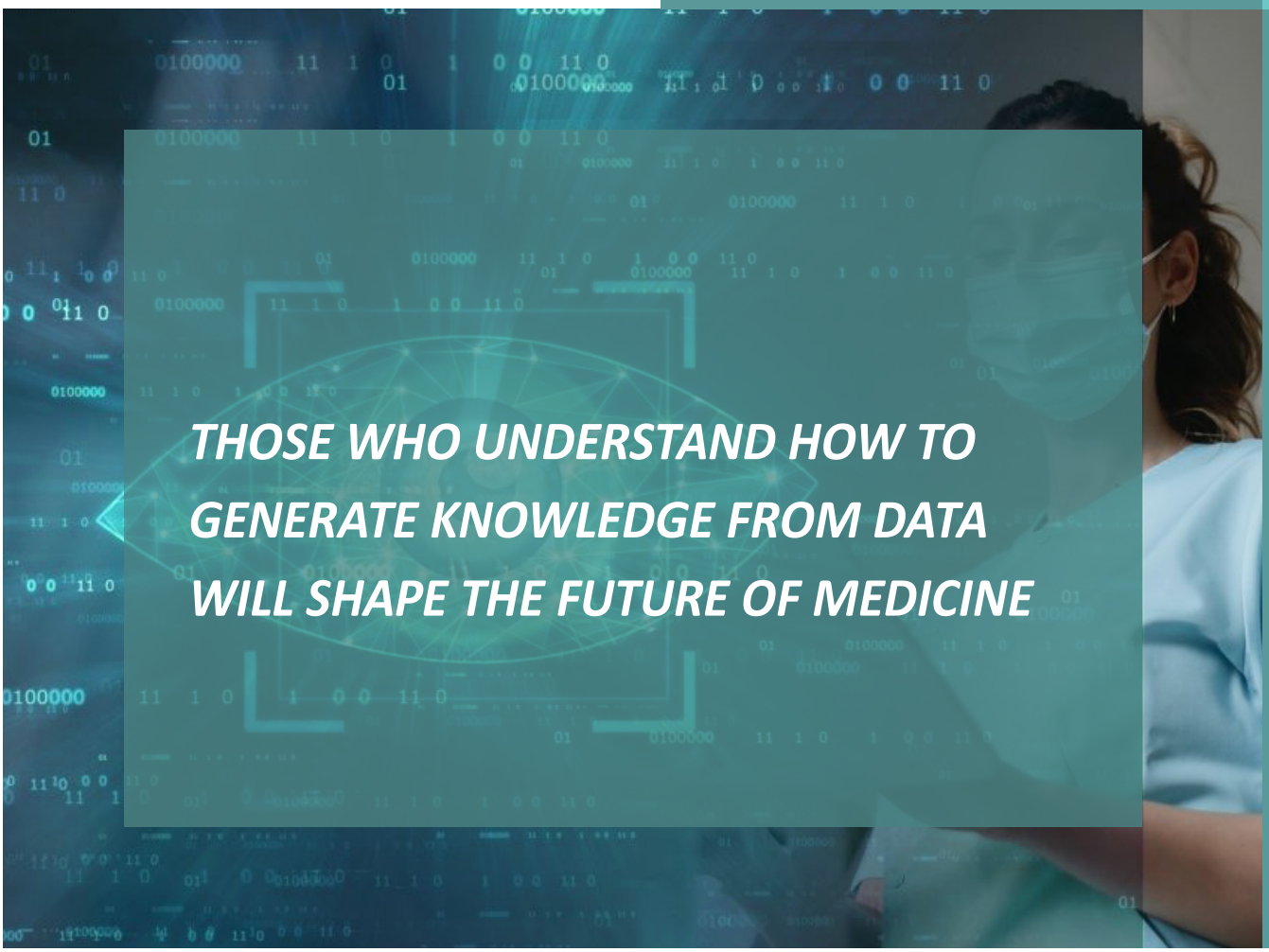
Xplain Data
Discover Causality

VISION PAPER



VISION4HEALTH – THE POTENTIAL OF CAUSAL DISCOVERY FOR THE FUTURE OF MEDICINE

Xplain Data GmbH
Grünlandstr. 27
85604 Zorneding
info@xplain-data.com
xplain-data.com

„EMPIRICAL KNOWLEDGE“ MEANS: DATA

Medicine is considered „empirical“, i.e. science based on experience and observation. Relationships are observed and used, even when the underlying processes are not known in detail. Pharmaceutical companies analyze the effects of drugs through empirical observations in control and treatment groups.

