

White
paper

Efficiently generating medical insights

Advanced analytics on
longitudinal health records

Watson Health™

IBM

IBM Explorys data

- Includes more than 55 million de-identified patients based in the United States and billions of data points that are updated weekly for near-real-time analysis.
- Provides longitudinal data for these patients, including demographic, diagnostic, and therapeutic information, creating a detailed clinical picture. Many patients are also linked with financial records.
- Grows larger via direct connections to an expanding array of large health system networks.

Summary

The increasing pressures faced by life sciences companies and the rise of big data have led to a new approach for the generation of medical evidence in recent years: mining longitudinal electronic medical records (EMRs). This change is brought to life here with two recent IBM projects carried out with leading European life sciences companies. In these projects, advanced analytics was applied to the IBM Explorys® data set, which consists of real-life data from millions of patients. Both examples illustrate the power of a novel method that allows medical evidence to be efficiently generated when mining EMRs: Agile Advanced Analytics (AAA). The technique helps pave the way towards personalized and outcome-based healthcare.

Generating medical evidence with advanced analytics has become essential

The healthcare industry is in the midst of a value-and patient-driven transformation. Government and private payers are exerting pressure to restrain drug prices based on outcomes.¹ Subsequently, life sciences companies are increasingly being asked to justify their prices and to develop innovative models that take healthcare outcomes into account.

At the same time, growing costs and increasing complexity are associated with the development of medical products. Recent estimates of bringing a molecular entity to market reach up to USD 2.6 billion and about half of Phase III drug candidates still have to be dropped.^{2,3}

As part of the medical development process, modern clinical trials have been conducted for decades in order to study the safety and efficacy of medical treatments.⁴ However, in recent years a number of challenges typically associated with clinical trials have been exacerbated:

- Rising trial costs: Research has identified increases in the cost of trials that impact the cost of bringing new products to market.⁵
- Discrepancies between trials and real life: Ensuring that the conditions of a clinical trial are representative of those encountered in real life are challenging.
- Lack of exploratory design: Often design for purpose is in the foreground, hampering the ability to identify unmet medical needs.

As a result, real-world data science is rapidly growing as a complementary tool and has been establishing itself as a new standard with regard to generating medical evidence and exploring medical insights.⁶ In fact, advanced analytics in combination with large real-world data sets has opened up new fields. Whereas data analysis in healthcare has traditionally focused on outcomes and drug efficacy measurements, today predictive analyses based on real-world data sets support the industry in delivering innovative healthcare solutions and driving efficiencies.

This paper describes the application of advanced analytics to large sets of longitudinal EMRs using the example of two recent IBM client projects that used the IBM Explorys real-world data set (see summary box). These projects are exemplary with regard to the significant value such an approach can bring to life sciences companies. Both projects were run for life sciences clients in Europe using United States EMR data.

Benefits of client project 1:

- Robust view on the patient group's clinical picture (including demographics, diagnostics and laboratory results)
- Expected improvement of medication adherence
- Identification of supporting clinical evidence for the optimization of drug launch activities

Client Project 1: Population analytics generates deep insights and confident decisions

Population analytics carried out with longitudinal data allows patient populations to be dissected and understood. Actionable insights from population analytics yield significant gains in areas such as individual patient management, drug development, product strategy and therapy outcome assessments. Furthermore, it becomes possible to evaluate the impacts of regulatory changes and verify business assumptions.⁷

In this client project example, population analytics was applied to generate insights regarding treatment adherence and outcomes. Various demographic factors, which were assumed to have an impact on healthcare access and treatment outcome, were examined. Patients in care and patients dropped from care were compared in detail (Figure 1). Demographic

