Big data, bigger opportunities

It is likely that somewhere between 70–80% of the information in the electronic health record is pathology data (1). These data are relatively readily available and represent the collective experience of the pathology profession. The amount of this data is enormous. There are more than 500 million pathology tests performed each year in Australia alone (2). We are now in what Sikaris (3) has called the third phase of medical learning [(I) masters; (II) journals; (III) databases) and we have the capability of turning this enormous resource of data into information and knowledge.

Pathology is the study and diagnosis of disease and the new database mining tools combined with the pathology database allow us to use this data in a variety of ways including determination of reference intervals and new quality control techniques, diagnostic algorithms, defining the usefulness of newer and current tests, guiding treatment, and perhaps most excitingly in knowledge discovery. Pathology data can be used to determine many relationships between data sets including anomaly detection, association, clustering, classification and regression.

Pathology (medical) data mining has unique problems related to the heterogeneity of the data classification, the ethical and legal issues of dealing with patient information, and the statistical philosophy of large data sets with missing elements (4).

Recognising the importance of pathology data mining, the Journal has introduced a new category of manuscript defined as "Clinical Database in Laboratory Medicine Research Column" which will highlight hot topics and major advances in the field. The Editors are willing to support articles that lead to greater insights into the issues described earlier. Papers that provide broad practical results that assist the readership in applying techniques to their data are sought. The column will serve as a source of advice and direction for laboratory staff and informaticians to exchange ideas and results to further this exciting move towards the widespread use of big laboratory data mining .

References

1. Hallworth MJ. The '70% claim': what is the evidence base? Ann Clin Biochem 2011;48:487-8.
2. RCPA. Pathology the Facts. (Accessed 20 June, 2017). Available online: https://www.rcpa.edu.au/Library/Fact-Sheets/Pathology-The-Facts/Docs/Path-Fcts-Booklt
3. Previous Discovery in Pathology Data Mining. (Accessed 20 June, 2017). Available online: https://www.slideshare.net/informaoz/ken-sikaris-melbourne-pathology
4. Cios KJ, Moore GW. Uniqueness of medical data mining. Artif Intell Med 2002;26:1-24.

Dr. Tony Badrick, Chair to the Clinical Database in Laboratory Medicine Research Column.

Dr. Tony Badrick
Chief Executive,
The Royal College of Pathologists of Australasia Quality Assurance Programs (RCPAQAP),
St Leonards, NSW, USA.
(Email: Tony.Badrick@rcpaqap.com.au)
doi: 10.21037/jlpm.2017.07.02
Conflicts of Interest: The author has no conflicts of interest to declare.
View this article at: http://dx.doi.org/10.21037/jlpm.2017.07.02