

## The role of simulator-based assessments in physician competency evaluations

Edward Crosby, MD

Published online: 30 April 2010  
© Canadian Anesthesiologists' Society 2010

*“...competence involves continuing professional development ... that is the really crucial thing, not just passing an examination.” - Colette Bowe*

Members of a self-regulating profession such as medicine have an obligation to their professional body to ensure that their performance meets the expectations of a competent professional.<sup>1</sup> The professions and the regulatory bodies that govern them have an obligation to the public to ensure that the members possess the necessary competencies for safe practice and to make certain that they are engaged in developing and updating these competencies. In the past, it was assumed that physicians in practice participated in professional development through a process of self-assessment and self-directed learning. Unfortunately, the evidence suggests that many physicians made little effort to do so. Perhaps many of them perceived no deficiencies in their knowledge base or practice patterns. However, people tend to hold overly favourable views of their own knowledge and abilities, and the least competent tend to have the most inflated self-assessments.<sup>2</sup> As well, there is a poor correlation between physicians' self-assessments and those provided by external reviewers, indicating a limited ability on the part of physicians to self-determine their true personal need for education.<sup>3,4</sup> Physicians with more experience are generally believed to have accumulated knowledge and skills during their years in practice, and such experience is thought to be equated with enhanced care. In fact, there may be an inverse relationship between the number of years that a physician has been in

practice and the quality of care that is provided.<sup>5</sup> Choudry *et al.* performed a systematic review of studies that compared the medical knowledge of physicians and the health care they provided with their age and years in practice, and they concluded that the quality of care delivered may deteriorate rather than improve with increasing physician age and duration of practice.<sup>5</sup>

Obviously, a mechanism is needed both to ensure continuing competency among practicing physicians and to meet the public's insistence that this process be realized. The question is not whether periodic practice assessments and re-validation will become a fact of life for medical practitioners; they are already. However, the metric will no longer be the number of hours of continuing medical education that a practitioner can lay claim to, as this approach was in the early iterations of maintaining competency programs. As well, the simple and subjective self-assessments that have been the staple of many competency assurance programs are likely to give way in the near future to programs that include norm-referenced self-audits and assessments by peer reviewers. To achieve the goals of public accountability, practice assessments must be carefully structured and externally audited, with considerable rigor given to the establishment of appropriate norms for benchmarking. To be fair to physicians, efforts must be made to ensure that expectations for practice are measured against accurate norms. Arbitrary or unvalidated benchmarks for care, even if agreed to by self- or externally appointed reviewing bodies, serve no one well in this process. In addition to being fundamentally fair and affording the participants the expectation of an appropriate process for assessment, a system of review will be required that is valid, reliable, and reasonable.

In this issue of the *Journal*, Borges *et al.* reported that experienced anesthesiologists demonstrated major

---

E. Crosby, MD (✉)  
Department of Anesthesiology, University of Ottawa,  
The Ottawa Hospital – General Campus, Suite 1401,  
501 Smyth Road, Ottawa, ON K1H 8L6, Canada  
e-mail: ecrosby@sympatico.ca

deviations from the American Society of Anesthesiologists' (ASA) guidelines for management of the difficult airway during a simulated cannot intubate cannot ventilate (CICV) airway scenario.<sup>6</sup> Major deviations were defined as: making more than two laryngoscopy attempts, making use of a fiberoptic bronchoscope (FOB), not using a laryngeal mask airway device (LMAD), failing to call for anesthetic help, and failing to call for surgical help. Despite the fact that the literature would support limiting the number of direct laryngoscopy attempts to two or fewer, the ASA guideline is actually silent on the number of attempts to be made before laryngoscopy is abandoned. Let alone being a major deviation from the guideline, making more than two attempts is not a deviation at all. Further, once the CICV scenario is recognized, the ASA guideline suggests that LMAD insertion *be considered*. Again, it is questionable whether substituting another intervention at this point would be considered a major deviation. As CICV is recognized and the ASA emergency pathway is entered, the recommendation is for "Emergency Non-Invasive Airway Ventilation" to be considered. The options outlined in the ASA algorithm for non-invasive ventilation are options that most anesthesiologists would likely not have had experience or a degree of comfort performing (e.g., rigid bronchoscope, Combi-tube, transtracheal jet ventilation), and it could not be concluded that the options would be effective. None of the common alternatives to the direct laryngoscope, including the FOB, is specifically identified at this juncture of the algorithm. However, the literature, much of which was published since the ASA guideline, would support making an attempt with a practiced alternative to the direct laryngoscope while awaiting arrival of help and while preparing to establish an infraglottic airway. In many instances, the literature would suggest that the chances of this intervention being successful would be far more likely than not, if made with a practiced technique.<sup>7</sup> Obviously, the imperative is to ventilate the lungs and oxygenate the patient urgently, and it would be difficult to defend multiple attempts with an alternative device or employing a non-practiced alternative.

