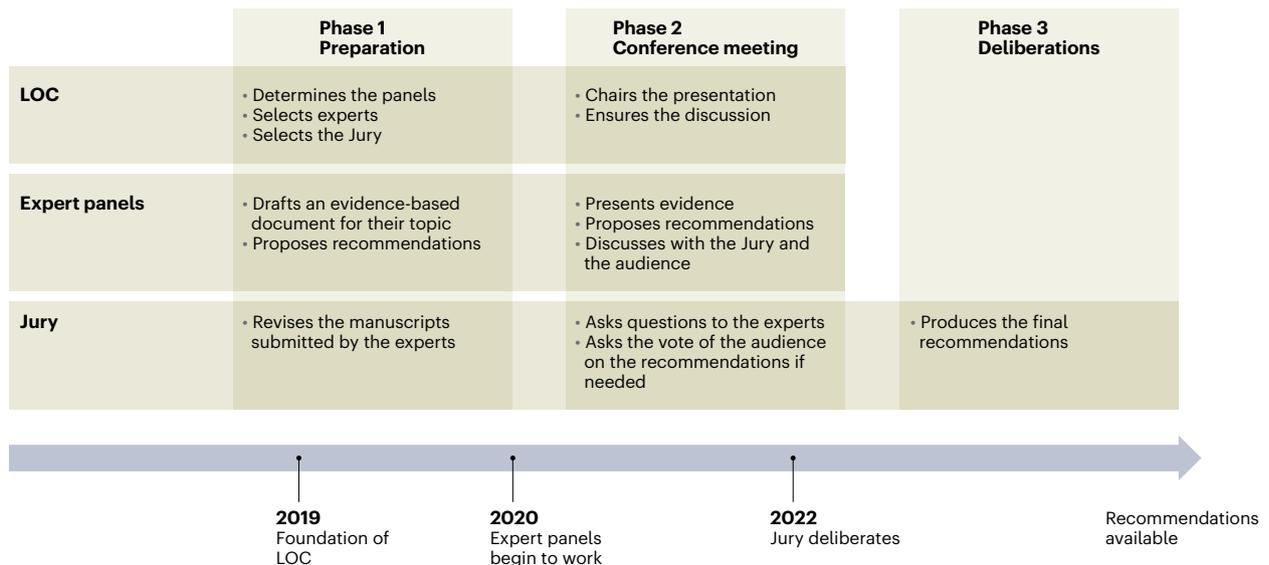


Fig. 1 | The Zurich–Danish consensus conference. The three phases of the Zurich–Danish consensus conference, an independent jury-based consensus conference model for the development of recommendations in medico-surgical practice. Adapted from Lesurtel et al.⁶

own deliberations^{6–9}. To our knowledge, our consensus conference is the first attempt to include a broad range of perspectives from various stakeholders affected by the quality of surgical outcomes. The resulting recommendations are intended to provide a general framework for surgical outcome assessment that can be adapted by researchers and health care providers for specific patient populations and medical interventions.

Methodology

Zurich–Danish model for consensus building

The Zurich–Danish model⁶ aims at producing evidence-based, internationally valid and unbiased recommendations that consider the perspectives of many stakeholders, including patients and health care providers, as well as payers or governments. We have previously used this approach to develop a consensus in the area of liver transplant for hepatocellular carcinoma⁷ and treatment options for neuro-endocrine liver metastases⁸, as well as for the selection of an academic chair in Medicine⁹.

The principle relies on a clear distinction between those who provide the evidence (the experts) and those who draw the final recommendations (the Jury). The Jury consists of individuals with sufficient background knowledge to cover the perspectives of a wide and important range of stakeholders, without being directly involved within their professional spheres in the topic under evaluation. The organizing committee, the experts and the Jury interact in three phases (Fig. 1) – that is, the preparation phase, the in-person consensus conference and the Jury deliberations. Each panel of experts addresses their specific question in the year-long preparation phase and then proposes evidence-based recommendations at the conference meeting. The answers to the questions are communicated to the Jury in writing at least three weeks before the meeting, with possible interaction in the interim between Jury members and panel chairs. At the conference, the Jury and the audience challenge these recommendations by asking questions and offering comments. Based on all the presented information, the Jury finalizes the consensus recommendations, which are then made available to the public.

Expert and Jury recruitment

Experts were recruited through four channels: (1) invitation to senior authors of relevant publications identified by the Local Organizing

Committee (LOC), (2) invitation to experts recommended by panel chairs or members (snowballing technique), (3) call to patient- and scientific organizations to participate in one of the panels and (4) consideration of any expert contacting the LOC directly, after validation of their expertise.

The constitution of the Jury by the LOC started with a list of perspectives and priorities to be covered, with consideration of geographic and gender balance. We also favored individuals previously involved in consensus conferences with a similar format; for example, we recruited Carmen Walbert as President of the Jury owing to her active participation in a previous consensus conference on how to select an academic chair⁹. To secure proper patient representation, we contacted the organization EUPATI (European Patients’ Academy on Therapeutic Innovation), which issued a call to their members to send us resumes and letters of intent to participate. To prevent any conflict of interest, Jury members were not directly involved in surgical or medical outcomes research.

Topics and panels

An initial list of topics was prepared by the LOC and experts with relevant publications or opinion leaders in their respective fields were invited to participate. Topics were subsequently discussed and modified by invited faculties. Nine panels composed of four to five recognized international experts with different perspectives and geographic origin were proposed by the LOC, selected to provide a wide breadth of expertise. To maximize the relevance of the topics, the respective panel chairs and members could adjust their specific questions as needed to better cover their respective topics. Panels one through five focused on the various stakeholders’ perspectives, while panels six through nine concentrated on specific aspects of outcome measurement, analysis and interpretation. Each panel had the task of answering three to five questions on a specific topic (Box 1). The full list of the panel chairs, panel members and jury members can be found at the end of the text.

Recommendations

Standardized time points for outcome assessments

First, the Jury recognized that outcome assessment is a dynamic process that requires standardized time points of observations. There is currently no agreement as to when outcomes should be captured and an urgent need to move away from historically collected discharge or 30-day data only^{10,11}. Through the panel presentations and discussions

BOX 1**Panel topics and specific questions****Panel 1. Outcomes after surgical interventions from a patient perspective**

- Which outcomes are typically not represented in PROMs and which additional items should be incorporated?
- Which PROMs should be used in clinical practice and in research?
- How important are PREMs? (Including time spent with patients, trust in care team.)

Panel 2. Outcomes after surgical interventions from a health care provider perspective (physicians, nurses)

- How can complications in clinical practice and research be measured?
- How can complication data be recorded and audited?
- How can the patient–health care provider relationship, in terms of patient satisfaction, be measured?
- How can mortality and morbidity conferences be implemented to improve patient outcome?

Panel 3. Outcomes after surgical interventions from a payer perspective

- Is there a relation between cost and quality and how would you measure it?
- How would you avoid unnecessary treatment?
- Which cost-relevant outcomes should be reported to whom and how?
- How would you measure the endpoints of value-based medicine?
- Who is accountable for assessing outcome?

Panel 4. Outcomes after surgical interventions from a state and/or government and/or society perspective

- What should be measured in outcomes to evaluate regulatory policies?
- On the basis of what data should regulatory decisions be made?
- How should outcomes influence regulatory policies?
- Which outcome should influence policies?
- What outcome should be used as an indication of whether to centralize more or less?
- What outcome could be used to measure differences in terms of outcomes in health care systems around the world?