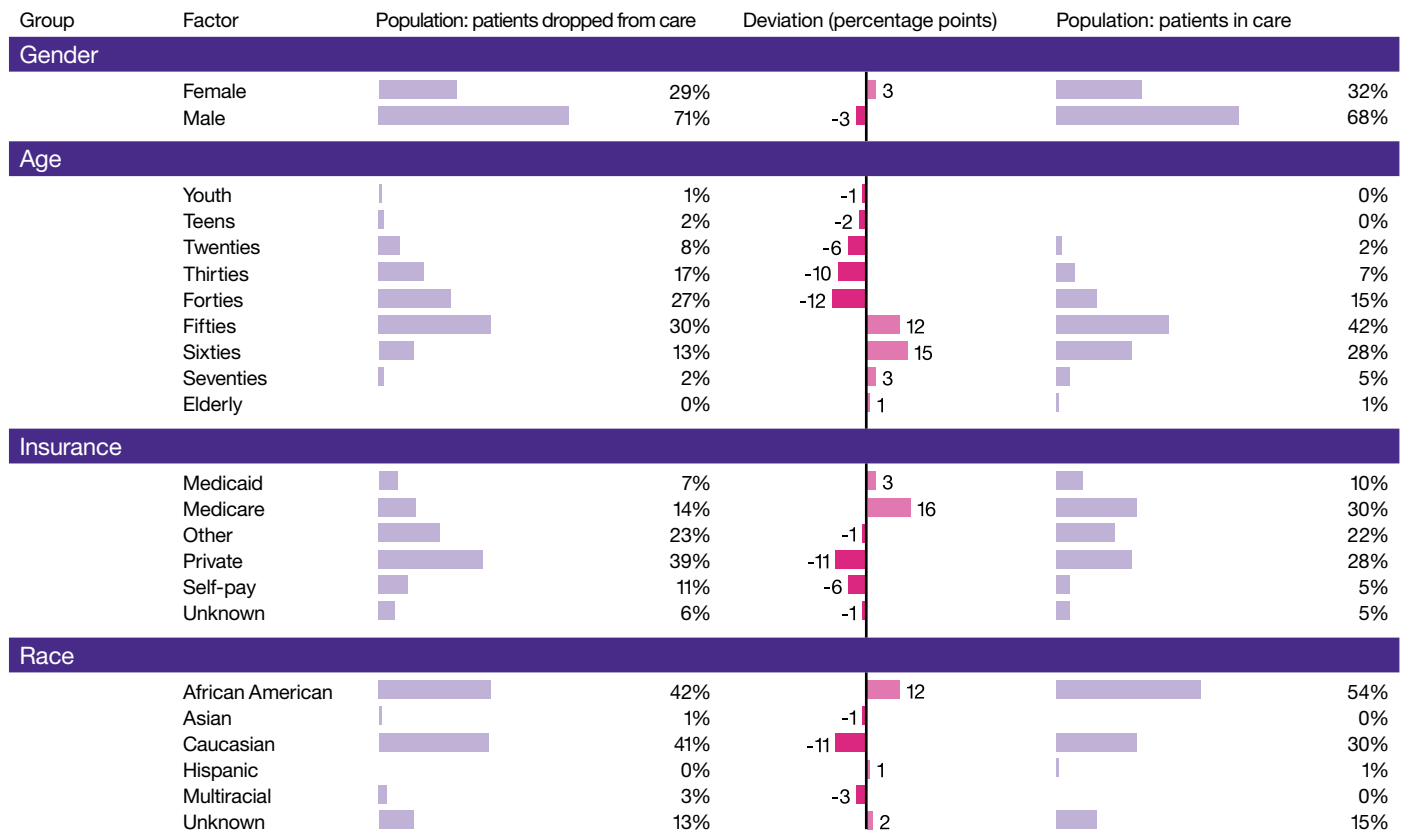


Figure 1:
Illustrative example
of population group
comparison. The purple
bars show the prevalence
of the factors within the
population group and
the deviations (the
pink bars) illustrate the
characteristics of
the different sub-
population samples.

characteristics that are more prevalent in certain sub-populations were identified and further analyzed. Among other results, the study revealed significant markers for treatment adherence, allowing pre-emptive measures to be taken. The strongest indicator suggested a potential improvement of 300 percent in medical adherence.

