



Big Data in healthcare

Definition, application,
and challenges

Betere gezondheid
door betere informatie

Nictiz 

Author

Stefan Ottenheijm

Editor

Karin Oost

Advice

Johan Krijgsman
Lucien Engelen
Tom van de Belt
Roel Schenk

Design

Media&More

Graphics

Charlotte Schreuder

Translation

Alinea Lingua

Nictiz**Visiting Address**

Oude Middenweg 55
2491 AC The Hague
The Netherlands
T 070 – 3173450
F 070 – 3207437

Postal address

P.O. Box 19121
2500 CC The Hague
The Netherlands

Copyright

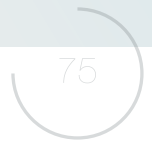
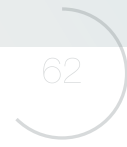
Nictiz, 2017

Contents

1. Introduction	4
2. Big Data: a definition	8
2.1 From data to wisdom	9
2.2 More, faster, better	9
2.3 More complex = more valuable?	10
2.4 A definition	11
3. Big Data: application in healthcare	12
3.1 Opportunities for research	14
3.2 Prevention is better than cure	15
3.3 High-quality, safe and affordable healthcare	16
3.4 Personalised healthcare	18
3.5 Population management and public health	18
3.6 Fraud detection	19
4. Big Data: many opportunities, many challenges	20
4.1 Technology	21
4.2 Standardisation	21
4.1 Access to data	21
4.2 Privacy	22
5. Conclusion	23
References	23

1.

Introduction



1.

Introduction

In August 1854, there was an outbreak of cholera in the Soho area of London. Within a month, hundreds of inhabitants had died. As the cause of this sudden outbreak was unclear, doctor and scientist John Snow decided to plot the victims on the basis of their address details, on a map of the district. He quickly noticed that most of the victims lived close to a water pump on Broad Street.

Thanks to the smart combination of information, Snow was able to find the source of the outbreak and the municipality could take action. The water pump was closed and the epidemic soon came to an end. Furthermore, measures were taken to prevent the same problem in other districts and other cities.



Figure 1. The map of the spread of cholera in London's Soho, created by John Snow in 1854. Ordnance Survey data © Crown.

The way in which Snow combined the data to explain events and achieve new, unexpected insights, is an example of big data 'avant la lettre'. More than 150 years later, we still try to apply these principles in healthcare. A sector that has collected a lot of data since day one, but is a long way from succeeding in extracting all the potential knowledge hidden within it. An increasing supply of data from countless (new) sources and progress in analysis technologies together allow us to make better and better use of this potential.

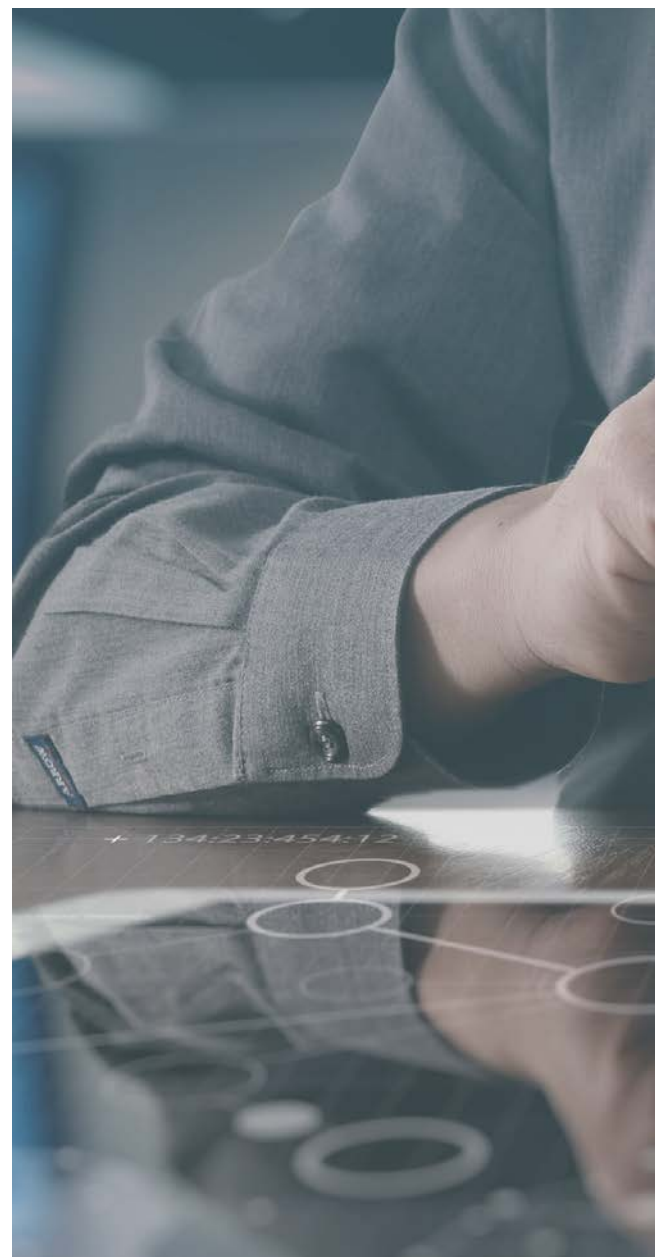
The concept of big data plays a role in almost every sector, and policy-makers, administrators, directors and managers are asking what they can do with it. Work is already being carried out successfully with it in many sectors. Data is collected and analysed on a large scale to create customer profiles, optimise the product range, put together customised offers and make service more personal. For example, shops can better determine where their customers' interests lie at a particular time of year by using purchasing habits and data from social media. The product range, stock and prices can be matched to this. In general, these developments result in increased customer satisfaction, more efficient use of resources and/or more revenue. But do these opportunities also apply to healthcare?

