Surgeons outperform normative controls on neuropsychologic tests, but age-related decay of skills persists

Zackary Boom-Saad, Ph.D.\textsuperscript{a}, Scott A. Langenecker, Ph.D.\textsuperscript{b}, Linas A. Bieliauskas, Ph.D.\textsuperscript{b}, Christopher J. Graver, Ph.D.\textsuperscript{c}, Jillian R. O’Neill, Ph.D.\textsuperscript{d}, Angela F. Caveney, Ph.D.\textsuperscript{b}, Lazar J. Greenfield, M.D.\textsuperscript{a}, Rebecca M. Minter, M.D.\textsuperscript{e}\textsuperscript{*}

\textsuperscript{a}Department of Surgery, University of Michigan, 1500 East Medical Center Dr., Taubman Center TC2920H, Ann Arbor, MI 48109-0331, USA
\textsuperscript{b}Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA
\textsuperscript{c}Department of Neuropsychology, Western State Hospital, Lakewood, WA, USA
\textsuperscript{d}Department of Neuropsychology, University of Illinois, Chicago, IL, USA
\textsuperscript{e}Department of Medical Education, University of Michigan, Ann Arbor, MI, USA

Manuscript received April 25, 2007; revised manuscript November 7, 2007

Presented at the Association for Surgical Education Annual Meeting, Washington, DC, April 12, 2007

Abstract

Background: The present study was undertaken to determine if psychomotor and visual-spatial abilities improve as a result of surgical training or are enhanced at baseline in those individuals choosing a surgical career.

Methods: Medical students entering a surgical field and practicing surgeons performed a series of neuropsychologic tests. Performance was compared between surgeon groups, as well as with normative aged-matched controls.

Results: An age-related decline was noted in the performance of all exercises, with the medical student group outperforming the midcareer surgeons, who in turn outperformed the senior surgeons. Interestingly, however, all 3 groups significantly outperformed their normative control groups on some or all tasks.

Conclusions: Improved visual memory and psychomotor performance compared with normative controls appears to be present at baseline rather than resulting from surgical training. Decline in performance with age is observed, however, and this should be considered when an older surgeon is learning new visually complex procedures. © 2008 Excerpta Medica Inc. All rights reserved.

Keywords: Cambridge Neuropsychological Test Automated Battery (CANTAB); Medical students; Surgeons; Aging; Attention; Psychomotor performance

Inquiries into the nature of surgical ability, especially with regard to its implications for surgical training, abound in the medical literature [1–3]. However, the relationship between physician age and certain components of surgical and procedural learning ability such as manual dexterity or visual spatial memory is a relatively new line of inquiry. Previous studies involving the general population have shown that cognition, manual dexterity, sustained attention, and visual-spatial ability decline with age [4–8]. However, when physicians are compared with the general population with respect to changes in cognitive and psychomotor ability associated with aging, the data are conflicting. Prior research using a neuropsychologic battery of tests, known as MicroCog (Harcourt Assessment, Inc., San Antonio, TX), compared various cognitive factors such as intelligence quotient, attention span, and visual-spatial recognition between physicians and nonphysician controls, and showed a significant decline in cognitive ability among both groups associated with aging [9]. However, a longitudinal study of aging surgeons that was conducted by Greenfield [10] failed to show any significant decrement in performance on MicroCog tests. This prompted a more extensive study of psychomotor and visual-spatial skill retention among aging surgeons using a potentially more sensitive battery of tests known as the Cambridge Neuropsychological Test Auto-
Practicing surgeons aged 45 to 75 years old (n = 308) performed a series of neuropsychologic tests (CANTAB) at the American College of Surgeons Clinical Congress, 2001–2004. To evaluate trainees who will be pursuing surgical careers but who have not yet begun their training, fourth-year medical students at the University of Michigan who had matched into General or Plastic Surgery or Urology residencies in 2005 and 2006 (n = 21) performed the same tests as the practicing surgeons. For the purposes of comparison, the practicing surgeons were divided into 2 groups: age 45 to 60 (n = 139) and age 61 to 75 (n = 169). As would be expected, the percentage of women in each group increased significantly as the age groups declined, ranging from 1.8% in the 61- to 75-year-old surgeon group, to 18% in the 45- to 60-year-old group, to 28% in the medical student group. All aspects of the study were approved by the University of Michigan Institutional Review Board and informed consent was obtained from all subjects.

