

Editorial

Editorial: Large Database Studies—What They Can Do, What They Cannot Do, and Which Ones We Will Publish

Jonathan N. Grauer MD, Seth S. Leopold MD

Most clinicians have come to imagine that scientific evidence has levels, and at the top of the pyramid are randomized clinical trials (RCTs) and meta-analyses [12]. But we all have read some

poorly designed randomized trials, and sometimes a humble case series can substantially influence thought about a technique or implant, particularly if it describes failures of treatment. Nevertheless, as the Centre for Evidence-Based Medicine analogizes [6], if clinical evidence were stored in boxes marked “randomized trial,” “cohort study,” “case series,” and the like, and we needed to answer an important question, we might first open the box stamped “randomized trial” to see if there is anything inside.

All the same, there are some important kinds of questions that RCTs and meta-analyses will never answer. Surgical RCTs generally are small to medium-sized, and as such, are tooled to assess efficacy more than safety. Randomized trials have almost no ability to detect rare events, evaluate resource utilization in normal practice, or compare any but the most common complications. Likewise, premature failures of commonly used interventions probably will not be identified first by RCTs, but rather by studies of other designs. To assess real-world and important endpoints like these, we will need to look elsewhere for evidence.

Registry studies and analyses drawn from large-system databases represent two robust approaches to answering these kinds of questions.

Clinical Orthopaedics and Related Research® is proud to support the excellent, thought-provoking work being done using the world’s registries. We publish it frequently [10, 11], and we have promoted it in Spotlight commentaries and interviews on these pages [13]. *CORR*® will publish selected proceedings from the most-recent meeting of the International Society of Arthroplasty Registries this summer, and we will cover the important role registries play in clinical research in an editorial in the coming months.

We also believe that important, pressing questions can be answered efficiently and accurately using large databases, including the Nationwide Inpatient Sample (NIS), National Surgical Quality Improvement Program (NSQIP), National Hospital Discharge Survey (NHDS), National Trauma Databank (NTDB), and Medicare administrative databases, among others. In fact, there has been something of a bloom of orthopaedic papers from these

The authors certify that they, or any members of their immediate families, have no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

The opinions expressed are those of the writers, and do not reflect the opinion or policy of *CORR*® or the Association of Bone and Joint Surgeons®.

J. N. Grauer MD
Department of Orthopaedics and Rehabilitation, Yale School of Medicine, New Haven, USA

S. S. Leopold MD (✉)
Clinical Orthopaedics and Related Research, 1600 Spruce Street, Philadelphia, PA 19103, USA
e-mail: sleopold@clinorthop.org

Editorial

kinds of sources lately [2–4]. With sample sizes that run from thousands to millions of patients, these studies allow us to tackle questions that likely will never be answered any other way.

To make best use of this material, readers need to know just what kinds of questions can be asked of databases, how the same question may elicit different answers from different databases, and how slipshod or carelessly presented work from databases can mislead us. Finally, authors should know how editors for *CORR*® evaluate this relatively new type of research.

We believe studies drawn from these databases best serve clinicians when they focus on what they are uniquely suited to do: Correlate less-common adverse events with modifiable and previously unidentified risk factors [1], identify adjustable provider- or hospital-level variables associated with readmissions or complications [7], and compare resource utilization for common interventions across diverse geographic regions or practice settings [5]. As with any other study design, the more novel the questions (and the more unexpected the answers), the more important these studies are. They do not make such good reading when they tell us what we already know.

Readers also should realize that the same question asked of different databases may result in different answers [2, 4]. The most obvious reason

for this is that each database surveys a specific population and may gather different data in different ways. Some databases, like the NIS, are inpatient-only samples: The moment a patient leaves the hospital, (s)he no longer is followed. As a result, studies drawn from inpatient-only samples that ask about complications that occur both before and after discharge—such as surgical site infection—will badly underestimate event rates. By contrast, prospective surgical registry-type databases (such as NSQIP) follow patients for 30 days or longer after surgery, and so are better tooled to find problems that occur after discharge. What do they lose in the exchange? Sample size. Prospective registries are smaller. There are other differences as well. Some registries focus on particular types of patients. The NTDB, for instance, includes patients with higher-energy injuries. Clearly, matching the study population to the question being asked is crucial.

Additionally, each database uses different data elements. Many databases, including NIS, are built from ICD-9 codes. Since these codes' primary use is billing, they have lower sensitivity to answer the clinical questions in which most readers are interested [8, 9]. By contrast, other databases—including NSQIP and parts of the NTDB—use chart-abstracted data, which offer more of the details clinicians seek. Readers

should also recognize that missing data limits all databases to some degree, and any study drawn from them should comment on the degree to which this might influence its findings. We believe the most important failing of orthopaedic studies from these sources is that these databases generally do not capture enough patient-level diagnostic information or patient-reported outcomes data. Efforts are underway to address many of these limitations. But at present, questions calling for that level of detail probably are better studied using single or multicenter designs that capture more complete patient-level data of this sort.

