Postoperative Cognitive Decline: The Unsubstantiated Phenotype

To the Editor:
Correlation of a clinical outcome (phenotype) with a biomarker or genotype requires accurate phenotyping and genotyping. We question the accuracy and reliability of phenotype (determination of postoperative cognitive dysfunction [POCD] at 1 yr) by McDonagh et al.1 POCD lacks consensus diagnostic criteria, and disparate methods have been employed for its detection.2 There are several method-

References

9. Ansgar M. Brambrink, M.D., Ph.D.,* Alex S. Evers, M.D., Michael S. Avidan, M.B.B.Ch., F.C.A.S.A., Nuri B. Farber, M.D., Catherine E. Creeley, Ph.D., John W. Olney, M.D. *Oregon Health & Science University, Portland, Oregon. brambrin@ohsu.edu

Correspondence

Ansgar M. Brambrink, M.D., Ph.D.,* Alex S. Evers, M.D., Michael S. Avidan, M.B.B.Ch., F.C.A.S.A., Nuri B. Farber, M.D., Catherine E. Creeley, Ph.D., John W. Olney, M.D. *Oregon Health & Science University, Portland, Oregon. brambrin@ohsu.edu

References


(Accepted for publication July 28, 2010.)
ological issues pertaining to the approach taken by McDonagh et al.\(^1\) that substantially weaken the measurement of POCD: (1) meticulously-matched control groups are imperative if cognitive decline is attributed to anesthetic/surgical events rather than underlying patient comorbidity\(^3\) (there was no control group in this study); (2) clinically meaningful, objective outcome measures, such as incident dementia,\(^4\) are interpretable and comparable across studies, as opposed to the artificial composite cognitive outcome measures used in this study; (3) the arbitrary statistical thresholds for POCD diagnosis employed in this study are likely inappropriate and of questionable clinical relevance;\(^5\) and (4) preoperative mild dementia should have been assiduously identified, because mild dementia, although difficult to detect, is a potent confounder in relation to POCD.\(^5\)

McDonagh et al.\(^1\) reported a 46% 1-yr incidence of POCD and mention that other studies have found different results. However, no studies by other groups have found rates of persistent POCD even remotely approaching the alarming rate found in this study. Some investigators have found little evidence for persistent POCD attributable to a surgical event,\(^1,3,5\) whereas others have found rates between 1 and 5%.\(^6,7\) The investigators suggest that patient, surgical, and anesthetic factors may all be implicated in the causes of POCD.\(^1\) However, they specifically provide evidence against anesthetic factors in showing no difference in POCD between regional and general anesthesia.\(^1\) Could it be that surgery and anesthesia were coincidental and that patient factors alone determined POCD? Might an appropriate control group have shown equivalent cognitive decline?

McDonagh et al.\(^1\) employed a principal component analysis to derive four components representing four cognitive domains. No factor loadings on individual tests were reported. Because most principal components contain positive and negative loadings because of the orthogonality requirement, these loadings are important to provide the direction of cognitive score. Furthermore, assuming that a decline in a component score is associated with POCD, McDonagh et al.\(^1\) defined POCD as a decline of 1 SD or more in at least one of the four cognitive domains. This definition has two major drawbacks: (1) it depends heavily on the number of principal components and (2) it ignores the time interval after the surgery. By definition, the probability (\(P\)) of POCD increases with the number of principal components used. To illustrate this, consider a hypothetical nonsurgical control group. Let \(q\) be the probability of individuals in this group not meeting the definition of POCD for a specific cognitive domain (i.e., not declining more than 1 SD). If we assume that the extents of change from baseline for the four cognitive domains follow the same normal distribution and are independent, then the probability in the control group of meeting the definition of POCD would be 1 − \(q^4\). Even with the relatively conservative assumption that 90% of the hypothetical controls would not meet the diagnostic threshold for POCD on each of the cognitive domains (i.e., \(q = 0.9\)), the apparent incidence of cognitive decline by the McDonagh et al.\(^1\) criteria would still be 34% (i.e., 1 − 0.9\(^4\)) in this hypothetical group. It would therefore be useful if the investigators would clarify what incidence of POCD they expected purely by chance with their model and what their incidence of postoperative cognitive enhancement was, using more than 1 SD improvement in any cognitive domain as a threshold criterion. It would also be helpful to know how the investigators classified patients if they declined by more than 1 SD in one uncorrelated test and improved by more than 1 SD in another.

