Among the earliest systematic observations of the physiologic effects of anesthetic agents was John Snow’s description, in 1847, of the various stages of ether anesthesia. Although the focus has evolved somewhat, our interest in measures of the depth of anesthesia has persisted. Although the concern was initially largely one of avoiding the hazards of overdose, we have added a greater interest in the prevention of “underdosage.” There is considerable interest in preventing potentially hazardous hemodynamic and movement responses and in preventing recall. The latter concern applies most particularly to the patient who has received neuromuscular blocking agents. The contemporary literature also indicates an interest in using depth of anesthesia monitors as a means of controlling cost. The hope is that precise titration of anesthetic agents, as guided by a monitor of anesthetic depth, can serve to avoid wastage of expensive anesthetics and expedite postanesthesia care unit or hospital discharge, or both.

There have been several thoughtful discussions of the monitoring of depth of anesthesia1–3 and of the problem of awareness.4 Table 1 lists many of the techniques or devices that have been proposed or tested as methods for determining depth of anesthesia. A thorough discussion by Heier and Steen1 published in 1996 reviews the status of all but the most recent of those techniques. Briefly, the review leads to the conclusion that, although several techniques allow one to identify statistically significant differences in depth of anesthesia among defined anesthetic conditions for populations of patients, none has the sensitivity and specificity to allow the clinician to draw certain conclusions about depth of anesthesia in the individual patients for whom he or she treats. The then-available (1996) devices served as trend monitors of varying reliability but did not permit conclusive statements about depth of anesthesia in individual patients.

The purpose of this review is to summarize the developments that postdate the articles cited previously.1 That progress has involved principally two depth-of-anesthesia monitoring methods: the Bispectral Index, known by the trademarked acronym BIS (Aspect Medical Systems Inc., Newton, MA); and the middle latency auditory evoked response (MLAER). The BIS is an empirically derived index that is dependent on a measure of the “coherence” among components of electroencephalography.5 The MLAER uses measurements of the amplitude and latency of the early cortical components of the auditory evoked response. This discussion will focus on developments related to those two methods. In addition, because of the interest on the part of the media, patients, practitioners, and investigators regarding the topic of awareness during anesthesia, the issue to which this review gives greatest attention is: Can the available monitors be used to prevent the occurrence of awareness during anesthesia?
I preface this review by highlighting two difficulties that pervade the literature about depth of anesthesia. The first is the heterogeneity of the end points that have been used in the evaluation of the various monitors. Frequently used end points include hemodynamic responses (heart rate, blood pressure) to noxious stimulus, movement in response to stimulus, response to command, and recall. Recall is further subdivided into explicit (conscious) and implicit (subconscious) memory. These various end points do not appear to be part of a continuum and can occur independently of one another. A monitor that has some effectiveness with respect to predicting one end point may not predict others. With respect to movement, the explanation may be that this response can occur as a spinal level reflex. Thornton and Sharpe\(^6\) suggest that this may explain the relatively poor correlation between cerebrally derived parameters and movement response. The second difficulty is that of a lack of precision with respect to nomenclature. In using the term “consciousness,” investigators and, as a result, the public, sometimes do not distinguish between the ability to respond to command and the ability to form consolidated memory with subsequent recall of intraanesthetic events. Some investigations have equated the ability to respond to command and the ability to form explicit (conscious) and implicit (subconscious) memory. These various end points do not appear to be part of a continuum and can occur independently of one another.

For a monitor of depth of anesthesia to be valuable to the clinician, two conditions should be met. First, not only must the average values yielded by the device in two distinct states (e.g., hemodynamically responsive vs. nonresponsive; aware vs. oblivious) be statistically different, but also the range of values seen in those two states should not overlap. That is, in the ideal, there should be 100% sensitivity and specificity. At a minimum, if what clinicians seek is a specific numeric threshold that can be interpreted to mean “not aware,” it is essential that there be very high reliability in detecting the event of interest, \(i.e.,\) essentially 100% sensitivity (no false-negatives). A monitor used in that manner to detect an event that occurs with a very low incidence might cause more events than it prevents if its sensitivity to the event of interest is not approximately 100%. In addition, a reasonably low rate of false-positives, \(i.e.,\) a high specificity, will also be necessary for the instrument to be practical. As noted previously herein, many of the techniques studied have been limited by insufficient sensitivity or specificity. Second, the critical threshold values that distinguish depth-of-anesthesia states of interest should not be influenced by choice of anesthetic agent or by patient physiology, including coincident disease states and long-term use of medications. That is, critical thresholds should be constant (or at least should vary very predictably) from patient to patient and anesthetic to anesthetic. What follows leads to the conclusion (for this reviewer) that, although neither technique completely meets the two preconditions for the identification of probable patient awareness, it is the MLAER that comes closest.

