PERSONAL VIEW

Big data in medical education – Are we ready?

Heng-Wai Yuen[1,2,3] & Abhilash Balakrishnan[2,3,4]

[1]Department of Otolaryngology-Head & Neck Surgery, Changi General Hospital, Singapore; [2]Duke-NUS Medical School, Singapore; [3]National University of Singapore, Yong Loo Lin School of Medicine, Singapore; [4]Department of Otolaryngology, Singapore General Hospital, Singapore

I. INTRODUCTION

Big data (BD) involves aggregating and melding large and heterogeneous datasets, allowing searches and cross-referencing, and deriving insights and meaning from them. It has tremendous potential for application in medical education (ME) where the massive amounts of data that are generated and collected about learners, their learning, and the organisation of their learning can be analysed and interpreted to provide meaning and insights into various aspects of ME. This article briefly introduces BD, potential areas of application, and highlights the pitfalls and challenges surrounding the use of BD in ME (BDME) from the authors' perspectives.

II. BIG DATA IN MEDICAL EDUCATION (BDME)

The concept of BD has its origins in commercial industries, and also academic and technical disciplines (e.g., astronomy and genomics) where enormous amounts of complex data and information are routinely collected, managed and analysed (Ellaway, Pusic, Galbraith, & Cameron, 2014; Schneeweiss, 2014). This information possesses characteristics denoted by the four Vs: high Volume, Variety, Velocity, and Veracity (validity); conventional database software tools are unable to fully capture, store, process, or analyse them (Ellaway et al., 2014). BD is relatively new in clinical medicine and applying BDME has been slow and limited (Cook, Andriole, Durning, Roberts, & Triola, 2010; Ellaway et al., 2014) Nonetheless, in the last few years, there are increased efforts to apply BD to ME (Chahine et al., 2018; Ellaway et al., 2014). To this end, ME is well suited for BD application as a massive volume of complex data is generated and collected constantly from different programs and educational institutions, and from

multiple sources, both structured and unstructured: e.g., electronic medical records, assessment results and test scores, evaluation and feedback information, as well as curriculum and program evaluation (Chahine, et al., 2018; Cook et al., 2010). By harnessing the power of BDME, information and data can be aggregated, integrated, and analysed, then interpreted and acted on if necessary (Ellaway et al., 2014; Schneeweiss, 2014).

III. POTENTIAL APPLICATIONS OF BDME

The potential of BDME includes both practical (e.g., program and curriculum assessment and evaluation) and research applications. Depending on the purpose and/or research question, the data mining may be on a broad, systems-level or a personalised small-scale basis. BDME application organises and crystallises the data to enable a better understanding of and insight into what happened, and what is currently happening. This may occur through various different ways of analyses including prospective longitudinal analysis, trend discovery, pattern recognition and predictive analytics. Hence, predictions or extrapolations might be made in regards to what may yet happen in curriculum, programs and educational practices (Chahine et al., 2018; Cook et al., 2010; Ellaway et al., 2014).

For instance, BDME can facilitate decision-making in undergraduate ME, e.g., entry selection of medical students, or readiness of a medical student to graduate. In postgraduate ME, BDME can provide insights into data on learners' experience and exposure, feedback information, as well as assessment data within and across programs (Chahine et al., 2018; Ellaway et al., 2014). This allows personalised feedback and individualised

learning plans (Chahine et al., 2018), and facilitates the implementation of entrustable professional activities (EPA). Learning gaps and teaching lapses can also be identified to support improvement or changes to certain practices or contents. Applying BDME on these educational and other data (such as demographics, admission criteria or educational practices) in a longitudinal and cross-sectional manner allows benchmarking and accountability across different cohorts, programs, and institutions. This is vital for continuous quality assurance and improvement of ME practices (Chahine et al., 2018; Cook et al., 2010; Ellaway et al., 2014), or for evaluation of upstream policies (Chahine et al., 2018; Schneeweiss, 2014). These same processes can also be performed across countries to inform ME from international or cross-cultural perspectives.

Another potential application of BDME is to investigate the (hitherto assumed) link between ME and patient care. Drawing on combined data from educational and clinical information repositories (e.g., correlating patient outcomes from hospital and clinic health information systems with different models of educations within and across institutions), one would be able to evaluate if, and to what extent, educational practices translate into improved health care outcomes for patient and society (Chahine et al., 2018). One example is the Jefferson Longitudinal Study of Medical Education (Callahan, Hojat, Veloski, Erdmann, & Gonnella, 2010) whereby data on 8000 students who were tracked over 40 years showed that MCAT examination performance is a valid predictor of medical school and residency performance. This and other studies confirmed the feasibility and utility of applying BD to inform current medical educational practices, and to bridge the gap between pedagogical theory and practice. Further, by enabling a longitudinal view of physicians' progression and development through their education, and the career choices made, BDME can provide information and evidence to facilitate recommendations for important strategic policies and decisions, e.g., manpower planning or speciality development. These are subjects of interest for policy-makers, regulatory authorities, medical educators and researchers.

IV. POTENTIAL OBSTACLES AND PITFALLS

Whilst there are many potential fruitful applications of BDME, some challenges and issues must be critically addressed before the widespread adoption of BD into mainstream ME practice.

