EDITORIAL SERIES ON METHODOLOGY I

Perioperative database research: possibilities and pitfalls

P. M. Fleischut, M. Mazumdar and S. G. Memtsoudis

1 Department of Anesthesiology, New York-Presbyterian Hospital, Weill Medical College of Cornell University, 525 East 68th Street, M-308, New York, NY 10065, USA
2 Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University, 402 East 67th Street, New York, NY 10065, USA
3 Department of Anesthesiology, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA
* Corresponding author: Department of Anesthesiology, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA. E-mail: memtsoudiss@hss.edu

The utilization of large databases for perioperative and anaesthesia-related research has gained increasing interest among the anaesthesiology community over the last few years. Advances in information technology, research methodology, and the realization that large data collection constructs, even if designed primarily for administrative purposes, could provide valuable clinical information, have been fueling this trend. However, despite these trends, the vast majority of anaesthesiologists remain unfamiliar with the value and pitfalls associated with database research.

The goal of this editorial is therefore to provide an overview of this line of research, and its utility for various types of investigations, especially translational science research. As an introduction to the topic, we aim to point out some of the major advantages and limitations associated with large database research.

A variety of large databases amenable to medical research are available. Many are accessible to the public free of charge (e.g. the National Hospital Discharge Database), others are only commercially available (e.g. Premier Inc.) or are restricted in their accessibility (e.g. Veterans Affairs Administrative Database). Most require formal applications before use and in order to protect against misuse, many contain de-identified data, and require the completion of a data use agreement. Because of the lack of identifiable information in most non-institutional databases, work with these data constructs is often excused from review by research ethics boards, however exemptions should still be sought before work commences.

The choice of a specific database for research will depend to a large extent on the question to be answered and the availability of relevant information within the data collection construct. In this context, the number of variables collected is usually limited and varies in completeness and accuracy. Further, a clear understanding of the primary purpose of the database (administrative or clinical), the number and the type of contributing sites, and also the database design, and the data collection and verification process are some of many factors to be considered when deciding to work with a particular dataset. For example, researchers need to consider if they need nationally representative data (usually provided by specially designed weighting procedures) collected from a large number of sites or if institutional data from select hospitals are sufficient. An important consideration before deciding to proceed with this type of research is the availability of an adequate infrastructure that includes programmers, statisticians, epidemiologists, and methodologists, and also equipment suitable to handle large datasets.

The attractiveness of large databases for medical research lies largely in the ability to perform many types of studies with the same dataset, including but not limited to descriptive, epidemiologic and trend analytical investigations, various types of outcomes, and comparative effectiveness research, and also risk factor evaluations. The availability of information from large numbers of patient or case entries allows for the study of rare events of otherwise difficult to study outcomes in subpopulations of patients. As data are not subject to the strict inclusion and exclusion criteria of randomized control trials, they represent an overview of actual practice among the participating institutions, thus affording results a high level of external validity. The characteristics mentioned above have therefore led to publications on topics that have previously eluded study using traditional designs, such as randomized control trials. In this context, the comparison of outcomes of various types of anaesthetics on perioperative complications did not seem feasible because of the rarity of specific adverse events that require extremely large sample sizes. Testimony of the value of large databases was provided by a number of recently published studies ranging from the study of the epidemiology of malignant hyperthermia, to the study of risk factors for cerebrovascular accidents, to perioperative outcomes after joint arthroplasty.

The value of large databases can also be seen in the field of translational science. Specifically, hypotheses gained from basic science research and proof of concept clinical trials can be put to the test by using well-designed database methodologies. An example represents our recently finished series of studies evaluating the differential perioperative outcomes of patients undergoing bilateral vs unilateral total knee