Notwithstanding the shortcomings of large-database studies, we at *CORR*® believe that they have an important role to play. But since these studies are not always easy reading, there must be a payoff for the reader who toughs it out—at the very least, a finding that changes the way the reader thinks about an important problem or treats his or her patients. Going forward, *CORR*® will consider a study that draws from a large database if it either presents a genuinely counterintuitive descriptive finding, or provides a specific suggestion to improve clinical care, practice management, or public policy. We will publish papers meeting one or both of those criteria when they are methodologically robust. As with all of the important studies we publish in *CORR*®, we will

Editorial

continue to make every attempt to accompany these studies with *CORR*[®] Insights commentaries or Editor's Spotlight features.

Acknowledgements The authors thank Matthew B. Dobbs MD, Mark C. Gebhardt MD, Terence J. Gioe MD, Paul A. Manner MD, Clare M. Rimnac PhD, and Montri D. Wongworawat MD for their guidance, which determined the Journal's editorial approach as described at the end of this essay.

References

1. Belmont PJ, Jr., Goodman GP, Kusnezov NA, Magee C, Bader JO, Waterman BR, Schoenfeld AJ. Postoperative myocardial infarction and cardiac arrest following primary total knee and hip arthroplasty: Rates, risk factors, and time of occurrence. *J Bone Joint Surg Am.* 2014;96:2025–2031.
2. Bohl DD, Basques BA, Golinvaux NS, Baumgaertner MR, Grauer JN. Nationwide Inpatient Sample and National Surgical Quality Improvement Program give different results in hip fracture studies. *Clin Orthop Relat Res.* 2014;472:1672–1680.
3. Bohl DD, Grauer JN, Leopold SS. Editor's spotlight/Take 5: Nationwide inpatient sample and national surgical quality improvement program give different results in hip fracture studies. *Clin Orthop Relat Res.* 2014;472:1667–1671.
4. Bohl DD, Russo GS, Basques BA, Golinvaux NS, Fu MC, Long WD, 3rd, Grauer JN. Variations in data collection methods between national databases affect study results: a comparison of the nationwide inpatient sample and national surgical quality improvement program databases for lumbar spine fusion procedures. *J Bone Joint Surg Am.* 2014;96:e193.
5. Bozic KJ, Kurtz SM, Lau E, Ong K, Chiu V, Vail TP, Rubash HE, Berry DJ. The epidemiology of revision total knee arthroplasty in the United States. *Clin Orthop Relat Res.* 2010;468:45–51.
6. Centre for Evidence Based Medicine. The 2011 Oxford CEBM levels of evidence: Introductory document. Available at: <http://www.cebm.net/2011-oxford-cebm-levels-evidence-introductory-document/>. Accessed Feb 16, 2015.
7. Forte ML, Virnig BA, Swiontkowski MF, Bhandari M, Feldman R, Eberly LE, Kane RL. Ninety-day mortality after intertrochanteric hip fracture: does provider volume matter? *J Bone Joint Surg Am.* 2010;92:799–806.
8. Golinvaux NS, Bohl DD, Basques BA, Fu MC, Gardner EC, Grauer JN. Limitations of administrative databases in spine research: A study in obesity. *Spine J.* 2014;14:2923–2928.
9. Golinvaux NS, Bohl DD, Basques BA, Grauer JN. Administrative database concerns: accuracy of international classification of diseases, ninth revision coding is poor for preoperative anemia in patients undergoing spinal fusion. *Spine (Phila Pa 1976).* 2014;39:2019–2023.
10. Greene ME, Rolfson O, Nemes S, Gordon M, Malchau H, Garellick G. Education attainment is associated with patient-reported outcomes: Findings from the Swedish Hip Arthroplasty Register. *Clin Orthop Relat Res.* 2014;472:1868–1876.
11. Jansen E, Eskelinen A, Peltola M, Makela K. High early failure rate after cementless hip replacement in the octogenarian. *Clin Orthop Relat Res.* 2014;472:2779–2789.
12. OCEBM Levels of Evidence Working Group. The Oxford 2011 levels of evidence. Available at: <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>. Accessed February 16, 2015.
13. Troelsen A, Malchau H. Editor's spotlight/Take 5: A review of current fixation usage and registry outcomes in total hip arthroplasty: The uncemented paradox. *Clin Orthop Relat Res.* 2013;471:2047–2051.