It is this author's opinion that Borges *et al.* have applied an overly rigid interpretation of a guideline document as a benchmark with which to evaluate the practices of experienced anesthesiologists in a simulator setting. This interpretation was made despite the authors' intention—when the guideline was published in 2003—that it not be overtly prescriptive. To paraphrase the introductory statements by the guideline authors regarding their recommendations: They were intended to assist the practitioner in making decisions and could be adopted, modified, or rejected according to clinical needs and constraints. Further, they were not intended to be applied as standards or absolute requirements, and their

application would not guarantee any specific outcome. Thus, in the opinion of the authors, they could be considered at the time that they were published to describe a reasoned and prudent approach to the management of a patient with a difficult airway. It is the author's opinion that the guidelines would not be considered the only acceptable strategy for management of the difficult airway, either at the time of publication (2003) or thereafter. In fact, they should be currently evaluated with respect to the evidence published in the last decade regarding the salvage of the difficult or failed intubation.

Borges *et al.* draw our attention to three elements that are likely to figure prominently in physician competency assurance programs and are being established on multiple frontiers in the medical system:<sup>6</sup> 1) the establishment of appropriate norms or benchmarks with which to evaluate care; 2) the role of guidelines in the establishment of those norms; and 3) the role of simulation in the evaluation of physician's practices. Who will set the norms and standards for the instruments with which physicians' practices will be assessed and physician competency determined? The provincial and territorial regulatory authorities have the legislated responsibility to ensure public safety by ensuring physicians practicing in their domains have the required competencies to ensure safe and appropriate practice. Many of the regulatory authorities already operate both targeted and routine peer assessments and use instruments derived by both the assessors and the colleges. The university training programs and the certifying national colleges (e.g., Royal College of Physicians and Surgeons of Canada and the Canadian College of Family Practice) have the task of ensuring that graduating physicians possess the necessary competencies to enter practice; however, so far they have had little involvement in the assessment of physicians in practice. The role of the national specialty societies, such as the Canadian Anesthesiologists' Society, is less clear. Many have been silent in this regard, although they are likely in a position to offer commentary on safe practice for their members. Obviously, a coherent strategy is required to ensure that physicians are not subjected to multiple evaluation programs with varying objectives, assessment schemes, and remediation agendas.

The standard that a newly trained specialist must meet is that of a generalist; that is to say, they are expected to have a broad range of knowledge even if that knowledge is somewhat limited in certain subject domains. A similar motivation behind assessing the practice of experienced physicians pertains to assurance of competency. However, it should be recognized that many practicing anesthesiologists have evolved practices that commonly may exclude subspecialty domains, e.g., pediatric, obstetrics, chronic pain. The assessment scheme employed to assess their practice rightly needs to recognize this reality while preserving

evaluation strategies for domains that are common to practice in general. It is important to assess the competency of physicians with respect to their actual practice and to ensure they meet the standard within the care domains in which they operate rather than to a generalized practice. Ultimately, a process is needed that weighs, balances, and protects the interests of the stakeholders in the process, i.e., the patients who want access to safe health care, the physicians who want to maintain their ability to practice, the hospitals that want to ensure the provision of quality care and to avoid financial liability while maintaining a safe environment, and the governments who fund the enterprise and are ultimately held responsible for its integrity.<sup>8</sup>

Clinical practice guidelines are intended to improve quality of care by translating evidence into practice and ensuring the conformity of practice with the evidence available. Ideally, the guidelines promote interventions of proven benefit, discourage ineffective ones, and improve the overall consistency of care. Although guidelines might describe a reasonable and prudent pattern of practice in a given scenario, there are a number of issues that would encourage caution in promoting a guideline as a benchmark for care. Ideally, guidelines should be developed using established methodological standards and a solid evidence base.<sup>9</sup> However, guidelines published in the peer-reviewed medical literature often do not adhere well to these ideals.<sup>10</sup> In fact, many guidelines have been generated in the absence of evidence to support the authors' recommendations; they may be largely or entirely dependent on expert opinion, yet they are offered under the guise of evidence-based practice guidelines. The longevity of guidelines is also a particularly relevant concern, particularly when the underlying supporting evidence is absent, weak, or new and not affirmed. Shekelle *et al.* asked members of the original guideline panels to assess the validity of 17 current clinical practice guidelines published by the US Agency for Healthcare Research and Quality using six criteria and to determine whether revisions to the guidelines were required.<sup>11</sup> Thirteen of the seventeen (76%) required either a major or minor revision, and half of the guidelines became outdated within 5.8 years (95% confidence interval, 5-6.6 yr) from the time of publication.