The "Nationale DenkTank" (Dutch National Think Tank), made up of young academics, thinks it does. In its research into big data in the public domain in 2014, it attributed major opportunities to the use of big data in healthcare. In fact, according to the authors there is no other sector that has as much to gain as healthcare, because healthcare can be provided more efficiently, in a more customised way, with fewer errors and therefore more cheaply. In addition to the mountains of data collected by healthcare professionals, the growing quantity of data that citizens gather (whether or not consciously) through wearables, sensors, emails, tweets, photos, Facebook messages etc., also represents fertile ground for this potential. The quantity is increasing exponentially.

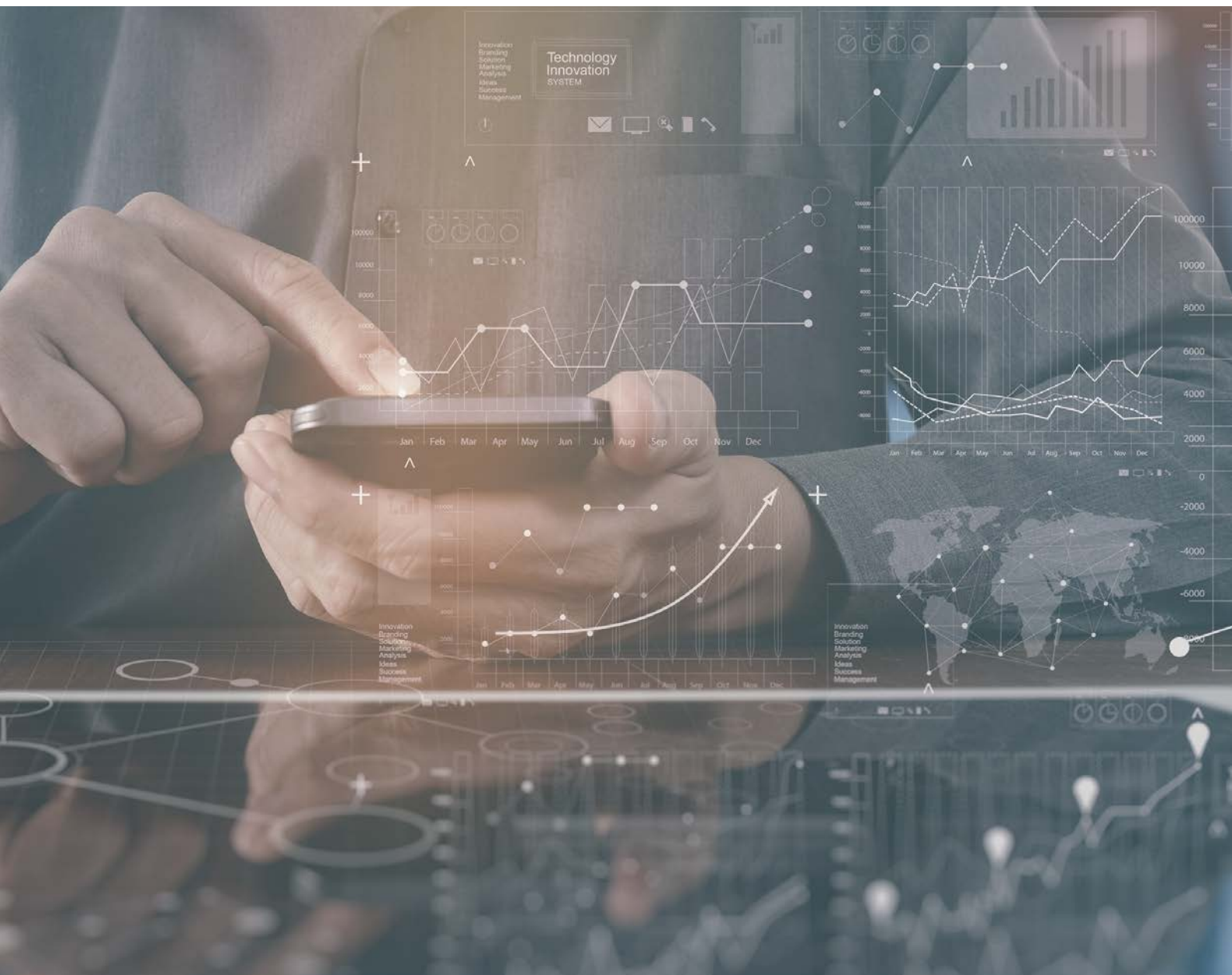
Big data in healthcare is currently mainly in the hype phase: it is talked about a lot, but practical application is still elusive. Or as Dan Ariely of Duke University put it: "*Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.*" The Rathenau Institute made a somewhat less racy

