

MEETING ABSTRACT

Open Access

# Big data versus small data analysis towards personalized medicine practice

Mira Marcus-Kalish

From EPMA-World Congress 2013  
Brussels, Belgium. 20-21 September 2013

The new era in medical research and development, based on new approaches and advanced technologies, enabled profound insights, in various levels and degrees of sensitivity, on the human body physical and mental functioning. The identification of these micro and macro environmental factors created a real need for new analysis tools and translational models. The “big data” created and harvested in health care, from the molecular to the physiological, imaging and environmental data, etc, created an inevitable growing need for sophisticated analysis tools. “In 10 years each individual will be surrounded by a virtual cloud of billions of data pointsP4 medicine” 1. More than that there is a need to provide a broad understanding on the all factors interplay, based on combining all relevant data sources and utilizing all available analysis tools. New technologies are developed to harvest, verify, control, transfer, align and transform data, as well as new innovative analysis and predictive targeted tools to meet these challenges, both in academia and industry. In parallel there is a growing need to new innovative and sophisticated tools to analyse “small data”. That is to enable reliable and responsible preventive, predictive and personalized medicine, in most cases where there is more data than equivalent records. New philosophical approaches, as well as innovative analysis tools, specially developed to meet these challenges, must be taken, based on the coupling of computer sciences, complexity theory, non linear dynamics, logic theory, etc. That is in compliance and in parallel to the statistical models such as generalized linear models or classification trees that are further endowed by new regularization methods, as well as recent developments in FDR (False Discovery Rate) testing emphasizing the hierarchical structure.

A targeted data mining tools will be presented to comply with the real need to deal with all interplay of all factors originated in different sources, such as: molecular, genetics, images, physiology, etc, presented in different scales. More than that the goal is to analyze small sets of data containing many parameters, including missing values. The presented patented rule discovery and prediction data mining tool, enables the simultaneous analysis of multilevel, multisource data (imaging, signals, categorical, numerical and descriptive data) as is, with no data manipulation, such as normalization, or neglecting part of the data. The goal is to relate to the whole data set as is, while handling missing data - which is one of the major needs in providing the whole system approach. The algorithm reveals the underlying rules while attaching a level of confidence to each rule, identifies the unexpected rules and detects the unexpected cases. Furthermore the tool was found to be less sensitive to over-fitting (small sets with many parameters).

In summary the Data Mining and prediction tool is proven to reveal *all If Than* rules, *all If than Not* rules and a subset of *If and Only If* (necessary and sufficient conditions) tools, with no prior assumptions. Three detailed examples will be presented. The evaluation of the Epoetin adverse effects to assess long term risks and advancing towards better Epoetin driven treatment modalities, based on the survival analysis of dialysis patients and cardiovascular patients treated with EPO. The second example will deal with dementia and neurodegenerative combined data taken from the hospital practice (MCI, AD and normal patients) to evaluate and optimize the various treatments and provide accurate predictions. The third example deals with the analysis of the progression of disease in ALS patients based on the patient s current disease status. The data mining tool was already successfully applied to other applications such as: autoimmune diseases, drug development,

Correspondence: [miram@post.tau.ac.il](mailto:miram@post.tau.ac.il)  
Tel Aviv University, Israel

personalized skin treatment, skin genetic data and biomarkers analysis, etc.

Published: 11 February 2014

#### Reference

1. Hood L: The methods of fractal analysis of diagnostic images. Initial clinical experience. *Nat Biotechnol* 2011, **29**(3):191.

doi:10.1186/1878-5085-5-S1-A52

**Cite this article as:** Marcus-Kalish: Big data versus small data analysis towards personalized medicine practice. *EPMA Journal* 2014 **5**(Suppl 1):A52.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