Guidelines may be generated that have little or no evidence base and are dependent almost entirely on the opinion of experts. Alternatively, they may be supported by evidence that is new and innovative yet inconclusive and not yet ratified, even as its application to care is being promoted. A recent example was the enthusiastic and widespread endorsement of the administration of perioperative  $\beta$ -blockers in high-risk patients and the criticism, both implicit and explicit, of physicians who did not adhere to guidelines promoting their use. However, many non-adherent practitioners expressed reservations about the

strength and quality of the supporting evidence. The POISE study then revealed the potential for harm of the unrestrained administration of  $\beta$ -blockers in this population.<sup>12</sup> Some innovations have a tendency to engage practitioners to a greater degree than their scientific support might warrant. Ioannidis *et al.* assessed how frequently popular and highly cited studies (> 1,000 citations) were either ultimately contraindicated or less convincing results were found when the studies were replicated.<sup>13</sup> They reported that 16% of such studies were eventually contraindicated, and 16% found clinical effects larger than those reported by subsequent studies; only 44% were replicated and 24% remained unchallenged. The findings of non-randomized studies and those involving a smaller number of study subjects were more likely to be eventually challenged. Montori *et al.* reviewed the epidemiology and reporting quality of 143 randomized controlled trials—published predominantly in high-impact peer-reviewed journals—that were discontinued early for benefit.<sup>14</sup> They concluded that, in many of the trials, there was a failure to report relevant information about the decision to stop early; there was an incentive to stop early after relatively few patients experienced the endpoint driving study termination, and there were implausibly large treatment effects reported. Although it is important that guidelines have an adequate evidence base on which to support their recommendations, it is equally important that the evidence base be comprised of literature with integrity.

The use of guidelines to establish a benchmark for the evaluation of physicians' patterns of practice is especially problematic when those recommendations are established in the absence of evidence. They may not be the only pattern of practice acceptable to prudent practitioners, and they may drive practice on the basis of liability avoidance rather than on the basis of good patient care. Since guidelines cannot consider all possible clinical scenarios or patient fact sets, there may be clinical situations where it is prudent and reasonable to deviate from a guideline, especially one that is not supported by affirmed evidence. Finally, guidelines may be cited in a critical audit of a physician's practice, when, in fact, there is no evidentiary basis for the criticism.

Medical simulators have been studied extensively for the purpose of skills training, particularly among novice practitioners.<sup>15</sup> Simulation may allow novices to move more quickly through the stages of motor skill acquisition so as to achieve a higher level of technical competence over shorter time periods. Automation of the technical aspect of procedure may then allow the novice practitioner to devote greater attention to other aspects of clinical performance, such as new knowledge acquisition and clinical decision-making when the technique is subsequently applied in a clinical scenario. Lynagh *et al.* performed a systematic

review of the published literature and concluded that medical skills laboratories do lead to improvement in procedural skills acquisition.<sup>16</sup> However, the authors also reported the lack of well-designed trials addressing the issue of transferring the acquired skills to clinical practice and retaining the skills over time. Kudavalli *et al.* concluded that simulation-based training significantly improved performance in the management of the anticipated difficult airway, but the performance enhancement was short-lived.<sup>17</sup> There is emerging evidence that simulation-based training may also influence patient outcomes as well as early skills performance, although the literature in this regard is novel and limited. For example, Barsuk *et al.* reported that a simulator-based educational intervention in central venous catheter insertion significantly reduced the incidence of catheter-related bloodstream infection and improved patient outcomes.<sup>18</sup>

Although there is considerable evidence to support a role for simulators in novice skills and practice training, their role in assessing performance and determining competence of more experienced practitioners remains unclear. Both Devitt *et al.* and Murray *et al.* reported that simulator-based evaluation techniques assessing clinical performance were able to discriminate the level of training of anesthesiologists and could differentiate between categories of physicians.<sup>19,20</sup> However, additional studies are needed to determine whether these simulation-based assessments are valid measures of clinical performance. Also, the elements of the assessments need to be rigorously reviewed; otherwise their relevance might be subject to criticism. For example, Kurrek *et al.* tested the hypothesis that the adherence to Advanced Cardiac Life Support (ACLS) guidelines during resuscitation of ventricular fibrillation as assessed in a simulator environment would be poorer by anesthesiologists not trained in ACLS compared with those who had received training.<sup>21</sup> They reported that a greater proportion of subjects without previous ACLS training demonstrated major deviations from protocol than did subjects who had received training. However, the most common major deviation from the ACLS protocol cited was the failure of the participant to discontinue the anesthetic (turn off the vaporizer), an intervention on which the ACLS guidelines were actually silent. In the absence of circulation, some might argue that whether the vaporizer was on or off was a moot point; the imperative was rapid defibrillation, and any task that delayed that intervention was of secondary importance. Some might consider it excessive to criticize the practice of the anesthesiologists on the above mentioned basis.

Practice review for the purposes of competency validation is here and now. It is not clear how the different tiers in the medical system will work together, but it is abundantly clear that regulatory authorities must be in a position

to demonstrate to the public and to government that competency evaluations are assured. A coherent reliable valid strategy is required to ensure that the process is fair, efficient, and effective and that the norms used to assess practice are appropriate. Guidelines are now an integral part of medical practice, but their use in formulating the norms used to assess clinical behaviour should be restrained. Efforts should be made to ensure that the guidelines used in this fashion are relevant, current, and based on evidence that has been affirmed. The role of simulators for the purpose of practice evaluations has yet to be determined. The proponents' enthusiasm for this application is appreciated, but it remains in the investigative phase, and further research is both desirable and encouraged. Excessively critical assessment schemes are unlikely to enhance the credibility of simulator-based evaluations among the profession.