Panel 5. Outcomes after surgical interventions from a legal and ethical perspective

- What outcomes should be measured to assess outcome quality from a normative (ethical and legal) perspective?
- How can we determine what a good, a bad and/or an acceptable outcome is?
- When something goes wrong, how can we determine the presence and degree of culpability?
- What would be an ethically appropriate response to culpable mistakes?
- What would be an ethically appropriate response to non-culpable harms?
- How can the potential benefits and risks as a basis for informed decision-making be presented?

Panel 6. Approaches for benchmarking of surgical procedures

- What is the goal of 'benchmarking' in surgery?
- Which techniques should be used for benchmarking?
- What are the gold standards for databases used for benchmarking?
- How are feedback and communication strategies best used to communicate results of benchmarking in surgical care?

Panel 7. High-risk patients

- What do we mean by 'high-risk' patients?
- How do we identify and measure high risk?
- What is the future for risk prediction in high-risk patients?

Panel 8. Public reporting of surgical outcome data

- In terms of sharing patient outcomes (involving patients in decision-making), how can patients give a validated informed consent?
- In terms of auditing data integrity (statistical testing, adverse selection, case-mix), how should data be validated and made available?

Panel 9. Consideration of cultural and demographic differences in outcome interpretation

- Should surgical outcomes be interpreted in the context of cultural and demographic factors?
- How can culture and demographics be integrated into selection and interpretations of these outcomes?

at the consensus conference, the Jury saw a need for standardized time points of outcome assessment to ensure comparability of outcomes. They proposed five fixed time points (Box 2). The first assessment should capture the pre-disease state – meaning a time before the patient had their condition – and should include some information on their quality of life (T0), followed by a recording of disease state and related symptoms before the intervention (T1), outcomes during the early postoperative phase (T2), mid-term (T3) and long-term (T4) (providing data five years after intervention).

Pre-disease conditions or QoL are difficult to assess retrospectively, but information on employment status, exposure to risk factors or other health behaviors are relatively reliable. T1 should include information collected a few days before the intervention. T2 and particularly T3, referring to the length of long-term follow-up, should be disease-, procedure- and context dependent. Ideally, the optimal

length of mid-term follow-up should be defined by research, as it can vary greatly among individual procedures; for example, three months for liver resection¹², six months for pancreatic resections¹³ and over one year for liver transplantation¹⁴. The T4 assessment should be carried out five years after the intervention (open ended from then on) and very long-term follow-ups should also be considered when appropriate.

Outcome assessment goes beyond mortality

The health care providers', mostly physicians', perspective has been the only (or predominant) view on outcomes for a long time, typically by just reporting on short-term, for example, 30-day, mortality rates. With the dramatic decrease in perioperative mortality rates following most procedures, the focus has turned toward postoperative morbidity. However, reporting on complications has been inconsistent and notoriously lacking information on the severity of the respective events and

BOX 2

Jury recommendations

Timepoint of outcome assessment

- T0: Pre-disease state
- T1: Before intervention
- T2: Early postoperative phase (short-term)
- T3: Mid-term follow-up (intervention and disease specific)
- T4: Five-year follow-up

Outcome assessment for postoperative complications

- When measuring postoperative complications use at a minimum:
 - Clavien–Dindo classification (severity of complications)
 - Comprehensive Complication Index (CCI) (total morbidity)
 - Failure to rescue
 - Conduct regular, interdisciplinary ‘morbidity and mortality conferences’ in clinical practice and include discussions of successful cases to reinforce effective behavior.

Patient-centered outcome assessment

- Patient education and empowerment are critical for them to take on responsibility for their care.
- Incorporate patient-centered outcome measures such as PROMs and PREMs into routine clinical care.
- Use, at a minimum, one standard global life satisfaction measurement (for example, EQ-5D) to quantify change over time.

Benchmarking

- Benchmarking is a prerequisite to assess and improve quality of care.
- Make benchmarking mandatory for all institutions, irrespective of their size.
- A robust methodology to create benchmark values includes low-risk patients treated at expert, high-volume centers.

Risk assessment

- Make preoperative assessment and postoperative reporting of high-risk patients mandatory and disease specific.
- Consider multiple factors such as patient-, physician- and procedure-related factors as well as the context in which patients live (for example, socio-demographic factors, social determinants of health).

- Involve ‘informed’ patients in the establishment of risk profiles and in defining their expectations of the intervention.

Data management

- Governments and relevant sectors of society must engage in establishing nationwide health care databases, working toward global standardization of data collection, curation, storage and validation.
- Hire/appoint a ‘data quality guarantor’ at every institution to ensure data accuracy and completeness.
- Involve patients in database creation and data collection so that outcome data can be shared with patients in a simple, lay-language manner.
- Inform the public about the benefits of well-designed research under strict privacy protections.
- Call on the WHO and the G20 and their global responsibility to master the political process and make the creation of anonymized centralized data centers a priority.

Approaches to secure adequate treatments

- Implement initiatives to avoid over and undertreatment, such as:
 - ‘Choosing Wisely’ initiatives
 - Binding guidelines
 - Removing financial incentives for low-value interventions

Dealing with cultural and demographic factors

- Define and incorporate cultural and demographic factors in the interpretation of outcomes after surgical interventions.

Strategies to deal with unwarranted outcomes

- Shift from a culture of blame to a culture of collaborative and collective learning.
- Develop and establish new systems and procedures to mitigate the consequences of unwarranted outcomes (for example, compensation at institutional, regional and/or national level).
- In the handling of medical errors, apply truthful disclosure and the TRACK principle: Transparency, Respect, Accountability, Continuity and Kindness.

their time of occurrence, making the evaluation of surgical procedures a ‘comic opera’¹⁵. Reliable comparisons of postinterventional outcome are only possible if results are uniformly and comprehensively reported. The ideal outcome measures should be relevant to most procedures, collected in ways that minimize bias, and interpretation must eventually be widely accepted to generate a universal language.

Postoperative negative events can be divided into three categories¹⁶. First, failure to cure – indicating that the objective of an intervention was not achieved (for example, no curative resection of a malignant tumor); second, sequelae when the negative event is inherent to the procedure (for example, amputation of a leg inevitably leads to invalidity); and third, complications covering all other events. Terms like major, severe, minor, serious, mild and intermediate must be avoided unless clearly defined.

A few systems to classify complications have been proposed, including the Clavien–Dindo^{17,18}, Accordion¹⁹ and Memorial Sloan Kettering cancer center classifications – all of which are based on an inaugural proposal made in Toronto, Canada, in 1992 to critically assess the

introduction of laparoscopic cholecystectomy¹⁶. To secure a universal language, there is a need to select the best system, which should be precise, reproducible, intuitive and quantitative, and should minimize biases in data collection. Irrespective of the system, data collection is best done independently by dedicated staff, instead of surgical interns or residents in training²⁰. The Clavien–Dindo classification (Table 1), which has been applied to most fields of surgery, fulfills these criteria best and has been widely utilized.

A limitation of the Clavien–Dindo and other classifications is that the full description of complications must be tabulated; therefore, they are difficult to use for outcome comparisons. Additionally, most studies only capture the most severe complications while omitting complications of lesser degree²¹. To address this limitation, the Comprehensive Complication Index (CCI), based on the Clavien–Dindo system, was developed to assess overall morbidity by capturing all complications in a single patient^{22,23}. The patients’ perspective was explicitly considered in the development of the CCI by allotting weights from the patient view to the respective complications. The CCI expresses

Table 1 | The Clavien–Dindo classification^{17,18}

Grade	Description
1	Any deviation from the postoperative course without the need for pharmacological treatment or surgical, radiological or endoscopic intervention. Acceptable therapeutic regimens are drugs as antiemetics, antipyretics, analgesic, diuretics, electrolytes and physiotherapy. This regimen also includes wounds opened at bedside
2	Requiring pharmacological treatment other than allowed for grade I. Blood transfusions, antibiotics and total parenteral nutrition are also included
3	Requiring surgical, endoscopic or radiological intervention
3a	Intervention under regional or local anesthesia
3b	Intervention under general anesthesia
4	Life-threatening complication requiring intensive care unit management
4a	Single-organ dysfunction
4b	Multi-organ dysfunction
5	Patient demise

the cumulative burden with a single normalized metric ranging from 0 (no complication) to 100 (death) and accounts for both the number and severity of the complications. The CCI has been validated in several independent patient cohorts^{23–25}, correlates highly with cost^{26,27} and has proven to be a highly sensitive endpoint for randomized trials^{28,29}. A web application (<https://www.cci-calculator.com>) is available for the calculation of the CCI.

