Background and Aims: The lack of consensus on how to define and grade adverse postoperative events has greatly hampered the evaluation of surgical procedures. A new classification of complications, initiated in 1992, was updated 5 years ago. It is based on the type of therapy needed to correct the complication. The principle of the classification was to be simple, reproducible, flexible, and applicable irrespective of the cultural background. The aim of the current study was to critically evaluate this classification from the perspective of its use in the literature, by assessing interobserver variability in grading complex complication scenarios and to correlate the classification grades with patients’, nurses’, and doctors’ perception.

Material and Methods: Reports from the literature using the classification system were systematically analyzed. Next, 11 scenarios illustrating difficult cases were prepared to develop a consensus on how to rank the various complications. Third, 7 centers from different continents, having routinely used the classification, independently assessed the 11 scenarios. An agreement analysis was performed to test the accuracy and reliability of the classification. Finally, the perception of the severity was tested in patients, nurses, and physicians by presenting 30 scenarios, each illustrating a specific grade of complication.

Results: We noted a dramatic increase in the use of the classification in many fields of surgery. About half of the studies used the contracted form, whereas the rest used the full range of grading. Two-thirds of the publications avoided subjective terms such as minor or major complications. The study of 11 difficult cases among various centers revealed a high degree of agreement in identifying and ranking complications (89% agreement), and enabled a better definition of unclear situations. Each grade of complications significantly correlated with the perception by patients, nurses, and physicians (P < 0.05, Kruskal-Wallis test).

Conclusions: This 5-year evaluation provides strong evidence that the classification is valid and applicable worldwide in many fields of surgery. No modification in the general principle of classification is warranted in view of the use in ongoing publications and trials. Subjective, inaccurate, or confusing terms such as “minor or major” should be removed from the surgical literature.

The absence of a definition and a widely accepted ranking system to classify surgical complications has hampered proper interpretation of surgical outcome data for a long time. Terms, such as minor, moderate, major, or severe complications, have been inconsistently used among authors, centers, and over time periods. A number of attempts have been made in the 1990s to classify surgical complications, but none of them have gained widespread acceptance.

In 1992, a novel approach was presented to rank complications by severity based on the therapy used to treat the complication, and differentiated 3 types of negative outcome after surgery: (a) complication, (b) failure to cure, and (c) sequelae. Although this system was used by few investigators, we revisited this grading system in 2004, after its routine use for more than 12 years. We developed a new 5-scale classification system with the aim of presenting an objective, simple, reliable, and reproducible way of reporting negative events after surgery (Appendix A). This classification was further validated before publication through a large cohort of patients, who underwent a variety of surgical procedures. The new proposal was also tested for its simplicity and “interobserver” variation in 10 centers around the world.

Similar to the initial classification, this new system was based on the type of therapy required to treat the complication. The rationale to preserve this approach was to eliminate subjective interpretation of serious adverse events and any tendency to down-grade complications, because it is based on data that are usually well documented and easily verified. To further avoid subjectivity and imprecision in complication reporting, we purposely avoided qualitative terms such as “minor” or “major” to grade the complications. Compared with the 1992 system, we also eliminated hospital stay as a criterion and increased the weight of life-threatening complications involving organ failure. We gave more emphasis to the patient perspective by introducing the notion of disability indicating the need for further follow-up, which could be added to each type of complication. Finally, the classification offered the possibility to combine grades of complications to simplify its use depending on the focus and the patient cohort that is analyzed.

This classification has been used in many centers as a tool for quality assessment in audits and every day practice, and it is increasingly used in the surgical literature. It has also been endorsed by societies and study investigators; for example, by the Transplantation Society to record living-related liver transplantation in the United States. We are also aware of many randomized trials, which have included this system in their end points. Every complication has been recorded in the prospective quality database at the University Hospital of Zurich since the beginning of its application.