## Le rôle des simulateurs pour l'évaluation des compétences d'un médecin

*« ...la compétence implique le développement professionnel continu ...c'est là l'élément le plus important; il ne s'agit pas simplement de passer un test. » - Colette Bowe*

Les membres d'une profession auto-réglementée telle que la médecine ont, vis-à-vis leur ordre professionnel, l'obligation de s'assurer que leur rendement correspond à la définition d'un professionnel compétent.<sup>1</sup> Les professions et les organismes de réglementation qui les gouvernent ont, quant à eux, l'obligation envers le public de s'assurer que leurs membres possèdent les compétences nécessaires pour pratiquer en toute sécurité et qu'ils s'emploient à perfectionner ces compétences. Par le passé, on présumait que les médecins pratiquants se perfectionnaient sur le plan professionnel par l'intermédiaire d'un processus d'autoévaluation et d'apprentissage autonome. Malheureusement, tout porte à croire que plusieurs médecins ne faisaient que très peu d'efforts en ce sens. On peut présumer que plusieurs d'entre eux ne décelaient aucune lacune dans leur base de connaissance ou modèle de pratique. Toutefois, l'être humain a tendance à entretenir une image exagérément favorable de ses propres connaissances et habiletés, et les personnes les moins compétentes sont souvent celles qui se perçoivent de la manière la plus optimiste.<sup>2</sup> En outre, il n'existe qu'une faible corrélation entre l'autoévaluation du médecin et celle produite par un évaluateur externe, ce qui indique que le médecin possède une capacité limitée à déterminer

ses propres besoins en termes de formation.<sup>3,4</sup> On croit généralement que les médecins plus expérimentés ont accumulé davantage de connaissances et d'habiletés au fil des ans, et que cette expérience se traduit par des soins de meilleure qualité. En réalité, il pourrait y avoir un rapport inverse entre le nombre d'années de pratique d'un médecin et la qualité des soins offerts.<sup>5</sup> Choudry *et coll.* ont examiné de manière systématique des études comparant l'âge et le nombre d'années de pratique des médecins à leurs connaissances médicales et aux soins de santé qu'ils étaient en mesure d'offrir. Ils concluent que la qualité des soins peut se détériorer, et non s'améliorer, avec l'âge et le nombre d'années de pratique du médecin.<sup>5</sup>

Il apparaît évident que la mise en œuvre d'un processus est nécessaire, tant pour s'assurer du maintien des compétences des médecins pratiquants que pour répondre à l'insistance du public. La question n'est pas de savoir si l'évaluation périodique des pratiques et la revalidation des qualifications deviendront une réalité concrète pour les praticiens; elles le sont déjà. Or, le nombre d'heures de formation médicale continue déclarées par le médecin ne sera plus employé à titre de mesure, puisqu'il a été utilisé dans les premières itérations des programmes de maintien des compétences. De plus, dans un avenir rapproché, les auto-évaluations simples et subjectives, qui forment actuellement la base de plusieurs programmes de vérification des compétences, laisseront probablement place à des programmes composés d'auto-examens et d'auto-évaluations effectués par des pairs selon des normes établies. Pour atteindre les objectifs établis en termes de responsabilité à l'égard du public, les évaluations de pratique devront être soigneusement structurées et vérifiées à l'externe, en plus de se montrer très rigoureuses quant à l'établissement des normes appropriées aux fins d'analyse comparative. Pour faire preuve d'équité envers les médecins, des efforts devront être déployés afin de s'assurer que les attentes en matière de pratique sont mesurées selon des normes adéquates. Les références arbitraires ou non validées en termes de soins, même celles approuvées par le médecin concerné ou par un organisme d'évaluation externe, n'aideront personne dans ce processus. En plus d'être fondamentalement équitable et de permettre aux participants de s'attendre à un processus d'évaluation adéquat, le système d'évaluation devra être valable, fiable et raisonnable.

Dans la présente édition du *Journal*, Borges *et al.* démontrent que les anesthésiologistes expérimentés dévient considérablement des lignes directrices établies par l'*American Society of Anesthesiologists* (ASA) en ce qui concerne la prise en charge de l'intubation difficile dans une simulation au cours de laquelle l'intubation et la ventilation sont impossibles (« cannot intubate cannot ventilate » - CICV).<sup>6</sup> Les écarts majeurs sont définis en