comparison and argued that big data is still in the 'gold-rush phase': "*Everyone sees something in Big Data but, at the same time, it is unclear where exactly the value is to be found and how this can be exploited.*" (Rathenau Institute, 2015)

This publication provides insight into the world of big data in healthcare: how do we define big data, what is its potential, and what are we doing already?

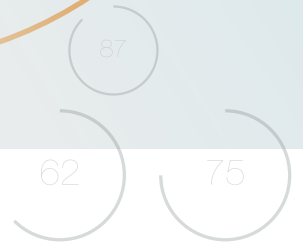


“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”



2.

Big Data: a definition



2.

Big Data: a definition

2.1 From data to wisdom

In order to prevent misunderstandings, it is important first of all to define the difference between the concepts of data and information, knowledge and wisdom. Alavi & Leidner (2001) clarified these concepts in their 'knowledge pyramid', see Figure 2. The 'data' concept forms the bottom layer of this pyramid and consists of unprocessed factual data, which represents reality. Information, one level up, is meaningful, processed data, which has been interpreted. Knowledge consists of information that has been processed to become meaningful and is seen as information at an aggregated level, which can lead to a change in behaviour or understanding (Huysman, 2006). Finally, the top of the pyramid is formed by the concept of wisdom, consisting of meaningful, processed knowledge.

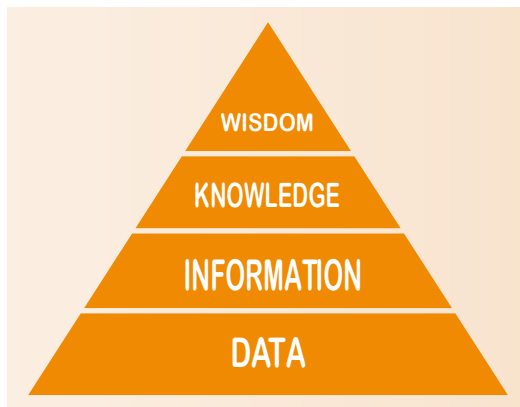


Figure 2. The difference between data, information, knowledge and wisdom (Alavi & Leidner, 2001).

2.2 More, faster, better

Big data relates to more than just large quantities of data. A description of big data that is often used, which provides a practical approach, comes from Mark Beyer and Douglas Laney (2012) at research agency Gartner and consists of the 'three Vs': *volume*, *velocity* and *variety*. The first characteristic is the large quantities of data (*volume*). By way of illustration: every day, we create 2.5 trillion gigabytes of data. This is increasing so rapidly that 90% of all worldwide data was created in the last two years. The speed with which data is generated, changes and is distributed is

also characteristic (*velocity*). For example, 2.5 million emails and 10,000 twitter messages are sent each second. The different types of data are also important (*variety*). In addition to documents full of figures and letters, there are numerous other forms of data such as audio, video, and photos. These days, an important fourth V is added to the previous three: *veracity* (*truthfulness*). In other words, the reliability or integrity of all this data, on which analysis is unleashed. After all, it is not known how a lot of data was generated and by whom.