Experience also plays a central role in daily practice:

An experienced General Practitioner is in high demand. Faced with complex cases, a prudent doctor will consult a colleague or specialist to tap into their experience and know-how.

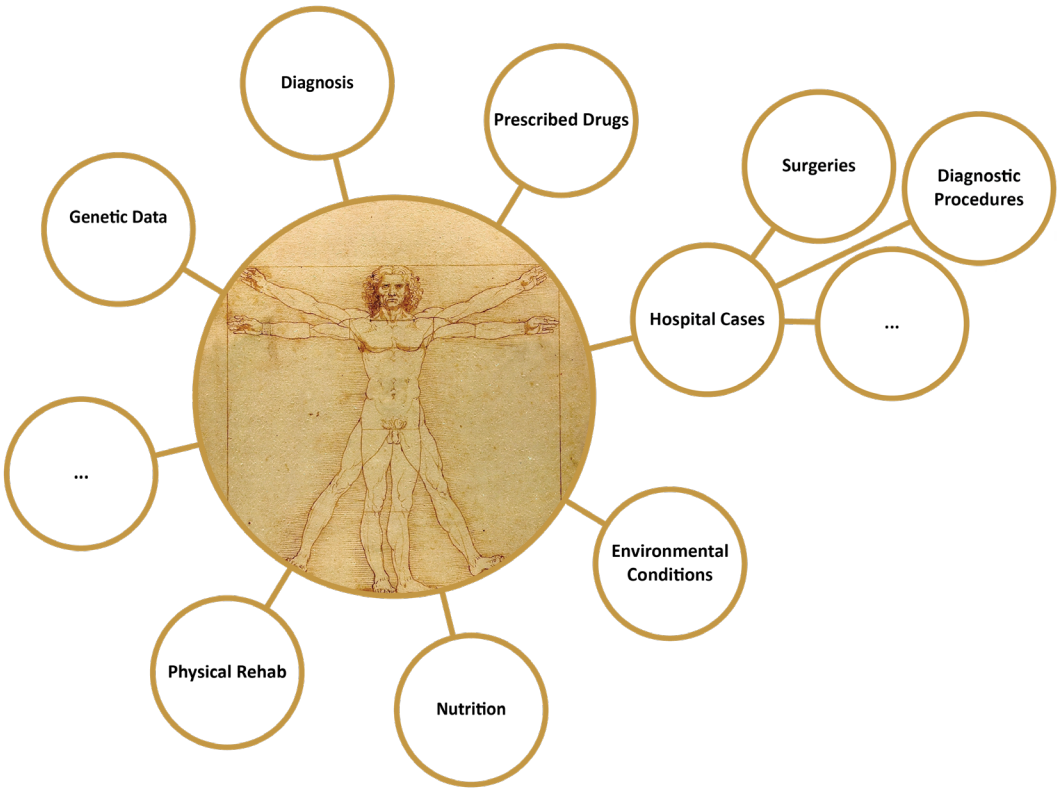
Imagine that our doctor has treated 100 million patients and gained a wealth of experience. We would certainly feel that we are in good hands, especially when faced with difficult or rare diseases.

Of course, a single doctor cannot have the accumulated experience of treating 100 million patients.

But the doctor could have direct access to this knowledge.

This knowledge is implicitly contained in patients' data. Let's assume we had comprehensive medical information of 100 million patients: all symptoms, diagnoses, medications administered, dietary habits, sporting activities, professional background, social environment, genetic data, laboratory data, MRI and CT scans and proteomic information. Then add information on medical interventions and their long-term consequences – in short: 100 million comprehensive medical CV's.

Such data would represent an enormous reservoir of medical know-how. Even a fraction of this would constitute an enormous amount of knowledge.



DESPITE ITS GREAT POTENTIAL: DATA BY ITSELF IS NOT SUFFICIENT

The goal is to transform data into knowledge – in the form of general associations or concrete recommendations. This remains a challenge in which we are still at the very beginning.