fonction des critères suivants : plus de deux tentatives de laryngoscopie, utilisation d'un bronchoscope à fibres optiques (BFO), non-recours au masque laryngé (ML), défaut de demander de l'aide sur le plan anesthésique et défaut de demander de l'aide sur le plan chirurgical. Bien que la littérature privilégie une limite au nombre de tentatives de laryngoscopie directe à deux ou moins, les lignes directrices de l'ASA ne font aucune mention du nombre de tentatives recommandées avant l'abandon de cette méthode. En effet, le fait de tenter deux laryngoscopies ou plus n'est pas considéré comme un écart, encore moins un écart majeur. De plus, une fois le scénario CICV reconnu, les lignes directrices de l'ASA recommandent de *considérer* la pose d'un ML. Une fois encore, on peut se demander si, à ce stade, la substitution par une autre intervention serait considérée comme un écart majeur. Lorsque le scénario CICV a été reconnu et que l'approche d'urgence de l'ASA a été mise en œuvre, il est recommandé d'évaluer la possibilité d'une « ventilation non invasive d'urgence des voies aériennes ». Les options décrites par l'algorithme de l'ASA en matière de ventilation non invasive sont des options avec lesquelles la plupart des anesthésiologistes sont susceptibles d'être inexpérimentés ou de ne pas se sentir à l'aise (p. ex., bronchoscope rigide, Combitube, Jet-ventilation transtrachéale), et il est impossible de conclure que ces options seraient efficaces. Aucune des solutions de rechange habituelles à la laryngoscopie directe, y compris le BFO, n'est spécifiquement identifiée à ce point de l'algorithme. Toutefois, la littérature, dont la majeure partie a été publiée après les lignes directrices de l'ASA, appuie la tentative d'une solution de rechange connue à la laryngoscopie directe en attendant de l'aide et en préparant l'établissement d'une voie infraglottique. Dans plusieurs cas, la littérature porte à croire que les probabilités de réussite de cette intervention sont très bonnes si l'on utilise une technique connue.<sup>7</sup> De toute évidence, puisqu'il est impératif de ventiler les poumons et d'oxygéner le patient de toute urgence, il serait difficile de justifier de multiples tentatives à l'aide d'un autre dispositif ou l'emploi d'une solution inconnue.

Nous sommes d'avis que Borges *et coll.* ont utilisé une interprétation beaucoup trop rigide des lignes directrices à titre de référence pour évaluer les pratiques des anesthésiologistes expérimentés dans le cadre d'une simulation. Cette interprétation a été utilisée malgré l'intention initiale des auteurs (lorsque la ligne directrice a été publiée en 2003), c'est-à-dire d'éviter l'établissement d'un caractère exagérément normatif. Voici une version paraphrasée des déclarations liminaires des auteurs de la ligne directrice en ce qui a trait à leurs recommandations: elles sont destinées à aider les praticiens à prendre des décisions et peuvent être adoptées, modifiées ou rejetées en

fonction des besoins et des restrictions cliniques. En outre, elles n'ont pas pour but d'être appliquées à titre de normes ou d'exigences absolues, et leur application ne permet pas de garantir un résultat spécifique. Par conséquent, selon les auteurs, au moment de leur publication, elles auraient pu, au moment de leur publication, être interprétées comme la description d'une approche raisonnée et prudente pour la prise en charge d'un cas d'intubation difficile. Nous sommes d'avis que les lignes directrices ne seraient pas considérées comme l'unique stratégie acceptable pour la prise en charge de l'intubation difficile, que ce soit au moment de leur publication (2003) ou par la suite. En fait, elles devraient être mises à jour en fonction des données probantes publiées au cours de la dernière décennie en ce qui a trait à la prise en charge de l'intubation difficile ou ratée.

Borges *et coll.* soulignent trois éléments susceptibles d'apparaître de manière prédominante dans les programmes de vérification des compétences des médecins et qui, à l'heure actuelle, sont mis en place sur plusieurs plans au sein du système médical:<sup>6</sup> 1) l'établissement de normes ou de références appropriées pour évaluer les soins offerts; 2) le rôle des lignes directrices dans l'établissement de ces normes; et 3) le rôle de la simulation dans l'évaluation des pratiques. Qui établira les normes et références relatives aux instruments par l'intermédiaire desquels les pratiques et compétences des médecins seront évaluées? Les autorités de réglementation provinciales et territoriales ont la responsabilité légale de garantir la sécurité du public en s'assurant que les médecins pratiquant dans leurs domaines possèdent les compétences essentielles à une pratique sécuritaire et adéquate. Parmi les autorités de réglementation, nombreuses sont celles qui emploient déjà des évaluations par les pairs ciblées et régulières, ainsi que des instruments dérivés tant par les évaluateurs que par les collègues. Les programmes de formation universitaire et les collèges nationaux de certification (p. ex., le Collège royal des médecins et chirurgiens du Canada et le Collège des médecins de famille du Canada) sont chargés de s'assurer que les médecins diplômés possèdent les compétences nécessaires pour pratiquer; hors, jusqu'à maintenant, ils ne se sont que très peu impliqués dans l'évaluation des médecins pratiquants. Le rôle des sociétés nationales de spécialistes, comme la Société canadienne des anesthésiologistes, est moins clair. Plusieurs sont demeurées silencieuses à cet égard, bien qu'elles soient probablement habilitées à formuler des commentaires sur la pratique sécuritaire auprès de leurs membres. De toute évidence, une stratégie cohérente est nécessaire pour s'assurer que les médecins ne sont pas soumis à de multiples programmes d'évaluation se fondant sur divers objectifs, schémas d'évaluation et critères de correction.

