Why Serial Assessments of Cardiac Surgery Patients' Neurobehavioral Performances are Misleading

Julian R. Keith, PhD, Dale J. Cohen, PhD, and Len B. Lecci, PhD

Memory Assessment and Research Services, and Department of Psychology, University of North Carolina Wilmington, Wilmington, North Carolina

In 1995, a statement of consensus was published in *The Annals of Thoracic Surgery* recommending that researchers who assess neurobehavioral outcomes after cardiac surgery adopt a standard set of core procedures [1]. The purpose of the present editorial is to evaluate statement 3 of the *Statement of Consensus* that reads: "The individual change in performance from baseline to a time after operation is essential to any evaluation of the impact of the operation or any intervention associated with it" [1].

We vigorously endorse standardized and rigorous neuropsychological methods and commend the efforts of the researchers and clinicians who contributed to the *Statement of Consensus*. Nevertheless the soundness of each statement of the *Statement of Consensus* rests on the validity of the assumptions from which they are derived. Presently our focus is on the validity of the assumptions underlying statement 3.

See page 374

Blumenthal alluded to the assumptions underlying statement 3 in the published commentary that accompanied the *Statement of Consensus* when he noted, "part of the rationale for statement 3 is to address the issue that these patients may be abnormal preoperatively. Therefore it is important to examine change in scores from baseline" [1]. Thus statement 3 assumes that (1) patients' neurobehavioral performances may be abnormal preoperatively, (2) preoperative neurobehavioral performances are valid measures of neurobehavioral competence, and (3) one can use differences between measures obtained preoperatively versus postoperatively to isolate an effect of surgery or treatment, or both, on neurobehavioral fitness. The validity of assumption 1 that patients' neurobehavioral performances may be abnormal preoperatively is convincingly demonstrated in the published literature [2]. However a considerable body of evidence runs counter to the second and third assumptions. As follows we examine assumptions 2 and 3 and consider the impact of the violations of these assumptions on analyses of the neurobehavioral outcomes of cardiac surgery.

Address correspondence to Dr Keith, Department of Psychology, University of North Carolina, Wilmington, 601 South College Rd, Wilmington, NC 28403-5612; e-mail: keithj@uncw.edu.

Our evaluation of statement 3 can be delineated as a statistical model of a standard, within-subjects, preoperative and postoperative surgical research design typically used in research on neurobehavioral outcomes in cardiac surgery patients. For simplicity we restrict all nonsubject factors in the models to dichotomous levels (ie, either present or absent), we only include the terms in the model that are present, and therefore we omit subscripts.

$$y_{i2} = y_{i1} + p + e \quad (1)$$

Equation 1 models performances of noncardiac surgery control subjects for the second assessment (ie, corresponding to the postoperative assessment time point for cardiac surgery patients). In Equation 1, y_{i2} is subject i's performance on the second assessment, y_{i1} is subject i's performance on the first (baseline) assessment, p is the effect of practice (change associated with repeated administration of neurobehavioral assessment instruments and all other factors associated with systematic changes between baseline and follow-up assessments), and e is variance associated with random error. This is the simplest, plausible model, whereby control subjects' performances on the second assessment are the sums of their performances on the first assessment and any effect of practice.

$$y_{i2} = y_{i1} + p + s + p * s + e \quad (2)$$

Equation 2 is the predicted model for the second assessment of patients who undergo cardiac surgery. Equation 2 includes several factors that do not affect the noncardiac surgery controls. In Equation 2, s is the effect of cardiac surgery on neurobehavioral competence (presumably negative, but not necessarily so), and $p * s$ is any effect associated with the interaction between practice and cardiac surgery. The important point to notice here is that the (presumably negative) effect of the cardiac surgery can not be isolated simply by subtracting out the initial baseline (y_{i1}). Indeed subtracting initial baseline scores from postoperative scores overlooks the combined effects of surgery and practice. Thus if one were to use a preoperative and postoperative design to isolate the effects of surgery, one would have to assume that (1) there is no effect of practice, and (2) there is no statistical interaction between practice and cardiac surgery. However a substantial body of data demonstrates that both of these assumptions are false.