A further concern is that the reported incidence rate of POCD seems inconsistent with the reported change scores from baseline on the continuous cognitive index, which is the average of four component scores.\(^1\) Among the entire sample, including subjects with and without the APOE4 allele, the 6-week change in cognitive score from baseline had a mean of 0.05–0.07 (i.e., very mild improvement compared with baseline) with an SD of approximately 0.27–0.28. Given these results, the estimated incidence of a decline of at least −0.5 (the decline in the cognitive index that the authors defined as clinically meaningful) would be roughly 2%, much smaller than the reported 6-week POCD incidence of 54.3%.

For a longitudinally observed function such as cognitive decline, discrete analyses (i.e., incidence of POCD at 6 weeks and 1 yr) might not be optimal. It is crucial to understand the time course of POCD. A longitudinal analysis was not provided but is required to assess the entire longitudinal change from baseline to 6 weeks and then to 1 yr. A simple incidence rate of POCD at each individual time point does not offer much information on those who initially declined fast and then recovered or those who did well initially and then declined fast. The investigators should comment what the extent of overlap was between those who met their POCD diagnostic criteria at 6 weeks and those who met their criteria at 1 yr.

The findings of McDonagh et al.\(^1\) could have paradigm-changing implications. If 46% of patients undergoing major elective surgery were really to experience clinically meaningful persistent POCD, it is likely that informed patients would choose to forego many such procedures. If, on the other hand, considerably fewer than 46% were to have persistent POCD, the reported phenotyping would be inaccurate, and correlation with the genotype (APOE4) would be meaningless. Let’s hope that the latter is true.

Michael S. Avidan, M.B.B.Ch., F.C.A.S.A.,* Chengjie Xiong, Ph.D., Alex S. Evers, M.D. *Washington, University in St. Louis, St. Louis, Missouri. avidannm@wustl.edu

References
We have seen a significant association between our defined POCD and cognitive difficulty reported by patients themselves in a 39-item assessment of perceived problems in memory, concentration, attention, and psychomotor coordination. The questionnaire includes items such as “I forget errands I planned to do” and “I fail to recognize people I know.” A higher score is worse. Patients we classified as having POCD had a mean increase [mean (SD)] of +1.64 (22.1) in difficulties at 6 weeks after surgery, whereas no-POCD patients had a mean decrease of −2.54 (19.4), \( P = 0.02 \). A trend toward increased mortality at 1 yr in patients with POCD (3.16 vs. 1.25%, \( P = 0.23 \)) was also reported in our article. Thus, our choice of a 1 SD decline does have some clinical relevance. Nevertheless, because any threshold is arbitrary in the absence of a criterion standard, we have also examined a continuous change score, which is calculated by subtracting the baseline from the follow-up cognitive index (mean of the four domain scores). As with the dichotomous outcome, there is no association between APOE4 and cognitive change, once again validating our finding of no significant effect of APOE4 genotype.

Third, the issue of baseline “mild dementia” was raised. As Avidan and colleagues have demonstrated, baseline mild cognitive impairment or dementia may be confounders when assessing long-term POCD. However, this cognitive impairment can be in single or multiple cognitive domains and can have a variety of clinical manifestations. The International Working Group on Mild Cognitive Impairment published their first efforts toward a consensus in this arena in 2004, but only near the completion of our study. Gauthier et al. have more recently highlighted the ongoing uncertainty and debate in this field, noting in particular that “the prognosis in terms of progression to dementia is more heterogeneous in population studies than in the setting of specialized clinics (such as an Alzheimer’s Disease Research Center) and is driven by the nosological and exclusionary criteria being used in either setting.” At the very least, further studies defining and validating the effects of baseline dementia are still needed. In our study focused on short-term cognitive dysfunction, we compared postoperative cognition with preoperative baseline values to make each patient his/her own “control” while acknowledging the inherent limitations to this approach.

Fourth, the etiology of POCD remains incompletely ascertained at this time, as Dr. Avidan points out. We do think, however, that it is inappropriate to conclude from our study that anesthetic factors have no effect on POCD because our study was not designed to examine that hypothesis. Thus, we made no conclusions regarding the etiology of POCD. We simply found no association between APOE4 genotype, as well as a panel of serum biomarkers, and postoperative cognitive decline, assessed either as a dichotomous or a continuous measure.

Fifth, Avidan et al. requested greater detail on the factor loadings used in our cognitive analyses; these are provided in Table 1. We did not publish this table because we believed it would not be meaningful or interpretable to most readers. Trails-Making B scores were reversed to...