There are other metrics such as the ‘textbook outcome’ approach, which refers to the proportion of patients without any negative events or just minimal deviation from the optimal clinical course³⁰. For example, in pancreatic surgery, a textbook outcome is defined as a patient without any pancreatic fistula, bile leak, severe complications or readmission after being discharged³¹. Readmission rate is another frequently used parameter, as is length of hospital stay, days alive out of hospital (DAOH) and treatment costs. In this context, DAOH represents a more global outcome that includes all reasons for hospitalization (medical issues, adjuvant therapy, and so on) and is therefore more patient-centered³². An extension of this concept, known as ‘failure to rescue’, has been developed as a new indicator of quality to highlight the ability of superior centers to recognize complications at an early stage, and therefore to properly treat those complications, thus minimizing the risk of death. Based on the extensive literature, expert panels’ assessments, and thorough discussions at the consensus conference, the Jury proposed that, as a minimum, the Clavien–Dindo classification, the CCI and failure to rescue should be used to assess outcomes when it comes to postoperative complications (Box 2).

The Jury agreed that proper assessment from the surgical perspective must include regular, interdisciplinary morbidity and mortality conferences with the intent to reflect and learn from adverse patient outcomes in real-world practice, and to find solutions to reduce the risk for adverse outcomes in the future. Good outcomes in challenging cases should also be discussed at morbidity and mortality conferences, to better understand the ‘favorable’ factors affecting outcome³³.

Patients at the center of their outcome assessment

While most metrics, except the CCI, were developed exclusively from the health care providers’ perspective, modern medicine has begun to reset its focus on the most central stakeholder, the patient, to deliver more patient-centered and holistic care.

For patients, many of the data recorded by their physicians may seem abstract. They also may give more value to their functional status after an intervention than to the quality of the non-medical services

provided, such as quality of food or comfort of the hospital room. Patient-reported outcome measures (PROMs) allow for quantitative measurement and continuous improvement of these outcomes. PROMs should be used to ensure that the patients’ voice is heard and incorporated into clinical decisions, such as in shared decision-making, which is a cornerstone of patient-focused medical practice. The incorporation of PROMs into the clinical care pathway not only highlights patients’ perception of their treatment but can also change how patients think about their condition and can even improve survival rates, as shown, for example, in lung cancer studies³⁴. PROMs can also improve the quality of interventions by considering outcomes that are inadequately represented by metrics relating only to a short-term interventional perspective; rather PROMs should (and often do) include questions relating to the entire care pathway and its integration, including the transitions of care. The Jury decided to recommend the use of PROMs in routine clinical care and research but refrained from making recommendations on the use of specific instruments. The choice of psychometrically validated PROM instruments depends on the patient population, intended use in clinical practice, and on the time frame of outcome measurement and comparability or quality improvement efforts by others.

While it is standard in some health care systems to inform and engage patients in decisions about their treatment options, patient passivity remains an issue. The challenge lies in how to effectively engage patients and to accompany them in understanding the process and benefits of shared decision-making with their physicians. Playing an active role in decision-making can be challenging for some patients, owing to the high cognitive and emotional burden requested³⁵. To truly empower patients in the process of shared decision-making, coaching patients and supporting them in self-management is of utmost importance. Offering access to adequate information on the disease, treatments and outcomes allows the patient to understand what to expect in the future and adapt to living with their disease, which is highly relevant in the case of chronic disorders. For the health care provider, this means tailoring the information presented to the individual patient, for example, using well-developed decision aids^{36,37}. While the patient is at the center of the conversation, it is important to also include loved ones such as family members or caretakers, as surgical interventions and their outcomes will also influence the people close to the patient and the relationships between them. Mutual trust between patients and their health care providers is fundamental to optimize care, and can be achieved based on empathy, kindness and a positive patient-centered environment.

Creating a trustworthy and empathic environment and listening to the patient must be part of any inclusive outcome assessment, making patient-reported experience measures (PREMs) a relevant metric for optimal patient care. Unlike PROMs, which assess patients’ health status, PREMs evaluate patients’ personal experience of receiving care; they should be monitored by means of questionnaires (such as the EQ-5D quality of life instrument) and interpreted by independent staff members (for example, study nurses).

The Jury found that the need for more patient-centered assessment and treatment is immense. They recommend internationally standardized outcome measures like PROMs and PREMs to facilitate holistic patient management and emphasize the importance of communication between health care providers and patients (Box 2). Health care providers must communicate clearly with patients to ensure that they fully understand their condition and the potential consequences of an intervention. To achieve this goal, providers may benefit from formal training. However, through the process of engaging in shared decision-making and being truly empowered, patients themselves also take on a share of the responsibility for their outcome.

Comparisons of outcomes

Credible and relevant comparisons of specific procedures across hospitals, competing therapies and over time are requested by most stakeholders

within health care systems, foremost by patients and their families. The goal is not limited to a ranking in the quality of care, but rather continuous improvement at each level of care delivery, including physicians and other health care personnel, hospitals and even health care systems.

Benchmarking. Benchmarking is a quality improvement and monitoring approach originally used in business to compare the performance of an organization to the ‘best in class’. It differs from conventional quality improvement efforts where the aim is to reach the ‘average’ result across a range of institutions. To assess the value of benchmarking in surgery, a Delphi consensus study suggested specific steps to inform the benchmarking process³⁸. Benchmarking targets are usually validated outcomes, rather than process measures for a specific operation. These outcomes are measured among ‘best case patients’ who have minimal risk factors and undergo the operation at designated ‘best centers’. These centers should have (1) a high caseload, (2) a specialized multidisciplinary team including non-surgical disciplines and (3) be part of or responsible for a national and/or international registry. This approach avoids debates about ambiguous risk adjustment and sets a target that is the best achievable result, to inspire and motivate physicians and the whole health care team. The targets require complete and accurate granular clinical data that meet source data verification standards.

By referring to a point of reference (the benchmark value), health care teams can better assess their strengths and weaknesses and strive for the best possible results. The actions taken to reduce or close the gap between an institution’s performance and the benchmark have great potential to improve outcomes. The CCI – which quantifies overall morbidity – has often been used as the main benchmarking outcome^{12,13,39,40}. Additional markers should include adverse outcomes that are relevant to specific surgical interventions (for example, anastomotic leak in colorectal surgery or graft failure after transplantation)³⁸ or textbook outcomes (as described above)^{30,41}. Despite being an important pillar of patient-centered care, PROMs have rarely been addressed in surgical benchmarking initiatives, possibly because these measures often lack context-specific validity⁴² and are not universally agreed upon or collected in surgical databases.

The Jury (Box 2) recommended comparing standardized and reproducible outcomes through benchmarking, regardless of the size of hospital or standing of the individual department. Everyone should start with benchmarking because everyone should strive to improve. The Jury calls on editors of medical journals to ensure that authors referring to benchmarking relate it to the best possible result and not just average outcome.

