The Overpowered Mega-study Is a New Class of Study Needing a New Way of Being Reviewed

To the Editor:
The recently published Memtsoudis et al.’s retrospective “mega-study” reviewed electronic billing data of 382,236 patients who had undergone primary hip and knee arthroplasty surgery. A study of this size has the potential to detect very tiny between-group differences for very rare events. Specifically, Memtsoudis et al. observed a statistically significant mortality reduction of 0.08% in the group receiving neuraxial blocks versus the group who received general anesthesia for total knee arthroplasty patients.

The huge number of patients studied here nearly represents the equivalent of the entire 40-yr careers of 40 full-time orthopedic anesthesiologists, assuming they perform 1,000 anesthesia cases per year, with 40% of cases being for primary hip and knee arthroplasty procedures. This represents 1,000 individual anesthesia practice years. The observed mortality difference would represent about one added 30-day death every 5 yrs per anesthesiologist administering only general anesthesia. Although the death of any individual patient is tragic, the size of the “treatment effect” as well as the retrospective database-derived nature of the study should prompt us to ask whether or not the results justify a change in anesthesia practice?

Huge studies such as this are unquestionably valuable, because they CAN detect differences in the incidence of rare events—differences that could never be detected in prospective, randomized trials—largely because performing such trials would be prohibitively difficult. However, such retrospective studies, unlike prospective trials, can never define causality, only association, and the inherent problems produced by missing data, miscoded information, and unrecognized (and hence unincorporated) covariants may be large enough to influence the reliability of any conclusions particularly when differences between groups are very small (perhaps regardless of statistical significance).

A recent editorial by Collins et al., commenting on a 10-million patient database study, recognized such observational mega-study limitations and emphasized the need to develop tools and consensus-based guidelines for authors, editors, and readers to better study and understand the deeper meanings and limitations of such observational analyses.

What factors (e.g., missing covariates) might have confounded the work by Memtsoudis et al.? We believe that two critical questions are (1) why was neuraxial anesthesia chosen for any patient and (2) how was neuraxial anesthesia conducted?

There are always some subtle (and perhaps not so subtle) variations in patient’s comorbidities, individual anesthesiologist and surgeon training, skills, and experience and decision-making processes and institutional resources of anesthesia drugs, equipment, and patient care facilities. Another recent mega-study on 367,796 patients examining viewing general surgical mortality showed patients being operated within one unitary healthcare system, but in a different hospital, could experience a significantly 30-day mortality 200% difference between best and worst scoring hospitals and this correlated with the number of intensive care unit beds available.

The decision to use a regional anesthesia technique on an arthroplasty patient is often decided by a surgeon’s idiosyncratic likes or dislikes for regional anesthesia, similar idiosyncrasies of the anesthesiologist, the time available to perform the regional anesthetic, and finally the personal fears and preferences of the patient. Thus three parties commonly contribute to the decision to use neuraxial anesthesia or not and only one of those three parties is trained in anesthesia. Anesthetic considerations in choosing an anesthesia plan for an individual patient may be overshadowed by unscientific covariates around the anesthesia plan decision process which may in turn influence mortality directly, if only slightly. Such factors could easily influence small mortality differences in a mega-study—but would almost certainly be impossible to incorporate as covariates in the analysis.

It could be also speculated that the increased use of neuraxial anesthesia is only a marker for the fact that neuraxial blocks are more likely performed by anesthesiologists more skilled in

(Accepted for publication September 26, 2013.)
neuraxial blocks, and perhaps additionally more skilled at the management of orthopedic patients. That alone could explain a tiny patient mortality improvement associated with use of neuraxial anesthesia. Conversely, it could be speculated that decreased use of neuraxial anesthesia may be influenced by slower surgeons with reduced surgical skills, causing increased patient surgical stress, blood loss, and sepsis and who may prefer (or their anesthesiologists prefer) general anesthesia.

Based on such factors, we are reluctant to conclude that the use of regional anesthesia per se is “superior” to general anesthesia in terms of patient 30-day mortality. Although neither do we dispute the observed difference in the mortality of PATIENTS having regional versus general anesthesia nor do we disagree with the importance of this difference, we cannot know whether the choice of anesthetic itself was the causative factor. We conversely observe that, at least, neither the metaanalysis performed by Rodgers et al. in 2000 (which was perhaps the “mega-study” for that era) nor the Memtsoudis et al.’s study on 382,236 patient records suggest neuraxial blocks are detrimental to the patient.

The Memtsoudis’s study it is exceptionally useful for the debate it raises. But as noted by both Memtsoudis et al. and Neuman and Brummet6 in the accompanying editorial, this study is unlikely to be the last word. Further study is clearly needed—but the challenge will be to find improved ways to actually perform (and analyze) such studies.

Robert M. Raw, M.D., Michael M. Todd, M.D., Brad J. Hindman, M.D., Rashmi Mueller, M.D. University of Iowa, Iowa City, Iowa (R.M.R.). rob-raw@uiowa.edu

In Reply: I thank Drs. Raw, Todd, Hindman, and Mueller for their interest in our study1 and their thoughtful comments. The authors raise important points representing different perspectives regarding the use of large databases for research in general and for anesthesia outcomes in particular. Despite the widespread use of administrative databases by various specialties over the last decades, perioperative physicians have only relatively recently been engaged in this type of population-based research. Therefore, traditional approaches to review and interpret studies may seem inadequate for the evaluation of resulting data that often use complex methodologies. However, I believe that it is time for our specialty to embrace these type of studies as many other specialties have done, accepting their strengths AND weaknesses. Although it must be stressed that outcome studies of this kind are not meant to and cannot be the final word on a research question, they do represent population-based observations on issues using information from actual practice that then require further inquiry.

Raw et al. discuss the fact that the outcome of mortality may be of limited significance during the career of an individual practitioner. However, viewed through the eyes of a public health researcher, one has to consider that between 1990 and 2004 over 23,000 patients died in the United States alone after hip and knee arthroplasty. At current rates of 1 million joint arthroplasties performed annually, a difference of 0.08% in mortality rate (which is the difference found between cases using neuraxial vs. general anesthesia) relates to 800 lives saved if all cases were performed under neuraxial anesthesia versus a general approach.

In addition, mortality was only one of the outcomes studied in our analysis, and many other more frequently occurring complications were positively affected in the setting of neuraxial anesthesia. This is important because although individual practitioners may experience very few fatal outcomes during their career, certainly many of us will observe serious complications on many more occasions. It is becoming increasingly clear that although causal relationships cannot be established with database analyses, the association found in our and other studies between outcomes and anesthetic practice needs to be further explored.1,3

There is no question that database studies lack the ability to provide explanations of mechanisms associated with observations and are unable to account for many important cofounders. In this context, Raw et al. are correct that we cannot establish why neuraxial anesthesia was chosen and it is imperative that readers take this fact into consideration when interpreting the study results. However, these results should be reason for pause and at the very least lead to the rethinking of long-held beliefs, reevaluation of practices, and the generation of hypotheses that should be pursued in a quest for answers regarding potential mechanisms of action. Even if it turns out that, for example, neuraxial anesthesia use by itself is not the reason for better

References

(Accepted for publication August 23, 2013.)