Les critères auxquels un spécialiste nouvellement formé doit se conformer sont ceux d'un généraliste, ce qui signifie qu'on attend de lui qu'il possède une base de connaissances très large, même si ces connaissances sont quelque peu limitées dans certains domaines. Une motivation comparable se cache derrière l'évaluation de la pratique des médecins expérimentés se rapportant à la vérification des compétences. Il convient toutefois de reconnaître que de nombreux anesthésiologistes pratiquants ont développé des pratiques qui, en règle générale, peuvent exclure les domaines de sous spécialité, par exemple, la pédiatrie, l'obstétrique et la douleur chronique. À cet effet, le schéma utilisé pour évaluer leur pratique doit reconnaître cette réalité, tout en préservant les stratégies d'évaluation associées aux domaines communs de la pratique dans son ensemble. Il est important d'évaluer les compétences des médecins en ce qui a trait à leur pratique réelle, ainsi que de s'assurer qu'ils répondent aux normes établies dans leur domaine de soins, et non à celles d'une pratique généralisée. En fin de compte, le processus nécessaire doit pondérer, équilibrer et protéger les intérêts de ses parties prenantes, à savoir les patients qui veulent accéder à des soins de santé sûrs, les médecins qui veulent conserver leur capacité à pratiquer, les hôpitaux qui veulent garantir la prestation de soins de qualité et éviter les obligations financières tout en maintenant un environnement sûr et les gouvernement qui financent l'entreprise et sont, ultimement, tenus responsables de son intégrité.<sup>8</sup>

Les lignes directrices de la pratique clinique ont pour but d'améliorer la qualité des soins en intégrant les données probantes dans la pratique et en garantissant la conformité de cette pratique aux données probantes disponibles. Idéalement, les lignes directrices privilégient les interventions dont les bienfaits ont été établis, dissuadent les médecins d'utiliser des interventions inefficaces et améliorent la cohérence globale des soins. Bien que les lignes directrices puissent décrire un modèle de pratique raisonnable et prudent dans un scénario donné, il existe nombre de situations incitant à la prudence quant à la promotion d'une ligne directrice en tant que référence en matière de soins. Idéalement, les lignes directrices doivent être élaborées à l'aide de normes méthodologiques établies et de preuves indiscutables.<sup>9</sup> Cependant, il arrive souvent que les lignes directrices publiées dans la littérature révisée par les pairs ne correspondent pas à ces idéaux.<sup>10</sup> En fait, de nombreuses lignes directrices ont été créées en l'absence de données probantes appuyant les recommandations des auteurs; elles peuvent être largement ou entièrement dépendantes de l'opinion des experts, mais demeurent offertes sous le couvert de lignes directrices fondées sur des données probantes. La longévité des lignes directrices est également une préoccupation particulièrement pertinente, surtout lorsque les données

probantes sous-jacentes sont inexistantes, faibles, ou nouvelles et non confirmées. Shekelle *et coll.* ont demandé aux membres des comités d'origine responsables de l'établissement des lignes directrices d'évaluer la validité de 17 lignes directrices actuelles de la pratique clinique publiées par la *US Agency for Healthcare Research and Quality* en se basant sur six critères, ainsi que de déterminer s'il était nécessaire de les réviser.<sup>11</sup> Treize des dix-sept membres d'origine (76 %) ont répondu qu'une révision majeure ou mineure était requise et que la moitié des lignes directrices étaient devenues désuètes au cours des 5,8 années (intervalle de confiance à 95 %, 5 à 6,6 ans) suivant la date de publication.

Il est possible de générer des lignes directrices basées sur des données probantes minimales ou inexistantes et presque entièrement dépendantes de l'opinion des experts. D'un autre côté, elles peuvent s'appuyer sur des données probantes novatrices à la fois non concluantes et non confirmées, et ce, même si leur application pour les soins est privilégiée. À titre d'exemple récent, nous pourrions citer la promotion enthousiaste et généralisée de l'administration de  $\beta$ -bloquants chez les patients à haut risque, et les critiques, tant implicites qu'explicites, des médecins qui n'adhéraient pas aux lignes directrices privilégiant leur utilisation. Cependant, plusieurs praticiens non adhérents ont exprimé des réserves quant à la solidité et la qualité des données probantes. L'étude POISE a ensuite révélé les éventuels dangers liés à l'administration non limitée de  $\beta$ -bloquants dans cette population.<sup>12</sup> Certaines innovations présentent une tendance à engager les praticiens au-delà du niveau justifié par les preuves scientifiques qui les soutiennent. Ioannidis *et coll.* ont évalué la fréquence à laquelle les études populaires et souvent citées (> 1 000 citations) étaient ultimement contredites ou présentaient des résultats moins convaincants lors d'études similaires.<sup>13</sup> Ils ont indiqué que 16 % des études de ce type étaient éventuellement contredites, et que dans 16 % des cas, des études subséquentes avaient révélé des effets cliniques plus importants; seulement 44 % d'entre elles ont été reproduites, et 24 % n'ont pas été contestées. Les résultats des études non randomisées et des études impliquant un plus petit nombre de sujets étaient plus susceptibles de se voir éventuellement contestées. Montori *et coll.* ont examiné l'épidémiologie et la qualité des rapports de 143 essais contrôlés et randomisés (publiés principalement dans des revues à grand facteur d'impact avec comité de lecture) qui ont été interrompus de manière précoce à cause d'un bénéfice établi.<sup>14</sup> Ils ont conclu que dans plusieurs des études, des renseignements importants concernant la décision d'interrompre l'essai de manière précoce n'avaient pas été rapportés; il existait un incitatif à interrompre l'étude de manière précoce après qu'un

