



---

## Volume 16 - Issue 1, 2016 - Cover Story

### Big Data and Analytics in Healthcare



**[Prof. Dr. med. Mathias Goyen](#)**

Industry Adviser

\*\*\*\*\*@\*\*goyen.de

Professor of Diagnostic  
Radiology - Hamburg  
University, Germany  
Chief Medical Officer - GE  
Healthcare, Europe

[LinkedIn](#) [Twitter](#)

---

*X-ray, computed tomography, blood tests, genetic diagnostics, health apps and wearables: we are producing a whole host of health-related data. What opportunities and risks do “Big Data” provide for patients and their care?*

□

After increasing importance in the financial industry, logistics and production, Big Data and analytics have rapidly started to gain momentum in healthcare as well. The potential of this emerging, IT-driven wave of innovation is huge. Big Data is supposed to make medicine more affordable and better. Data transparency and analysis are the drivers towards value-based medicine, where the basis of payment is not the resulting medical expense, but the success of treatment for the individual patient.

Much hope is associated with Big Data in medicine: reduced treatment costs, shortened hospital stays, enabled or optimised individualised therapy plans and combatting submission of false claims. Accordingly, the market potential is huge. The Big Data and analytics market in the U.S. is valued at \$80 billion for the year 2020 (SNS Telecom 2015). These prospects are increasingly drawing in companies such as Google, Apple, IBM or Salesforce in addition to medical technology companies native to the healthcare market. It will be fascinating to see whether and how these new players assert themselves in the jungle of healthcare systems.

Drivers of “Big Data” in Medicine

The explosion of “Big Data” in medicine is based on three factors:

1. Exponential growth of digitised information: due to the increasing digitalisation of formerly analogue media (images, reports, lab results, etc.) as well as the continuous optimisation of diagnostic laboratory and imaging sensors. Cutting-edge CT systems allow the recording of several thousand projections in only 1 second. Within 10 seconds, the entire body is depicted in 2mm thin layers. Add to this new patient data such as imaging data or molecular or genetic information. Furthermore, growing amounts of data are generated by increased monitoring with sensors of all kinds. Such monitoring is by no means limited to patients any longer, but also encompasses healthy people in ever greater numbers. Smartphones, activity trackers such as Fitbit and so on continuously produce huge floods of data. Correspondingly structured, semi-structured and unstructured data from a variety of sources coexist today, which jointly constitute “Big Data”.

2. The biological processes underpinning modern medicine are increasingly better understood in their heterogeneity. The cause of this improved understanding is the deciphering of the human genome, through which genetics determines whether and how people respond to different therapies. The associated differentiation is the basis of individualised medicine, to which prognostic risk stratification is inherent. Diagnosis and treatment are thus growing exponentially complex, and increasingly overwhelm any given physician. Big Data analytics are needed to filter out relevant, differentiating information from the resulting flood of data and, based on this, to help make correct, individualised treatment decisions.

3. The technological development of cloud-based IT solutions has cleared important thresholds in several key areas. Cloud computing allows data from different sources and of different quality, modality and structure to be technically unified into a whole “Big Data”. Transparency and networking coupled with mobility make data available independently of time and space. Moreover, available data are useful for the development of novel algorithms (eg, closest neighbour analysis, deep machine learning). This in turn allows conclusions to be drawn for individual patients (decision support) as well as for whole populations (population health).

## **Big Data Challenges**

Together, the above-described drivers are causing an explosion of “Big Data” in medicine. However, there are also significant challenges. Technically, it is important to master the three “Vs” in the domain of Big Data: huge amounts of data (volume), great variety of data (variety) and high speed of data generation and processing (velocity). In this regard, technical development is making rapid progress.

Apart from technical challenges, there are other challenges that hinder the development of Big Data in medicine. Of these, the issue of data security certainly takes up the biggest space. Any health cloud has to meet specific standards of safety and protection. There are two aspects to data security in cloud computing: first, the so-called security of personal data and secondly, the security of corporate data from unauthorized access or from loss. European Union-wide or global standards need to be developed here.

### **1. Dashboard / transparency:**

Big Data technologies combine different data sources and create transparency by sorting

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to [copyright@mindbyte.eu](mailto:copyright@mindbyte.eu).

unstructured data. Thus relationships are visualised, which improves the quality of medicine at the same time as reducing costs. This networking is not limited to hospital or physicians' practice data.

Rather, even patient-specific data from sensor watches such as pulse, blood pressure or blood sugar can be wirelessly transmitted and analysed in real time. Without delay, abnormalities in the readings can be identified and corrective action taken in this way. Most applications of Big Data are currently at this stage.

## 2. CAD: Computer-Assisted Detection

Computer-assisted detection (CAD ) combines elements of artificial intelligence and of digital image processing with radiological imaging. The typical field of application already is tumour diagnosis, especially breast and lung diagnosis. CAD supports screening mammography, which has been employed for the early detection of breast cancer for many years. CAD , established mainly in the U.S. and the Netherlands, serves diagnosticians as a second opinion to their own analysis. In lung cancer diagnostics, computed tomography (CT) has established itself as the gold standard, thanks to special three-dimensional CAD. Here, a volumetric data set of up to 3,000 individual images is processed and analysed. Nodules (outbreak, metastases and benign changes) can be detected from 1mm in size. Thus the strain on physicians is relieved and cognitive errors are reduced.