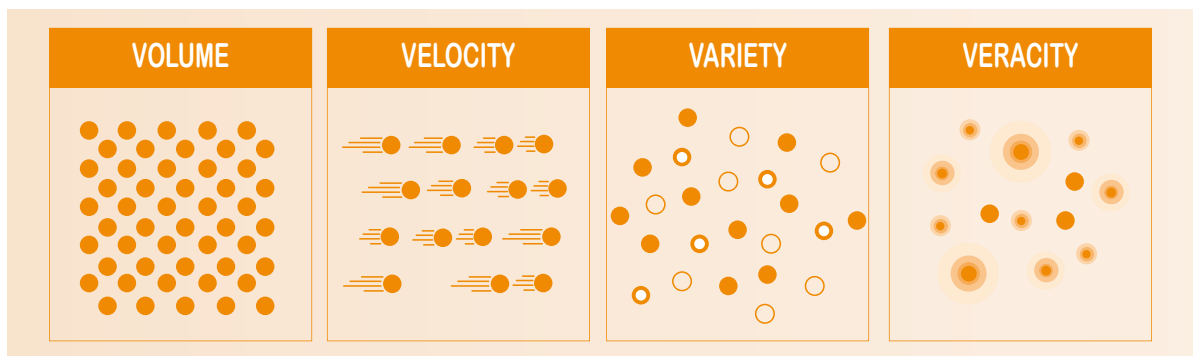


Figure 3. Characteristics of big data (Schenk, 2015).

The different characteristics of various types of data are also reflected in the data sources for big data analysis within healthcare. According to the Institute for Health Technology Transformation (2013), a distinction can be made between five different categories or information flows:

1. Internet and social media: clicking and browsing behaviour on the internet and interaction data from social media such as Facebook, Twitter, LinkedIn, etc.
2. Machine to machine: data from sensors and other measurement equipment, such as data from a medication dispenser.
3. Transaction data: declaration data and other data from the financial information flows in healthcare.
4. Biometric data: X-ray photos and other visual material, fingerprints, genetic information, iris scans, etc.
5. Human-generated data: data entered by people from electronic health records (EHRs), notes, emails and paper documents.

The data in these sources can also be divided into structured and unstructured data. Data is structured if it has already been processed in some way, into categories or according to specific logical structures. For instance, declarations with a fixed structure produced by an information system or the data that a blood pressure monitor or set of weighing scales consistently generates and presents in the same way. It is estimated that around twenty percent of all data is structured. So eighty percent of all data is unstructured (Grimes, 2008). Unstructured data includes, for example, medical images from cardiology or radiology and (hand) written reports in natural text. With regard to the volume of medical images as unstructured data: thirty (!) percent of the total worldwide data storage is used for medical images.

2.3 More complex = more valuable?

In the study “*The ‘big data’ revolution in healthcare*” (2013), McKinsey described the distinct types of big data analysis that can be identified. They introduce a model that consists of two axes, along which the following types are placed: the added value and/or impact on one side and the technological complexity on the other. According to the authors, the change from data, via information and knowledge, to wisdom, results not only in increased added value but also in increased technological complexity in order to reach this value. In concrete terms, this model makes a distinction between the following consecutive levels:

- **Reporting:** what happened?
Examples at this level consist, for example, of 'simple' databases, which can be used to produce (internal) reports or which can answer questions such as the number of operations that took place in a hospital during a particular period. This often concerns data that can be presented as meaningful information through 'dashboards' (to managers, healthcare professionals and healthcare consumers, for example).
- **Monitoring:** what is happening now?
Examples at this level can resemble those in the level above (reporting), but they are more extensive as they also make it possible to compare the current situation with a benchmark or a desired situation. Use is made of both recent and real-time data. For example, to alert healthcare professionals to contraindications during prescription or provision of medication. Other examples include monitoring (complex) stock management and the OR planning and OR staffing in a hospital.
- **Data mining and evaluation:** why did it happen?
Data mining and evaluation concern the analysis of data with the objective of discovering certain correlations, in order to obtain valuable information. An example of an application could

be the discovery of an (unexpected) relationship between contraction of an infection by patients and the hospital room in which these patients were located. In short: data mining is the identification of correlations between data and evaluation is the integration of the outcomes in the testing of hypotheses to establish correlation(s) and/or causality.

- **Prediction and simulation:** what will happen?
The most complex level in this model concerns the ultimate use of big data: being able to predict what might happen in the future on the basis of big data. A (known) example of technology that analyses data at this level with the aim of processing it to knowledge and wisdom is IBM's Watson. This supercomputer can process personal data for specific patients and compare it with the most recent scientific literature, comparable situations in other patients and the effect of their treatment, for example. This eventually results in a customised diagnosis and possibly a treatment recommendation along with the reliability margin of this prediction.

