Decision Support Alerts

Importance of Validation

Daniel I. Sessler, M.D.

Electronic anesthesia records are now ubiquitous in the United States and are rapidly becoming standard worldwide. Once electronic records are implemented, it is relatively easy to add decision support functions. There has thus been a proliferation of systems that interpret anesthetic variables and provide guidance to clinicians.

In this issue of Anesthesiology, Kheterpal et al. present an evaluation of a locally developed decision support system called AlertWatch. The system guides clinicians to avoid hypotension, limit tidal volume, and give appropriate amounts of fluid. They compared a cohort of patients in whom clinicians chose to activate decision support with: (1) a control cohort from the 22 months before the system was available, and (2) a contemporaneous cohort in which clinicians did not activate decision support. In addition to process measures, the investigators evaluated outcomes including myocardial and kidney injury, hospital length of stay, and mortality.

Comparing results in patients from before and after decision support became available has distinct limitations. There are three major sources of error in before-and-after studies. The first is time-dependent confounding. In general, health care improves over time as new methods and understanding are incorporated into the clinical routine. Outcomes thus improve over time consequent to many factors, some of which are subtle, unrecognized, and unquantified. There is little basis for selecting one change (introduction of decision support, in this case) and attributing all benefits to that single intervention.

The second major bias is the Hawthorne effect, the process by which investigator or organizational interest improves performance by directing attention to outcomes of interest. Finally, regression to the mean can falsify before-and-after studies, especially when an intervention is activated because of an apparent problem that may be a random perturbation. (This bias is common in studies of surgical site infection, which are often started during an apparent spike in infection rates.) Over a long observation period, such as the study by Kheterpal et al., regression to the mean is less likely to falsify results than time-dependent confounding or the Hawthorne effect. For additional detail about each source of error, see recent reviews.

Because all three sources of bias make interventions appear more effective than they actually are, and by unquantifiable amounts, before-and-after designs are weak and often simply invalid. Had that been the only comparison, this report would have been considerably less valuable. Fortunately, the authors also included a contemporaneous control group.

Before-and-after designs are never randomized. The investigators’ second comparison between contemporaneous patients could have been randomized, but it was not. Instead, they compared patients in whom clinicians chose to activate decision support to patients in whom decision support was not used. While this is a stronger design than a historical control, it lacks the protection against selection bias that would result from randomization or other investigator-allocated use of decision support systems. For example, better clinicians may have routinely activated the system, whereas weaker clinicians chose not to, or the system may have been used for sicker patients. Multivariable modeling, including propensity score adjustment as used by Kheterpal et al., helps, but only to the extent that potential
confounding factors are known and accurately quantified—which is never completely the case. That the apparent benefit was much larger in historical, as opposed to contemporaneous, comparison indicates how misleading before-and-after results can be. In fact, the historical comparison indicated that decision support markedly improved process measures and outcomes by amounts that were both clinically important and highly statistically significant. In contrast, the more reliable contemporaneous comparison showed small effects on the process measures and no significant effect on outcomes. The remaining question, of course, is whether even the relatively small benefit for process measures observed in the contemporaneous cohort would remain had the use of decision support been allocated by the investigators rather than been self-selected.

The usual approach to limiting selection bias and confounding is individual patient randomization. But, a conventional randomized trial would likely have been impractical for Kheterpal et al. because the required sample size would have been large. There are, however, other approaches that are more efficient yet preserve the benefits of investigator-allocated treatment. For example, a recently developed approach is alternating intervention.\(^4\) In such trials, an intervention—such as access to a decision support system—is used for a period and then discontinued for a period, with this cycle repeated many times. Because it is highly unlikely that surgical scheduling will be determined by intervention, potentially confounding time-dependent practice changes will be comparable during each cross-over period. The Hawthorne effect is also likely to both dissipate and average out over time. A limitation of alternating intervention trials is that they are inherently unblinded; however, in practice, participating clinicians could hardly have been blinded to decision support in Kheterpal et al., and primary outcomes were largely objective. Another approach is to use the decision support system itself to randomize patients who meet certain thresholds to either activate alerts or not.\(^5\) Both approaches require waived consent and are thus most suitable for interventions that are not yet standard, likely to prove beneficial, and extremely unlikely to be harmful.

