Ensuring that surgeons can safely and effectively perform complex procedures is crucial for patient safety and optimal outcomes, yet it is widely believed that case volume is an inadequate proxy for competence. Petersen and colleagues are to be congratulated on a major contribution to the field of competency-based medical education (CBME) in thoracic surgery. The development of a straightforward, reliable tool for the objective assessment of competency is a significant accomplishment, and the sheer effort required to review hundreds of hours of unedited surgical footage must be recognized.

However, validation of this tool falls prey to a recurring problem in CBME research: How does one demonstrate validity other than by correlating scores against case volume or years in training? The central tenet of CBME is that volume and time are not sufficient to ensure competency, yet we use these numbers to validate our tools. As we move to higher-stakes issues of certification and credentialing, we must think carefully about how to best define competence. As Birkmeyer and colleagues demonstrated, technical skill correlates with patient outcomes. One way of validating a competency measure, then, might be to set a minimum competence score associated with a maximum level of complications.

Using the contrasting groups method, the authors determined a cutpoint separating “beginners” from “experts,” categorized by case volume. However, if the goal is to determine competence, then what is needed is a group of incompetent surgeons to contrast against competent ones, instead of low-, intermediate-, and high-volume groups. Likely, most if not all of the practicing surgeons in the study were competent to perform video-assisted thoracoscopic surgery (VATS) lobectomy, but perhaps at varying levels of proficiency. Furthermore, the inclusion of residents in the “beginner” group may introduce bias, because of how their scores were handled. If the attending physician took over any part of a resident physician’s case, it was automatically scored 1 out of 5, possibly artificially depressing the “beginner” scores and augmenting the difference between groups.

Although the authors present the cutpoint as a pass-or-fail marker, they appropriately caution against the use of the tool as the sole determinant in high-stakes assessments. Indeed, it is not clear that the false-negative and false-positive results should be interpreted as such. Do the 2 so-called failures performed by expert surgeons mean that they are incompetent? After all, it is possible that a surgeon who has performed 49 VATS lobectomies can generate a “competent” score and a surgeon who has done 501 procedures could score in the “incompetent” range on an individual case.

A VATS assessment tool is an important first step toward assessing competence in VATS lobectomy, and the researchers’ use of unbiased, objective observers in a clinical setting and the inclusion of all thoracic departments in the authors’ country provides strength to their findings. Petersen and colleagues join a growing movement in thoracic surgery, including work by Ferguson and colleagues and Meyerson and colleagues, developing various, complementary strategies for assessing competence. Although more evidence is needed before any single tool should be used as the basis of pass-fail decisions, VATS lobectomy assessment tool has potential to transform thoracic surgery CBME.
References