The Importance of Validation Studies in Perioperative Database Research

Mark D. Neuman, M.D., M.Sc.

O

VER the past decade, administrative data have come to play a major role in research on perioperative care. Beyond providing a basis for studies focusing on the organization and delivery of care for patients undergoing surgery and anesthesia, insurance claims and other administrative records have been used to examine fundamental questions of clinical epidemiology and comparative effectiveness research. Yet, despite the growing number of investigators who seek clinical and policy insights from administrative data, efforts to validate such data for identifying specific diagnoses remain relatively uncommon although important examples do exist within\textsuperscript{1} and outside of\textsuperscript{2,3} the perioperative setting. In this issue of Anesthesiology, McIsaac et al.\textsuperscript{4} offer quantitative insights that can help readers and investigators make better sense of the opportunities and challenges involved in using administrative data for perioperative research.

McIsaac et al. examine the validity of administrative data for obstructive sleep apnea (OSA). OSA is a widespread public health problem that may increase the risk of adverse events after surgery,\textsuperscript{5,6} and to date, research using administrative data has played a major role in highlighting it as an important issue for perioperative medicine.\textsuperscript{7-9} Using a large clinical database from one major Canadian teaching hospital, McIsaac et al. identified 4,965 surgical patients who underwent a major surgical procedure between 2003 and 2012 and who also had undergone a polysomnogram at that same hospital before surgery. The authors linked these clinical data to national and provincial health administrative databases that included claims for inpatient, emergency department, and day-surgery care, as well as for physician services and durable medical equipment. They obtained a successful linkage for 4,353 patients, or 88% of their original cohort, and found 56% of linked patients to have a confirmed diagnosis of OSA based on polysomnogram results or documentation by a sleep medicine physician.

The authors used this linked dataset to study the accuracy of a range of case-ascertainment algorithms for obstructive OSA; these algorithms used codes from the 9th and 10th revisions of the International Classification of Diseases (ICD), including ICD codes that have been previously used to identify patients with sleep apnea in administrative data, as well as claims for polysomnography or positive airway pressure devices. Several of the algorithms that the authors assessed demonstrated high specificity (i.e., a high proportion of people who truly lacked OSA who were correctly identified by the algorithm); however, most of the methods assessed, and particularly those that relied exclusively on ICD codes, showed low sensitivity when tested against a standard of a diagnostic polysomnogram or documentation of OSA by a sleep medicine physician (i.e., a low proportion of people who truly had OSA who were correctly identified by the algorithm). Extrapolating their findings to a general surgical population, the authors estimate that only 40% of patients identified as having OSA based on an ICD-9 or ICD-10 diagnosis code in their administrative database could be expected to actually have the condition, whereas 10% of patients without such a code could still be expected to have OSA.
The study by McIsaac et al. has limitations of its own. Their use of data from Canada potentially limits the relevance of their work to studies that use administrative data sources from the United States because coding practices could potentially differ across countries due to variations in the organization and reimbursement of care. Its single-center design may also limit generalizability because coding practices may vary from institution to institution. Although they present findings for a range of ascertainment algorithms, these algorithms overlap incompletely with methods used in recent large database studies focusing on outcomes for surgical patients with and without OSA, thus limiting direct comparisons.

Yet this work offers important lessons for consumers of perioperative database research as well as for investigators—myself included—who use ICD codes to identify patients with specific health conditions. Essentially all empirical studies—regardless of the source of the data used—suffer from some degree of misclassification; in the present context, the low sensitivity of administrative claims for OSA creates the possibility that patients who actually have OSA might be incorrectly classified by common administrative data algorithms as lacking the disease. As McIsaac et al. note, such misclassification is often cited as being likely to bias study results toward the null hypothesis or a finding of no effect. In a study aiming to compare outcomes among patients with and without OSA, the failure to identify OSA in subjects who actually have the disease might make the two comparison groups appear more similar due to the presence in both groups of patients who actually have OSA. Such misclassification would work to diminish the differences in outcomes between groups that would be attributed to the presence or absence of OSA. Study results affected by such bias might be argued to be “conservative” because the true differences between groups would be expected to be larger than those that such a study would find.

Such an argument assumes that misclassification occurs independently of other important factors (i.e., “nondifferential”). Yet if other factors that might impact outcomes also influence an individual’s likelihood of being classified as having the condition under study, such “differential” misclassification can also work to bias study findings away from the null hypothesis, potentially yielding spuriously large results.10

The work by McIsaac et al. suggests that the misclassification of OSA in administrative data may indeed be related to other important factors; administrative diagnosis codes were more likely to correctly identify OSA in patients who also had other major perioperative risk factors, such as heart disease and diabetes, and who moderate or severe versus mild OSA. Such considerations create the potential that studies relying on administrative data to identify patients with OSA could potentially overstate the overall impact of this condition on postoperative outcomes if those patients classified in the study as having OSA tended to be sicker or to have more severe disease than patients with OSA overall.

To their credit, authors of studies that have used administrative data to examine the impact of OSA on postoperative outcomes have been frank about the limitations of such data sources for health research.9,11 Rather than invalidating these past studies, the work by McIsaac et al. provides an opportunity to add nuance to their interpretation. In particular, it offers new quantitative data that can help readers and researchers to better gauge the potential extent and nature of misclassification of OSA within administrative data. Future investigators might draw lessons from this work, as well as other investigations12,13 that have shown substantial variability in the coding of common diagnoses over time, as to what insights (regarding OSA or other conditions) might or might not be obtainable from administrative data sources. This investigation also offers useful information that can help practitioners and researcher more correctly interpret the results of past studies by newly quantifying and explicating limitations of administrative data sources in this context that have been suspected and debated, yet not fully described. Most importantly, this work offers a strong example of the importance of well-done validation studies in perioperative database research, not only for guiding the types of questions that we choose to investigate through the lens of retrospective data but also for making better sense of what insights such investigations might be able to convey.

Acknowledgments

Support was provided from institutional/departmental sources and from the National Institute on Aging, Bethesda, Maryland (grant no. K08AG043548).

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. Neuman: neumannm@mail.med.upenn.edu

References

2. Carnahan RM, Moores KG: Mini-Sentinel’s systematic reviews of validated methods for identifying health outcomes using administrative and claims data: Methods and lessons learned. Pharmacoepidemiol Drug Saf 2012; 21(suppl 1):82–9
Moonflowers: Heart-racing Deliriants behind Hanaoka’s Mafutsusan (Year 1804, Part 1)

For much of the past two millennia, as successive masters of herbal medicine trekked eastward from India to China, Korea, and Japan, they hand-carried or harvested plants of the Datura genus, the name of which was derived from dhatur (Sanskrit: “white thornapple”). With pale flowers (left) opening at twilight, these mysterious “moonflowers,” or “Indian thornapples,” bore spiny fruit (right) filled with high concentrations of tropane alkaloids, such as scopolamine, atropine, and hyoscyamine. Ingesting trace amounts could produce rapid heart rates and anticholinergic delirium. Such disruptions in awareness and memory were initially exploited for religious experiences by some, and for medical use by a few of anesthesia’s pioneers, including China’s Hua Tuo (c. 140–c. 208) and Japan’s Seishū Hanaoka (1760–1835). The anticholinergic slurry of such “moonflowers” provided the “brain dulling” portion of a primitive balanced anesthetic. (Copyright © the American Society of Anesthesiologists, Inc.)

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