Consensus-Derived Coronary Anastomotic Checklist Reveals Significant Variability Among Experts

Ara A. Vaporciyan, MD, MHPE, Vid Fikfak, MD, Matthew C. Lineberry, PhD, Yoon Soo Park, PhD, and Ara Tekian, PhD, MHPE

Department of Thoracic and Cardiovascular Surgery, Division of Surgery, MD Anderson Cancer Center, University of Texas, and Department of General Surgery, Houston Methodist Hospital, Houston, Texas; Department of Health Policy and Management, Zamierowksi Institute for Experiential Learning, University of Kansas Medical Center, Kansas City, Kansas; and Department of Medical Education, College of Medicine, University of Chicago, and Department of Medical Education, College of Medicine, University of Illinois at Chicago, Chicago, Illinois

Background. **Surgical skill assessment tools frequently reflect the opinions of small groups of surgeons. That raises concerns over their generalizability as well as their utilization when applied broadly. A Delphi approach could engage a broad group of experts to identify key elements for a checklist assessing coronary anastomotic skill, improving generalizability.**

Methods. **Expert surgeons in North America (10 or more years in practice, actively teaching coronary artery surgery) were contacted randomly to participate. Consenting surgeons first provided items they believed were mandatory when performing a coronary artery bypass. These were then entered into a three-round Delphi. Positive consensus was reached when 75% or more of participants ranked an item mandatory.**

Results. **Sixteen faculty consented to participate. Each participant provided 25 ± 10 items. The 407 items provided were condensed, resulting in 146 items in the final list, divided into six sections based on the conduct of the operation. Twenty-three items reached consensus in the first round, 14 in the second, and 3 in the third. These 40 items represented only 27% of the initial 146 items. Agreement within sections varied widely, from 0% for "management of assistants" to 47% for "testing and final steps."**

Conclusions. **A randomly selected group of experts using a Delphi approach can generate a checklist to assess construction of a coronary artery bypass. Considerable disagreement among experts regarding what steps are mandatory calls into question the generalizability of any locally developed checklist.**

(Ann Thorac Surg 2017;■:■–■)

Education in cardiothoracic surgery, as with all surgical disciplines, requires the trainee to acquire both cognitive and motor skills. Whereas assessment of cognitive skills is addressed with a number of tools well suited to their evaluation (eg, multiple-choice question examinations), motor skills have very few assessment tools currently in use. To address this gap, new tools for many procedures have been developed, including checklists, objective structured assessments of technical skills, virtual reality simulators, and so forth [1, 2]. Unfortunately, important outcome measures are missed by many of these as the majority have been developed for basic surgical skills assessment, rather than advanced open surgical skills—the dominant skill in cardiothoracic surgery.

With roughly 400,000 CAB surgeries performed annually in the United States alone, coronary artery bypass (CAB) surgery is the most common procedure in cardiothoracic surgery [3]. A key component of that procedure is the construction of a CAB anastomosis, which is a multiple-step process. Considering that each CAB surgery requires an average of three anastomoses, roughly 1.2 million anastomoses are performed annually. In addition to its frequency, the procedure is also of short duration and can be easily simulated with both high and low fidelity simulators, making it an ideal procedure for assessment tool development.

Direct observation, which is the most commonly used method for assessment of surgical motor skills in CAB creation, suffers from poor reliability, poor compliance, and inaccuracy [4, 5]. Checklists and behaviorally anchored global assessments, originally created for

Accepted for publication July 17, 2017.

Presented at the Poster Session of the Fifty-third Annual Meeting of The Society of Thoracic Surgeons, Houston, TX, Jan 21–25, 2017. Winner of the Blue Ribbon as the top Cardiothoracic Education Poster.

Address correspondence to Dr Vaporciyan, Department of Thoracic and Cardiovascular Surgery, Division of Surgery, MD Anderson Cancer Center, University of Texas, 1515 Holcombe Blvd, Box 1489, Houston, TX 77030; email: avaporci@mdanderson.org.

The Supplemental Material can be viewed in the online version of this article [http://dx.doi.org/10.1016/j.athoracsur.2017.07.029] on http://www.annalsthoracicsurgery.org.

