

The rise of data science in public healthcare and commercial pharma

How public healthcare's ramp-up of data science solutions is an opportunity for well-prepared commercial pharma organisations

The Danish healthcare system is digitalising on all fronts. With the current state of big data, artificial intelligence (AI), and machine learning, the question is not whether these tools will be used to prioritise medication, but when and how. Managers responsible for future sales need to be aware that this provides a real competitive advantage for pharmaceutical companies that are ready to integrate data science in sales, marketing, market access, and medical affairs. This white paper outlines how to begin this transition and presents cases explaining how data science adds value to business operations.

Contents

Introduction	3
The main data-driven knowledge domains in pharma	3
The proliferation of big data and data science among pharma's stakeholders in Denmark	6
Case 1 - Measuring brand attention and loyalty of prescribers via big data analysis	9
The traditional organisation of commercial pharma	8
Case 2 – Building a treatment-dynamic model via machine learning	11
Implementing data science capabilities in commercial pharma	12
Case 3 – Identifying key opinion leaders via natural language processes and network analysis	13
Step 1: Add value to all business areas in the commercial section of pharma	14
Step 2: Be ready to use big data and data science methods to challenge local decisions	15
Conclusion	17
References	18
Authors	20

Introduction

Commercial pharma has been at the forefront of modern developments in the use of data to analyse and track markets and stakeholders for some time. However, recent growth in the application of data science by healthcare stakeholders – in big data, machine learning, AI and data advanced analytics – is about to transform core commercial business processes in pharma. Governments, healthcare authorities and medical professionals are increasingly turning to data science solutions in their treatment guidance, choice of medicines, predictions of safety risks for patients and patient engagement (see the three Case Studies in this white paper).

The question is: What preparations should commercial pharma be making to ensure that it engages effectively with a healthcare system that utilises data science solutions in the prioritisation of treatment, choice of medicines and patient support? And how can data science capabilities begin to add real business value to commercial pharma?

This white paper will answer these questions by explaining how:

1. Data science *already* influences the Danish healthcare system and is shaping its future development
2. The proliferation of data science in healthcare is an *opportunity* for pharmaceutical companies that are prepared to engage with authorities, medical professionals and patients on data science solutions
3. Data science solutions *add value* to pharma organisations today, as we will illustrate with concrete cases
4. The *integration* of data science capabilities in areas such as medical affairs, safety, market access, market research and sales/marketing will shape the future of the commercial pharma sector

The main data-driven knowledge domains in pharma

Big data, AI, and machine learning are no longer mere buzzwords signifying rather different resources from industry to another. They are converging into a real, value-creating business discipline under the umbrella term “data science”. Worldwide, revenue for big data and analytics companies has grown from \$130.1 billion in 2016 to an anticipated revenue of more than \$203 billion in 2020.¹ Yet, to a large extent data science specifically, and big data in general, remain uncharted territory for pharma companies, making it difficult for those companies to use data science for value generation. Thus, we need to understand how big data, and data science, is a part of the fundamental data knowledge domain in pharma.

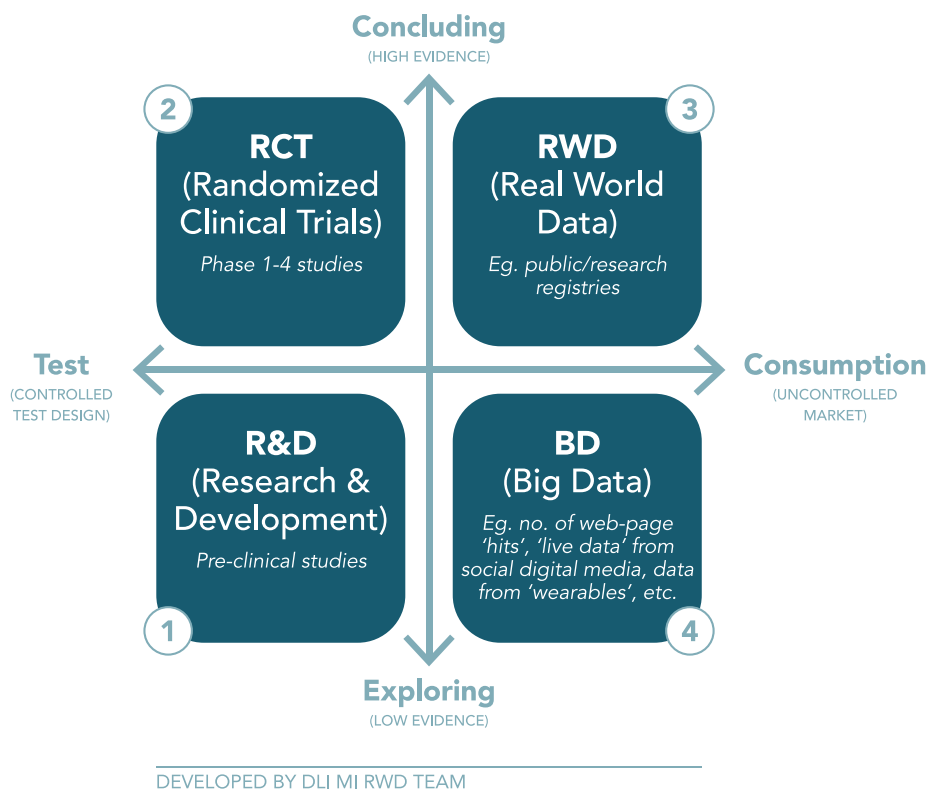
In a previous white paper, we introduced a heuristic model of the four main data-driven domains in pharma (Fig 1.).² The model describes real world data (RWD) and big data (BD) in relation to research and development (R&D) and randomised clinical trials (RCT). R&D explores the medical potential of compounds through pre-clinical studies in a controlled environment. Compounds found to have medical potential are then tested in RCT studies in order to produce evidence of efficacy and safety.

¹ Press, “6 Predictions For The \$203 Billion Big Data Analytics Market.”

² Micheelsen, “Beyond Real World Evidence: How to Improve Utilisation of Real World Data throughout an Organisation.”

RCT is still the gold standard for evidence of efficacy and safety, and among medical research communities, health authorities and priority bodies, it is currently the only widely accepted study design for documenting causal-effects. However, increasingly RCT studies are being challenged on the basis that they have a number of drawbacks. They are costly and take time to conduct. They can also suffer from poor external validity, because treatments in real-world settings vary, and because the study populations they use are not necessarily representative of the populations that will come to use the medicine in a real-world setting.³

Fig 1. The four data-driven knowledge domains in the pharma industry



As a result, public authorities are increasingly requesting evidence that safety, cost and treatment outcomes in the real-world setting are in line with the evidence produced by RCT⁴. To this end, access to, and analysis of, real world data has become ever more important. In Denmark, public authorities such as the Medicines Council, the national hospital tender organisation Amgro, the Danish Medicines Agency and the Reimbursement Committee use real world studies when evaluating, for example, treatment cost, safety and drug utilisation.⁵ Public authorities and pharma are well aware that the structured data found in registries of treatment and outcomes (e.g. treatment quality and research registries) as well as all the other structured data from the healthcare system (e.g. patient journals, prescription registries, etc.) can be used to argue for or against the real world value of a medicine.⁶

³ Bothwell et al., "Assessing the Gold Standard — Lessons from the History of RCTs."

⁴ "FDA's Woodcock: The Clinical Trials System Is 'Broken' | RAPS."

