See Article page 2232.

Commentary: Goldilocks and the three bears

Sergio A. Carrillo, MD, FACS, FACC

Although designed to garner information on coding and billing, administrative databases are useful in clinical research as they are easily accessible, inexpensive, and, more importantly, contain large groups of patients, thus generating enough sample size and power for analysis.¹ Administrative data optimize reimbursement and can be hamstrung by the absence of critical clinical variables, complications and comorbidities coding accuracy, as well as completeness. However, their use in thoracic surgery,² in particular, congenital heart surgery,³ is not uncommon. In contradistinction to administrative databases, clinical databases contain detailed disease-specific information and are intended primarily for outcomes analysis but may lack data concerning other diagnoses or procedures, resource use, use of medications, laboratory and imaging studies, hospital charges, and often do not contain longitudinal follow-up data. In essence, each is more relevant than the other depending on the research focus, as the 2 contain complimentary outcome data.

Important (and relevant) clinical questions, require accurate—often specific—clinical information and administrative databases are thus, suboptimal. With that in mind, Zafar and coworkers⁴ in this issue of the *Journal* collaborated to merge individual efforts to develop their collective *pièce de résistance* statistical code to link and crosswalk the Pediatric Health Information System (PHIS) and the Society of Thoracic Surgeons–Congenital Heart Surgery (STS-CHS) databases. They did this by using the *International Classification of Disease, 10th Revision* diagnosis and procedure codes to identify the 10 congenital heart surgery benchmark procedures as defined by the STS. With their novel algorithm, the authors were able to capture all

Sergio A. Carrillo, MD, FACS, FACC

CENTRAL MESSAGE

The use of administrative and clinical databases in congenital heart surgery is common. An algorithm linking the two would provide more comprehensive information than either of them in isolation.

10 benchmark operations with a >91% sensitivity and >98% specificity using the PHIS database. Sensitivity was the lowest for off-bypass coarctation of aorta repair. Furthermore, in their schema crosswalk, inpatient mortality was similar (and not statistically different) when the algorithm was applied to both databases.

In the current climate of transparency and accountability, it becomes incumbent to use pertinent and comprehensive datasets as foundation to outcomes research. This novel algorithm seems to have “just the right amount of code”—fulfilling the Goldilocks principle, which is commonly used by a range of disciplines such as psychology, astrobiology, medicine, and statistics—to maximize the extraction of relevant data points from administrative database that perform satisfactorily as seen by reaching the same conclusions when the clinical database was used. This in turn, allows the resultant data to be appropriately analyzed and presented in the context that is intended and preempt the tendency to overreach the use of administrative data for outcomes or marketing research, especially by those around the fringes of health care.⁵

Notwithstanding the observed advantages seen by authors when applying their proprietary but publicly available model to administrative databases, administrative and clinical database linking has been the focus of research by other investigators.^{6,7} The study by Pasquali and colleagues⁶ was the first to described linking methodology between the PHIS and the STS-CHS databases. In that landmark study, indirect identifiers such as date of birth, date of admission,

From Department of Cardiothoracic Surgery, The Heart Center, Nationwide Children's Hospital and the Division of Cardiac Surgery, Department of Surgery, The Ohio State University, Columbus, Ohio.

Disclosures: The author reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Received for publication Oct 25, 2021; revisions received Oct 25, 2021; accepted for publication Oct 25, 2021; available ahead of print Oct 29, 2021.

Address for reprints: Sergio A. Carrillo, MD, FACS, FACC, Department of Cardiothoracic Surgery, The Heart Center, 700 Children's Dr, Columbus, OH 43205 (E-mail: sergio.carrillo@nationwidechildrens.org).

J Thorac Cardiovasc Surg 2022;163:2240-1

0022-5223/\$36.00

Copyright © 2021 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2021.10.039>

date of discharge, sex, and center were used to link the 2 databases, whereby a proportion of 87.4% of exact matches was seen with the lowest unmatched cohort accounting in the lower surgical risk categories.