### The Bispectral Index

The BIS, for several end points, and for several anesthetic regimens, yields the best combination of sensitivity and specificity of any commercially available depth-of-anesthesia monitoring device. In particular, during propofol-induced hypnosis, it is highly predictive of depth of sedation, as judged by responsiveness of the patient to command and tactile stimulation.\(^8\)\(^\text{--}\)\(^1^\text{1}\) It also is largely unaffected by the electroencephalographic pseudoarousal (activation) phenomenon that occurs with some anesthetics during the initial stages of anesthesia.\(^9\) A significant limitation, however, is that BIS thresholds do not appear to be independent of the combinations of anesthetic agents administered.\(^1^\text{2}\)\(^\text{--}\)\(^1^\text{5}\) That is, comparable BIS values achieved with different combina-
tions of agents do not represent the same depth of anesthesia. Mi et al. brought the BIS to similar levels by use of propofol infusion with or without fentanyl before tracheal intubation. With intubation, BIS increased similarly in both groups, but there were greater increases in blood pressure and heart rate in the propofol-only patients. Vernon et al. correlated preincision BIS with movement or nonmovement in patient groups anesthetized using isoflurane–alfentanil or propofol–alfentanil. The mean BIS of movers and nonmovers was significantly different within anesthetic groups (though the standard deviations overlapped, revealing sensitivity and specificity < 100%). However, BIS values did not distinguish isoflurane–alfentanil nonmovers from propofol–alfentanil movers, i.e., a BIS value that predicted unresponsiveness in one group predicted movement in the other. They concluded that “BIS values may not be independent of the anesthetic agent used.” Mi et al. observed that loss of response to command and loss of lash reflex occurred at higher BIS levels in patients who had received fentanyl in combination with propofol than in patients who had received propofol alone. Practitioners who use this device to titrate to an “adequate” depth of anesthesia must be aware that, for a particular anesthetic regimen, the BIS value that indicates “adequate” anesthesia will be vary among patients and that the range of “adequate” values will vary among anesthetic regimens.

Several evaluations of the BIS have sought to identify threshold values for responsiveness to command or stimulation or for the ability to form memory. The precise relation between the ability to respond to command and the ability to form memory is not certain. It is clear that, in some instances, subjects who respond to command do not have memory of the event. However, it has been observed that patients who could repeatedly respond to command formed explicit memory. Accordingly, response to command may be a conservative surrogate for awareness, by identifying those who may soon become capable of formulating recall. Flaishon et al. using the isolated-arm technique, administered induction doses of thiopental or propofol using muscle relaxant and followed the BIS until response to command recovered. Patients were not otherwise stimulated. No patient was responsive at a BIS of less than 58. Glass et al. correlated BIS with responsiveness to voice in volunteers who received incremental doses or concentrations of propofol, midazolam, or isoflurane. The lowest BIS at which nonresponsiveness to voice occurred was 40, and 95% of subjects were unconscious at a value of 50. Gajraj et al., in an investigation in which propofol was administered by infusion to patients undergoing joint replacement surgery with use of regional anesthesia, recorded the BIS value immediately after the transition to unresponsiveness to command. The lowest BIS value noted was 40. Lubke et al. also found evidence of implicit memory formation in trauma patients anesthetized with isoflurane and fentanyl, with BIS values between 60 and 40.

