Commentary: In search of a data utopia

Felix G. Fernandez, MD, MSc

Subramanian and colleagues\(^1\) share their considerable expertise on databases used for observational research in this edition of the *Journal*. Databases, both clinical and administrative, have been appropriately used in many well-designed studies to advance the field of cardiothoracic surgery. These studies have benefited our patients with respect to improvements in quality of care and outcomes.\(^2\)

Equally as prevalent, however, has been the inappropriate use of such data sources to test hypothesis for which they are not well suited, or worse, for data-mining exercises that are not hypothesis-driven at all. The result is that our literature has been populated with studies that are riddled with unmeasured and often-unrecognized confounding that fail to inform our practice.

The authors do an exemplary job of illustrating the strengths, limitations, and appropriate uses of both administrative and clinical databases. By design, administrative data are collected for billing purposes, whereas clinical databases use observational study methods to collect uniform data on a defined population for a prespecified purpose. Administrative datasets are well suited to examine health care use and associated costs, in assessing the impact of health care policy implementation, and for longitudinal analyses. The granular detail in clinical databases, in contrast, can be leveraged for risk prediction and comparative effectiveness analyses. This requires that the informed researcher carefully consider their study question and then select the most appropriate dataset with which to conduct that particular study.

Frustrating for the outcomes researcher, however, is the lack of a comprehensive single data source that can be used to examine both quality and cost outcomes in cardiothoracic surgery. An ideal database, or data utopia, would belong to a single-payer system under which all patients are cared for and followed. Penetrance would be complete and the data broadly generalizable. Clinical and cost/resource use data would be available on all patients. Further longitudinal events, such as survival and readmissions or reinterventions, and quality of life data would be tracked. Clearly, such data do not exist.

The Society of Thoracic Surgeons (STS) National Database, however, continues to evolve to develop the attributes of an ideal data source. The STS Adult Cardiac Surgery Database and Congenital Heart Surgery Database already have greater than 95% penetrance in the United States. The Adult Cardiac Surgery Database, as well as the General Thoracic Surgery Database, has been linked to Centers for Medicare and Medicaid Services Medicare data for information on long-term survival and other cost and resource use outcomes.\(^3,4\) This linkage creates a comprehensive clinical–financial database for those patients who are Medicare beneficiaries. Another ongoing initiative is focused on linking the entire STS National Database, more than 8 million records, to the National Death Index maintained by the Centers for Disease Control. Linkage to the National Death Index, as well as ongoing
internal linkage of STS records, will provide longitudinal survival and reoperation data on all patients in the database. In cardiothoracic surgery, the search for a data utopia remains a challenging but worthy pursuit.

References