Moreover, population analytics was used to identify and characterize relevant patient sub-populations so that the launch activities for a new drug can be targeted to an optimal patient population. The population was dissected based on patients' treatment pathways (Figure 2) and characterized according to different demographic and clinical factors. Sub-populations were identified and classified by their health state. The results effectively illustrate potential economic implications for the healthcare system if no treatment is carried out. For instance, this analysis provided critical insights for the success of the drug launch with regard to the payer environment and market estimate. In addition, there is potential to conduct Phase IV clinical trials on the basis of analyses of the type described. The entire project was completed in five weeks, demonstrating the speed with which actionable insights can be generated by advanced analytics.

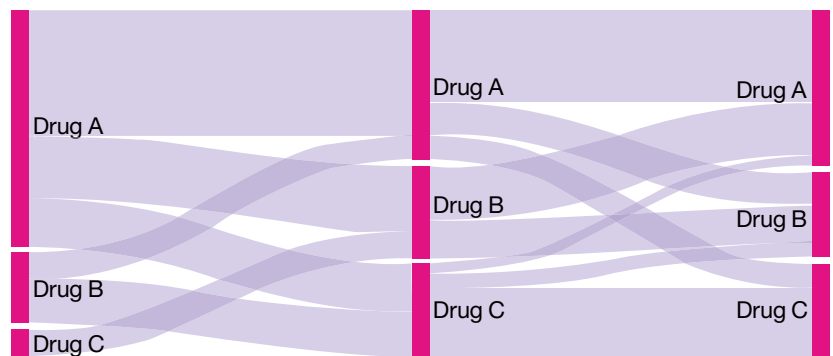


Figure 2:
Illustrative extract of
patient treatment pathway
population analysis to
understand patients and
sharpen further study
questions.

Benefits of client project 2:

- Rapid generation of predictive insights within weeks
- Identification of new correlations that would likely have not been found in clinical study data
- High potential to optimize healthcare costs triggered by chronic disease complications

Client Project 2: Predictive analytics and machine learning allow the medical development of patients to be forecast

In a second recent IBM client project carried out over three months, the longitudinal EMRs of millions of patients were analyzed to demonstrate the predictive power of advanced analytics and machine learning. Patients with chronic diseases often face related complications, which leads to anxiety among patients and may cause extensive healthcare costs. Predictions of these complications based on conventional medical methods are often complex due to the diversity of the influencing factors.

Therefore, a large set of patient characteristics were extracted and advanced analytics techniques were applied to predict whether a given patient is likely to experience complications in future years. Compared with existing prediction baselines and risk calculators, a significant increase in predictability was achieved as measured by, for instance, area under the receiver operating characteristic curve (AUC) and precision (Figure 3). In addition, a powerful new combination of biomarkers was discovered. Advanced analytics on longitudinal EMRs thus allowed for the space of patient characteristics to be explored, assessed and prioritized exhaustively to help allow more accurate predictions. Moreover, this approach was found to be especially powerful for long-term complications.

While the predictive outcomes were derived from large sets of de-identified patient data, the insights yield immediate benefits for individual patients and life sciences companies, including:

- Individual patient management
- Efficient cohort selection for further studies
- Treatment guidance for provider consideration

It is evident that the predictive power of analytical models ushers in a new era in which clinicians are enabled to provide personalized medicine at the level of an individual.

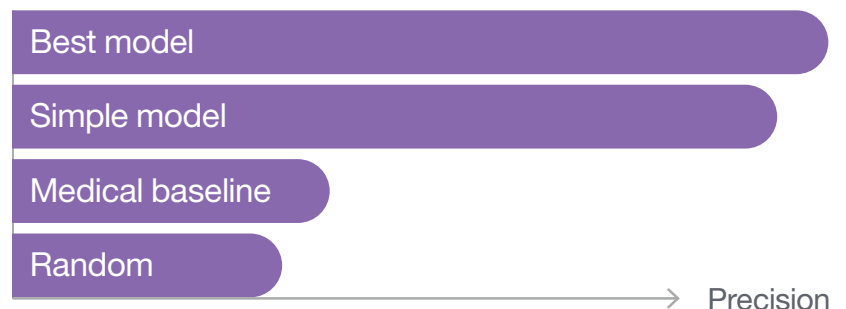


Figure 3:
Advanced analytical
models can significantly
outperform reference
measures.