Risk assessment. To properly compare surgical outcomes across patient groups and institutions, it is crucial to include risk profiles of all patients in the analysis and reporting. Failure to account for risk profiles leads to behaviors of avoiding interventions on higher-risk patients, potentially decreasing these patients’ access to care. This is also counterproductive for expert health care institutions because centers involved in the management of the most complex or high-risk patients, which consequently have a lower proportion of ‘benchmark’ (that is, low-risk, straightforward) cases, hence disclose better outcomes when risk status is accounted for^{12–14,43–50}. A high ratio of complex cases can positively impact the outcomes of all patients, as they logically enhance the capability of the surgeons and the center. Additionally, patients’ expectations can be adjusted and better understood when their potential individual risk is incorporated into discussions, thereby also improving outcomes, particularly in terms of QoL⁵¹. To ensure that fair and accurate comparison of outcomes between institutions is possible, the Jury recommended mandatory reporting of standardized risk profiles of patients, taking into account not just patient factors but also physician- and procedure-related factors, for example, surgical volume or high-risk procedures like pancreatic resection (Box 2).

Data management. Another recommendation of the Jury relates to the collection, verification and management of health care data. The need for reliable data, collected through secure channels and available for research projects and quality control, was widely acknowledged during the conference. The Jury concluded that there must be a position within every institution for a ‘data quality guarantor’, who would be responsible for data collection, management and storage. The role of this person would be to not only oversee and validate data collection, but also to train personnel and be the contact person for any official or governmental site overseeing quality and data. Beyond individual health care facilities, the role of governments and regulatory bodies was also seen as crucial by the Jury and experts.

Other perspectives relevant to society

The consensus conference included panels and discussions on the perspectives of payers, governments and society at large. This perspective does not only include outcomes per se, but also the resources invested to achieve certain outcomes. Wise spending of resources requires carefully developed, evidence-based guidelines with definitions of indications for surgical interventions. For example, the ‘Choosing Wisely’ campaign (an initiative of the American Board of Internal Medicine) seeks to address these challenges for specific indications and treatments by advancing a national dialogue with all stakeholders, focusing on shared decision-making with patients as partners to define ‘wise choices’. Its goal is to avoid unnecessary medical tests, treatments and procedures, especially in areas with limited resources⁵². High-quality outcome data are central to defining indications for surgical interventions and achieving the best possible outcomes for the resources invested. Therefore, governments have a vested interest in fostering the standardization of robust outcome data.

To this end, government regulatory agencies should follow a legislative mandate to promote and protect public and individual health, assuring fidelity to that mission by carrying out monitoring and measuring outcomes. They should clarify which metrics are most appropriate for addressing the required quality priorities, particularly those that can feasibly be collected using agreed-upon definitions (for example, long-term quality in care, PROMs, PREMs and standardized assessment of postoperative complications). This should involve qualified personnel to oversee, collect or validate the data, therefore requiring appropriately relevant and accurate data sources. The adage of ‘garbage in, garbage out’ is one that all efforts should prevent, otherwise wasted efforts and resources will ensue. Rather than basing measurements of care quality on a minimum number of procedures (for example, highly specialized procedures in general surgery), regulatory bodies should instead look at the quality and accuracy of recorded data and ways to improve centralized clinical expertise, such as multidisciplinary treatment of complex diseases (for example, by the intensive care unit and surgical department). Although high hospital volume (that is, the number of times a specific procedure is done at a facility per year) was shown to be an excellent tool to improve care quality in many domains, overconcentration can also have potentially harmful effects by creating a monopolistic market and less willingness to invest in novel procedures and adequate education and training^{53–57}. Furthermore, global equity with fairness in financial contribution should be addressed, including basic coverage for everybody.

The Jury concluded that governments should be responsible for overseeing data collection, storage, management and access among researchers. Nationwide data collection enhances trust among patients, health care providers and the public. To secure the most appropriate interventions and treatments, the Jury acknowledges the importance of second opinions and removal of financial incentives – that is, monetary motivation to conduct procedures without evidence of benefit, which may lead to harm and exaggerated costs – as well as the importance of implementing initiatives like Choosing Wisely^{58,59} (Box 2).

Cultural and demographic differences in outcome interpretation

Differences in outcomes after medical interventions may occur when unjust and avoidable systemic differences exist in health care delivery that cannot be attributed to the disease, clinical indication for surgical procedures or type of procedure performed. These differences can arise from structural health systems or societal barriers to care^{60,61}. Cultural factors also impact the way patients participate in their own care after a medical intervention. How we perceive, experience and cope with disease is based on representations regarding causes and consequences of sickness, which are shaped by cultural factors, our social positions and systems of meaning⁶². Cultural issues also play a major role in patient adherence and partnership with the health care team⁶³.

While there are standards for the collection and evaluation of some specific social and demographic factors, such as employment and insurance status, there are no standards for the collection of information regarding cultural attitudes and social norms that can have an impact on the health of the individual or performance of a health system^{64–66}. Additional information on social determinants (for example, poverty, food insecurity, discrimination and unsafe housing) and cultural and demographic factors (including gender identification, religion and others) would facilitate interpretation of outcomes after medical interventions.

The Jury concluded that cultural and demographic factors might have an extensive, although so far poorly assessed, effect on outcomes and outcome assessment. The Jury suggested incorporating cultural and demographic factors into the evaluation of outcomes, through cultural adaptation of outcome measures themselves and/or consideration of socio-cultural determinants of health when interpreting outcomes. Socio-demographic data should be collected in a consistent way – for example, by defining a minimal dataset in large national databases – and should be interpreted in the context of specific cultural and demographic backgrounds (Box 2).

A new culture in dealing with unwarranted outcomes

When something goes wrong during or following a surgical intervention, it is usually the result of multiple systemic factors, rather than a single cause⁶⁷. While cases of gross negligence, recklessness or intentional harm call for an assignment of individual culpability and disciplinary action, it is critical to avoid treating the care provider as solely liable in cases of unintentional errors. Attitudes toward medical errors must focus on improving the overall process of care delivery – providing professional safety tools, training and support to clinicians so that they can express empathy, and where appropriate, apologize^{68,69}. The best lesson is to offer transparent and honest disclosure to patients and families⁷⁰, which appears to be the best modality to prevent more suffering. In line with this, the Jury recommended that health care facilities foster a shift from a culture of blame to one of collaboration and collective learning (Box 2).

The Jury also addressed the need for clearly defined systems and procedures to mitigate the consequences of unwarranted outcomes. From an ethical and legal standpoint, outcomes should be evaluated according to the consequences of the intervention (clinical outcome) and whether all required conditions were met (procedural outcome, such as compliance with the law, or informing patients about the risks). They should include benefits and harms jointly identified by care practitioners, individual patients and experts, and assessed against a standard of a decent or flourishing life, beyond just biological and psychological functioning⁷¹, as well as the process of health care journey. Developing such a standard requires further research. Discussion of clinical outcomes should be supported by evidence-based decision aids as part of shared decision-making and advance care planning.

Discussion

This consensus conference delivered Jury-based recommendations on how to assess outcomes of surgical interventions using a rigorous

BOX 3

Jury recommendations: the seven final statements

- (1) Record outcome parameters at standardized time points.
- (2) Routinely use PROMS and PREMS in clinical care.
- (3) Record individual and global morbidity according to the Clavien–Dindo classification and by using the CCI.
- (4) Define benchmark values and compare results.
- (5) Conduct routine interdisciplinary mortality and morbidity conferences.
- (6) Appoint a ‘data quality guarantor’ at every institution.
- (7) Follow the TRACK principle in case of unwarranted outcomes: Transparency, Respect, Accountability, Continuity and Kindness must be applied.

format designed to minimize biases and conflicts of interest. The Jury was composed of independent members including key stakeholders of the society, from economy, industry, psychology/psychiatry, science and patient advocates. The Jury’s recommendations were mostly based on the work of nine panels of international and multidisciplinary experts, who succeeded in delivering their responses to specific questions to the Jury well in advance of the consensus meeting.