Equipment and methodology

The series of neuropsychologic tests performed were selected from a battery of CANTAB (the parallel version), which is a computerized neuropsychologic assessment tool that has been used extensively to evaluate a variety of cognitive and psychomotor domains including visual spatial memory, sustained attention, reaction time, and efficiency of motion [12]. The CANTAB battery was chosen for the assessments performed as well as the simplicity of the computer interface, thus minimizing the effect of prior computer experience, which could be a significant potential confounder in this cross-sectional study. The CANTAB interface requires a subject to simply use a touch screen and touchpad sitting approximately 8 inches in front of the screen for completion of all exercises (Fig. 1), and a training phase is completed by all subjects before testing. The parallel version of the CANTAB tests were used in the initial design of the practicing surgeon study because this was a longitudinal evaluation of cognitive/skill changes in practicing surgeons with age (results of which will be reported elsewhere). The parallel version of the test allows for equivalent forms of the test to be administered over time to assess for interval change in cognitive functioning. Because the parallel version was used initially to evaluate the practicing surgeons, this same battery of tests was used to evaluate the medical student group. Although there is some variability in tasks and metrics between the clinical and parallel versions of CANTAB, the clinical version provides a large normative control database that consists of more than 2,000 patients who can be used for normative comparison for several of the CANTAB tasks on the parallel version as well.

The specific tests chosen from the array of tests available on CANTAB were the Reaction Time test (RTI), the Rapid Visual Information Processing test (RVIP), and the Visual Paired Associates Learning test (VPAL). These 3 CANTAB tasks (RTI, RVIP, and VPAL) were chosen specifically by the investigators because it was believed that efficiency of motion, reaction time, sustained visual attention, and visual-spatial memory were critical for optimal psychomotor performance by surgeons.

For the RTI task, a white circle appears in the middle of the screen and at random times a yellow spot appears within the confines of this circle. Each subject sits in front of the screen with his/her hand on the pedal. When the dot appeared, the subjects released the pedal and touched the screen. The metrics on the task include release time, which is the time that elapses between the appearance of the visual stimulus and release of the pad, and time to touch the target, which measures the time between pad release and touching the screen. The latter metric can be thought of as efficiency of motion. The normative data provided from the clinical version of the test were used for comparison of performance on this task. This may result in slight inaccuracies in measures of variance when making surgeon-normative comparisons on the parallel version of the test, but was believed to be appropriate because the test presentation and response parameters were the same and the instructions were identical for both test versions (clinical and parallel).

In the RVIP task, a white box appears in the center of the computer screen, inside which digits from 2 to 9 appear in a pseudorandom order at a rate of 100 digits per minute.
Subjects are required to look for a specific number and press a button every time that number appears. This continues for a set number of numeric sequences. Response time, which is the time that elapses between seeing the right number and pressing the button, and overall accuracy rate are measured. This task measures sustained visual attention. There is no reported deviation between clinical and parallel versions of CANTAB for this test, and as such normative data for the clinical version were used for comparison.

For the final task, VPAL, the subject is required to remember geometric patterns associated with different locations on the screen with increasing difficulty. During the test phase, as each pattern is presented, the subject must point to the appropriate location on the screen that is associated with that pattern. For example, a box appears in the right upper corner of the screen and then disappears. The subject then must point to the location of the right upper corner of the screen to the location where the box appeared. Another box then is shown in another location and the subject identifies the correct location of this box when it disappears. After 3 correct sets with a single pattern (1 box), the patterns become increasingly difficult, requiring the subject to identify a sequential pattern of 2 boxes in different locations for 2 sets, followed by 3 boxes for 2 sets, and finally 6 and 8 box sequences for 1 set each. This task measures episodic visual-spatial memory and is an index of rapid learning. Unfortunately, normative data were not available for this test because the number of stimuli and difficulty between clinical and parallel versions was substantial, precluding using normative data from the clinical version. As such, data from the VPAL was used only to test the hypothesis of learned excellence in visual-spatial functioning in surgeons.

