Excerpta Medica

The American Journal of Surgery

The American Journal of Surgery 189 (2005) 412–415
Surgical education

Surgical skill is predicted by the ability to detect errors

Simon Bann, B.Sc. (Hons), M.B., B.S., M.D., F.R.C.S.*,
Mansoor Khan, B.Sc. (Hons), M.B., B.Ch., A.F.R.C.S.,
Vivek Datta, B.Sc. (Hons), M.D., F.R.C.S.,
Ara Darzi, M.D., F.R.C.S., F.R.C.S.I., F.A.C.S., K.B.E.

Department of Surgical Oncology and Technology, Imperial College of Science, Technology and Medicine, 10th Floor QEQM Wing, St. Mary's Hospital, London W2 1NY, UK

Manuscript received February 4, 2004; revised manuscript July 16, 2004

Abstract

Background: Objective analysis methods of surgical performance are now available so comparison between surgeons is available. One such method is by direct observation using the Objective Structured Assessment of Technical Skills (OSATS), but this is a time-consuming process; therefore, a simple screening tool for the ability to detect errors (previously validated) was analyzed and considered as a predictor of qualitative performance.

Methods: Thirty-eight volunteer surgeons were recruited to the skills laboratory to undertake 3 exercises. Two were bench-top surgical tasks that were scored using the global rating of the OSATS technique. The third task was the ability to detect simple errors in 22 synthetic models of common surgical procedures, some of which contained purposefully made errors. $P < .05$ was deemed to be statistically significant.

Results: The scores (interquartile ranges in parentheses) for the 3 sections were excision of sebaceous cyst = 21 (19,24), closure of small bowel enterotomy = 23 (21,27), and identification of errors = 31 (27,34). Three scorers blinded to the operative models exhibited an interobserver reliability of .9 and .91 for the video tasks, respectively. Spearman's rank correlations between the error examination and performance on the 2 tasks were both statistically significant at .69 (cystectomy) and .54 (enterotomy).

Conclusions: The ability to detect simple surgical errors is a predictor of technical skill and performance of bench tasks. What must be answered is whether the use of such models and principles can shorten the qualitative surgical learning curve.

Keywords: Objective assessment; Technical skills; Error

The objective assessment of surgical performance is highly desirable. Methods with which to achieve these aims are becoming recognized and validated both qualitatively and quantitatively. Quantitative methods include motion analysis [1–3] and virtual reality-based simulators [4–8]. Qualitative methods include observational methods such as the Objective Structured Assessment of Technical Skills (OSATS) [9–11], developed in Toronto, and—to some extent—virtual reality-based simulations.

Common sense suggests that technical skill affects outcome and it is well recognized that variability exists regarding outcomes between different surgeons [12–14]. However, the link between outcome and skill has not yet been shown. This will require acceptable forms of standardization that permit the meaningful comparison of patients. However, with standardized simulations and tasks, this should be possible in the skills laboratory.

Technical skill consists of dexterity and judgement based on knowledge [15]; this technical knowledge will increase with experience. Therefore, methods for assessing technical knowledge will be strongly related to operative performance and outcome. Previous attempts at correlating academic achievement and surgical skill, however, have been shown either not to correlate or to correlate negatively [16,17].

A novel method for assessing knowledge is to measure the ability to detect errors. Certain errors are knowledge

* Corresponding author. Tel.: +1-44-20-886-6119; fax: +1-44-207-886-1810.
E-mail address: s.bann@ic.ac.uk

0002-9610/05/$ – see front matter
doi:10.1016/j.amjsurg.2004.07.040

based. Reason defines an error as "the failure of planned actions to achieve their desired goal" [18]. For surgery, this may start at a very basic level, even down to the simple tasks of knot tying and suturing, where even a simple error may have profound consequences. The aim of this study was to examine the relationship between performance on a novel error analysis, which had previously been validated [19], and actual surgical performance on 2 standardized bench-top tasks.

Methods

Subjects

Thirty-eight volunteer surgeons were recruited from the department and from other local hospitals. All had completed the Basic Surgical Skills Course of the royal College of Surgeons of England as a prerequisite; they were either Basic Surgical trainees (postgraduate years 2 through 4) or Higher Surgical trainees (postgraduate years 3 through 9) (see reference [20] for an explanation of British training).

The surgical error examination

This examination consisted of high-quality synthetic models that were constructed to demonstrate either what is considered good surgical practice or a purposefully made technical error. These models have previously been shown to exhibit face and construct validity [19]. This section was divided into 7 models of simple interrupted skin closure, 4 models of continuous subcutaneous skin closure, 4 models of elliptical excision of a skin lesion, 3 models of small-bowel enterotomy closures (using the principles taught on the Basic Surgical Skills Course), 2 models of smallbowel anastomoses (using the principles taught on the Basic Surgical Skills Course), and 2 models of arteriotomy closures.

The models were created to test the surgical knowledge of more junior trainees. Because of the limitations of the synthetic models and possibility of wear and tear, they were kept relatively simple. They errors constructed were deemed to be obvious to careful inspection.

Questionnaire

A questionnaire protocol, with each model listed, was devised. The trainees were asked to state whether they saw an error. If an error was present, the error was to be named or described for the second mark. A mark was given for each correct response (there was no negative marking). Twenty-two questions, based around the 6 areas, yielded a maximum score of 38. Marking was carried out in a blind with the subjects' experience only being revealed on entry to the database.

The surgical exercises

The 2 models used were excision of a sebaceous cyst [21,22] and closure of a small-bowel enterotomy [21,22]. These models were chosen to reflect the experience of more junior trainees. The model of the sebaceous cyst (Limbs and Things, Bristol, UK) consisted of a multilayered foam pad surrounding a polymer capsule containing a yellow oily liquid such that it had the potential to burst if incised. The subjects were asked to excise the sebaceous cyst and to close the incision with interrupted instrument-tied sutures. The sebaceous cyst model required planning, preparation, incision, dissection and tissue handling skills, closure or wound approximation, and appreciation of cosmesis and was therefore considered a more technical task. The enterotomy closure was based on a 2-cm incision on a section of synthetic small bowel (Limbs and Things). The subjects were asked to close using techniques taught by the Royal College of Surgeons of England as taught in their Basic Surgical Skills Course. This involved the insertion of 2 stay sutures and the use of interrupted, hand-tied serosubmucosal sutures placed 5 mm apart and >3 mm from the edge of the bowel. All procedures were recorded and digitized using a Sony DVCAM digital videocassette recorder (Sony Corporation, Tokyo, Japan). These were then scored independently by 3 experienced scorers who were blinded to the subjects' identities.

The OSATS scoring methods

The OSATS scoring methods were developed in Toronto [9–11] and consist of a task-specific checklist and global rating scale. The global rating scale consists of an 8-category, 5-point Likert scale anchored by behavioral descriptors. The 8 categories are respect for tissue, time and motion, instrument handling, suture handling, flow of the operation, knowledge of the procedure, overall performance, and quality of the final product.

Statistical analysis

Statistical analysis was undertaken using the Statistical Package for Social Science (SPSS, Chicago, Illinois). Correlations were calculated using Spearman's correlation coefficients. The interobserver reliability was calculated using Cronbach's alpha. A probability value $<.05$ was deemed significant.

Results

Thirty-eight trainees were recruited to the study (postgraduate years 2 through 9). Three scorers independently reviewed the digitized performance of the enterotomy closure and sebaceous cyst excision. The median scores (interquartile range in brackets) for the 3 stations were 31