⁵ Hirsch and Micheelsen, "From Evidence to Consequence – the New Reality for Prioritisation of Medicines in Denmark."

⁶ Hirsch and Micheelsen.

“ The clinical trials system is “broken” ”
*JANET WOODCOCK, DIRECTOR OF FDA'S CENTER FOR DRUG EVALUATION AND RESEARCH*⁷

Indeed, authorities are beginning to question whether RCT is the optimal method for documenting a medicine's treatment benefits. Janet Woodcock, director of FDA's Center for Drug Evaluation and Research, stated in 2017 that the clinical trials system is “broken”, and that we need find new ways to collect and utilise real world patient data in future clinical trials.⁸ Similarly, the European Medicines Agency (EMA) has established a taskforce assigned to explore “the potential applicability and impact of big data on medicines regulation” and to evaluate “applications for marketing authorisations or clinical trials”.⁹ And EU's Horizon 2020 programme, together with commercial partners, is sponsoring big data research projects such as the Innovative Medicine Initiative's (IMI) Big Data for Better Outcomes programme with funds up to €93 million.¹⁰ In other words, key authorities are actively pursuing ways to include real world data and big data in the development, testing, and introduction of medicines to patients.

There is, however, a certain amount of unclarity about what count as real world data and big data in pharma. “Big data” are usually thought of as the ever-growing, often unstructured, body of data produced by digital technologies, coupled with data science methods (AI and machine learning), which aim to give insights into, and algorithms for, e.g. hidden patterns in data, future predictions and decision-making. Wearables, mobile phones, sensors, online media, apps, and similar technologies are all contributing to the production of data volumes that are too large (hence big data) to be stored on local computers, but which can in theory be harnessed in innovative statistical analyses of the kind exploited in machine learning and artificial intelligence (i.e. data science).¹¹

“Real world data”, on the other hand, are usually defined as structured data of the sort often found in registries produced by the healthcare system and other public institutions, and methods of working with such data (e.g. in epidemiology) designed to produce medical evidence (i.e. real world evidence).¹²

It is important to point out that data science methods can just as readily be used on real world data as on big data, although the aims and methodologies are different. Real world evidence, like RCT, often seeks to test hypotheses, whereas big data combine hypothesis generation and validation.¹³ Real world evidence and data science are best thought of as scientific paradigms rather than as the actual data involved.

Key takeaways

- Pharma companies must think of evidence in broad terms going beyond RCT studies in order to accommodate demand from key stakeholders
- Big Data consist of unstructured and ‘found’ data that are not generated for specific purpose
- Real World Data provide more structured data with intentional data collection methods and higher validity

⁷ “FDA's Woodcock: The Clinical Trials System Is ‘Broken’ | RAPS.”

⁸ “FDA's Woodcock: The Clinical Trials System Is ‘Broken’ | RAPS.”

⁹ “Heads of Medicines Agencies: HMA/EMA Joint Task Force on Big Data”; “European Medicines Agency - News and Events - Use of Big Data to Improve Human and Animal Health.”

¹⁰ “Big Data, Medicines Safety and Respiratory Disease Focus of €93 Million IMI Call for Proposals | IMI Innovative Medicines Initiative.”

¹¹ Kitchin, “Big Data, New Epistemologies and Paradigm Shifts.”

¹² Micheelsen, “Beyond Real World Evidence: How to Improve Utilisation of Real World Data throughout an Organisation.”

¹³ Kitchin, “Big Data, New Epistemologies and Paradigm Shifts.”

The proliferation of big data and data science among pharma's stakeholders in Denmark

Data science methods and solutions have already affected the Danish healthcare system, and all of the evidence indicates that big data, machine learning and AI will have a significant impact in terms of improved care and outcomes, and the cost of treatment in the future.

"We will optimise the use of data from registers for [safe] signal generation as well as for evaluating and validating risk minimisation measures. We will work strategically with new data sources and develop a plan for the use of big data for pharmacovigilance."

DANISH MEDICINE AGENCY, DRUG MONITORING STRATEGY 2017 (AUTHORS TRANSLATION)¹⁴

EMA's taskforce is chaired by the Danish Medicine Agency, which in turn has raised the issue of using big data in the debate among Danish stakeholders.¹⁵ This may pave the way for swift adoption of the taskforce's recommendations in Denmark.

The Danish government, regional bodies (*Danske regioner*) and Local Government Denmark (KL) have also all made plans and launched strategies with a view to ensuring that, with the ongoing digitalisation of public healthcare, data science methods and solutions are used to improve treatment outcomes, patient safety and research, and reduce treatment costs.¹⁶ Already, big data projects combining data science skills with domain specific expertise have emerged in public healthcare. Among other things, these have improved treatment in intensive care units, led to better predictions of which patients will need acute hospitalisation, and helped to shape tailored treatment regimens and personalised medicine more generally.¹⁷

The development of data science solutions in the healthcare sector runs in parallel with the centralisation and digitalisation of public health, and data science will thus become an integral part of the healthcare system and treatment in general. It is anticipated that by 2019 patients visiting their general practitioner (GP) will have their own app ("My Doctor") which, among other things, will include mail and video consultation, consultation booking and lab results.¹⁸ Healthcare practitioners (HCPs) will soon have access to an intelligent medicine-decision tool to guide them in prescribing safe medicine with reference to each patient's medical history.¹⁹ Emergency call operators will be supported by machine learning algorithms which can detect in up to 9 out of 10 cases whether a patient is in cardiac arrest.²⁰ New data sources are also being built and implemented. The upcoming hospital medicine registry will collect data on all medicine usage in Danish hospitals at patient level, new electronic patient journal systems are being implemented (although the process has not been without difficulty) and patient reported outcomes (PROs) are beginning to be collected.²¹

As a consequence, health authorities will soon have the infrastructure to conduct their own real-time real world outcome studies of new medical treatment, and connect outcomes from these studies with, for example,

¹⁴ Lægemiddelstyrelsen, "Strategi for Lægemiddelovervågning i Lægemiddelstyrelsen."

¹⁵ "Heads of Medicines Agencies: HMA/EMA Joint Task Force on Big Data."

¹⁶ "SIRI-Kommissionen | IDA.Dk"; Sundheds- og Ældreministeriet et al., "Et Sikker Og Sammenhængende Sundhedsnetværk for Alle: Strategi for Digital Sundhed 2018-2022"; KL, "Bedre Sammenhæng for Borgere Og Virksomheder - Udspil."

¹⁷ "Tværspor - et Tværsektorielt Projekt i Patientens Spor - Regionshospitalet Horsens"; "Sundheds- Og Ældreministeriet Nationalt Genom Center - Sum.Dk"; "Big Data Skal Udfylde Huller i Viden Om Intensivpatienter."

¹⁸ "Regeringsudspil Bygger På Data Redder Liv."

¹⁹ "Trifork Og pro.Medicin.Dk Vinder Nationalt Udbud."

²⁰ "112-Personale Får Hjælp Af Kunstig Intelligens."