Collectively, these studies indicate that responsiveness to command or the formation of memory, at least implicit memory, may occur at BIS values as low as 40. Does this mean that to ensure that no paralyzed patient will be aroused in response to a noxious stimulus that the BIS must be kept less than 40 in all patients? That would probably result in unnecessarily and perhaps hazardously deep anesthesia in many patients. Furthermore, the peer-reviewed data base is inadequate, both in terms of the number of investigations and of the spectrum of physiologic and pharmacologic circumstances in which studies have been performed to allow identification of critical BIS thresholds. With respect to the spectrum of study conditions, it has been common to exclude subjects with a history of neurologic disease, medication affecting the central nervous system, age younger than 18 or older than 75 yr, and alcohol or drug abuse. Although this may be reasonable for initial investigations, it does not constitute a “real-world” test of the effectiveness of the device being studied. Similarly, many investigations have entailed single-anesthetic-agent or carefully prescribed anesthetic formulas rather than the polypharmacy of contemporary practice. Although investigations have been performed in real-world conditions, they have been limited in number. One such study was performed by Sleigh and Donovan. Those investigators observed BIS and spectral-edge frequency (SEF) at the time of various events during induction of and emergence from anesthesia while clinicians blind to monitor output applied “standard clinical practice.” The patients were 26 women, American Society of Anesthesiologists physical status I or II, undergoing “minor surgery” (a restricted corner of the real world). The anesthetics consisted of variable doses of midazolam and fentanyl before a slow induction of anesthesia with use of propofol and maintenance of anesthesia with isoflurane and nitrous oxide. The BIS-related observations are displayed in figure 1, which is reproduced from that report. The figure reveals the same problem of sensitivity and specificity for the BIS (the results were qualitatively similar for the SEF) that has been apparent with other monitors of depth of anesthesia. Although the population mean

Anesthesiology, V 93, No 3, Sep 2000
values for BIS (and SEF) during surgery differed significantly from the mean values at the time of first gag or first response to command, there was a substantial overlap of the ranges. Some values recorded during apparently adequate surgical anesthesia were within the range of values seen in awake patients, revealing again incomplete sensitivity and specificity.

The BIS monitor may provide useful trend information in individual patients. However, if the objective is to choose a single BIS threshold that would ensure nonresponsiveness and prevention of awareness in all patients, that number would have to be sufficiently low to result in unnecessarily deep anesthesia in a significant portion of the population. In addition, I am concerned that a dependence on a predetermined numerical threshold, as the primary determinant of the adequacy of depth of anesthesia, will result in occasional patients being inadequately anesthetized. Many of the investigations that have suggested suitable numerical thresholds have been performed using homogenous populations of relatively healthy patients who receive a standardized anesthetic. Although some of these investigations may have suggested substantial specificity and sensitivity for specific BIS output numerical thresholds, it is not confirmed that these thresholds will be as robust in real world heterogeneity of the varying combinations of anesthetic agents used by individual practitioners; nor is it confirmed that intercurrent disease states and, perhaps more importantly, medications will not influence the behavior of the coherence measures on which the BIS algorithm depends. What effect do anticonvulsants (diphenylhydantoin, phenobarbital, carbamazepine, antidepressants of many different classes, sedatives and anxiolitics, and long-term administered analgesics have on the numeric parameter derived by the BIS algorithm? How do the relevant thresholds vary with hypothermia, as may be used in either intracranial aneurysm or cardiac surgery? And, as a final caveat, practitioners must appreciate that variation in electrode montage will also alter the derived BIS values.20 There is much to learn before clinicians should contemplate adopting specific BIS parameters as thresholds that will reliably prevent adverse events, including movement, hemodynamic changes, or awareness and recall.

Middle Latency Auditory Evoked Response

Signal averaging of the electroencephalogram recorded from a mastoid–vertex electrode montage after repeated click stimulation of the ear yields a highly reproducible sequence of wave forms. These wave forms arise, in sequence, from the brain stem, the auditory radiation, the auditory cortex, and association areas of the cortex (fig. 2). The brain stem auditory evoked response (BAER) is resistant to the effects of anesthetic agents. The wave forms that follow the BAER are increasingly sensitive to anesthetics, with the characteristic

Anesthesiology, V 93, No 3, Sep 2000
pattern of change being increases in latency and decreases in amplitude in response to increasing drug concentrations. The early cortical responses, notably Pa and Nb, vary in a dose-dependent and consistent manner in response to the administration of inhaled and intravenous anesthetics. However, opioids and midazolam have less pronounced effects than do the inhaled agents and the various intravenous induction agents. Ketamine has no effect. The information regarding the use of MLAER in the detection of awareness was thoroughly reviewed by Thornton and Sharpe in 1998. Investigations preceding and following that review have focused on the latencies and amplitudes of Pa waves and Nb waves and on the general morphology of the three-wave Pa-Nb-P1-Nc-P2 complex (fig. 2).