The new method of AAA applied to longitudinal EMRs efficiently generates medical evidence

A new approach, namely the application of AAA to EMRs (Figure 4), enabled the two projects outlined above. The Cross Industry Standard Process for Data Mining (CRISP-DM) and IBM's agile methodology inform this methodology.⁸ AAA emphasizes the study questions that should be continuously reprioritized to help ensure that the most value-adding and tangible outcomes can be generated quickly. The need to adapt existing methodologies arose due to the greater focus on managing the backlog of study questions.

AAA begins with a backlog of prioritized questions. Data processing and verified modeling results allow these questions to be further refined and prioritized in an iterative nature. Discovering potential findings or limitations early on minimizes later re-work. Finally, documented and tangible research results and predictive insights are generated quickly with suitable data science techniques. Effective communication is then required when sharing the insights with stakeholders. Due to the broad applicability of this novel methodology in the life sciences, we expect that AAA on EMRs will soon become widely adopted across the industry.

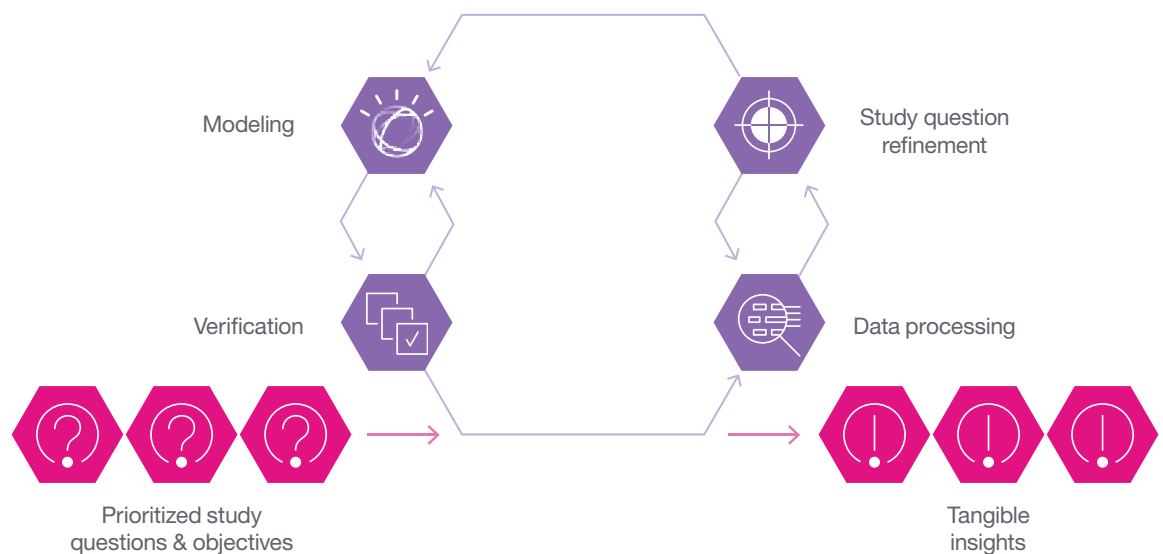


Figure 4:
AAA, the novel
methodology for mining
longitudinal EMRs.

Recommendations

- The discovery of novel insights in the study of EMRs is enabled by multidisciplinary teams combining medical expertise, excellence in advanced analytics and execution discipline.
- Even simple machine learning models can outperform the status quo when the focus is on patient outcomes and the research questions drive the modeling techniques (and not vice versa).
- The above can be significantly accelerated with AAA, which promotes collaboration, quick learnings and a focus on results.

Life sciences companies should act now to take advantage of the enormous benefits achieved through mining EMRs

The described examples showcase how multi-faceted advanced analytics can be applied to longitudinal EMRs and how significant value can be created for patients as well as healthcare and life sciences organizations by doing so. It is easy to see how vital questions in the industry can be addressed, such as:

- What do **patient and population behavior** and responsiveness look like?
- What are **product and therapy differentiators** or challenges?
- How do **diseases progress**, and what are **early indicators**?
- What is the **impact of provider/payer behavior** and **treatment regimens**?
- Which **population segments** should a new drug be **marketed** to?

