Video-based assessment (VBA) is gaining traction as an adjunctive, and potentially alternative, method to more traditional assessments of surgical technical skills during residency training and may someday play a role in professional certification. In this issue of *JAMA Network Open*, Ketel et al report the use of a previously developed, video-based, minimally invasive esophagectomy (MIE) competency assessment tool (CAT) to examine the association between the intraoperative technical quality of how MIE is performed at 15 Dutch hospitals and postoperative complications. Similar to the initial seminal work reporting an association of technical quality with better patient outcomes, the authors used masked reviewers with expertise in MIE and found that hospitals with better overall technical quality in the performance of the operation had lower intraoperative and severe postoperative complication rates and better performance during the anastomosis phase was associated with lower leak rates.

This important work supports the concept that superior technical skill is associated with better patient outcomes. It also builds on prior work by suggesting that expertise within the team (rather than just the individual) may be similarly critical. The association between better performance during the creation of anastomosis phase and lower postoperative leak rates is particularly compelling because it suggests that experts can accurately identify technical quality during the most critical steps of a procedure. However, as with anything, the devil is in the details. Therefore, we must consider the nuances of these data to fully understand their meaning and how they might eventually be operationalized. For example, hospitals that performed in the highest quartile of MIE-CAT score had a significantly lower rate of severe postoperative complications compared with those in the lowest quartile (18.7% in the highest quartile vs 39.2% in the lowest quartile). We should exercise caution, however, in interpreting this association because many postoperative complications may be affected by numerous other relevant patient-, process-, and hospital-level factors above and beyond the technical quality of the intraoperative care delivered.

Another important nuance is the stratification of performance into quartiles for the purposes of this analysis. While there is certainly precedent for using this type of stratification and it is often necessary for presenting data in a cohesive and interpretable manner, our natural tendency is to equate lower-ranked performance with worse outcomes. However, this type of oversimplification can be problematic because technical ability spans a range (rather than a static rank), and increments within that range may not necessarily translate into clinically meaningful differences. Furthermore, performance may be associated differently with different outcomes, some of which may or may not be expected to be directly associated with better performance. For instance, in this study, R0 resection, lymph node yield, and 30-day mortality were not clearly associated with hospital MIE-CAT performance. While intraoperative technical quality may not necessarily directly affect postoperative mortality, it might have been reasonable to expect an association with R0 resection and lymph node yield. Because they weren’t, this finding begs the question of why intraoperative and postoperative complications were associated while these other outcomes were not. It may simply be that these findings are a reminder that the goal is not, and should not be, numeric performance alone but, rather, clinically relevant competence. Because skill level represents a spectrum and a variety of skill levels may achieve the same desired result, as VBA becomes more common and takes on a larger role in evaluating surgical competence, it will be imperative to continue to anchor technical skill (as measured with these types of tools) to clinically relevant patient outcomes.
A critical challenge highlighted by this study is the relatively labor-intensive nature of creating footage and completing assessments as part of VBA. That the authors cite an unfeasible workload as the main reason they could only use 2 representative videos to create a hospital-level assessment speaks directly to this concern. While the smaller team sizes at the Dutch hospitals included in this study may have facilitated a limited number of videos providing a generalizable representation of each hospital’s performance, we cannot know what we haven’t seen. With only 30 total videos assessed from the 15 hospitals that treated the 970 patients included in the study, potential explanations for these findings may be that (1) the authors have identified a tool that generalizes perfectly (with minimal data) to all the surgical team members and patients undergoing MIE at a given hospital (seems unlikely); (2) some hospitals have identified a means of using only surgeons and teams that are highly technically adept at intraoperative and postoperative care of these patients while other hospitals only employ less technically adept team members (seems even more unlikely); and (3) the MIE-CAT score may be a surrogate for other factors or postoperative care processes that were associated with the lower complication rates.

Perhaps the most critical question to be addressed is one of transparency and whether data from these types of operative assessments should be disseminated to interested stakeholders. As elaborated previously, lower-scoring performers may in fact achieve clinically adequate performance, and the correlation between performance and outcome rates may be strongly predicated on, or vary with, the outcome being measured. Patients will invariably want to be treated by the most technically adept surgeon, and hospitals will want to employ surgeons who are most likely to have the best outcomes. So, while transparency is certainly an aspirational goal, there are important risks that must be considered. Specifically, complications may not always be preventable and can have a negative effect on surgeon well-being.6 As such, without proper context or a well-adjudicated means of interpreting the data, the information could easily be misconstrued. Until these types of questions are answered, VBA should primarily be used for internal quality improvement rather than regulatory purposes.

Every surgeon wants to become as technically adept as possible to deliver the best intraoperative and postoperative care they can for patients. Video-based assessment will undoubtedly play an important role in the way we train the next generation of surgeons and how each surgeon engages in ongoing professional development and continuous demonstration of clinical competency. In an age when surgical videos for nearly every type of surgical procedure are readily available on a variety of online platforms, the surgical profession should look forward and take an active role in scripting how we incorporate new approaches into our training and practice paradigms. For example, a substantial proportion of open operations are still performed; for esophagectomy, a minority of patients in the US undergo MIE.7 How does (or will, or can) VBA translate and be operationalized for the assessment of technical skill in open surgery? While VBA is here and likely to stay, if we learn anything from the thought-provoking work by Ketel et al.,2 it is that we still have a lot to learn. In the future, the old surgical adage of “see one, do one, teach one” may have to place greater emphasis on “see one.”
REFERENCES