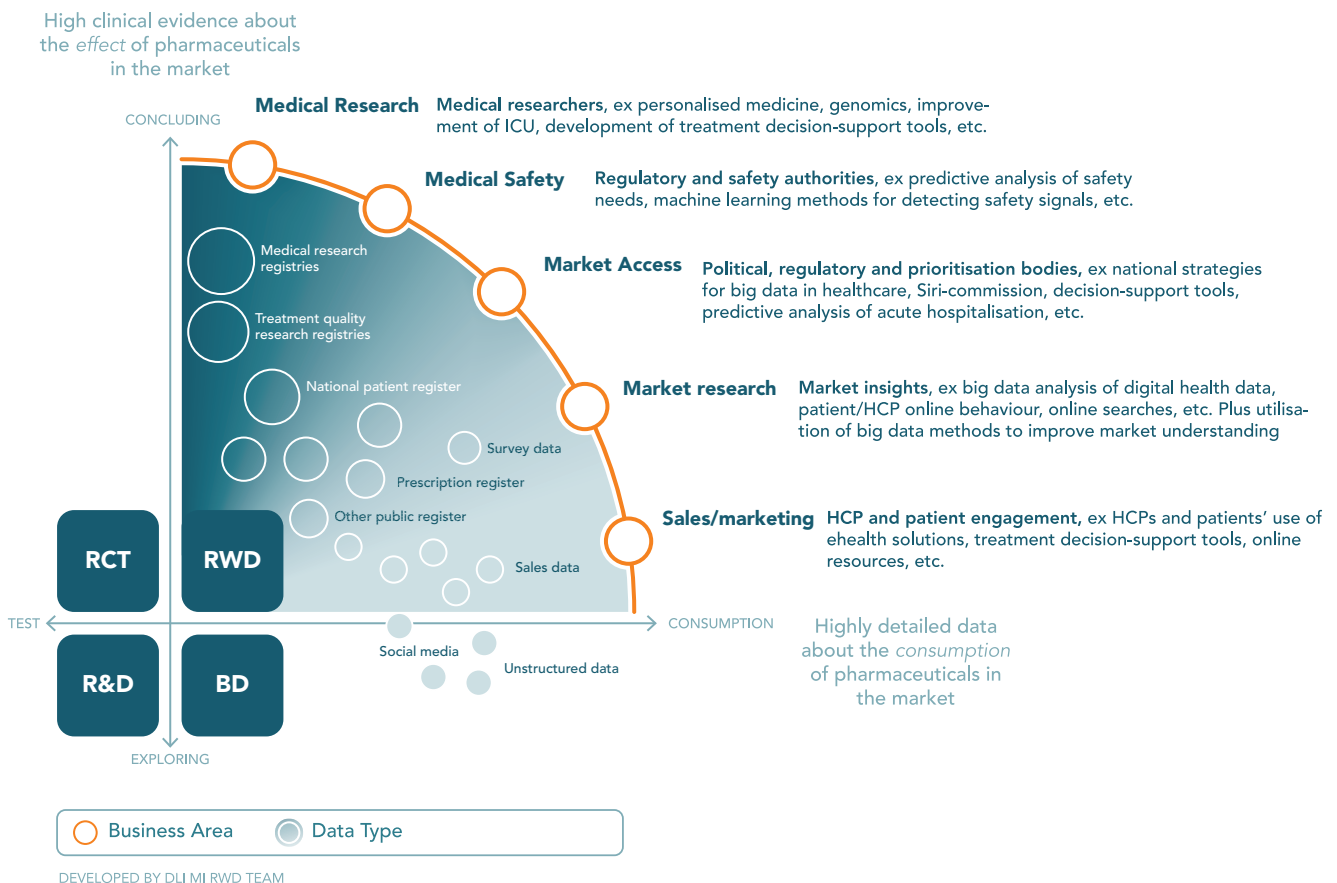
²¹ "Sundhedsplatformen - Én Samlet Patientjournal"; "PRO - PRO"; "Økonomaftale Sikrer Nyt Nationalt Register for Sygehusmedicin - Dagens Medicin."

treatment costs and PRO data. Such “commissioning through evaluation big data solutions” will inevitably strengthen the position of priority bodies when negotiating tenders for medicine.²²

Patients appear to expect health authorities to develop digital solutions based on their willingness to use online solutions such as apps and email in their interaction with HCPs. ²³ Indeed GP email consultations have risen from 856 consultations per day in 2007 to 5,139 per day in 2017.²⁴ In other words, significant political capital, taxpayers’ money, public willingness and scientific labour are already invested in the application of data science methods to improve current treatment and cut costs by harnessing the potential of digitalisation in healthcare (Fig 2.).

“ The [Danish] government, together with parties in the healthcare sector in regions and municipalities, want to make the right framework for investment and utilisation of future innovative technologies in Denmark, such as artificial intelligence, ‘big data’ and personal medicine”
DATA SAVES LIVES – GOVERNMENT INITIATIVE, 2018 (AUTHORS TRANSLATION) ²⁵

Fig 2. The proliferation of data science among pharma’s stakeholders



²² Andersen and Vilhelmsen, Notat - Danske Regioner: Model for vurdering af lægemidler.
²³ “MindApps”; DST, “Personer Med Kontakt Til Læge Mv. Med Offentlig Tilskud Efter Område, Ydelsesart, Alder, Køn Og Socioøkonomisk Status - Statistikbanken - Data Og Tal.”
²⁴ DST, “Personer Med Kontakt Til Læge Mv. Med Offentlig Tilskud Efter Område, Ydelsesart, Alder, Køn Og Socioøkonomisk Status - Statistikbanken - Data Og Tal.”
²⁵ “Regeringsudspil Bygger På Data Redder Liv.”

Pharma companies may be inclined to view these developments as a challenge for the future. However, given that the FDA and EMA, as well as the Danish Medicine Agency, all levels of Danish governmental bodies, hospitals, medical researchers and HCP have invested in data science solutions, they, and local HCPs and public health authorities, will soon challenge pharma companies based on the insights and solutions gained by big data and data science.

To illustrate the point: since HCPs will come to use AI-driven decision-support solutions for medical treatment, the underlying machine learning algorithms steering treatment options will need to be integrated in treatment guidelines, assessments of treatment cost, safety assessments, and so on. Thus, machine learning algorithms will become an integral part of treatment, and pharma companies will need to understand the implications of the deployment of big data and data science methods in order to meet the challenges presented to them by local HCPs and public health authorities.²⁶

Key takeaways

- Market access and sales have been increasingly centralized and decision power is moving from clinical practice to central administration
- Recent changes in the stance of key regulatory bodies towards a broader evidence base for the prioritisation of treatments mean that pharma companies will need to employ data science to stay competitive

The traditional organisation of commercial pharma

The proliferation of big data in public healthcare is a two-edged sword for pharma. On the one hand, if pharma companies begin utilising big data sources and data science methods, there is a real opportunity for them to strengthen their position in their market by gaining a new understanding of the ways in which their medicines are used, perceived and improve patient health. On the other hand, however, pharma companies must continue to work on developing their organisational frameworks and methods of producing evidence in order ensure they successfully exploit the new data science capabilities.

In our last white paper *From Evidence to Consequence* we described how each business area within the commercial section of a pharma organisation has its own distinctive knowledge paradigm (Fig 3).²⁷

²⁶ Jiang et al., "Artificial Intelligence in Healthcare: Past, Present and Future"; Contreras and Vehi, "Artificial Intelligence for Diabetes Management and Decision Support: Literature Review"; "Machine Learning Techniques Show Promise for Supporting Medical Decisions - American College of Cardiology."

²⁷ Hirsch and Micheelsen, "From Evidence to Consequence – the New Reality for Prioritisation of Medicines in Denmark."