Generating knowledge from data means, in particular, understanding the causal relationships contained in the data, e.g. the factors that causally lead to the development of a disease such as multiple sclerosis (MS) or MS relapses. MS is one of many diseases whose exact cause is still unknown, hence understanding the causal factors will help in so many ways: It helps us to identify high risk patients. It helps us to develop new drugs to break the chains that cause the disease, and it helps us carry out intelligent interventions to change the course of the disease in the desired direction.

UNDERSTANDING CAUSALITY: BASIS FOR INTELLIGENT INTERVENTION

Achieving a goal requires an understanding of the causal factors involved - only then can we develop intelligent strategies to accomplish that goal.

A new field of artificial intelligence (AI) to understand cause-and-effect relationships based on observational data is emerging under the term "Causal AI". In the *Gartner Hype Cycle for Artificial Intelligence*, it was listed for the first time in 2022.

New cause-and-effect relationships can only be identified with very comprehensive and multi-faceted data. If the causal factors are missing in the data, only indirect indications can be obtained at best. To unveil those factors, all contextual information (so-called confounders) must also be included in the analysis. We therefore need a comprehensive picture of a patient, made possible by extensive patient data. Those who have access to such data, as well as the algorithms and expertise to derive causal insights from it, will play a critical role in shaping the future of medicine.

THE CHALLENGE: THE COMPLEXITY OF DATA

When we speak of comprehensive patient information, we refer to their complete electronic medical record. Patient data has hundreds of different tables used to store a whole range of information. This data schema is extremely large and complex. Take 100 million patients over a period of 10 years, and this results in a data set containing up to 10 to 30 billion prescriptions and 50 billion diagnoses – a total of several hundred billion entries.

THIS COMPLEXITY CLASHES WITH TODAY'S AI LIMITS

How can we obtain empirical knowledge, specifically causal insights from such complex patient data? There is still a long way to go. Many of us have performed statistical analysis and thereby used data in a flat spreadsheet. Machine learning or artificial intelligence techniques also expect data in a flat or simple structure – far off from the complexity of the data contained in an electronic patient file.

THE WORLD: FLAT ON A SMALL SCALE, ROUND ON A LARGE

To enter new horizons and move beyond flat data structures, we need technologies that can handle the complexity and volume of real-world data (RWD). Xplain Data is on the way, having developed patent-protected methods and tools. Without prior assumptions, these tools sift through extensive and complex data to identify potential causes related to a target. Numerically complex algorithms retrospectively analyze millions of factors from the target event (e.g. MS diagnosis) in order to identify potential cause-and-effect relationships. This analysis goes far beyond trivial correlations. It takes into account a large number of influencing factors as well as many "context variables" (confounders) in order to arrive at relevant hypotheses for causalities and uncover entire causal chains.

OBJECTANALYTICS: LOOKING AT THE BIG PICTURE

In order to be able to analyze extensive and complex data from many different angles, Xplain Data has developed ObjectAnalytics, a specific form of data storage.

Unlike traditional relational databases, which divide a patient into its entities and store different parts in different tables, Xplain Data ObjectAnalytics enables access to all information of an analyzed object as a whole. Without expensive joins between tables, all information belonging to a patient can be analyzed in relation to each other.

Exploiting this 360° view, sophisticated machine learning algorithms carry out a causal search. Despite conducting the search at high speed, this can take several hours to days, depending on the amount of data and hardware available. To take advantage of the performance on many processors, these algorithms are designed to be used on large computer systems.

These methods are already being used in many fields: For example, they have successfully identified causal factors for machine failures or quality defects in the manufacturing industry. In addition, ObjectAnalytics technology has also been effectively applied to very large volumes of healthcare data. Initial assessments have been carried out in this area, but there is still a lack of systematic implementation.

What if ...

- 30 million longitudinal patient data records over the last 10 years – an estimated 50 billion diagnoses, prescribed drugs, laboratory values and
- a computer farm with 10 servers à 128 cores, on which state-of-the-art causal discovery processes are implemented plus
- a team of medical experts from various disciplines,
- supported by a data science team

was at your disposal?