0003-4975/$36.00
http://dx.doi.org/10.1016/j.athoracsur.2017.07.029

objective structured assessments of technical skills [6], have been adapted for cardiothoracic surgery to improve the reliability and validity of direct observations of simulation exercises. The majority of these instruments, however, simply used the global rating component of the original objective structured assessments of technical skills tool and frequently dismissed the checklist component [7]. That was most likely the consequence of several studies showing that when compared with a checklist, the global rating scale provided better interstation reliability, better construct validity, and better concurrent validity [8, 9]. There was also no evidence that when added to the global rating component, checklists improved the reliability or validity of the global rating scale alone [10]. The limitation of all of these observational assessments tools, however, is that they lack reliability and validity at the specialist or higher trainee level [11] and exhibit a so-called "ceiling effect." That may be attributed to the design of the simulation models [12]; however, in assessment of live surgery, the lack of a model implies that any observed ceiling effect must be attributable to the assessment tool itself. Therefore, a more detailed checklist may increase its sensitivity and make it a more reliable and valid tool.

Another explanation for the limited sensitivity among existing checklists is the variability between institutions regarding the specific steps used. As highly complex procedures can be performed in a variety of ways with similar outcomes, single institution checklists may be populated with items that are institution-specific and detract from the overall sensitivity of the instrument. Unfortunately, the majority of checklists and GRS that have been developed lack broad participation [6, 13]. We hypothesized that a consensus building exercise that includes input from a randomly selected broad pool of clearly defined experts will produce a checklist that could address these needs. We intend to explore two specific elements of this hypothesis. First, is this approach feasible? And second, what is the degree of variability that exists between experts with regard to the items that eventually reach consensus?

Material and Methods

Selection of Consensus-Building Technique

After examining the existing consensus-building techniques, the Delphi method was found to be the best fit for our study owing to several unique characteristics. The entire Delphi can be done online, and therefore allows for global access to experts; the panel size requirements are modest, making the pool of experts queried manageable; and finally, the flexible design of the Delphi allows any number of follow-up interviews.

Selection of Experts

Experts were defined as North American cardiothoracic surgeons with at least 10 years' working experience after initial board certification, actively involved in performing coronary surgery, and teaching at an accredited cardiothoracic training program. A database of 314 North American experts who fit those criteria was created. Based on similar Delphi studies, 15 to 20 experts were sought for consensus building and were randomly invited to participate [14, 15]. When we had a minimum of 15 participants, the selection process was closed and no additional experts were invited to join. The final number of participants was 16 as more than one participant accepted the invitation simultaneously. Two participants were from the same institution, whereas the other 14 were from different institutions.

Delphi Method

Specific instructions were sent to selected participants asking them to provide items they believed were mandatory for competent performance of a CAB anastomosis. The instructions sent to each participant focused on four key elements. First, they were clearly instructed that any items they provided were "mandatory for the competent performance of a CAB anastomosis." These are the steps that "just have to be there otherwise it will not be a safe and well-constructed anastomosis." Second, the instructions defined what constituted a "CAB anastomosis," for example, when it began and when it was completed. Next, instructions on how to construct a checklist item were provided and included examples of well-constructed and poorly constructed items. Finally, to help organize the items and remind experts to address all aspects of a CAB anastomosis, the procedure was divided into six sections (dissection of the target vessel, creation of the arteriotomy, preparation of the graft, management of assistants, performance of the anastomosis, and testing or any additional manipulation of the graft). Experts were asked to provide as many items they believed were necessary for each section. As many as five reminders were sent to encourage submission of their initial self-created checklist.

When all the items were received, they were analyzed by the principal investigator and one uninvolved local expert, and similar items were grouped together. If any significant change in wording of the items was required, they would be sent back to the original participant to ensure that the true meaning of the item was preserved. Finally, after all revisions were approved, these condensed and edited items constituted a "master items list." This list was used for the Delphi consensus process.

In round one, all the items contained in the master items list were sent to the participants, who were asked to rank each item on a four-point scale (1 = not necessary, 2 = desirable, 3 = important, 4 = mandatory). After all the scores were obtained, a mean item score was calculated for each item. Consensus was defined separately for accepted and dropped items. We deemed that an item should be accepted (ie, "positive consensus") into the final checklist when 75% of experts ranked it as mandatory [3]. Conversely, an item was dropped (ie, "negative consensus") from the final checklist when its mean score was less than 2 ("desirable"). Items that fit neither of those criteria were advanced to round two.

In the second round, the participants were again asked to rank each item using the same four-point scale. This time, the items were accompanied with descriptive data derived from round one, including the minimum and