nombre relativement faible de patients ait atteint le critère d'évaluation conduisant à l'annulation de l'étude; et les effets constatés du traitement avaient une importance peu vraisemblable. Bien qu'il soit important que les lignes directrices soient fondées sur des données probantes adéquates en vue d'appuyer leur recommandation, il est tout aussi important que les données probantes soient composées de littérature intègre.

L'usage des lignes directrices en vue d'établir une référence pour l'évaluation des modèles de pratique des médecins est particulièrement problématique lorsque ces recommandations sont établies en l'absence de données probantes. Il est possible qu'elles ne constituent pas un modèle de pratique acceptable pour les praticiens prudents et incitent à une pratique fondée sur l'évitement de tout litige, et non sur la prestation de soins de qualité aux patients. Étant donné que les lignes directrices ne peuvent tenir compte de tous les scénarios cliniques possibles et de tous les faits concernant les patients, il peut exister des situations cliniques au cours desquelles il serait prudent et raisonnable de s'écarter d'une ligne directrice, surtout d'une ligne directrice non appuyée par des données probantes confirmées. Finalement, des lignes directrices peuvent être citées lors d'une vérification critique de la pratique d'un médecin lorsque, en réalité, il n'existe aucune base probante pour une telle critique.

On a beaucoup étudié les simulateurs médicaux utilisés à des fins de formation axée sur des compétences, surtout chez les débutants.<sup>15</sup> La simulation peut permettre aux débutants de compléter plus rapidement les stades de l'acquisition des aptitudes motrices afin d'atteindre un niveau plus élevé de compétence technique en moins de temps. L'automatisation de l'aspect technique de la procédure peut ensuite permettre au praticien débutant de se concentrer davantage sur d'autres aspects de son rendement clinique, comme l'acquisition de nouvelles connaissances et la prise de décisions cliniques, lorsque la technique est, par la suite, appliquée dans un scénario clinique. Lynagh *et coll.* ont effectué un examen systématique de la littérature publiée et conclu que les laboratoires d'acquisition de compétences médicales entraînent une amélioration réelle des compétences en cours d'intervention.<sup>16</sup> Cependant, les auteurs ont également indiqué le manque d'études bien conçues portant sur la question du transfert des compétences acquises à la pratique clinique et le maintien de ces compétences au fil du temps. Kudavalli *et coll.* ont conclu que la formation basée sur des simulations améliorerait de manière considérable le rendement lors de la prise en charge d'une intubation difficile anticipée, mais que cette amélioration n'était que de courte durée.<sup>17</sup> Il existe de nouvelles données probantes indiquant que la formation basée sur des simulations pourrait également influencer sur les

résultats constatés auprès des patients, ainsi que sur le rendement des compétences initiales, et ce, même si la littérature à cet égard est nouvelle et limitée. Par exemple, Barsuk *et coll.* ont indiqué qu'une intervention éducative par l'intermédiaire d'un simulateur pour l'insertion d'un cathéter veineux central diminuait considérablement l'incidence des infections hématogènes associées au cathéter et améliorait le sort des patients.<sup>18</sup>

Bien qu'il existe d'importantes données venant appuyer le rôle des simulateurs dans le cadre de la formation des débutants en termes de compétences et de pratique, leur rôle quant à l'évaluation du rendement et à l'identification des compétences des praticiens plus expérimentés demeure vague. Devitt *et coll.* et Murray *et coll.* ont démontré que les techniques d'évaluation du rendement clinique par un simulateur étaient en mesure de distinguer le niveau de formation des anesthésiologistes et permettaient de classer les médecins dans différentes catégories.<sup>19,20</sup> Toutefois, d'autres études sont nécessaires afin de déterminer si ces évaluations par simulateur constituent des mesures valables du rendement clinique. En outre, les éléments de l'évaluation doivent être rigoureusement examinés; autrement, leur pertinence pourrait être critiquée. Par exemple, Kurrek *et coll.* ont testé l'hypothèse selon laquelle l'adoption des lignes directrices en termes de technique spécialisée de réanimation cardio-respiratoire (Advanced Cardiac Life Support – ACLS) au cours de la réanimation de la fibrillation ventriculaire telle qu'évaluée dans un environnement de simulation serait plus faible chez les anesthésiologistes qui n'ont pas été formés en fonction des lignes directrices ACLS que chez ceux qui ont été formés.<sup>21</sup> Ils ont indiqué qu'une plus grande proportion de sujets n'ayant jamais reçu de formation ACLS se sont davantage éloignés du protocole que les sujets ayant reçu une formation. Cependant, l'écart majeur le plus courant au protocole ACLS était le défaut des participants d'interrompre l'anesthésie (éteindre le vaporisateur), une intervention à propos de laquelle les lignes directrices ACLS ne se prononçaient pas. En l'absence de circulation, certains pourraient s'accorder pour dire que l'état du vaporisateur était un sujet de controverse, sans intérêt pratique; la clé de l'intervention était la défibrillation rapide, et toute tâche venant retarder cette intervention était d'une importance secondaire. Certains pourraient juger qu'il est superflu de critiquer la pratique des anesthésiologistes en s'appuyant sur les faits mentionnés ci-dessus.