Published by Elsevier Inc

Ann Thorac Surg 2007;83:370–3 • 0003-4975/07/$32.00
doi:10.1016/j.athoracsur.2006.06.062

The neuropsychological tests generally used in studies of neurobehavioral outcomes in cardiac surgery patients are notoriously susceptible to practice effects [3, 4]. For example, a careful examination of practice effects by Raymond and colleagues [5] recently revealed large and statistically significant practice effects in every cognitive domain examined in patients tested prior to surgery and immediately before discharge from the hospital, with more than 65% of patients improving by a standard deviation or more (see Raymond and colleagues [5], Fig 1, Table 2). This documentation of practice effects in coronary artery bypass graft patients is consistent with the well-established observation of practice effects in other patient samples [6]. The presence of a practice effect precludes the possibility of isolating the effects of cardiac surgery in a single group, within-subjects, preoperative and postoperative research design in which all patients are subjected to surgery. Some researchers attempt to address this problem by estimating the effects of practice and thereby removing practice effects statistically from the model that predicts the second assessment of the cardiac surgery patients by including a noncardiac surgery control group [2, 7, 8]. However it is essential to note that statistical methods for removing the practice effect using the information gleaned from a noncardiac surgery control group are valid *only if* the practice effect does not interact with the cardiac surgery effect. However the evidence that we review as follows suggests that there is a clear interaction between practice effects and cardiac surgery. Thus conclusions based on analyses that assume the practice times cardiac surgery interaction is zero are necessarily invalid.

A practice times cardiac surgery interaction is observed when the magnitude of the practice effect is controlled by factors that are not equivalent across experimental groups (eg, cardiac surgery vs nonsurgery controls). In cardiac surgery studies that include nonsurgery control groups, there are many factors that are not equivalent across the groups, two of which will be highlighted here. First, cardiac surgery candidates' preoperative anxiety and depression levels are high (ie, clinically significant) during the days leading up to surgery and drop to moderate levels within the first postoperative week [7, 9, 10]. Elevated anxiety and comorbid depression, particularly when in the clinical range, significantly interfere with neuropsychological performance* [11, 12]. Second, in some studies, differences between the experimental groups have arisen from the fact that cardiac surgery patients were tested preoperatively in settings (eg, hospital rooms) that are not conducive to optimal neuropsychological performance. Indeed, Selnes and colleagues [13] reported that coronary artery bypass graft patients who were tested in the hospital environment (ie, approximately one-third of their cardiac surgery study sample) performed significantly poorer preoperatively on neuropsychological tests than those participants who were tested as outpatients. Therefore preoperative assessments may underestimate patients' true neurobehavioral competences, and postoperative practice effects observed in cardiac surgery patients may improve due to decreases in anxiety, depression, and shifts in the setting in which postoperative testing takes place (eg, the nonhospital setting). Alternatively, elevated anxiety, depression, and distraction during preoperative baseline testing may result in poorer learning of the tasks, which will translate into proportionately smaller practice effects in cardiac surgery patients than in nonsurgery control patients. The main point is that although a practice times surgery interaction is likely, one cannot predict a priori the magnitude or even the direction of the effect.

It is impossible to assess the presence of an interaction between practice effects and surgery effects from existing data without making further assumptions. If one assumes that (1) cardiac surgery causes neuropsychological decline beyond that which was present preoperatively in surgery candidates, and (2) the practice times surgery interaction is zero, then the differences between a group of patients who underwent surgery and a nonsurgery control group should be even greater postoperatively [5, 7, 14, 15]. However, recently published studies [2, 5, 15] that presented detailed neuropsychological assessment data from cardiac surgery patients observed just the opposite effect; between-groups postoperative differences were smaller postoperatively than they were preoperatively (ie, cardiac surgery patients improved more during postoperative testing than predicted by the nonsurgery control group's practice effects). Furthermore, when either of two widely-used criteria for defining significant change from baseline in postoperative test scores [9] (ie, one that requires a standard deviation change, and another that requires a 20% change, on two or more of the nine tests that comprised the battery) are applied to individual patient data, approximately 70% of cardiac surgery patients tested at discharge (ie, 3 to 10 days after surgery) actually significantly improved their performances. If one assumes the absence of a practice times surgery interaction, then one has to accept that cardiac surgery improves cognitive performance both in the short (ie, 3 to 10 days after surgery) [9] and long-term (ie, 3 years after surgery) [2, 15]. Unfortunately research based on within-subjects designs for the past 35 years has not led to agreement among researchers on the issue of whether cardiac surgery improves or diminishes neurobehavioral fitness. Cardiac surgery exposes the brain to many conditions that are known to be harmful to the brain (eg, emboli, inflammation, hypoperfusion, and so forth). Therefore we interpret the disproportionate postoperative neurobehavioral improvements observed in cardiac surgery patients as evidence of a practice times surgery interaction. Moreover, recently published data indicates that the magnitude of the practice effect and practice times surgery interaction is at least one-half

