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

References

1. Bennett TD, DeWitt PE, Dixon RR, et al. Development and prospective validation of tools to accurately identify neurosurgical and critical care events in children with traumatic brain injury. Pediatr Crit Care Med. 2017; XX:YYY–ZZZ.

2. Faul, M., Likang, X., Wald, M., et al. Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths 2002–2006. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010. (https:// www.cdc.gov/traumaticbraininjury/pdf/blue_book.pdf). Accessed January 9, 2017

3. The Learning Health Care System in America. Washington, DC: Institute of Medicine; 2012. (http:// www.iom.edu/Activities/Quality/LearningHealthCare.aspx). Accessed January 9, 2017

4. Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? JAMA. 2015; 314:767–768. [PubMed: 26305643]

5. Iezzoni LI. Assessing quality using administrative data. Ann Intern Med. 1997; 127:666–670. [PubMed: 9382378]

6. Mohammed MA, Stevens A. The value of administrative databases. BMJ. 2007; 334:1014–1015. [PubMed: 17510106]

7. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. J Clin Epidemiol. 2011; 64:1054–1059. [PubMed: 21474278]

8. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol. 2012; 65:126–131. [PubMed: 22075111]

9. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. Obstet Gynecol. 2010; 116:1018–1019. [PubMed: 20966682]

10. Grimes DA. Epidemiologic research with administrative databases: red herrings, false alarms and pseudo-epidemics. Human Reprod. 2015; 30:1749–1752.

11. Richesson, RL., Smerek, M., editors. NIH Health Care Systems Research Collaboratory. Electronic health records-based phenotyping. Rethinking Clinical Trials: a living textbook of pragmatic clinical trials. (https://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/). Accessed January 9, 2017

12. Kochanek PM, Carney N, Adelson PD, et al. Guidelines for the acute medical management of severe traumatic brain injury in infants, children, and adolescents—second edition. Pediatr Crit Care Med. 2012; 13(Suppl 1):S1–S82. [PubMed: 22217782]

13. Mo H, Thompson WK, Rasmussen LV, et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. J Am Med Inform Assoc. 2015; 22:1220–1230. [PubMed: 26342218]

14. Richesson R, Smerek M, Cameron CB. A framework to support the sharing and re-use of computable phenotype definitions across health care delivery and clinical research applications. EGEMS. 2016; 4:1232. [PubMed: 27563686]

15. Miller AC, Odenkirchen J, Duhaime AC, et al. Common data elements for research on traumatic brain injury: pediatric considerations. J Neurotrauma. 2012; 29:634–638. [PubMed: 22091862]