Why everyone should care about “Computable Phenotypes”

Robert C. Tasker, MBBS, MD, FRCP
Departments of Neurology and Anesthesia (Paediatrics), Harvard Medical School, Boston
Children’s Hospital, Boston, MA

Keywords
traumatic brain injury; comparative effectiveness; outcome; pediatric

In this issue of the Journal, Bennett et al.(1) report a prospective validation of administrative, billing and trauma registry coding data carried out with the purpose of generating a “computable phenotype” for most neurosurgical and critical care events in children with traumatic brain injury (TBI). This paper is important and potentially transformative for the whole field of pediatric critical care. If you don’t know why it is important, then please read and consider it thoughtfully because this article is the future.

Everyday technologically-modern healthcare systems produce huge amounts of electronically stored data. This resource amounts to information, on average, for about 100 new pediatric TBI hospitalizations in the United States (US), every day.(2) There is the possibility of using such data, within these systems, to advance patient care and improve outcomes.(3) We may also aspire to harnessing these so-called “big data” in real-time.(4) In principle, identifying and translating new knowledge about best treatment by “nesting” high-quality randomized controlled trials within the electronic health record (EHR). There is, however, a major problem that makes this aim aspirational for the time being: the EHR and administrative data that are generated from it were not designed as instruments for clinical research. They have different purposes.

Medical coding professionals transform healthcare diagnoses and procedures described within the EHR into universal medical alphanumeric codes (https://www.aapc.com/certification/medical-coding-certification.aspx). These diagnoses and procedures are taken from physician notes, laboratory and radiology results, etc., and should tell the whole story of the patient’s encounter with the physician, as well as being as specific as possible. In the US, since October 2015, the Center for Medicare and Medicaid Services has required that administrative coding use the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM), instead of version 9 (ICD-9-CM). Hospitals in the US – and, for our purposes, the pediatric intensive care unit (PICU) – rely on these codes for case-mix acuity indices, assessment of medical necessity for procedures, services and admissions, and for reporting of disease to public health departments. Medical billing professionals also use codes, but their task is to process claims sent to health insurance companies for reimbursement of services provided. In some hospitals the medical coder and medical biller may be the same person. Both of these administrative systems use guidelines and standards but their aims, however, differ: one is primarily focused on description of the
patient in compliance with government or other regulators; the other is about the requirements of individual payers. The American Academy of Professional Coders has more than 150,000 members. So, what has coding got to do with research?

A code enables access into the EHR; it is a flag in the administrative database and the way of identifying cases for large observational studies or comparative effectiveness research (CER). There are concerns about the potential for bias and lack of validity when key exposures or outcomes in a research question depends on the accuracy of diagnostic or procedural codes (5–8) – even going so far as to equate any such research as “garbage in, garbage out” (9), and being only capable of generating “red herrings, false alarms and pseudo-…” (10).

Bennett et al. (1) introduce us to the idea of “computable phenotypes”. That is, the ability to first define a condition, disease, patient characteristic or clinical event using only data processed by computer (11). Second, to use a model or algorithm to identify a population of patients with a condition of interest. Consider, as an illustration, our current state of knowledge about the decision to use decompressive craniectomy for refractory intracranial hypertension in pediatric severe TBI: we do not have any Class II evidence to guide management (12). This intervention is undoubtedly being used, but how could we harness information about its use in an EHR-based CER study? Certainly not at the moment, but Bennett et al. (1) provide us with some of the tools. For example, they have used ICD-9-CM and other administrative database resources to develop and test an algorithm/model that would accurately identify individuals with TBI in the PICU who underwent intracranial pressure monitoring or craniotomy/craniectomy. What is evident in both of these examples is that a single ICD-9-CM code was not adequate for purpose. Rather, Bennett et al. (1) found that a combination of coding and administrative data sources was required. The fact that no single data source or code was 100% specific and 100% sensitive reflects the prior discussion about the primary purpose of coding, and who does it.

At the moment, Bennett et al. (1) may have tested their models in a number of patients, but our focus should be on the equally important issue of the number of centers used with their respective medical coding and billing administrations (i.e., raters) that generated the data for the sources examined. It was only two, and if multiple hospitals and PICUs are to be used in a future CER study, then each additional center will need to demonstrate consistency of phenotype definition – similar to an operating procedure for “case definition” or inclusions/exclusions of current research network studies. In this context, a recently published desiderata (Latin: “desired things”) for computable representations of EHR-driven phenotype algorithms, funded by the National Institute of General Medical Sciences, outlined ten recommendations (13). Central to all of these desires is the vision for shared phenotype definitions between research and healthcare activities in a structured framework (14). Without coordination there is potential for unintended proliferation of computable phenotypes, for similar conditions or clinical profiles.

In my view, the report by Bennett et al. (1) is transformative for our field because the approach taken in phenotyping, using PICU admissions with TBI as an example, has applicability to other populations such as those with sepsis or acute lung injury. Our field
will need to develop more phenotype definitions that conform to research standards, e.g., in TBI we have the common data elements(15). We will also need to collaborate and contribute to multidisciplinary phenotype libraries such as the Phenotype Knowledge Base (https://phekb.org) and the phenotype working groups of the US National Patient-Centered Clinical Research Network (http://www.pcornet.org). And, of course, we will need to update to ICD-10-CM. That said, the rewards of sharing across healthcare delivery and clinical research structures are potentially huge.

Acknowledgments

Copyright form disclosure: Dr. Tasker has received support for travel when he attended the Guidelines Specialist Panel committee for the Brain Trauma Foundation; he also receives support from the National Institutes for Health (U01 NS081041).

References