CASE 1 - MEASURING BRAND ATTENTION AND LOYALTY OF PRESCRIBERS VIA BIG DATA ANALYSIS

Background: It is increasingly difficult for pharmaceutical companies to target and influence prescribers. The traditional role of sales reps is being challenged in a more centralised healthcare system. However, with a tendency for more patients to be moved from specialists to GPs, it becomes increasingly important to engage in productive dialogue with GPs. While effective marketing and sales activities should eventually generate increased sales, it is difficult to see underlying shifts in the market, or to follow up on marketing and sales activities, by using sales data alone. We used a big data approach with the aim of measuring the dynamics of prescriber attention and prescriber loyalty.

Solution: In this case, we evaluated the market for five products through big data analysis of online behaviour data. We harvested aggregated non-personal data at a national level from promedicin.dk, which has more than 46 million site views each year. Promedicin.dk is consulted by health practitioners prescribing drugs and is the preferred site for Danish prescribers seeking medical information about pharmaceuticals.

Deep analysis of the data reveals that sales and brand attention develop in opposite direction for 37% of all products, emphasising that prescribers' brand attention (as exhibited in the searches) does not correspond with sales as a rule. Using big data analytics, we were able to identify the association between attention and sales to measure prescribers' loyalty. We used the number of search hits on brand web pages on promedicin.dk as a measure of prescriber attention to a brand. Then, using the heuristic relationship

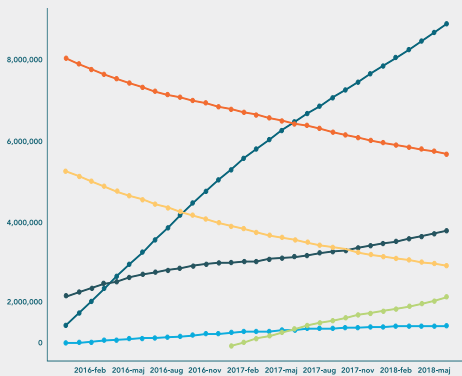
$$\text{Sales} = \text{Attention} \times \text{Loyalty}$$

we measured prescriber loyalty as Values sold/Number of hits. Below are graphs showing sales value, prescriber attention and prescriber loyalty from November 2015 to July 2018. The curves are smoothed to evaluate long-term trends.

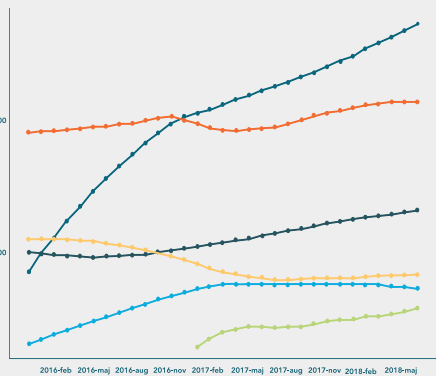
Insight: By utilising an online big data source, we were able to identify a shift in loyalty away from established products when new products entered the market. We were also able to observe prescriber attention levelling out long before sales. The solution presented thus gave an early indication of the long-term performance of the products.

Analysing the unsmoothed curves further allowed us to evaluate the effect of specific marketing and sales activities.

SALES (IN VALUE)



ATTENTION (NO. OF HITS)



LOYALTY (SALES/HITS)

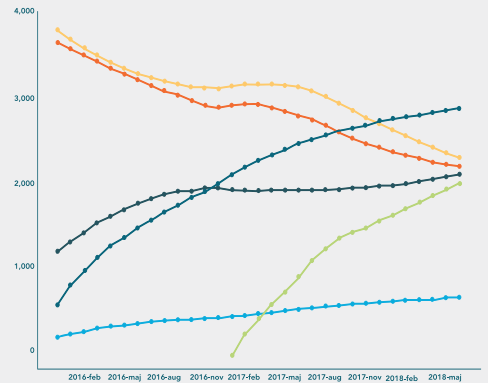
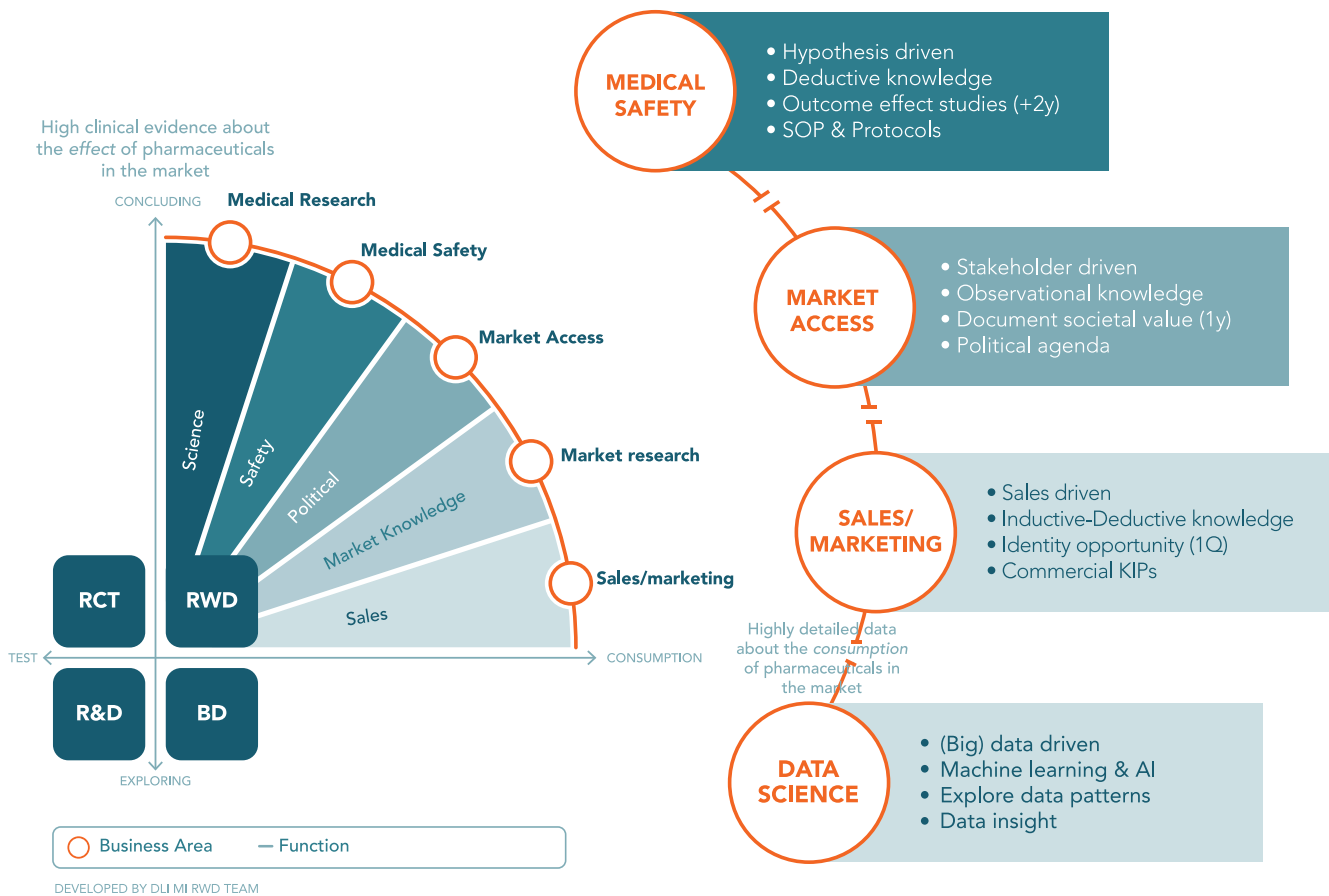


Fig 3. The risk of siloing data science in pharma organisations



Medical and safety groups’ disciplinary training encourage them to view evidence from RCT as the strongest type of knowledge. Knowledge, and the evidence that supports it, is produced by falsifying hypotheses via deductive reasoning that is governed by both scientific and regulatory thinking, with an emphasis on compliance with research protocols and Standard Operational Procedures (SOPs).