The statements of the Jury offer a better understanding of the various stakeholders, with a particular emphasis on patients, which are too often forgotten in the delivery of health care owing to overwhelming political and financial pressures. The Jury also highlighted the responsibilities in properly assessing the results of surgical interventions, which include not only the health care providers but also governments and, most importantly, patients themselves. Unfortunately, there is no single metric available covering all aspects, and likely there will never be. With these Jury-based recommendations, however, we provide a framework for outcome assessment that may be further developed by researchers and health care providers targeting specific patient populations and interventions. The most frequently recurring questions of the Jury to the panel chairs were ‘So what can be done better? What are the precise steps and actions you suggest for assessing outcomes more accurately after surgical interventions and thereby improving the quality of patient care worldwide?’ From the answers to these questions, we summarize seven priorities that emanate from the Jury’s recommendations to credibly report on surgical interventions (Box 3).

There are some limitations to the recommendations. While all attempts were made to minimize the risk of bias, each Jury member brought their own background and opinions. However, as previous consensus conferences, the Jury does not just accept expert statements but also rejects recommendations or strongly modifies them after a productive deliberation. Next, we faced the challenge of balancing specificity with broad applicability of the recommendations. We prioritized recommendations pertinent to a broad range of surgical interventions, with the need for further adjustments based on the specifics of interventions and underlying diseases.

A final aim of the consensus exercise was to highlight areas needing more research (Box 4). For example, while the use and implementation of artificial intelligence is currently widely discussed in the assessment of surgical interventions, studies measuring its precise benefit in clinical practice and research are yet to be conducted. Also, the influence of socio-demographic and cultural factors on outcomes after surgical interventions are only now being recognized. Measuring, recording and comparing such complex determinants of health will require specialized tools, which are still lacking today.

BOX 4

Jury recommendations on research priorities

- Definitions of PROMs and PREMS, specific for a condition and/or intervention, should be jointly developed together with patients.
- Evidence-based decision aids for patients should be developed.
- Minimum standards for ‘morbidity and mortality conferences’ should be defined.
- The applicability and value of artificial intelligence, for which the role and value for improving outcomes after medical interventions is still unclear, should be investigated.
- Tools for holistic risk assessment to quantify all risk (and protective) factors, including socio-demographic ones should be developed.
- Benchmarking tools to compare health care systems within and across countries and show their impact on health outcomes of individuals should be developed.
- Feasible, easy to administer, comprehensive and scientifically sound tools, to evaluate the complex construct of ‘social determinants’ of health on outcomes are needed. Three data elements could potentially be integrated in the monitoring of surgical outcomes: length of follow-up after an intervention, economic impact and cultural/demographic factors.
- Dashboards cocreated with patients that provide patients with access to their own data and entail visualizations of patient outcomes should be developed, implemented and evaluated scientifically.

The Jury underlined this challenge relevant to all culture and societies and made a call to the WHO and G20 to specifically address these issues with the aim of achieving a level of standardization that will enable credible comparisons and improvements in the delivery of health care worldwide. This will go a long way in facilitating accurate outcome expectations for patients and thereby achieving better results.

References

1. WHO. *WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives* (accessed 16 October 2022); <https://apps.who.int/iris/handle/10665/44185>
2. Ludbrook, G. Hidden pandemic of postoperative complications – time to turn our focus to health systems analysis. *Br. J. Anaesth.* **121**, 1190–1192 (2018).
3. Martin, R. C. 2nd, Brennan, M. F. & Jaques, D. P. Quality of complication reporting in the surgical literature. *Ann. Surg.* **235**, 803–813 (2002).
4. Vonlanthen, R. et al. The impact of complications on costs of major surgical procedures: a cost analysis of 1200 patients. *Ann. Surg.* **254**, 907–913 (2011).
5. Birkmeyer, J. D. et al. Hospital quality and the cost of inpatient surgery in the United States. *Ann. Surg.* **255**, 1–5 (2012).
6. Lesurtel, M. et al. An independent jury-based consensus conference model for the development of recommendations in medico-surgical practice. *Surgery* **155**, 390–397 (2014).
7. Clavien, P. A. et al. Recommendations for liver transplantation for hepatocellular carcinoma: an international consensus conference report. *Lancet Oncol.* **13**, e11–e22 (2012).
8. Frilling, A. et al. Recommendations for management of patients with neuroendocrine liver metastases. *Lancet Oncol.* **15**, e8–e21 (2014).
9. Clavien, P. A. & Deiss, J. Leadership: ten tips for choosing an academic chair. *Nature* **519**, 286–287 (2015).
10. Lawson, E. H. et al. A comparison of clinical registry versus administrative claims data for reporting of 30-day surgical complications. *Ann. Surg.* **256**, 973–981 (2012).
11. Parthasarathy, M. et al. Are we recording postoperative complications correctly? Comparison of NHS hospital episode statistics with the American College of Surgeons National Surgical Quality Improvement Program. *BMJ Qual. Saf.* **24**, 594–602 (2015).
12. Rossler, F. et al. Defining benchmarks for major liver surgery: a multicenter analysis of 5202 living liver donors. *Ann. Surg.* **264**, 492–500 (2016).
13. Sanchez-Velazquez, P. et al. Benchmarks in pancreatic surgery: a novel tool for unbiased outcome comparisons. *Ann. Surg.* **270**, 211–218 (2019).
14. Muller, X. et al. Defining benchmarks in liver transplantation: a multicenter outcome analysis determining best achievable results. *Ann. Surg.* **267**, 419–425 (2018).
15. Horton, R. Surgical research or comic opera: questions, but few answers. *Lancet* **347**, 984–985 (1996).
16. Clavien, P. A., Sanabria, J. R. & Strasberg, S. M. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery* **111**, 518–526 (1992).
17. Dindo, D., Demartines, N. & Clavien, P. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann. Surg.* **240**, 205 (2004).
18. Clavien, P. A. et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann. Surg.* **250**, 187–196 (2009).
19. Strasberg, S. M., Linehan, D. C. & Hawkins, W. G. The accordion severity grading system of surgical complications. *Ann. Surg.* **250**, 177–186 (2009).
20. Dindo, D., Hahnloser, D. & Clavien, P. A. Quality assessment in surgery: riding a lame horse. *Ann. Surg.* **251**, 766–771 (2010).
21. Clavien, P. A. & Puhan, M. A. Biased reporting in surgery. *Br. J. Surg.* **101**, 591–592 (2014).
22. Clavien, P. A. et al. The Comprehensive Complication Index (CCI(R)): added value and clinical perspectives 3 years ‘down the line’. *Ann. Surg.* **265**, 1045–1050 (2017).
23. Slankamenac, K. et al. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann. Surg.* **258**, 1–7 (2013).
24. Kim, T. H. 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* **21**, 171–181 (2018).
25. de la Plaza Llamas, R. et al. Clinical validation of the comprehensive complication index as a measure of postoperative morbidity at a surgical department: a prospective study. *Ann. Surg.* **268**, 838–844 (2018).
26. Staiger, R. D. et al. The comprehensive complication Index (CCI(R)) is a novel cost assessment tool for surgical procedures. *Ann. Surg.* **268**, 784–791 (2018).
27. Dell-Kuster, S. et al. Prospective validation of classification of intraoperative adverse events (ClassIntra): international, multicentre cohort study. *Brit. Med. J.* **370**, m2917 (2020).
28. Slankamenac, K. 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.* **260**, discussion 757–762 (2014).
29. Boxhoorn, L. et al. Immediate versus postponed intervention for infected necrotizing pancreatitis. *N. Engl. J. Med.* **385**, 1372–1381 (2021).