Data fragmentation, so common in healthcare systems, is a major obstacle to the widespread use of BDME (Ellaway et al., 2014; Schneeweiss, 2014). For a start,

electronic health or medical records (EMR) are frequently incompatible and heterogeneous across hospital systems that store the data (Chahine et al., 2018; Ellaway et al., 2014). Practice standards and vocabulary are also not standardised. Also, healthcare systems are not required (or willing) to exchange and share data with each other. In addition, organisational policies regarding security and confidentiality limit data accessibility (Chahine et al., 2018; Ellaway et al., 2014). Further, there are ethical and medicolegal considerations. For instance, most of the patient data captured on EMR was not originally intended for education purposes, and does not include informed consent in this respect. Even if the data can be anonymised with identifiers removed, questions remain on what data is collected, how the data is stored and protected, how it is used and shared – by whom, and with whom. These issues extend to ME data too; confidentiality issues and access restrictions to data collected on learners, programs and institutions can limit the quality, analysis and value of BDME.

Hence, government and health authorities, EMR companies, hospitals and training institutions must cooperate to improve medical data and information systems, and strengthen data exchange and integration across organisations (Chahine et al., 2018; Cook et al., 2010; Ellaway et al., 2014). Appropriate legislations or policies may be necessary. Investments in infrastructure, technologies and expertise to manage and protect data from different sources are also needed. The infrastructure and technological expertise (for collection, storage, processing and analysis) could be centralised in a 'data warehouse' – different institutions become data providers to this 'central' BD collective (Cook et al., 2010; Ellaway et al., 2014). It is likely that external partners (e.g., data science, informatics) will be involved to facilitate and optimise the use of BD. Under these circumstances, the governance, ownership of, and access to data are important issues to consider.

In using BDME to correlate training and clinical care outcomes, the challenge is being able to accurately link a learner's (or a cohort of learners') education and training with patient-level or system-level clinical outcomes (Chahine et al., 2018; Cook et al., 2010). Given that multiple healthcare providers (students, residents, practicing physicians) may be involved in the care of a particular patient, innovative data analytical algorithms or techniques will be necessary in order to identify or 'tag' different aspects of clinical care or patient encounters, and accurately attribute these to specific providers over prolonged periods of time, and across institutional, clinical and educational boundaries (Chahine et al., 2018; Cook et al., 2010). If successful,

this will provide unprecedented potential for performance assessment and evaluation.

The application of BDME also has inherent limitations and fallibility (Chahine et al., 2018; Ellaway et al., 2014). The interpretations and conclusions (and the subsequent decisions and actions) based on BDME must be made with extreme caution. The standards and rigours of academic and scientific research must be applied and met – in the collection methods, precision, representativeness of data. There is intrinsic bias in BD due to the fact that information that cannot (or simply are not) be captured may be undervalued or ignored. Predicting trends and judging current and future potential and success of individuals or programs must similarly be tempered with caution (Chahine et al., 2018; Cook et al., 2010). Major decisions (especially summative) must be based on time-honoured, empirically proven principles: multiple data points, from multiple sources (triangulation), at different time points (reiterative), and after considering the dynamic nature of learning and education in reality.

There are real risks to the individuals and systems if BDME is used out of context, or for unintended purposes. For instance, should BDME be used to alter a learner's (or a group of learners') career path or choice? Should we judge learners based on 'normal' patterns of learner behaviour derived from BDME? Also, from the faculty's perspective, it is tempting to use only those educational interventions that were 'shown to work' by BDME, at the expense of all others.

This article is not intended to propose solutions to the many issues surrounding the use of BDME. The permeation of BD into ME appears inexorable. It is time for the ME community to take the lead to critically appraise and shape the conversation surrounding BDME, so as to set the agenda and direction for the best use of BDME.

Notes on Contributors

Heng-Wai Yuen is an adjunct Associate Professor with the Duke-NUS Medical School and the Singapore University of Technology and Design (SUTD). He is the Director of Otology and Hearing Implants in the Department of Otolaryngology-Head & Neck Surgery, and the Deputy Director of Undergraduate Medical Education at Changi General Hospital.

Abhilash Balakrishnan is an adjunct Associate Professor with the Duke-NUS Medical School and Clinical Associate Professor with the Yong Loo Lin School of Medicine at the National University of Singapore. He is also the Deputy Head of Department (Education) in the Department of Otolaryngology at Singapore General Hospital.

Funding

Declaration of Interest

The authors declare no conflict of interest.

References

Callahan, C. A., Hojat, M., Veloski, J., Erdmann, J. B., & Gonnella, J. S. (2010). The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: A longitudinal study of 36 classes of Jefferson Medical College. *Academic Medicine*, *85*(6), 980-987. https://doi.org/10.1097/ACM.0b013e3181cece3d

Chahine, S., Kulasegaram, K., Wright, S., Monteiro, S., Grierson, L. E., Barber, C., … Touchie, C. (2018). A call to investigate the relationship between education and health outcomes using big data. *Academic Medicine*, *93*(6), 829-832. https://doi.org/10.1097/ACM.0000000000002217

Cook, D. A., Andriole, D. A., Durning, S. J., Roberts, N. K., & Triola, M. M. (2010). Longitudinal research databases in medical education: Facilitating the study of educational outcomes over time and across institutions. *Academic Medicine*, *85*(8), 1340-1346. https://doi.org/10.1097/ACM.0b013e3181e5c050

Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, *36*(3), 216-222. https://doi.org/10.3109/0142159X.2014.874553

Schneeweiss, S. (2014). Learning from big health care data. *The New England Journal of Medicine*, *370*(23), 2161-2163. https://doi.org/10.1056/NEJMp1401111

*Heng-Wai Yuen
Changi General Hospital,
2 Simei Street 3, Singapore 539889
Tel: +65 69366259
Email: yuen.heng.wai@singhealth.com.sg