Marketing and sales units, by contrast, seek to identify sources of business and short-term commercial opportunities. Market Access specialists are often driven by external stakeholders’ agendas, aiming to build strong arguments that support product claims in a societal context. Hence all of these knowledge domains – medical, safety, market access, sales, marketing, etc. – tend to have different views of what constitutes “knowledge”. In a similar way, of course, data scientists also have their own specific knowledge domain and assumptions about “knowledge” – e.g. assumptions revolving around big data, machine learning and AI. It can be seen, then, that the addition of data science capabilities to the organisation needs to proceed with the clear aim of supporting each business area and collaboration across silos.

Key takeaways

- Pharma companies will need to rely less heavily on RCT evidence if they are to appropriate data science capabilities
- This will require the implementation of an appropriate additional organisational framework and the development of existing practices

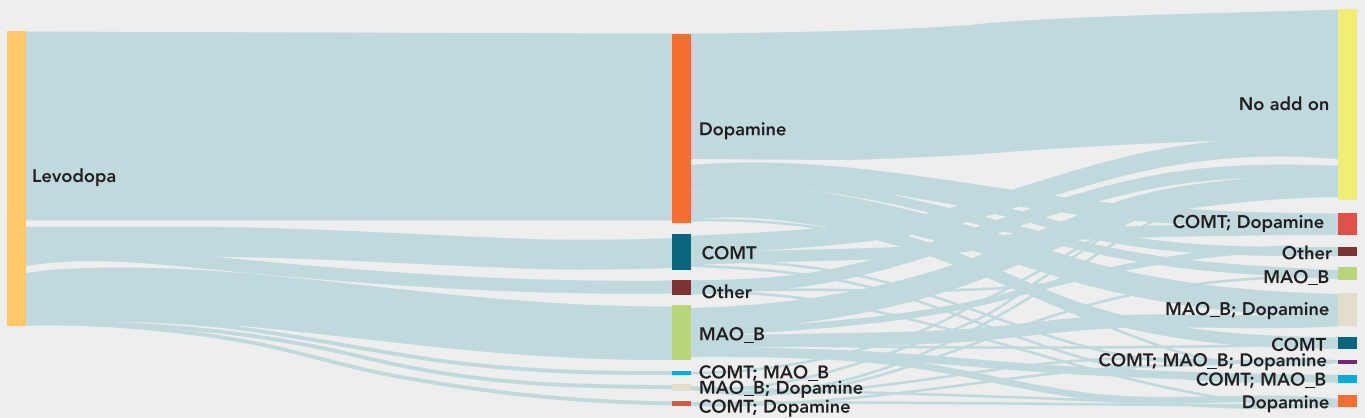
CASE 2 – BUILDING A TREATMENT-DYNAMIC MODEL VIA MACHINE LEARNING

Background: Parkinson’s is a neurological disease affecting between 51 and 177 people per 100,000 worldwide.²⁸ Standard treatments include Levodopa, Dopamine agonists, COMT inhibitors and MAO B inhibitors. Whether a company is entering this market, trying to optimise sales, or negotiating with the healthcare system, a thorough understanding of patient dynamics is necessary. We aimed to understand the patient dynamics in one of the Nordic countries.

Solution: Prescription data were extracted for all patients diagnosed with Parkinson’s in 2012-2014, and the patients were tracked to the end of 2017. Parkinson’s is a progressive disease, and sufferers tend not to switch treatment. Furthermore, treatment guidelines were unchanged in the study period.

Insight: Below we present an empirical model of the patient dynamics for the pooled cohorts – including minute changes in treatment patterns. The most common add-on to Levodopa was a Dopamine agonist. More than 60% of all patients never tried a third drug within the study period. Among those who did, the most common combination was MAO B inhibitor and Dopamine agonist in addition to Levodopa.

Although the model below is only a starting point, it provides solid common ground for further collaboration between domain experts in the medical area, in market access, and in data science. Indeed, when it is enriched with a local understanding of demographics, treatment practices and budget priorities, the model can support machine learning extrapolations and answer questions about future treatment dynamics, sales, patient populations and the cost of treatment. Collaboration between domain experts and data scientists with solid programming and statistical skills is key in capturing the potential of patient data of this kind.



²⁸ Feigin et al., “Global, Regional, and National Burden of Neurological Disorders during 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015.”

Implementing data science capabilities in commercial pharma

Pharma is in many ways a data-driven industry, but the data are used primarily for evidence generation and reporting. To a lesser extent, data are used to support business functions and decision-making via advanced data science analytics. To be successful in a period of transition it is necessary to make changes in the organisational framework and re-evaluate pharma's reliance on RCT evidence. The question is: what steps should a pharma organisation consider if it wishes to begin using data science solutions and build its own data science capabilities, bringing benefits across the organisation?

With years of experience working with real world data and big data projects for clients in the Nordic region, DLiMI have identified two critical steps in the successful integration of data science capabilities in local businesses:

1. A first step in integration driven by adding **value** to all business areas by supporting each area with domain specific big data sources and data science methods
2. A second step in integration focusing on implementing data science capabilities in all business areas, enabling each area be ready to **challenge** the decisions that public health authorities, "key opinion leaders" (KOLs) and researchers will make based on big data and data science methods

These two steps can be difficult and time-consuming to accomplish. Before we move on to discuss the specifics of each step we outline some of the key factors in a successful transition.

- **Determine which kind of data scientist you need.** Do you need someone to upgrade your capabilities to work with patient data? Do you need someone to drive automation of KPI reporting? Or do you need someone to develop a new pricing algorithm? Depending on your needs, you will be looking for a different kind of data scientist with a different set of technical competencies. It is important to acknowledge that the data management area is diverse in terms of skills and competencies, and to be realistic about your needs.
- **Provide strong leadership.** Strong leadership is always important, but it can be particularly lacking when it comes digital transformations. In the data science area, one of the common mistakes is to assume that data scientists will be able to combine data sources, discover hidden insights, and perhaps even make predictive algorithms, without a clear vision being set by the commercial leadership on what they need. Ensuring that the data scientists are integrated in the relevant business areas is essential, and that require leaders who bring an understanding and vision of how and when data science methods are to be applied.
- **Work out how to obtain the right data.** Finding the right data for the right task is often a major if not impossible challenge. However, it is important to remember that a solution that gives half of the answers is better than none. Obtaining access to and collecting data can be costly and time-consuming, and in most cases it is merely a prerequisite of data science, leaving much of the "science" to be done. The importance of having a data strategy in place across all business areas should not be underestimated.

CASE 3 – IDENTIFYING KEY OPINION LEADERS VIA NATURAL LANGUAGE PROCESSES AND NETWORK ANALYSIS

Background: Prior to entering a new market, it is important to have good dialogue with key opinion leaders on the relevant therapeutic area. This is especially important if the drug is being evaluated in the Medicines Council. As pharmaceutical companies are not permitted access to members of the Decision Board at the Medicines Council, it is important to have a strategy on how to engage with the key opinion leaders.