30. Kolfschoten, N. et al. Focusing on desired outcomes of care after colon cancer resections; hospital variations in 'textbook outcome'. *EJSO* **39**, 156–163 (2013).
31. van Roessel, S. et al. Textbook outcome: nationwide analysis of a novel quality measure in pancreatic surgery. *Ann. Surg.* **271**, 155–162 (2020).
32. Huang, L. et al. Days alive and out of hospital after enhanced recovery video-assisted thoracoscopic surgery lobectomy. *Eur. J. Cardiothorac. Surg.* **62**, ezac148 (2022).
33. Verhagen, M. J., de Vos, M. S. & Hamming, J. F. Taking morbidity and mortality conferences to a next level: the resilience engineering concept. *Ann. Surg.* **272**, 678–683 (2020).
34. Denis, F. et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. *JAMA* **321**, 306–307 (2019).
35. Barello, S. & Graffigna, G. in *Patient Engagement* (eds Guendalina, G., Serena, B. & Stefano, T.) 78–93 (De Gruyter Open Poland, 2015).
36. Pope, T. M. & Lessler, D. Revolutionizing informed consent: empowering patients with certified decision aids. *Patient* **10**, 537–539 (2017).
37. Hostetter, M. & Klein, S. *Helping Patients Make Better Treatment Choices with Decision Aids* (accessed 23 September 2022); <https://www.commonwealthfund.org/publications/newsletter-article/helping-patients-make-better-treatment-choices-decision-aids>
38. Staiger, R. D. et al. Improving surgical outcomes through benchmarking. *Br. J. Surg.* **106**, 59–64 (2019).
39. Azoulay, D. et al. Defining surgical difficulty of liver transplantation. *Ann. Surg.* **277**, 144–150 (2021).
40. Vetterlein, M. W. et al. Improving estimates of perioperative morbidity after radical cystectomy using the European Association of Urology quality criteria for standardized reporting and introducing the Comprehensive Complication Index. *Eur. Urol.* **77**, 55–65 (2020).
41. Tsilimigras, D., Pawlik, T. & Moris, D. Textbook outcomes in hepatobiliary and pancreatic surgery. *World J. Gastroenterol.* **27**, 1524 (2021).
42. Fiore, J. Jr et al. How do we value postoperative recovery?: a systematic review of the measurement properties of patient-reported outcomes after abdominal surgery. *Ann. Surg.* **267**, 656–669 (2018).
43. Schmidt, H. M. et al. Defining benchmarks for transthoracic esophagectomy: a multicenter analysis of total minimally invasive esophagectomy in low risk patients. *Ann. Surg.* **266**, 814–821 (2017).
44. Gero, D. et al. Defining global benchmarks in bariatric surgery: a retrospective multicenter analysis of minimally invasive Roux-en-Y Gastric bypass and sleeve gastrectomy. *Ann. Surg.* **270**, 859–867 (2019).
45. Gero, D. et al. Defining global benchmarks in elective secondary bariatric surgery comprising conversional, revisional, and reversal procedures. *Ann. Surg.* **274**, 821–828 (2021).
46. Gero, D. et al. How to establish benchmarks for surgical outcomes?: a checklist based on an international expert delphi consensus. *Ann. Surg.* **275**, 115–120 (2022).
47. Breuer, E. et al. Liver transplantation as a new standard of care in patients with perihilar cholangiocarcinoma? Results from an international benchmark study. *Ann. Surg.* **276**, 846–853 (2022).
48. Schlegel, A. et al. A multicentre outcome analysis to define global benchmarks for donation after circulatory death liver transplantation. *J. Hepatol.* **76**, 371–382 (2022).
49. Abbassi, F. et al. Novel benchmark values for redo liver transplantation. Does the outcome justify the effort? *Ann. Surg.* **276**, 860–867 (2022).
50. Mueller, M. et al. Perihilar cholangiocarcinoma – novel benchmark values for surgical and oncological outcomes from 24 expert centers. *Ann. Surg.* **274**, 780–788 (2021).
51. Greenhalgh, J. et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *J. Patient Rep. Outcomes* **2**, 42 (2018).
52. ABIM Foundation. *Choosing Wisely Initiative* (accessed on 1 January 2023); <https://www.choosingwisely.org/> (2012).
53. Halm, E. A., Lee, C. & Chassin, M. R. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann. Intern. Med.* **137**, 511–520 (2002).
54. Luft, H. S. The relation between surgical volume and mortality: an exploration of causal factors and alternative models. *Medical Care* **18**, 940–959 (1980).
55. Or, Zeynep & Renaud, T. Impact du volume d'activité sur les résultats de soins à l'hôpital en France. *Économie publique/Public economics* **24–25**, 187–219 (2012).
56. Mesman, R. et al. Why do high-volume hospitals achieve better outcomes? A systematic review about intermediate factors in volume-outcome relationships. *Health Policy* **119**, 1055–1067 (2015).
57. Vonlanthen, R. et al. Toward a consensus on centralization in surgery. *Ann. Surg.* **268**, 712–724 (2018).
58. Levinson, W. & Huynh, T. Engaging physicians and patients in conversations about unnecessary tests and procedures: Choosing Wisely Canada. *CMAJ* **186**, 325–326 (2014).
59. Cassel, C. K. & Guest, J. A. Choosing Wisely: helping physicians and patients make smart decisions about their care. *JAMA* **307**, 1801–1802 (2012).
60. Healthy People 2030 U.S. Department of Health and Human Services. *Social determinants of health* (accessed on 30 April 2022); <https://health.gov/healthypeople/objectives-and-data/social-determinants-health>
61. Ndugga, N. and S. Artiga. Disparities in health and health care. 5 key questions and answers. *KFF*; <https://www.kff.org/racial-equity-and-health-policy/issue-brief/disparities-in-health-and-health-care-5-key-question-and-answers/> (2021).
62. Kleinman, A., Eisenberg, L. & Good, B. Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. *Ann. Intern. Med.* **88**, 251–258 (1978).
63. Luan, A. et al. Are we curing by cutting? A call for long-term follow up and outcomes research in global surgery interventions - perspective. *Int. J. Surg.* **87**, 105885 (2021).
64. Hayes, S. et al. The effect of insurance status on pre- and post-operative bariatric surgery outcomes. *Obes. Surg.* **25**, 191–194 (2015).
65. Rohlfling, M. L. et al. Insurance status as a predictor of mortality in patients undergoing head and neck cancer surgery. *Laryngoscope* **127**, 2784–2789 (2017).
66. Weyh, A. M., Lunday, L. & McClure, S. Insurance status, an important predictor of oral cancer surgery outcomes. *J. Oral Maxillofac. Surg.* **73**, 2049–2056 (2015).
67. Institute of Medicine (US) Committee on Quality of Health Care in America in *To Err is Human: Building a Safer Health System* (eds Kohn, L. T., Corrigan, J. M. & Donaldson, M. S.) (The National Academies Press, 2000).
68. Wu, A. W. et al. To tell the truth: ethical and practical issues in disclosing medical mistakes to patients. *J. Gen. Intern. Med.* **12**, 770–775 (1997).
69. Heaton, H. A. et al. In support of the medical apology: the nonlegal arguments. *J. Emerg. Med.* **51**, 605–609 (2016).
70. Levinson, W. Physician-patient communication. A key to malpractice prevention. *JAMA* **272**, 1619–1620 (1994).
71. Venkatapuram, S. Health, vital goals, and central human capabilities. *Bioethics* **27**, 271–279 (2013).

