

The race is on! Which team or company is able to gain sufficient expertise to extract data generated during medical practice and turn it into useful information? Although multiple attempts have been published over the last years, several facts impeding innovation are worth mentioning.

As patient privacy is a key topic in our business, data availability in comparison with other scientific domains and industries, is relatively limited. The analytics enterprise is driven by dozens of scientists located around the globe. With similarities of a hacking community, the top level scientists recognize each others qualities and form temporary teams to solve specific problems. In an environment where patient data is used, progression is limited by data privacy regulations. This requires new attention as efforts to anonymize or de-identify medical data could provide more available medical data open for exploration with similarities of data lakes. A huge effort in this context has been accomplished by the team behind the MIMIC database (<https://mimic.physionet.org/>) which is available for research.

Companies focusing on precision medicine are faced with the limitations of their own structure in the sense that they often lack sufficient affinity with daily practice to capture the priorities in medical questioning. As such they often limit their interest to a narrow domain where the real power of analytics is to be found in the analysis of multivariable, high dimensional time series data. Additionally, providing descriptive analysis of the logistics or activity of a hospital is not the high end predictive and prescriptive analytics the medical community is waiting for. Also the medical community should focus on the science of ontology as a significant amount of data is continued to be entered entirely unstructured. Although several scientific papers from the New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, Buffalo, NY, USA and others cover this topic, a sufficient level of data quality generated by the medical community is frequently lacking resulting in GIGO (garbage in-garbage out) bias.

It is fascinating to observe the way tools and platforms such as R, normally used in basic science and various industries, are adapted to the medical domain thanks to the incredible efforts of a relatively small group of dedicated data scientists being fully active as health care professional in a busy medical practice. Combining two jobs seems an essential requirement to fully understand the path between medical data and generating clinical useful information.

Sven Van Poucke, MD, PhD

Anesthesiologist, Emergency Physician,
Data Scientist Ziekenhuis Oost-Limburg, Genk, Belgium