In fact, in this paper we showed that advanced analytics on longitudinal EMRs not only helps understand such questions but also allows organizations to predict certain outcomes. This illustrates the trend that in the big data era companies are transitioning from “descriptive” to “predictive” and eventually to “prescriptive” insights.

Ultimately, the application of AAA to EMRs is more than just a complement or alternative to controlled study environments. This approach is a novel standard to explore new grounds in diverse areas of the life sciences, from research and development to medical affairs, product strategy and market access. This approach can enable the transition to personalized and outcome-based healthcare.

How IBM can help

IBM Explorys solutions maintain an ongoing connection to healthcare systems and draw on de-identified data that resides in a cloud-based, Health Insurance Portability and Accountability Act (HIPAA)-enabled, security-rich platform. Our clients can generate insights from the IBM Explorys data set and analyze study populations based on their defined selection criteria.

IBM Global Business Services® (GBS) consultants and data scientists help your organization run advanced analytics projects based on IBM Explorys and other real-world data sources. IBM GBS applies the AAA methodology and uses IBM Watson® technology in order to more quickly generate deep insights from real-world data for you. For further information, please contact Lars Böhm at lars.boehm@ch.ibm.com.

Authors

- Dr. Stefan Ravizza, Cognitive & Analytics Lead, IBM Global Business Services, Switzerland
- Dr. Frederik Flöther, Data Scientist, IBM Global Business Services, Switzerland
- Alexander Büsser, Data Scientist, IBM Global Business Services, Switzerland
- Lars Böhm, Life Sciences Account Lead, IBM Global Business Services, Switzerland

1. O'Connor, Rory J. and Vera C. Neumann. "Payment by results or payment by outcome? The history of measuring medicine." Journal of the Royal Society of Medicine. May 2006. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1457759/>
 Latkovic, Tom. "The Trillion Dollar Prize." McKinsey & Company. February 2013. <http://healthcare.mckinsey.com/sites/default/files/the-trillion-dollar-prize.pdf>
 National Audit Office. "Outcome-based payment schemes: government's use of payment by results." June 19, 2015. <https://www.nao.org.uk/wp-content/uploads/2015/06/Outcome-based-payment-schemes-governments-use-of-payment-by-results.pdf>
 The Economist Intelligence Unit. "Value-based Healthcare: A Global Assessment." 2016. http://vbhcgloballasessment.eiu.com/wp-content/uploads/sites/27/2016/09/EIU_Medtronic_Findings-and-Methodology.pdf
2. Avron, Jerry. "The 2.6 Billion Dollar Pill: Methodologic and Policy Considerations." New England Journal of Medicine. 2015;372:1877–1879.
3. Kola, Ismail and John Landis. "Can the Pharmaceutical Industry Reduce Attrition Rates?" National Reviews Drug Discovery. 2004;3:711–715.
4. Stolberg, Harald O., Geoffrey Norman and Isabelle Trop, "Randomized Controlled Trials," American Journal of Roentgenology, December 2004. www.ajronline.org/doi/full/10.2214/ajr.183.6.01831539
5. Collier, Roger. "Rapidly rising clinical trial costs worry researchers." CMAJ News. February 3, 2009. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2630333/>
6. Arimboor, Dhara Yesudas. "An Industry View Point: The Future of Leveraging Data & Technology in Clinical Trials." March 22, 2016. <https://www.linkedin.com/pulse/industry-view-point-future-leveraging-data-technology-arimboor>
 Vina, Gonzalo. "Big data promise exponential change in healthcare." Financial Times. November 29, 2016. <https://www.ft.com/content/1b614d10-9a03-11e6-8f9b-70e3cabccfae>
7. IBM. "IBM Explorays offerings for life sciences." Watson Health. November 2016.
8. Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer and Rüdiger Wirth, Cross Industry Standard Process for Data Mining 1.0. "Step-by-step mining guide," 2000. www.the-modeling-agency.com/crisp-dm.pdf

IBM Corporation
 Route 100
 Somers, NY 10589

Produced in the United States of America
 February 2017

IBM, the IBM logo, IBM Explorays, IBM Global Business Services, Watson, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/us/en/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