**Statistical analysis**

The medical student group (age, 20–35 y), midcareer practicing surgeons (age, 45–60 y), and the senior practicing surgeon group (age, 61–75 y) performance was compared using analysis of variance with post hoc analysis performed using least-squares difference. For comparison of the surgeon group, performance with age-matched normative controls analysis of variance was also used for comparison. CANTAB provides normative control data by age groups as well as intelligence quotient (IQ). In the present study, however, formal IQ testing of all subjects was not feasible, and it was assumed that any variability in IQ that existed should be similar between the 3 groups. We therefore chose the highest functioning normative group for comparison (IQ > 120) because this would be at least an equivalent, if not conservative, comparison group. All data are expressed as the mean ± standard error of margin, and statistical significance was set at a P value of less than .05.

**Results**

Performance of the 3 surgeon groups on the CANTAB tests evaluated, RTI, RVIP, and VPAL, showed age-related decay of skills across all domains assessed. Specifically, the medical student group (age, 20–35 y) outperformed the midcareer practicing surgeons (age, 45–60 y), who in turn outperformed the senior surgeons (age, 61–75 y) on the reaction time, rapid visual information processing, and the visual paired associates learning tasks (Table 1). This difference was particularly notable on the visual paired associates learning task—a measure of visual-spatial memory—in which there was a marked incremental decline in visual memory accuracy as age increased (Table 1: VPAL, total adjusted errors).

Interestingly, however, when surgeon group performance was compared with age-matched normative control performance, the surgeon groups performed significantly better than the nonsurgeon normative controls of the same age. This was most marked on the reaction time task efficiency of motion metric (total movement time), in which all 3 surgeon groups showed significantly improved efficiency of motion as compared with normative controls (Fig. 2). A similar trend was observed on the total reaction time metric on the RTI task, although the most senior surgeons’ group performance did not achieve statistical significance as compared with same-aged controls (P = .08) (Fig. 3). Lastly, on the RVIP task, the medical student (20- to 35-year-olds) and 45- to 60-year-old practicing surgeon groups also showed improved accuracy as compared with same-aged normative controls when asked to correctly identify a target number from a presented numeric sequence. The senior surgeon group (61–75 y) showed similar performance to controls on this test of sustained attention and working memory function (Fig. 4). Unfortunately, the normative reference values for the VPAL task were not available, allowing us to compare only performance between the surgeon groups for the visual-spatial memory metrics.

**Comments**

The present studies show that an incremental decline in reaction time, rapid visual information processing, and visual-spatial learning capabilities occurs in surgeons with
advancing age just as in the general population. However, despite this incremental decline, surgeons generally outperform their age-matched normative peer group, particularly for the younger and midcareer practicing surgeon groups. Contrary to our hypothesis, this improved performance as compared with normative controls appears to be present at baseline before initiation of surgical training rather than resulting from training. This is reflected by the significantly improved performance that our medical student group showed as compared with both the midcareer and senior practicing surgeon groups, as well as with their age-matched controls.

The implications of these results must be considered cautiously because we currently do not know if decline in visual-spatial memory and psychomotor performance as measured by neuropsychologic tests can function suitably as proxy measures for procedural learning capacity. Indeed, we are embarking on investigations with surgical simulators and neuropsychologic functioning to address this question. However, the present findings add to a body of literature that should not be ignored addressing the potential impact of advancing age on physician and surgeon performance. In the realm of primary care, prior research has shown that older physicians are less likely to prescribe appropriate medications, achieve lower scores on recertification examinations, and possess a less current fund of knowledge when compared with younger colleagues [13–15]. Within the surgical field, 2 studies found that mortality rates after coronary artery bypass grafting and carotid endarterectomy were associated negatively with advanced surgeon age as compared with their younger colleagues [16,17]. In contrast, a more recent study by Waljee et al [18] suggested that advanced age per se does not contribute to increased surgical mortality rates for most complex surgical procedures, but rather low surgical volume appears to be the primary predictive variable for outcome when age and volume are controlled. Similarly, looking at the other end of the experiential spectrum, Prystowsky [19] found that young surgeons within 5 years of completing training had increased morbidity and mortality rates after complex alimentary tract surgery as compared with more experienced surgeons, presumably as a result of inadequate technical and clinical experience. The conflicting nature of the findings from these retrospective studies, which are derived largely from administrative databases, shows the need for longitudinal studies on this important issue, so that appropriate strategies can be developed and used to address the potential impacts and influences of both aging and experience.