Experienced researchers can, with careful design and sophisticated statistical methodology, develop models based on administrative data for certain conditions that may be appropriate for profiling. Zafar and colleagues⁴ have gone one better with their linking methodology, which could allow researchers to capitalize on the strengths of both databases and expand the pool of data available for analysis. Furthermore, this novel algorithm may provide metrics of performance that are not accurately analyzed and/or presented by either of the databases in isolation. However, it must be tested and validated in a multi-institutional fashion and expanded to identify the vast majority, if not all, of the cardiac procedures within congenital heart surgery. Otherwise, this algorithm could end up like Goldilocks, in the classic children's book,⁸ as she runs into the forest, never to be seen again.

References

1. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med.* 1997; 127:1518-27.
2. Prasad A, Helder MR, Brown DA, Schaff HV. Understanding differences in administrative and audited patient data in cardiac surgery: comparison of the University HealthSystem Consortium and Society of Thoracic Surgeons databases. *J Am Coll Surg.* 2016;223:551-8.
3. Jenkins KJ, Gauvreau K, Newburger JW, Spray TL, Moller JH, Iezzoni LI. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg.* 2002;123:110-8.
4. Zafar F, Allen P, Bryant R, Tweddell JS, Najm HK, Anderson BR, et al. A mapping algorithm for International Classification of Diseases 10th revision codes for congenital heart surgery benchmark procedures. *J Thorac Cardiovasc Surg.* 2022;163:2232-9.
5. Torchiana DF, Meyer GS. Use of administrative data for clinical quality measurement. *J Thorac Cardiovasc Surg.* 2005;129:1223-5.
6. Pasquali SK, Jacobs JP, Shook GJ, O'Brien SM, Hall M, Jacobs ML, et al. Linking clinical registry data with administrative data using indirect identifiers: implementation and validation in the Congenital Heart Surgery population. *Am Heart J.* 2010;160:1099-104.
7. Hammill BG, Hernandez AF, Peterson ED, Fonarow GC, Schulman KA, Curtis LH. Linking inpatient clinical registry data to Medicare claims using indirect identifiers. *Am Heart J.* 2009;157:995-1000.
8. Southey R. *The Story of the Three Bears*. London: Longman, Rees; 1837.

See Article page 2232.

Commentary: What's in a name? The effect of misclassification in administrative data in the setting of congenital heart disease surgery

Osman O. Al-Radi, MD

When William Shakespeare wrote "What's in a name? That which we call a rose by any other name would smell as sweet,"¹ I believe he intended to say that names in of themselves don't have meaning or add value to the subject being named. However, I would contend that in our field of surgery on congenital heart disease (CHD), there is value in a name if it is self-descriptive and universally accepted.

From the Faculty of Medicine, Department of Surgery, King Abdulaziz University, Jeddah, Saudi Arabia.

Disclosures: The author reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Received for publication Nov 4, 2021; revisions received Nov 4, 2021; accepted for publication Nov 5, 2021; available ahead of print Nov 11, 2021.

Address for reprints: Osman O. Al-Radi, MD, Faculty of Medicine, Department of Surgery, King Abdulaziz University, PO Box 80215, Building 10, Room 1027, Jeddah 21589, Saudi Arabia (E-mail: oradi@kau.edu.sa).

J Thorac Cardiovasc Surg 2022;163:2241-2

0022-5223/\$36.00

Copyright © 2021 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2021.11.007>

Osman O. Al-Radi, MD

CENTRAL MESSAGE

Misclassification, one thorn in the side of administrative database research.

When it comes to disease nomenclature and procedure classifications in CHD, administrative data sources and clinical data sources use different nomenclature. Diagnoses and diagnostic complexes are not only classified using different terms, but the terms themselves are often not directly translatable and may differ in text, number, and