Several studies suggest that the MLAER has substantial potential to be an effective discriminator between the anesthetized and conscious states. In a study of cardiac surgery patients during near normothermia in the prebypass period, Schwender et al. studied the occurrence of implicit memory during the administration of several different anesthetic agents. They observed that implicit memory occurred only in patients in whom the latency increase in Pa was less than 12 ms. That threshold had a sensitivity of 100% for the detection of patients capable of forming implicit memory and a specificity of 77%, i.e., 23% of patients with Pa less than 12 ms did not form implicit memory. Thornton et al. compared ability to respond to command with the latency of the Nb wave. After morphine premedication, thiopental and a neuromuscular blocking agent were administered to permit tracheal intubation. Anesthesia was maintained with 70% nitrous oxide (N₂O), which was gradually reduced to 50%. An Nb latency of less than 44.5 ms provided 100% sensitivity for identifying patients capable of responding to command (isolated forearm). Specificity was not calculated. Newton et al. studied the effect on the ability of volunteers to form explicit memory during inhalation of sub-minimum alveolar concentrations (MAC) of isoflurane. An Nb latency of 47 ms separated, with 100% sensitivity and specificity, those who could and could not form explicit memory for words presented during inhalation. These investigations indicate that, in at least some circumstances, the MLAER can provide 100% sensitivity, albeit with imperfect specificity, to the occurrence of conscious perception or awareness in anesthetized patients. In addition, three more recent comparisons of MLAER derivatives confirm the high level of sensitivity and specificity in the prediction of consciousness that can be achieved with the MLAER. These studies suggest that the distinction between the anesthetized and awake states is sharper, i.e., there is less overlap in the ranges of conscious and unconscious values, with MLAER derivatives, than is the case with the BIS.

There is no commercially available device for the intraoperative monitoring of the MLAER for the purpose of depth-of-anesthesia evaluation. In addition, there is ongoing evaluation of which derivative of the MLAER, including the latencies of individual waveforms (usually Nb) and derived indices that incorporate information about the overall morphology of the MLAER complex, is optimal. Furthermore, the performance of the MLAER in a broad range of patients and anesthetic conditions remains to be explored. Although there is reason to believe that the MLAER may be useful in identifying situations with a risk of awareness, data suggest that it may be less effective in predicting movement in response to surgical stimulus. This needs clarification, as does application of the MLAER during cardiopulmonary bypass. Although MLAER is known to be relatively insensitive to temperature change within the ranges commonly used during cardiopulmonary bypass, opioids and benzodiazepines have relatively little impact on the MLAER. It is not known whether MLAER has suitable sensitivity and specificity for depth-of-anesthesia monitoring in cardiac surgery with principal use of these agents. As is the case with BIS, knowledgeable application of the MLAER will need investigation in a broader range of patient types and anesthetic conditions.

Summary and Additional Observations

There is considerable current interest in the issue of awareness. The concern that, in our patients, unnecessary anxiety about the risk of awareness and unrealistic expectations about the ability of the BIS monitor to prevent the phenomenon have developed has already been discussed in Anesthesiology. It has also been asserted that careful, prospective study with subsequent peer-reviewed publication will be necessary to establish the effectiveness of any putative awareness-prevention device. The peer-reviewed literature does not support the notion that any commercially available monitor can serve to prevent awareness, although it indicates that useful trend-mon-
itoring of depth of anesthesia and titration of depth of sedation can be accomplished with the BIS.\(^\text{10,11}\) Furthermore, even in the event of the development of a device that reliably identifies anesthetic states representing a high risk for awareness, episodes of awareness still may occur. The first reason is that depth of anesthesia at any moment is probably the sum of the effects of the anesthetic agents being administered and the prevailing degree of stimulus-related arousal. Even a monitor that meets the stringent sensitivity-specificity conditions suggested above might "fail," in the context of light anesthesia with minimal surgical stimulus, in the event of a sudden increase in the intensity of stimulus. The second is that there will continue to be situations in which the clinician is limited by failing hemodynamics from administering the anesthetic agents that are otherwise warranted. It is unrealistic to expect any monitor to be proof-positive against the occurrence of awareness.

**References**

19. Sleigh JW, Donovan J: Comparison of bispectral index, 95% spectral edge frequency and approximate entropy of the EEG, with changes in heart rate variability during induction of general anesthesia. Br J Anaesth 1999; 82(5):666–71
29. Gajraj RJ, Doi M, Mantzaridis H, Kenny GNC: Comparison of bispectral EEG analysis and auditory evoked potentials for monitor-
ing depth of anaesthesia during propofol anaesthesia. Br J Anaesth 1999; 82(5):672–678