Solution: The search for key opinion leaders in a given area can be approached from several angles. In this specific example, the primary focus was on key academic opinion leaders. By searching PubMed, we found 849 publications by 143 Danish authors in the given area between 2002 and 2017. By using natural language processing techniques, we connected all authors who had co-authored at least two publications to arrive at the network shown below. In the network, each node is an author and each link indicates co-authorship. The size of each node represents the number of publications, with special weight attached to first and last authorship.

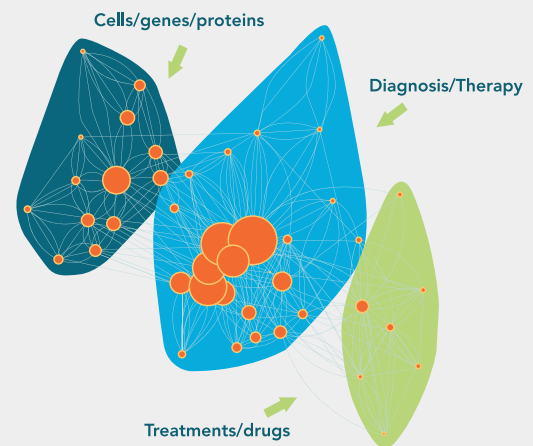
Insight: Data science analysis enabled us to rank the key academic opinion leaders. Analysing their rankings over time we were able to evaluate whether their influence was increasing or decreasing.

We further analysed the publications of the top opinion leaders to find out what their main research interest was and determine whether there were any groups of authors with similar interests. This served two purposes. First, it helped to identify key opinion leaders to be prioritised. Second, it enabled the market access and medical team to prepare effectively before engaging with each opinion leader.

RANKING OF STAKEHOLDER

Rank	Name	Importance	Interest
1	[Name Redacted]	Stable	Patients, treatment, immunology
2	[Name Redacted]	Increasing	Patients, treatment, diagnosis, bile
3	[Name Redacted]	Increasing	Elderly, observational studies, high risk population
4	[Name Redacted]	Decreasing	Cells, gene expressions, clinical studies
5	[Name Redacted]	Increasing	Surgery, complications, diagnosis
6	[Name Redacted]	Stable	Statistics, genetics, associations, analysis
7	[Name Redacted]	Decreasing	Patients, clinical studies, therapy, observational study
8	[Name Redacted]	Decreasing	Associations, gene, cells, pathology
9	[Name Redacted]	Increasing	Patients, comorbidity, population cohorts
10	[Name Redacted]	Stable	Patients, complications, therapy, quality of life

STAKEHOLDERS' RESEARCH INTEREST NETWORK



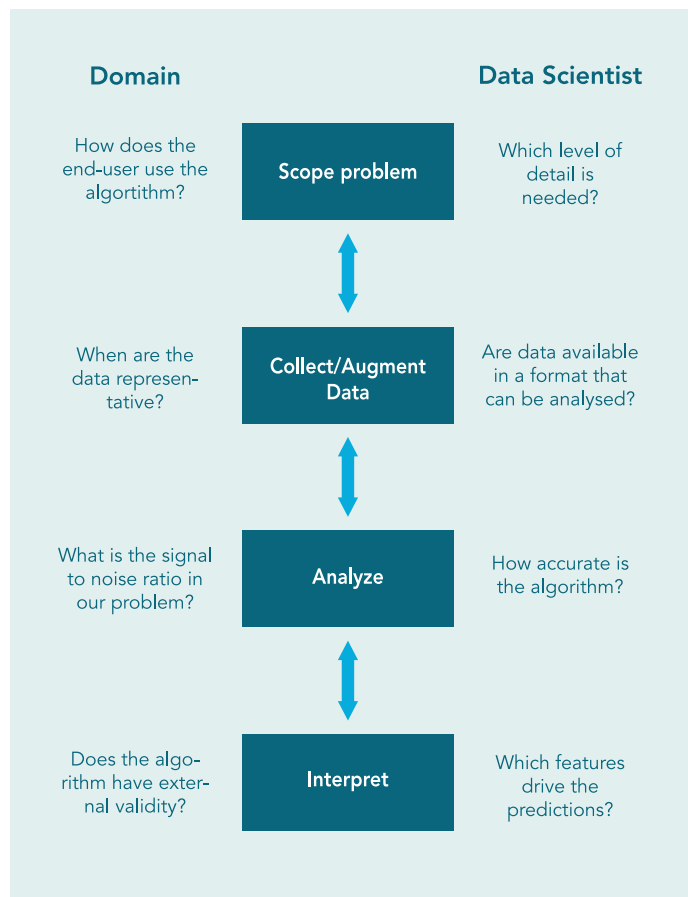
Step 1: Add value to all business areas in the commercial section of pharma

As the cases included in this white paper demonstrate, data science methods can add real value to different business areas within the commercial section of pharma, and this should be the driving force in Step 1. The key thinking in Step 1 is to make data science capabilities available in each area with the goal of adding value to concrete projects. Thus, data scientists will be used to analyse and solve specific challenges within each business area, and will therefore be dependent on the utilisation of local domain expertise. This way of thinking goes against a popular understanding of big data as a method in which data alone provide answers.²⁹ On the contrary, our experience at DLiMI shows that data science methods only add value when relevant domain experts have formulated the issue to be analysed and have been given the opportunity to use their expertise to enrich the big data analysis (the cases we have presented illustrate this).

Fig 4. **Step 1:** Collaboration between data science experts and specific domain experts, specifically on the development of a predictive algorithm.

Using the development of a predictive algorithm as an example, Figure 4 outlines some of the key questions the domain expert needs to address with the data scientist and be able to discuss with the healthcare system. While the questions on the right are usually those answered by the data scientist, and the questions on the left are usually those answered by a domain expert, it should be noted that these questions cannot be answered in isolation. It is important that both domain experts and data scientists have the knowledge and skill to thoroughly answer all the questions.

In the course of the collaboration, the data science experts are introduced to the local challenges and domain thinking used to solve these challenges in each business area, and domain experts are likewise introduced to data science methods and solutions. This approach does, however, mean that data science teams must be organised in the business areas they support and not in a centralised manner as a support function.



One example of this involves the analysis of medical publications databases to identify key opinion leaders, their collaborators and their areas of interest (see Case 3). This could be particularly relevant when a company is entering a new therapeutic area covering rare diseases or finding the best participants for an advisory board. In a healthcare system in which it is increasingly hard to know who the key decisions makers are, a structured approach, coupled with domain knowledge expertise, can be crucial in reaching the right stakeholders with the right message.

²⁹ Kitchin, "Big Data, New Epistemologies and Paradigm Shifts."

A full data-driven approach to this question would not be feasible. However, an approach in which domain expert knowledge and state of the art AI are combined may lead to new insights into how to make a successful application to e.g. the Danish Medicines Council without manually correlating thousands of medical publications with outcomes in several hundred decisions made by the Council.

Step 2: Be ready to use big data and data science methods to challenge local decisions

Following the first step – the integration of data science capabilities – each business area will gradually prepare to work with data scientists and understand the benefits and limitations of data science. The long-term goal, however, is to prepare each business area to engage with, challenge and improve the big data solutions and insights that external stakeholders will base their future decisions on.