© The Author [2013]. Published by Oxford University Press on behalf of the British Journal of Anaesthesia. All rights reserved. For Permissions, please email: journals.permissions@oup.com
arthroplasty. Clinical and basic science data suggested a dose dependent detrimental effect of pulmonary bone and cement embolization on lung catabolism and haemodynamics.\textsuperscript{10,11} However, despite the availability of small institutional studies on perioperative outcomes, the evidence on differential outcomes on a population-based level remained inconclusive. Utilizing data collected from hundreds of institutions on over half a million patients, we were able to show that bilateral total knee arthroplasties performed at the same time were indeed associated with increased risk for morbidity and mortality compared with unilateral procedures despite being performed in younger and healthier patients.\textsuperscript{9} An important aspect of this study was the ability to provide nationally representative data, thus taking into account various practice settings. A number of follow-up studies into the mechanisms and outcomes of these procedures in subpopulations of patients eventually led to the ability to translate our findings into evidence based guidelines for the performance of bilateral knee arthroplasty.\textsuperscript{12,13} Clearly, the ability to use this methodology provided the tools necessary, as randomized control trials would not have been feasible. Other examples of the use of databases to evaluate basic science principles in a clinical setting are available.\textsuperscript{14}

Despite the advantages and recent enthusiasm associated with database research, investigators need to consider a number of limitations.\textsuperscript{15} Some of the largest databases were created for administrative purposes with limited data fields in order to maintain feasibility and focus. Thus, detailed clinical information, including documentation of intraoperative events, is often missing. In this context, comorbidities and complications are usually based on the ICD-9-coding system with considerable risk for coding bias and the inability to discern the severity of diseases and events, and also information on whether the condition was present on admission. The latter has recently been addressed by a number of database administrators through inclusion of a ‘present on admission’ variable. The former may be at least partially addressed by a more detailed ICD-10-coding system to be adopted in the near future.

Further, many databases (e.g. National Inpatient Sample and National Hospital Discharge Survey) do not allow for the longitudinal analysis of patient outcomes as one cannot determine separate admissions for the same individual, either because no patient identifiers are available or because the number of participating institutions are limited, thus opening up the possibility that additional hospitalizations in a facility outside the hospital universe may not be captured. For the same reason, events after hospital discharge may also remain elusive as frequently only in-hospital events are captured. In this context, efforts to create databases with longitudinal data analysis capability are underway.\textsuperscript{16,17}

A particular limitation for our field is the fact that anaesthesia-related information is frequently not readily available and relies on the use of billing codes where available. Until recently, The American College of Surgeons’ National Surgical Quality Improvement Program’s database provided one of the most commonly utilized sources of information for perioperative outcomes research.\textsuperscript{18} However, with advances and expansion of anaesthesia information management and billing systems, recent initiatives, headed by national and institutional entities such as the American Society of Anesthesiology, sponsored Anesthesia Quality Institute, and the Multicenter Perioperative Outcomes Group, have created further databases specifically addressing the needs for information critical to our specialty.\textsuperscript{19,20} However, as is true for other databases, the completeness of information provided by participating institutions for various categories is variable, thus posing some limitations to the researchers’ ability to perform desired analyses. Thus, although of tremendous promise, the value of the above mentioned sources remains to be seen and the verification of data integrity before use remains vital.

Irrespective of the amount and type of data fields collected, the ability to determine causality is not possible with data analysed using large databases. Although large databases have been used for many decades for medical research, the increasing size and complexity of these constructs demand continued advances in computer software and statistical methodologies to allow for appropriate manipulation and interpretation of data.\textsuperscript{21} Unbeknown to most clinical and laboratory researchers, the field of biostatistics has expanded dramatically over the last few decades in order to provide cutting edge means to keep up with the growing needs for methodological tools. Collaboration with biostatisticians and epidemiologists is an essential component of perioperative database analysis in order to achieve a rigorous scientific process.\textsuperscript{22}

In summary, large database analyses as research tools are gaining increasing attention in the arena of perioperative medicine, providing possibilities for research encompassing trend analyses, comparative effectiveness research, epidemiologic evaluations, risk factor analysis, and outcomes research, especially of rare events and patient subpopulations. Examples of the use of databases for translational research are available and are increasing in number. It must be cautioned that despite access to large numbers and associated power, databases are burdened by limitations, which include the restricted amount of variables available and the inability to draw conclusions with respect to cause and effect. To address the lack of anaesthesia-related information in traditionally available databases, recent initiatives have taken place to fill this void, but quality and utility of these data sources remain unknown to date. Despite promising possibilities, researchers need to be aware of the pitfalls of large database research and the need for ongoing methodology development in order to address the important questions posed by our specialty in a rigorous and robust scientific way.

**Declaration of interest**

None declared.
References