Acknowledgements

We would like to thank each participant and expert who attended the consensus conference in June 2022. P.-A.C., M.A.P., D.L.B. and A.D. would like to give very special thanks to S. Saldivia-Schwab and I. Obrecht for their administrative work organizing all the framework for the consensus conference. We would like to express our particular gratitude to S. Gaal for her administrative support and to A. Anselmi from Lunge Zürich, who was the professional congress organizer of the consensus conference. We would also like to thank V. Dzau, President of the National Academy of Medicine, for his support. We want to thank the local organizing committee for their valuable input throughout the three years of preparation, consisting of H. Hengartner, L. Regli, F. Ruschitzka, D. Scheidegger, U. Schnyder, K. Slankamenac, G. Thalmann and C. Wolfrum. We want to thank the moderators who guided the Jury and the participants through the panel sessions, namely A. Perrier, G. Thalmann, C. Gerber, L. Regli, D. Hahnloser, S. Dell, I. Schmitt-Optiz, G. Senti, H. Bonaumeaux, B. Ure, E. Montalvo, R. Fritsch, M. Turina and V. von Wyl. We also thank Careum School of Health where the face-to-face consensus conference was hosted. This conference received funding from some organizations in the industry, namely Medtronic (major sponsor), Laubscher (major sponsor), Bethanien (major sponsor), Astellas, Baxter, Biotest, Corzo medical, EUSA Pharma, Helsana, Johnson & Johnson, Neovii, Roche, SGC, Schweizerische Ärztezeitung, SWICA and Sprüngli. This work was also supported by the LGID (Liver and Gastrointestinal Disease) Foundation (<https://www.lgidfoundation.ch/>) (major sponsor). These supporters did not have any influence on the planning, conduct or results of this conference nor on the recommendations or the preparation, review or approval of the manuscript, or the decision to submit the manuscript for publication. The OUTCOME4MEDICINE project was endorsed by the Swiss Federal Institute of Technology Zurich (ETH), the University Hospital of Zurich (USZ), the University of Zurich (UZH), the Swiss Academy of Medical Sciences (SAMW) and the Swiss Medical Network.

Author contributions

P.-A.C. and M.A.P. conceived the idea of an international consensus conference on outcomes after medical interventions and were the chairs of the local organizing committee. Together with P.-A.C. and M.A.P., A.D. and D.L.B. were the main scientific organizers of the conference. They led the local organizing committee and all their meetings to develop the different panels and refine the respective questions. P.-A.C., M.A.P., A.D., D.L.B. and C.W. served as members of the Writing Committee, which prepared the first draft of this report and participated in subsequent development. C.W. was the

president of the Jury and, together with the Jury, C.W., D.H., N.K.M., J.M., J.F.P., S.P.M., F.U., S.S. and J.W. created, wrote and refined the final recommendations. The panel chairs, C.A.M., H.-K.Y., T.Stamm, T.Z., S.H., J.B., M.P.W.G., T.Szucs and J.E.T., together with their panel members, developed the evidence-based drafts, which were the basis of the conference. All authors reviewed and approved the final manuscript.

Competing interests

C.W. is the Vice-President/Head of Medical Affairs Europe, Canada and Partner Markets at the biotechnology company Biogen. Her contributions to the conference as a Jury member and as an author are her own independent opinion and not as an employee of Biogen. P.A.C. developed the Clavien–Dindo classification. P.A.C. and M.A.P. developed the CCI and benchmark study designs. The LOC, panel and Jury members were aware of this prior work of P.A.C. and M.A.P. throughout the process and P.A.C. and M.A.P. did not influence the Jury recommendations. J.M. is chief editor at *Nature Medicine* and has excused himself of the peer-review and editorial process of this Consensus Statement. The views expressed in this manuscript are from the panel experts and Jury members, participants and the authors. These views do not necessarily reflect those of their institutions or funders. All recommendations presented by the Jury have no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Milo A. Puhan or Pierre-Alain Clavien.

Peer review information *Nature Medicine* thanks Antonia Pinna and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Karen O’Leary, in collaboration with the *Nature Medicine* team.

Reprints and permissions information is available at www.nature.com/reprints.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

The Outcome4Medicine consensus group

Jury

Dieter Heuskel⁴, Nancy Kwan Man⁵, João Monteiro⁶, Juan F. Perellon⁷, Sara Pérez Martínez^{8,9,10}, Soraya Seedat¹¹, Frank Urbaniok^{12,13}, Carmen Walbert³ & Jayson Woodbridge⁹⁰

Panel chairs

Jeffrey Barkun¹⁴, Michael P. W. Grocott¹⁵, Samia Hurst¹⁶, Christoph A. Meier^{17,18}, Tanja Stamm^{19,20}, Thomas Szucs^{21,22}, J. E. Tuttle^{23,24}, Han-Kwang Yang^{25,26} & Thomas Zeltner²⁷

Panel 1

Eyal Zimlichman²⁸, Florian Rüter²⁹, Jens Deerberg-Wittram³⁰, Lara Bloom^{10,31,32}, Marie-Pascale Pomey^{33,34,35,36} & Susan M. Webb^{37,38,39,40}

Panel 2**Felix Berth⁴¹, Jose M. Ramia⁴², Roxane D. Staiger⁴³, Kathy Sienko⁴⁴, Marit S. de Vos⁴⁵, Henrik Kehlet⁴⁶ & Philip C. Müller¹****Panel 3****Martin Fey⁴⁷, Susan O'Reilly⁴⁸ & René Vonlanthen⁴⁹****Panel 4****Andrew C. von Eschenbach⁵⁰, Jeffrey V. Lazarus^{51,52}, Zeynep Or^{53,54} & Clifford Y. Ko^{55,56}****Panel 5****Mia Aurelia Huber^{57,58}, Jonathan Ives⁵⁹, Tanja Krones⁶⁰, Robert D. Truog⁶¹ & Sridhar Venkatapuram^{62,63}****Panel 6****Jad Abou-Khalil⁶⁴, Julio F. Fiore Jr.^{14,65}, Cesar A. Hincapié^{66,67}, Norihiro Kokudo⁶⁸ & Timothy M. Pawlik⁶⁹****Panel 7****Mark Adams⁷⁰, Joel Armitage⁷¹, Anne Lübbecke^{72,73}, Patricia Sánchez-Velázquez⁷⁴, Reto Schuepbach⁷⁵, Alicja Zientara⁷⁶, Marit Kalisvaart¹, Daniel A. Hofmaenner^{75,77}, Jean-Daniel Chiche⁷⁸, Alexander I. R. Jackson⁷⁹ & S. Ramani Moonesinghe⁸⁰****Panel 8****Fariba Abbassi^{1,2}, Bernice Elger^{81,82}, Alexander Joeris⁸³, Fabia Rothenfluh⁸⁴, Felix Schneuwly⁸⁵, Yuki Seidler¹⁹ & Guru Sivaraman⁵⁸****Panel 9****Tala Ballouz², Ana Costa-Ramon⁸⁶, Rebecca A. Snyder⁸⁷, Mirriam Tyebally Fang⁸⁸ & Claudia M. Witt⁸⁹**