Within the context of surgical education, the findings from this study suggest that CANTAB may be a useful construct for identifying or evaluating surgical trainees who have relative deficiencies with respect to psychomotor and visual-spatial memory, which may impair their procedural learning ability. If this relationship can be established, then
early identification of these individuals and characterization of their particular deficiencies could allow for the development of individualized curricula specific to their needs, which should lead to more efficient training. This becomes particularly relevant in the era of restricted resident duty hours as well as increased demands for efficiency in the operating room. Within the realm of continuing medical education, interpretation of the present data suggest that perhaps acquisition of brand new, visually complex procedures such as learning to perform a complex laparoscopic or endoscopic procedure late in one’s career may be more challenging, perhaps requiring more training to achieve proficiency as compared with one’s younger colleagues. Further studies clearly are needed in this area, and the impact of clinical experience on the learning of new visually complex procedures must be evaluated as well.

Interpretation of the present results requires recognition of this study’s limitations. The most notable of these is the inability to assess the impact of surgical experience on procedural learning ability within this study. Obviously the CANTAB tests measure only one subset of skills relevant to procedural learning, and older surgeons may be able to compensate for the decline in raw psychomotor and visual-spatial memory skills by drawing on experience with similar cases. Certainly, other studies have reached similar conclusions when other visual-spatial tests have been used to address this question [1,3]. However, despite this important consideration, recognition of the fact that decline in physical and cognitive ability occurs with advancing age is paramount, to ensure that compensatory mechanisms are in place to prevent adverse outcomes. With respect to interpretation of the normative comparisons, it should be recognized that the use of the parallel version of the CANTAB limits the strength of inferences that might be drawn when comparing surgeons with a normative group collected on a highly similar, yet slightly different, version of the CANTAB tests. This possibility was controlled for by using a conservative comparison normative sample group, those with an IQ greater than 120. This normative control group provided a reasonable and conservative comparison group for the aspiring and practicing surgeons’ performance. An additional limitation of this study was the limited sample size and single institutional nature of the medical student group. Although statistically significant differences in performance were observed even with this small group relative to age-matched normative controls and the older practicing surgeons, expansion of this group to include subjects from multiple institutions would strengthen this study significantly. This also potentially would allow us to perform gender subset analyses within the medical student group, which could provide important and potentially interesting data regarding gender differences with respect to psychomotor and visual-spatial memory ability.

Despite these limitations, the present study adds to an important and growing body of literature focused on the impact of both aging and baseline skill level on procedural learning capability. Further longitudinal investigation along these lines potentially will help us to identify and correct specific deficits within individuals, which may impair their rate of proficiency acquisition or, more importantly, decrease patient morbidity and mortality after surgical procedures.

Acknowledgment
Supported by a research grant from United States Surgical (Norfolk, CT); the Association for Surgical Education Foundation’s Center for Excellence in Surgical Education, Research, and Training (Springfield, IL); and the Departments of Surgery and Psychiatry (Neuropsychology Section) at the University of Michigan.

References
[17] Hartz AJ, Kuhn EM, Pulido J. Prestige of training programs and single institutional nature of the medical student group. Although statistically significant differences in performance were observed even with this small group relative to age-matched normative controls and the older practicing surgeons, expansion of this group to include subjects from multiple institutions would strengthen this study significantly. This also potentially would allow us to perform gender subset analyses within the medical student group, which could provide important and potentially interesting data regarding gender differences with respect to psychomotor and visual-spatial memory ability.

Despite these limitations, the present study adds to an important and growing body of literature focused on the