It might seem obvious that decision support alerts can only be beneficial. After all, they are just recommendations that remind clinicians to consider various interventions. Presumably, clinicians will still use good judgement and respond appropriately. In some cases, the assumption of benefit is probably valid. Consider, for example, an alert that reminds clinicians to re-dose antibiotics during prolonged surgery. Other alerts remind clinicians to comply with various pay-for-performance measures, which presumably provide financial benefits at the very least.

But it is also true that alerts that seem obviously beneficial may not be. For example, one of the few randomized trials of decision support showed that alerts for hypotension, which included both anesthesia record screen and pager notifications, did not improve response times or the amount of hypotension.\(^5\) That alert, which many may assume could only be helpful, could potentially distract clinicians without providing any actual benefit. And that is the problem: decision support alerts can be harmful if they distract clinician from other tasks or induce alarm fatigue, which impairs response to subsequent alarms for truly serious events.

My concern is that decision support systems in development may easily include a hundred or more alerts. Presumably many will be helpful, but others may not. It would therefore seem reasonable to expect most decision support systems and alerts to undergo formal testing, just as we would expect any other drug or device to be properly validated. I congratulate Kheterpal et al. on formally testing their system.

Decision support significantly reduced the amount of hypotension, fluid administration, and tidal volume—but by amounts that were probably not clinically important. Consequently, there were no significant differences in predefined substantive outcomes, including myocardial and kidney injury, mortality, or duration of hospitalization. AlertWatch guidance, which may have been considered beneficial, in fact had little effect even on mediators, and no apparent effect on outcomes. The study of Kheterpal et al. is therefore an excellent example of why it is important to specifically test decision support alerts and guidance systems.

Lack of apparent outcome benefit does not necessarily mean that the system is useless. Indeed, it elegantly presents physiologic data in ways that may help clinicians manage subtle aspects of anesthesia that benefit patients, despite unchanged “hard” outcomes. However, these putative explanations remain to be confirmed. It is also possible that clinicians failed to respond appropriately to the guidance provided and will get more value from the system when they have greater experience with it.

In summary, Kheterpal et al. tested a decision support system with alerts that many would consider obviously beneficial. In fact, the guidance provided only modest benefit, which supports my assertion that alerts should be formally tested—just like other medical interventions. As with any drug or device, a variety of study approaches can be used. The easiest and most obvious, a before-and-after comparison, is the weakest design and should generally be avoided. Contemporaneous observational studies are better, but may be potentially weakened by selection bias. Alternating intervention designs is easier, faster, and much less expensive than conventional randomized trials. They require waived consent, but nonstandard “obviously beneficial” optional guidance is a strong basis for waived consent since routine monitoring will not be denied, patients are likely to benefit, and clinicians remain firmly in charge of care decisions. Randomized trials provide a high level of internal validity, but will often be too challenging to allow for routine testing of novel decision support systems. However, any study is better than uncritically accepting novel decision support systems under the assumption that they must be beneficial.
Competing Interests
The author is not supported by, nor maintains any financial interest in, any commercial activity that may be associated with the topic of this article.

Correspondence
Address correspondence to Dr. Sessler: DS@OR.org

References

ANESTHESIOLOGY REFLECTIONS FROM THE WOOD LIBRARY-MUSEUM

Birkett Advertises Hurd’s Apparatus for Anesthesia with Vitalized Air

Born in Rochester, New York, and raised in London, England, Charles H. Birkett (1856 to 1932) returned home in 1873 to learn dentistry. After 13 yr of dental practice in East Liverpool, Ohio, Birkett visited Cleveland dentist George H. Hurd (1829 to 1904) in 1886 to acquire one of the Clevelander’s early “Vitalized Air” vaporizers for adding trace amounts of alcohol and chloroform to unoxygenated nitrous oxide. Impressed with Hurd’s ideas, Birkett began advertising on trade cards, such as the one above, that Birkett’s was the only practice in the “vicinity licensed and using Hurd’s apparatus—a new and harmless process of administering Vitalized Air, pleasant, agreeable, and harmless to all.” (Copyright © the American Society of Anesthesiologists’ Wood Library-Museum of Anesthesiology.)

George S. Bause, M.D., M.P.H., Honorary Curator and Laureate of the History of Anesthesia, Wood Library-Museum of Anesthesiology, Schaumburg, Illinois, and Clinical Associate Professor, Case Western Reserve University, Cleveland, Ohio. UJYC@aol.com.