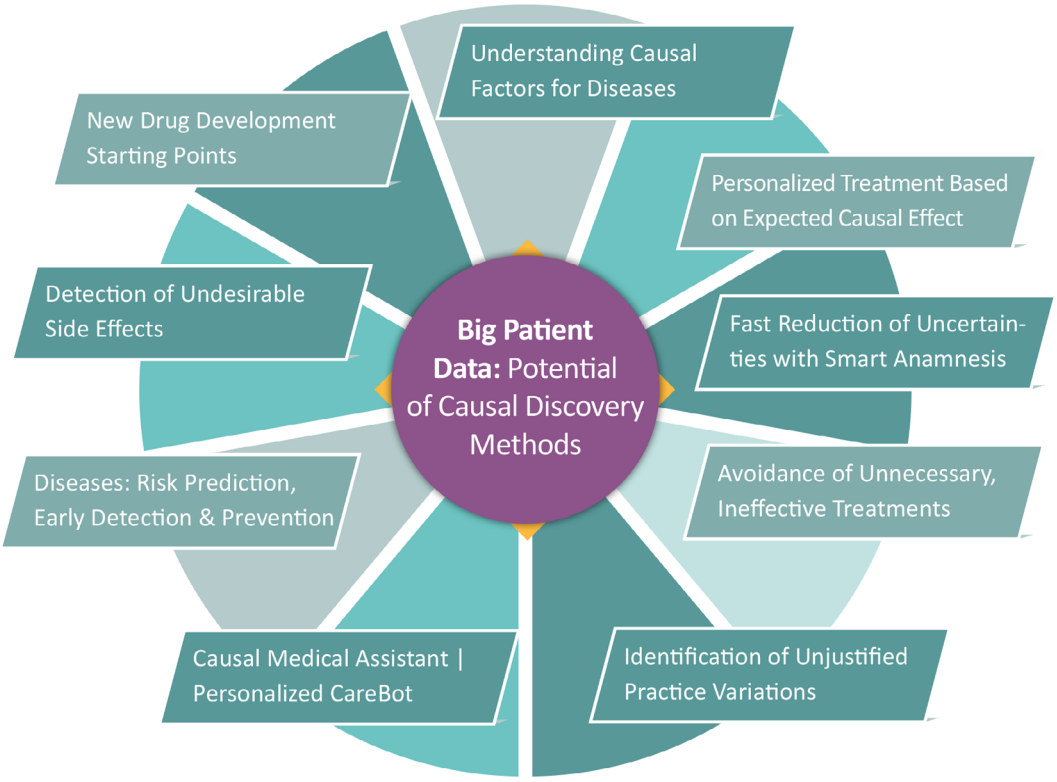
WE EXPECT A QUANTUM LEAP IN MEDICAL RESEARCH!

The combination of data, AI algorithms, and medical expertise has the potential to unleash groundbreaking developments in medical research and patient care.

It would not be long before the first discoveries took today's medical world by surprise. Not only is there a chance to uncover concrete medical results e.g. understanding complex autoimmune diseases such as Multiple Sclerosis.

Understanding causal chains offers a foundation for identifying novel therapies that can precisely target the underlying causal pathways of the disease in question. Causal AI based on Xplain Data ObjectAnalytics can present us with a new paradigm: "From data to therapy" with unlimited possibilities.

In addition to the development of new therapies, there is also an opportunity to bring this knowledge back to the direct point of patient care – to the doctors who are managing the specific case.



Causal models can be used to make predictions about a patient's likely diagnoses, especially rare diseases, for earlier detection. They facilitate the development of strategies for intelligent anamnesis processes to quickly minimize uncertainties about existing diseases. In addition, they can support the selection of relevant therapies with expected results in mind or the causal effect of the chosen therapy.

A „Causal AI-Medical Assistant" is within reach.

DATA AND INTELLIGENT ALGORITHMS ARE THE FUTURE OF MEDICINE

The future of medicine will be shaped by those who have access to comprehensive patient data and develop the ability to derive causal knowledge from it.

The crucial questions:

WHO WILL BE THE NEXT INNOVATION LEADER IN THIS NEW FIELD?
AND WHAT ARE WE STILL WAITING FOR?

Xplain Data
Discover Causality

Contact:
Xplain-Data.com
info@xplain-data.com