Big Data technologies go one step further. Deep machine learning can automate the evaluation of all image data. Deep learning arranges neural networks on to planes that use increasingly complex features, such as to recognise the content of an image. Thus masses of data can be sorted into categories. These networks, consisting of electronic nerve cells and the connections between them, are not explicitly programmed but trained using examples. It is only recently that the technology has become capable of simulating really complicated networks on the computer. With vast amounts of data, they can be trained.

□

The fascinating results of deep machine learning are known:

The IBM computer Watson, a semantic search engine that captures questions asked in natural language and finds appropriate facts and answers in a big data database in a short time, has not only beaten the human champion of the quiz show Jeopardy!, but is now also making medical diagnoses (IBM Watson n.d.). Now WATSON is learning how radiographs are diagnosed (IB M 2015). The first step is to differentiate between normal and pathological findings. This is followed by the structured analysis of abnormal findings that lead to a diagnosis. This technology will fundamentally change the reporting of all image data in radiology, pathology, and all other medical fields. Although the "Google Brain" can already simulate around 1 million neurons and 1 billion connections (synapses) (Dean and Ng 2012), we are currently only at the very beginning of this exciting technological development.

## 3. Decision Support

Decision Support Systems (DSS ) are software systems in medicine that are employed in hospital settings as instruments of knowledge management. They build on a knowledge base (of existing clinical pathways) maintained by experts and are able to autonomously reach conclusions, assessments and solutions of certain problems in complex treatment processes. They represent the next evolutionary stage of Big Data in medicine.

Expert systems are designed and implemented in such a way that they are integrated into existing hospital information systems in order to not only manage the information therein but also work independently with them. The need for this arises in situations where physicians have to rely on selecting and consultative end support in their daily medical decision-making processes, due to the increasing complexity of various treatment options. The set of rules for DSS systems is based on, eg guidelines adopted by professional bodies, so-called clinical pathways. Expert systems select and structure information along these pathways for specific treatment situations, thereby increasing the safety of patient treatment.

#### **4. Analytics: "Data Mining turns Big Data into Smart Data"**

After the capture, storage, visualisation and analysis of data, Big Data allows further analysis as a next step, by using cohort data to correlate various therapies as "inputs" with various "outcomes". What is crucial in this respect is the availability of validated data sets with well-documented patient histories. The right retrospective conclusions on the best therapeutic option can only be drawn with known outcomes, which require millions of records for statistical evaluation. The resulting correlative framework is enriched continuously through enrichment with new patient histories. The result is a self-optimising learning system.

The therapy optimisations that are identified based on such analysis then find their way into DSS systems. Clinical pathways are constantly being adapted to the latest findings. Thus, it is possible to develop improved algorithms and modify existing clinical care pathways. Optimal treatment is identified for newly diagnosed patients based on all available data.

Simultaneously, the analysis provides evidence regarding cost-effectiveness, by correlating outcomes with the cost of therapies. This is the basis of population health.

#### **5. Population Health**

Population health is about the optimal use of financial resources for maximum health for a larger cohort. Building on networking, storage and analysis, Big Data analytics provides the instruments to control healthcare services with the aim of optimizing outcomes for an entire cohort. Lessons learnt in this way are fed back into the design of individual patient pathways, thus enabling cost-effective use of medical therapies.

In Germany, population health evaluations are only just starting out. The National Cohort is a long-term study, funded by the Federal Ministry of Education and Research, which is running over a period of 20 to 30 years (German National Cohort Consortium 2014). 200,000 randomly selected participants aged 20-69 years from all over Germany will be medically examined and asked about lifestyle habits. In addition, blood samples are taken for all study participants and stored for later research projects in a centralised biobank. The aim is to acquire profound knowledge in terms of prevention and early detection of typical widespread diseases.

In the U.S., however, population health studies are already being employed by insurers today. Cohort studies from Finland and the U.S. have shown that the occurrence of type 2 diabetes in high-risk individuals (impaired glucose tolerance) could be lowered by more than half over

an average period of three years through lifestyle interventions (eg, reduction of excess weight, exercise). It will be crucial that, apart from costs, the outcome of treatment is adequately considered.

## Summary

Big Data and Analytics has become a top issue in the healthcare industry. The trend in medicine is moving from reactive treatment towards predictive and preventive medicine. Through early intervention, disease is to be prevented, or at least treated early and in an individualised way. The paradigm shift from generalised to individual, personalized medicine will help patients and reduce costs in the healthcare system.

Big Data and analytics can already point to impressive results in the medical field, but development is in its infancy. If it becomes possible to satisfactorily solve data protection issues in addition to technical challenges, broad societal acceptance of Big Data and analytics in healthcare can be expected.

Published on : Sat, 27 Feb 2016