Editorial

Int Neurourol J 2014;18:49
http://dx.doi.org/10.5213/inj.2014.18.2.49
pISSN 2093-4777 · eISSN 2093-6931

International Neurourology Journal
INJ

How to Make the Most of Big Data in the Era of Complexity

Seung-June Oh

Department of Urology, Seoul National University Hospital, Seoul, Korea
E-mail: sjo@snu.ac.kr

In this issue of International Neurourology Journal, we have a review article of 'Big Data Analysis Using Modern Statistical and Machine Learning Methods in Medicine' [1]. The primary objective of this article is to examine clinical data, gene-gene and gene-environment interactions using statistical approaches. The paper is divided into clinical data, gene expression data, single nucleotide polymorphism (SNP) data, and epigenetic data, and it concludes with the basics of Bayesian networks (BNs).

Typically, clinical data is analyzed using either linear or logistic regressions. Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more explanatory variables while logistic regression is widely used in outcome variables that has two outcomes. Nowadays, these clinical data has been extended to include genomic and environmental data.

Recent genome studies have discovered significant associations between complex diseases and SNPs. In conjunction with clinical data, SNP-SNP interactions play role in understanding the development of complex diseases. Algorithms that use different search mechanisms, different ranking criterion, and/or that are geared toward specific situations have been developed. One of the modern statistical models that enable us to combine these clinical, genomic, and environmental data, is BN. BN is a directed acyclic graph in which each node represents a variable and each arc represents a relationship. In BN, each arc is interpreted as a direct influence between a parent node and a child node. BN have basic sub-networks, i.e., converging, diverging, and serial, that provide ways to express causal interactions in more intuitive ways. BN model has been widely used to learn predictive models from data. BNs can model causality based on researcher's knowledge, data or both. It is also feasible to use two or more of the previous models in order to reduce errors while obtaining better results.

With ever growing data in medicine, we should be also aware of recent trend in statistical analyses to get better understanding of such data. In return we will obtain more comprehensive knowledge on pathophysiology and better idea on how to establish treatment strategies of complex diseases.

REFERENCE

1. Yoo C, Ramirez L, Liuzzi J. Big data analysis using modern statistical and machine learning methods in medicine. Int Neurourol J 2014;18:50-7.