L'évaluation de la pratique aux fins de validation des compétences est un fait d'actualité. Rien ne permet de prévoir comment les différents intervenants du système médical collaboreront, mais il ne pourrait être plus clair que les autorités réglementaires doivent occuper une position leur permettant de démontrer au public et au

gouvernement que les évaluations de compétences sont garanties. Une stratégie cohérente, fiable et valable est nécessaire pour s'assurer que le processus est équitable et efficace, et que les normes utilisées pour évaluer la pratique sont appropriées. Les lignes directrices font maintenant partie intégrante de la pratique médicale, mais leur usage en ce qui a trait à l'établissement de normes utilisées pour évaluer le comportement clinique devrait être restreint. Des efforts devraient être déployés afin de s'assurer que les lignes directrices utilisées à cet effet sont pertinentes, actuelles et fondées sur des données probantes confirmées. Le rôle des simulateurs aux fins d'évaluation de la pratique n'a toujours pas été déterminé. L'enthousiasme des promoteurs de cette application est apprécié, bien qu'elle demeure un domaine d'investigation, et la tenue d'autres recherches est à la fois souhaitable et encouragée. Il est peu probable que des schémas d'évaluation exagérément critiques améliorent la crédibilité des évaluations par simulation au sein de la profession.

**Competing interests** None declared.

## References

1. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *JAMA* 2006; 296: 1137-9.
2. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 1999; 77: 1121-34.
3. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006; 296: 1094-102.
4. Barnsley L, Lyon PM, Ralston SJ, et al. Clinical skills in junior medical officers: a comparison of self-reported and observed competence. *Med Educ* 2004; 38: 358-67.
5. Choudry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med* 2005; 142: 260-73.
6. Borges BC, Boet S, Siu LW, et al. Incomplete adherence to the ASA difficult airway algorithm is unchanged after a high-fidelity simulation session. *Can J Anesth* 2010; 57. doi:10.1007/s12630-010-9322-4.
7. Crosby E. The unanticipated difficult airway-evolving strategies for successful salvage. *Can J Anesth* 2005; 52: 562-7.
8. Van Tassel K. Hospital peer review standards and due process: moving from tort doctrine toward contract principles based on clinical practice guidelines. *Seton Hall Law Rev* 2006; 36: 1179-256.
9. Wollersheim H, Burgers J, Grol R. Clinical guidelines to improve patient care. *Neth J Med* 2005; 63: 188-92.
10. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281: 1900-5.
11. Shekelle PG, Ortiz E, Rhodes S, et al. Validity of the Agency for Healthcare Research and Quality Clinical Practice guidelines:



- how quickly do guidelines become outdated? *JAMA* 2001; 286: 1461-7.
12. *POISE Study Group; Devereaux PJ, Yang H, Yusuf S, et al.* Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008; 371: 1839-47.
  13. *Ioannidis JP.* Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294: 218-28.
  14. *Montori VM, Devereaux PJ, Adhikari NK, et al.* Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; 294: 2203-9.
  15. *Castanelli DJ.* The rise of simulation in technical skills teaching and the implications for training novices in anaesthesia. *Anaesth Intensive Care* 2009; 37: 903-10.
  16. *Lynagh M, Burton R, Sanson-Fisher R.* A systematic review of medical skills laboratory training: where to from here? *Med Educ* 2007; 41: 879-87.
  17. *Kudavalli PM, Jervis A, Tighe SQ, Robin NM.* Unanticipated difficult airway management in anaesthetised patients: a prospective study of the effect of mannequin training on management strategies and skill retention. *Anaesthesia* 2008; 63: 364-9.
  18. *Barsuk JH, Cohen ER, Feinglass J, McGahie WC, Wayne DB.* Use of simulation-based education to reduce catheter-related bloodstream infection. *Arch Intern Med* 2009; 169: 1420-3.
  19. *Devitt JH, Kurrek MM, Cohen MM, Cleave-Hogg D.* The validity of performance assessments using simulation. *Anesthesiology* 2001; 95: 36-42.
  20. *Murray DJ, Boulet JR, Avidan M, et al.* Performance of residents and anesthesiologists in a simulation-based skill assessment. *Anesthesiology* 2007; 107: 705-13.
  21. *Kurrek MM, Devitt JH, Cohen M.* Cardiac arrest in the OR: how are our ACLS skills? *Can J Anaesth* 1998; 45: 130-2.