Fig 5. **Step 2:** Examples of how data science will affect core business areas in commercial pharma

	Examples of focus points in the current healthcare system for each business area	Examples of new focus points in a data science healthcare
Medical affairs	<ul style="list-style-type: none"> • Communicate RCT evidence on the efficacy & safety of company brands to medical bodies • Engage KOL & researchers on medical guidelines & (RCT) research related to company brands 	<ul style="list-style-type: none"> • Communicate evidence on company brands from RCT combined with “evidence” derived from big data sources on efficacy & safety • Engage KOL & researchers on how company brands should be included in local AI-supported treatment decision systems & guidelines, and how machine learning algorithms for e.g. personal/precision medicine, treatment outcome studies, etc. should be designed and evaluated
Medical safety	<ul style="list-style-type: none"> • Monitor reported AE (adverse event) signals for company brands and ensure pharmacovigilance • Engage safety bodies & KOL on safety issues and local safety studies related to company brands 	<ul style="list-style-type: none"> • Uphold responsibilities within safety, pharmacovigilance and safety studies for company brands including reported AV signals derived from big data analysis • Engage safety bodies & KOL on safety signals derived from machine learning analysis of big data, potential safety issues based on predictive algorithms and local big data safety studies related to company brands
Market Access	<ul style="list-style-type: none"> • Communicate societal benefits of company brands in relation to cost of drug/treatment based on (generic) health-economic models • Access priority bodies and stakeholders in the healthcare sector on the value of company brands/treatment 	<ul style="list-style-type: none"> • Communicate societal benefits of company brands in relation to cost of drug/treatment based on health-economic algorithms that are built and trained on local (big) data to priority bodies and tender authorities • Engage prioritisation bodies, tender authorities and healthcare stakeholders on the design and evaluation of local health-economic algorithms
Market Research	<ul style="list-style-type: none"> • Analyse and identify sources of business based on sales data, business intelligence data, and qualitative/quantitative survey studies • Support internal commercial business functions with market insights, e.g. market dynamics, product perception, segmentation & targeting of HCP, etc. 	<ul style="list-style-type: none"> • Identify and predict sources of business based on a diverse set of online and offline data sources via machine learning algorithms • Support internal commercial business functions with evidence-based decision tools for market decisions, scenario analysis and predictive analysis
Sales/ Marketing	<ul style="list-style-type: none"> • Promote company brands to HCP, KOL and healthcare stakeholders with the aim of ensuring awareness and sales • Engage HCP on company brands value, efficacy, safety and patient value 	<ul style="list-style-type: none"> • Promote company brands to HCP with an understanding of how HCP and patients use online tools, the AI decision-support tool and online resources • Engage HCP with digital tools and resources on company brands’ value, efficacy, safety and patient value

Fig. 5 illustrates some of the new capabilities that will come to mark out commercial pharma operating in a healthcare system with integrated data science solutions.

The medical and pharmacovigilance departments in commercial pharma will need to be informed by a more nuanced understanding of the way, or rather ways, in which evidence is produced which includes machine learning and AI analysis of big data sources. As mentioned above, the sector, and especially its research base, currently relies heavily on evidence from RCT studies. However, this will change when medical research and treatment begin to fully utilise big data, machine learning and AI to make informed decisions about optimal treatments.

“ [...] machine learning may usher in a new era in digital health care tools capable of enhancing health care delivery by aiding routine processes and helping physicians assess patients’ risk”
THE AMERICAN COLLEGE OF CARDIOLOGY 2018 ³⁰

In a future healthcare system that utilises data science solutions, treatment guidelines will not need to rank drugs within a certain category as first-line, second-line, third-line, etc. Instead, numerous factors will be taken into account based on the data available – potentially, data on all known relevant scientific literature, individual patient characteristics (such as comorbidities, co-medication and disease severity) and the outcomes of all other locally treated patients within the healthcare system. Likewise, machine learning algorithms analysing big data from a variety of health data sources will be able to identify safety signals and predict potential safety issues within local patient populations.

For market access, understanding how prioritisation bodies design, train and use machine learning algorithms in their evaluations of medical treatments and tenders will be of great importance. Locally developed machine learning algorithms will thus replace (generic) health-economic models. However, as the implementation of data science solutions begins to support resource planning, budget management and other administrative decisions affecting treatment in general within the healthcare system, market access must also be in a position to challenge the resulting decisions.

For market research and sales/marketing, understanding how HCPs and patients use digital as well as data science solutions will be of paramount importance, since this will inform, for example, the way in which HCPs and patients make decisions about treatment and find information. A knowledge of these aspects of HCP and patient behaviour will help those in market research and sales to determine what information they need to engage with, and which digital footprints they need to analyse. Above, in Case 1, we used data from the Danish website prodmedicin.dk to measure the attention and loyalty of prescribers. This produced a nuanced picture of the current market situation and enabled effective follow up in market activities. Market research will thus be able to monitor, analyse and train predictive algorithms to support business decision-making.

In essence, pharma will need to be able to understand, engage with and challenge the use of data science solutions by researchers, medical authorities, HCPs, and priority and tender authorities. They need to be a po-

³⁰ “Machine Learning Techniques Show Promise for Supporting Medical Decisions - American College of Cardiology.”

sition to question and perhaps contest the decisions that these stakeholders make on the basis of the insights derived from data science.

Since machine learning algorithms and AI solutions need to be trained on local health data and/or by local specialists in order to ensure they result in insights with external validity, implementing a generic data science solution or algorithm is not an option.³¹ Healthcare systems vary tremendously from one region or country to another in terms of needs, resources and politics. Thus pharmaceutical companies must be able to engage with the healthcare system at a local level over the best data science solutions to select. Support from a global organisation will not be sufficient.

The questions outlined in Figure 4 are questions that local pharmaceutical companies must find some way of being in dialogue with the healthcare system about. This requires both domain experts with local expertise and data scientists with a deep and wide understanding of local business. Building this expertise takes time and practice.

Key takeaways

- Close collaboration between data scientists and domain experts is paramount to the successful utilisation of big data and real world data
- Data science must be integrated into all domains with an emphasis on market access and sales/marketing
- It is vital to challenge existing practices and infuse evidence-based decision-making in all business domains

Conclusion

The Danish healthcare system is digitalising on all fronts, and the quantity of healthcare data growing day by day. With the current acceleration in the use of big data, AI, and machine learning, it is not question of whether these tools will be used to prioritise medication and treatment, but rather when and how. This offers a real competitive advantage to pharmaceutical companies which integrate data science in all parts of their business operations. However, the integration must be achieved without creating a data science silo disconnected from sales, marketing, market access and medical. The best way to start the transition is by adding value to current business operations with concrete data science projects.

³¹ "IBM Pitched Watson as a Revolution in Cancer Care. It's Nowhere Close."

References

"112-Personale Får Hjælp Af Kunstig Intelligens." Accessed July 17, 2018. <https://sundhedspolitisktidsskrift.dk/nyheder/683-112-personale-far-hjaelp-af-kunstig-intelligens.html>.

Andersen, Thomas Birk, and Ann Vilhelmsen. Notat - Danske Regioner: Model for vurdering af lægemidler (2016).

"Big Data, Medicines Safety and Respiratory Disease Focus of €93 Million IMI Call for Proposals | IMI Innovative Medicines Initiative." IMI, 2015. <https://www.imi.europa.eu/news-events/press-releases/big-data-medicines-safety-and-respiratory-disease-focus-eu93-million-imi>.