*One recent study did fail to detect a statistically significant relationship between anxiety scores and preoperative neuropsychological performances in cardiac surgery patients (Tsuhima and colleagues, 2005). However, this study was limited by a small sample size and by the fact that the study did not examine preoperative and postoperative performances. Thus, this study was not sufficiently powered nor was it properly designed to evaluate a possible relationship between anxiety and practice effects.

standard deviation [2]. As such, this effect is quite large relative to the magnitude of the neuropsychological changes caused by cardiac surgery and, when ignored, could significantly undermine a researcher's ability to detect the effect of cardiac surgery.

In summary, the available evidence suggests both the presence of a practice effect and the presence of a practice times surgery interaction. The presence of the interaction renders it mathematically impossible to isolate the effect of cardiac surgery on cognitive performance in a within-subjects preoperative and postoperative research design, even if it includes both control and experimental groups. Thus by extension, statement 3 of the *Statements of Consensus* is invalid.

Despite the demise of statement 3, there are valid statistical and experimental methods to evaluate "the impact of the operation or any intervention associated with it" [1]. However, these methods evaluate the impact of the operation and intervention (eg, neuroprotective strategies) at the group level rather than on an individual subject level. Admittedly it is rare to come across circumstances in which a pure between-groups research design surpasses a within-subjects serial assessment design in terms of sensitivity to the effects of an intervention. However, for the reasons previously discussed, the study of cardiac surgery on neurobehavioral fitness seems to be an example of such a circumstance. Cross-sectional studies comparing cardiac surgery patients postoperatively with well-matched nonsurgery patients would permit one to estimate the effect of surgery. That is, eliminating the preoperative assessment will eliminate the contaminating effects of practice and the practice times surgery interaction. If there is an effect of cardiac surgery, the two groups' performances should differ. Between-groups designs would be particularly advantageous in studies aimed at assessing pharmacological or procedural (eg, temperature) neuroprotective strategies in cardiac surgery patients [16, 17]. Here, one simply, randomly assigns cardiac surgery patients to intervention (ie, hypothermia or drug) and nonintervention groups (ie, normothermic or placebo), and measures postoperative cognitive functioning. Any differences between groups will necessarily indicate the effect of the independent variable. There are some disadvantages associated with between-groups designs relative to within-subjects designs, such as reduced statistical power. Nevertheless, because within-subjects designs are not viable for the study of cardiac surgery effects on neurobehavioral fitness, researchers should explore the solutions to the between-groups design issues.

One goal of neuropsychological research in cardiac surgery patients has been to quantify the incidence of clinically significant cognitive decline that can be attributed to the surgery [5]. Therefore, one may argue that between-group designs can not tell us which individuals were negatively impacted by the surgical procedures. However, by using the large body normative data available for neuropsychological tests, postoperative neuropsychological assessments can enable researchers and clinicians to identify which and how many cardiac surgery patients satisfy standard criteria for being classified as having clinically significant neuropsychological impairments. Between-group comparisons of cardiac surgery patients with well-matched nonsurgery controls, such as the controls recently reported by Selnes and his colleagues [2, 8], should permit useful estimates of the incidence of neuropsychological impairment that are associated with cardiac surgery. Between-group comparisons are especially appealing because they provide a valid solution to the problem of how to classify neuropsychological change after cardiac surgery (a problem for which there has been no agreeable solution when using serial assessment methods, despite the 35 years of research on this topic [7, 14, 18, 19]).