⁴Boston Consulting Group, Düsseldorf, Germany. ⁵Department of Surgery, School of Clinical Medicine, LKS Faculty of Medicine, University of Hong Kong, Hong Kong, China. ⁶Chief Editor, Nature Medicine, New York, NY, USA. ⁷EHL Group, Lausanne, Switzerland. ⁸Department of Biological and Health Psychology, Autonomous University of Madrid, Madrid, Spain. ⁹Foundation for Biosanitary Research and Innovation in Primary Care (FIIBAP), Madrid Health Service, Madrid, Spain. ¹⁰EUPATI Fellow, EUPATI Foundation, Utrecht, The Netherlands. ¹¹Department of Psychiatry, South African Medical Research Council Unit on the Genomics of Brain Disorders, Stellenbosch University, Stellenbosch, South Africa. ¹²Universities of Konstanz, Konstanz, Germany. ¹³University of Zurich, Zurich, Switzerland. ¹⁴McGill University Health Center, McGill University, Montreal, Québec, Canada. ¹⁵NIHR Southampton NIHR Biomedical Research Centre, University Hospital Southampton & University of Southampton, Southampton, UK. ¹⁶Faculty of Medicine, University of Geneva, Geneva, Switzerland. ¹⁷Department of Internal Medicine, University Hospital of Zurich, Zurich, Switzerland. ¹⁸University of Geneva, Geneva, Switzerland. ¹⁹Institute of Outcomes Research, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria. ²⁰Ludwig Boltzmann Institute for Arthritis and Rehabilitation, Vienna, Austria. ²¹Helsana Group, Basel, Switzerland. ²²European Center of Pharmaceutical Medicine, University of Basel, Basel, Switzerland. ²³East Carolina Health, New Bern, NC, USA. ²⁴Fellow, Health policy at the Centers for Medicare and Medicaid Services Professor of Surgery and Critical Care, ECU Health, Greenville, NC, USA. ²⁵Seoul National University Cancer Hospital, Seoul, Korea. ²⁶Seoul National University College of Medicine, Seoul, Korea. ²⁷WHO Foundation, Geneva, Switzerland. ²⁸Sheba Medical Center, Ramat Gan, Israel. ²⁹University Hospital Basel, Basel, Switzerland. ³⁰RoMed Kliniken, Rosenheim, Germany. ³¹Ehlers-Danlos Society, New York, NY, USA. ³²Academic Affiliate Professor of Practice in Patient Engagement and Global Collaboration, Penn State College of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³³Department of Management, Evaluation and Health Policy, School of Public Health, Université de Montréal, Montréal, Québec, Canada. ³⁴Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada. ³⁵University of Montreal Hospital Center (CHUM), Université de Montréal, Montréal, Québec, Canada. ³⁶Centre of Excellence on Partnership with Patients and the Public (CEPPP), Montréal, Québec, Canada. ³⁷Department of Endocrinology, Hospital S Pau, Barcelona, Spain. ³⁸IIB-Sant Pau, Research Center for Pituitary Diseases, Barcelona, Spain. ³⁹Centre for Biomedical Network Research on Rare Diseases (CIBERER) Unit 747, Madrid, Spain. ⁴⁰Department of Medicine, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. ⁴¹Department of General, Visceral and Transplant Surgery, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany. ⁴²Department of Surgery, Hospital General Universitario Dr. Balmis, Alicante, Spain. ⁴³Department of Abdominal Surgery, Lucerne Cantonal Hospital, Lucerne, Switzerland. ⁴⁴Nursing Affairs, King Faisal Specialist Hospital and Research Centre, Jeddah, Saudi Arabia. ⁴⁵Directorate of Quality and Patient Safety, Leiden University Medical Centre, Leiden, the Netherlands. ⁴⁶Section of Surgical Pathophysiology, Rigshospitalet, Copenhagen University, Copenhagen, Denmark. ⁴⁷Department of Medical Oncology, University of Berne and Inselspital, Berne, Switzerland. ⁴⁸School of Medicine, Faculty of Health Sciences, Trinity College, Dublin, Ireland. ⁴⁹Vivévis, Clinic Hirslanden and Hirslanden Clinic im Park, Zurich, Switzerland. ⁵⁰Samaritan Health Initiatives, Cedar Park, TX, USA. ⁵¹CUNY Graduate School of Public Health and Health Policy (CUNY SPH), New York, NY, USA. ⁵²Barcelona Institute for Global Health (ISGlobal), Hospital Clinc, University of Barcelona, Barcelona, Spain. ⁵³Institute of Research in Health Economics (IRDES), Paris, France. ⁵⁴Université Paris Dauphine-PSL, Paris, France. ⁵⁵Schools of Medicine and Public Health, UCLA, Los Angeles, CA, USA. ⁵⁶Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL, USA. ⁵⁷Faculty of Law, University of Zurich, Zurich, Switzerland. ⁵⁸University Hospital Zurich, Zurich, Switzerland. ⁵⁹Centre for Ethics in Medicine, Bristol Medical School, University of Bristol, Bristol, UK. ⁶⁰Institute of Biomedical Ethics and History of Medicine, University Hospital/University of Zurich, Zurich, Switzerland. ⁶¹Center for Bioethics, Harvard

Medical School, Boston, MA, USA. ⁶²King's College London, London, UK. ⁶³Department of Philosophy and APK Institute for the Future of Knowledge, University of Johannesburg, Johannesburg, South Africa. ⁶⁴Department of Surgery, University of Ottawa, Ottawa, Ontario, Canada. ⁶⁵Department of Surgery, Faculty of Medicine, McGill University, Montreal, Québec, Canada. ⁶⁶EBPI-UWZH Musculoskeletal Epidemiology Research Group, Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich, Zurich, Switzerland. ⁶⁷University Spine Centre Zurich (UWZH), Balgrist University Hospital and University of Zurich, Zurich, Switzerland. ⁶⁸National Center for Global Health and Medicine, Tokyo, Japan. ⁶⁹Department of Surgery, The Ohio State University Wexner Medical Center & James Comprehensive Cancer Center, Columbus, OH, USA. ⁷⁰Newborn Research, Department of Neonatology, University and University Hospital Zurich, Zurich, Switzerland. ⁷¹University of Nebraska Medical Center, Omaha, NE, USA. ⁷²Division of Orthopedic Surgery, Geneva University Hospitals and University of Geneva, Geneva, Switzerland. ⁷³Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. ⁷⁴Division of Hepatobiliary and Pancreatic Surgery, Parc de Salut Mar Barcelona, Hospital del Mar Medical Research Institute (IMIM), Pompeu-Fabra University (UPF), Barcelona, Spain. ⁷⁵Institute of Intensive Care Medicine, University Hospital Zurich and University of Zurich, Zurich, Switzerland. ⁷⁶Royal Brompton & Harefield hospital NHS Foundation Trust, London, UK. ⁷⁷Bloomsbury Institute of Intensive Care Medicine, University College Hospital London, London, United Kingdom. ⁷⁸Department of Intensive Care Medicine, Centre hospitalier universitaire vaudois (CHUV), Lausanne, Switzerland. ⁷⁹Perioperative and Critical Care Theme, NIHR Southampton Biomedical Research Centre, University of Southampton, Southampton, UK. ⁸⁰University College London Hospitals NIHR Biomedical Research Centre and Central London NIHR Patient Safety Research Collaboration, London, UK. ⁸¹Institute for Biomedical Ethics, University of Basel, Basel, Switzerland. ⁸²Unit for Health Law and Humanitarian Medicine, Center for Legal Medicine, University of Geneva, Geneva, Switzerland. ⁸³AO Innovation Translation Center, Clinical Science, AO Foundation, Davos, Switzerland. ⁸⁴Navisano AG, Zurich, Switzerland. ⁸⁵Comparis.ch, Zurich, Switzerland. ⁸⁶Department of Economics & Jacobs Center for Productive Youth Development, University of Zurich, Zurich, Switzerland. ⁸⁷Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁸⁸Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland. ⁸⁹Complementary and Integrative Medicine, University of Zurich, Zurich, Switzerland. ⁹⁰Unaffiliated Jayson Woodbridge.