"Big Data Skal Udfylde Huller i Viden Om Intensivpatienter." Accessed July 17, 2018. <https://www.rigshospitalet.dk/presse-og-nyt/nyheder/nyheder/Sider/2016/April/big-data-skal-udfylde-huller-i-viden-om-intensivpatienter.aspx>.

Bothwell, Laura E., Jeremy A. Greene, Scott H. Podolsky, and David S. Jones. "Assessing the Gold Standard — Lessons from the History of RCTs." Edited by Debra Malina. *New England Journal of Medicine* 374, no. 22 (June 2, 2016): 2175–81. <https://doi.org/10.1056/NEJMms1604593>.

Contreras, Ivan, and Josep Vehi. "Artificial Intelligence for Diabetes Management and Decision Support: Literature Review." *Journal of Medical Internet Research* 20, no. 5 (May 30, 2018): e10775. <https://doi.org/10.2196/10775>.

DST. "Personer Med Kontakt Til Læge Mv. Med Offentlig Tilskud Efter Område, Ydelsesart, Alder, Køn Og Socioøkonomisk Status - Statistikbanken - Data Og Tal." www.statistikbanken.dk/SYGP. Accessed July 18, 2018. <http://www.statistikbanken.dk/statbank5a/selectvarval/define.asp?PLanguage=0&subword=tabssel&MainTable=SYGP&PXSid=110695&tablestyle=&ST=SD&buttons=0>.

"European Medicines Agency - News and Events - Use of Big Data to Improve Human and Animal Health." Accessed July 18, 2018. http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2017/03/news_detail_002718.jsp&mid=WC0b01ac058004d5c1.

"FDA's Woodcock: The Clinical Trials System Is 'Broken' | RAPS." Accessed July 18, 2018. <https://www.raps.org/regulatory-focusTM/news-articles/2017/9/fda-s-woodcock-the-clinical-trials-system-is-broken>.

Feigin, Valery L, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Foad Abd-Allah, Abdishakur M Abdulle, Semaw Ferede Abera, Gebre Yitayih Abyu, et al. "Global, Regional, and National Burden of Neurological Disorders during 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015." *The Lancet Neurology* 16, no. 11 (November 2017): 877–97. [https://doi.org/10.1016/S1474-4422\(17\)30299-5](https://doi.org/10.1016/S1474-4422(17)30299-5).

"Heads of Medicines Agencies: HMA/EMA Joint Task Force on Big Data." Accessed July 18, 2018. <http://www.hma.eu/509.html>.

Hirsch, Niels-Christian, and Arun Micheelsen. "From Evidence to Consequence – the New Reality for Prioritisation of Medicines in Denmark," 2017. https://www.dli-mi.dk/SiteCollectionDocuments/From_evidence_to_consequence_DLI_MI.pdf.

"IBM Pitched Watson as a Revolution in Cancer Care. It's Nowhere Close." *STAT*, 2017. <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>.

Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang.

"Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* 2, no. 4 (December 31, 2017): 230–43. <https://doi.org/10.1136/svn-2017-000101>.

Kitchin, Rob. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1, no. 1 (2014): 205395171452848. <https://doi.org/10.1177/2053951714528481>.

KL. "Bedre Sammenhæng for Borgere Og Virksomheder - Udspil," 2018.

Lægemiddelstyrelsen. "Strategi for Lægemiddelovervågning i Lægemiddelstyrelsen," no. December 2016 (2017).

"Machine Learning Techniques Show Promise for Supporting Medical Decisions - American College of Cardiology." American College of Cardiology Foundation, 2018. <https://www.acc.org/About-ACC/Press-Releases/2018/02/26/14/47/Machine-Learning-Techniques-Show-Promise-for-Supporting-Medical-Decisions>.

Micheelsen, Arun. "Beyond Real World Evidence: How to Improve Utilisation of Real World Data throughout an Organisation." Copenhagen, 2016. https://www.dli-mi.dk/SiteCollectionDocuments/Beyond Real World Evidence_DLI MI.pdf.

"MindApps." Psykiatrien i Region Syddanmark. Accessed July 17, 2018. <https://mindapps.dk/>.

"Økonomiaftale Sikrer Nyt Nationalt Register for Sygehusmedicin - Dagens Medicin." Accessed July 18, 2018. <https://dagensmedicin.dk/okonomiaftale-sikrer-nyt-nationalt-register-for-sygehusmedicin/>.

Press, Gil. "6 Predictions For The \$203 Billion Big Data Analytics Market." *Forbes*, 2017. <https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#62fe3b3e2083>.

"PRO - PRO." Accessed July 17, 2018. <https://sundhedsdatastyrelsen.dk/subsites/pro/da/pro>.

"Regeringsudspil Bygger På Data Redder Liv." Accessed July 17, 2018. <http://www.cphhealthtech.dk/nyheder/2018/regeringens-nye-strategi-sundhed-i-fremtiden>.

"SIRI-Kommissionen | IDA.Dk." Accessed July 17, 2018. <https://ida.dk/ida-star/siri-kommissionen>.

Sundheds- og Ældreministeriet, Finansministeriet, Danske Regioner, and KL. "Ét Sikkert Og Sammenhængende Sundhedsnetværk for Alle: Strategi for Digital Sundhed 2018-2022," 2018. https://sum.dk/Aktuelt/Nyheder/Digitalisering/2018/Januar/~~/media/Filer - Publikationer_i_pdf/2018/Strategi for digital sundhed/Strategi for digital sundhed_Pages.pdf.

"Sundheds- Og Ældreministeriet Nationalt Genom Center - Sum.Dk." Accessed August 25, 2018. <https://www.sum.dk/Temaer/Personlig-medicin/Nationalt-Genom-Center.aspx>.

"Sundhedsplatformen - Én Samlet Patientjournal." Accessed July 17, 2018. https://www.regionh.dk/sundhedsplatform/om-sundhedsplatformen/Sider/en_samlet_patientjournal.aspx.

"Trifork Og pro.Medicin.Dk Vinder Nationalt Udbud." Accessed July 18, 2018. <https://min.medicin.dk/Generelt/Nyheder/400>.

"Tværspor - et Tværsektorielt Projekt i Patientens Spor - Regionshospitalet Horsens." Accessed July 17, 2018. <http://www.regionshospitalet-horsens.dk/om-os/virksomhedsgrundlag/strategiske-indsatser/tvraspor/>.

Authors



Chief Market Researcher Arun Micheelsen is part of DLI MI's Real World Insight team. At DLI MI, Arun has conducted numerous RWD studies for the large pharmaceutical companies operating in the Nordic region. Arun holds a PhD in sociology and has extensive expertise in Real World Insight studies.



Data Scientist Lars Lynne Hansen is leading the data science activities at DLI MI. Lars has several years of experience working with drug development and market access in the industry and holds a MSc in Statistics.



Assistant Professor Jonas Valbjørn Andersen teaches research methods, evidence-based management and data analytics at the Department of Business IT, IT University of Copenhagen. Jonas has worked for major IT companies within the ERP and online collaboration industry and earned his PhD in information systems and management from Warwick Business School, UK.

About DLI MI

DLI Market Intelligence (DLI MI) is a leading provider of business intelligence and Real World insight to pharmaceutical and health-related companies operating in the Nordic region. We offer pharmaceutical sales statistics on a variety of platforms as well as market research, Real World Studies and consultancy services covering Denmark, Sweden, Norway and Finland.