In this commentary we have demonstrated that the assumptions on which statement 3 of the *Statement of Consensus* rests are invalid. In light of this, we advocate the use of between-groups designs when assessing the effects of cardiac surgery on cognitive performance and any intervention designed to mitigate the (presumably negative) effects of surgery. We hope that our discussion of some of the problems that are inherent in the serial assessment of cardiac surgery patients helps to promote a shift in thinking among researchers regarding the relative merits of between-groups versus within-groups research designs for assessing the neurobehavioral outcomes of cardiac surgery and interventions that may improve cardiac surgery outcomes.

References

1. Murkin JM, Newman SP, Stump DA, Blumenthal JA. Statement of consensus on assessment of neurobehavioral outcomes after cardiac surgery. Ann Thorac Surg 1995;59:1289–95.
2. Selnes OA, Grega MA, Borowicz LM Jr, et al. Cognitive outcomes three years after coronary artery bypass surgery: a comparison of on-pump coronary artery bypass graft surgery and nonsurgical controls. Ann Thorac Surg 2005;79:1201–9.
3. Benedict RH. Cognitive function after open-heart surgery: are postoperative neuropsychological deficits caused by cardiopulmonary bypass? Neuropsychol Rev 1994;4:223–55.
4. Arrowsmith JE, Grocott HP, Reves JG, Newman MF. Central nervous system complications of cardiac surgery. Br J Anaesth 2000;84:378–93.
5. Raymond PD, Hinton-Bayre AD, Radel M, et al. Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery. Eur J Cardiothorac Surg 2006;29:82–8.
6. Beglinger LJ, Gaydos B, Tangphao-Daniels O, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol 2005;20:517–29.
7. Keith JR, Puente AE, Malcolmson KL, et al. Assessing postoperative cognitive change after cardiopulmonary bypass surgery. Neuropsychology 2002;16:411–21.
8. McKhann GM, Grega MA, Borowicz LM Jr, et al. Is there cognitive decline 1 year after CABG? Comparison with surgical and nonsurgical controls. Neurology 2005;65:991–9.
9. Duits AA, Duivenvoorden HJ, Boeke S, et al. The course of anxiety and depression in patients undergoing coronary artery bypass graft surgery. J Psychosom Res 1998;45:127–38.
10. Rymaszewska J, Kiejna A, Hadrys T. Depression and anxiety in coronary artery bypass grafting patients. Eur Psychiatry 2003;18:155–60.

11. Bossema ER, Brand N, Moll FL, et al. Does carotid endarterectomy improve cognitive functioning? J Vasc Surg 2005;41:775–81; discussion 781.
12. Kizilbash AH, Vanderploeg RD, Curtiss G. The effects of depression and anxiety on memory performance. Arch Clin Neuropsychol 2002;17:57–67.
13. Selnes OA, Grega MA, Borowicz LM Jr, et al. Cognitive changes with coronary artery disease: a prospective study of coronary artery bypass graft patients and nonsurgical controls. Ann Thorac Surg 2003;75:1377–84; discussion 1384–6.
14. Keith JR, Puente AE. Deficiencies in the detection of cognitive deficits. Neuropsychology 2002;16:434–9.
15. Barry SJ, Zeger SL, Selnes OA, et al. Quantitative methods for tracking cognitive change 3 years after coronary artery bypass surgery. Ann Thorac Surg 2005;79:1104–9.
16. Arrowsmith JE, Harrison MJ, Newman SP, et al. Neuroprotection of the brain during cardiopulmonary bypass: a randomized trial of remacemide during coronary artery bypass in 171 patients. Stroke 1998;29:2357–62.
17. Regragui I, Birdi I, Izzat MB, et al. The effects of cardiopulmonary bypass temperature on neuropsychologic outcome after coronary artery operations: a prospective randomized trial. J Thorac Cardiovasc Surg 1996;112:1036–45.
18. Keith JR. The reliability of neuropsychologic changes after cardiac surgery and a deeper problem posed by practice effects. J Cardiothorac Vasc Anesth 2001;15:666–70.
19. Collie A, Darby DG, Falleti MG, et al. Determining the extent of cognitive change after coronary surgery: a review of statistical procedures. Ann Thorac Surg 2002;73:2005–11.