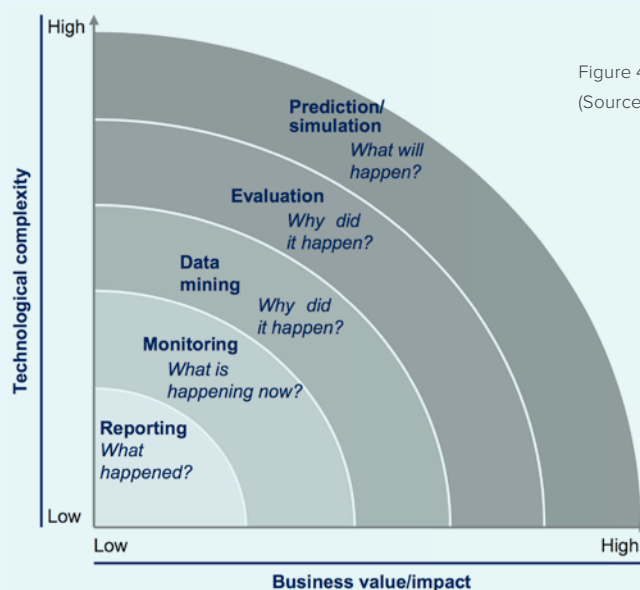


Figure 4. Gradations in technological complexity
(Source: McKinsey Business Technology Office, 2013)

2.4 A definition

It is difficult to give a comprehensive and generally accepted definition of big data. The concept is used not only to indicate the quantity, or complexity, of data. In practice, the term is mainly used as a container concept concerning the development regarding the acquisition of new knowledge and wisdom from this data. We therefore propose the following definition: *Big data refers to the ability to monitor, explain and predict events through the smart combination and analysis of complex data sets from various sources.*

3.

Big Data: application in healthcare



87

62

75

3.

Big Data: application in healthcare

Sectors such as retail, hospitality, tourism and energy, and financial institutions are already making use of big data analysis to improve the service they provide. Organisations in these sectors all try to use the available data to map customers' behaviour and wishes, to estimate and predict risks and to better match their services and prices to this. But how can this be done in healthcare and where is it happening already?



Within healthcare, big data analyses are carried out for various objectives, which can roughly be divided into two applications. Data that is used for business operations is known as business intelligence or business analytics. This includes, for example, digital dashboards with information presented in such a way as to provide insight to hospital managers and administrators. Data that is used for the care of patients and scientific research falls under the heading of medical intelligence, which includes decision-support software for doctors or perhaps even performing predictive analyses on treatment outcomes.

Below are a number of application areas based on examples from the Netherlands and abroad. These range from applications already used in practice to conceptual applications.

3.1 Opportunities for research

The use of big data can advance science if it is used in the organisation and implementation of research. In traditional scientific research, a research question is formulated first, after which data is collected through random sampling to test the hypothesis. Using as a rule a relatively modest amount of data, researchers look for a causal effect between two characteristics, for example. In modern 'big data research' this is the other way around. Data is gathered first, often much larger data sets, which are subsequently searched for expected and unexpected connections. This results in findings that would never have been made through random sampling.

This can also work in medicine, for instance in research into the effect of medication. In a study with a traditional design, medication is tested and approved on the basis of the average results in a select group of patients. This study group includes patients for whom the medication was effective but usually also patients for whom it was not. Nevertheless, the medication is eventually prescribed for all patients, if the study group benefited on average. Conversely, other medication does not make it to the market because a study group did not benefit on average, whilst individual patients may have. This traditional research method pays too little attention to these differences between individual patients. The use of more data, which maps the specific characteristics of each individual patient, may help to gain better insight into why medication works for certain patients and not for others.

New initiatives such as 23andme, Apple's ResearchKit, Google Genomics and Patientslikeme demonstrate how information and internet technology have the potential to extend the possibilities offered by conducting research. Although the initiatives listed differ in design and objective, they are all driven by data supplied by a large number of individual users. Using these organisations' services, scientists can attract a large number of participants for their study in no time and conduct quicker and better research. That is why many large pharmaceutical companies want to cooperate with these initiatives.



Collecting data within LifeLines (University Medical Centre Groningen)

LifeLines was set up in December 2006 as a large-scale population study among more than 165,000 inhabitants of the northern Netherlands (residing in the provinces of Groningen, Friesland and Drenthe). In this study programme, participants from three generations are followed for at least thirty years, to gain insight into how people can grow old healthily and the factors that are important in the occurrence and development of chronic conditions.

The participants will be invited for a screening once every five years. This will involve measurements and tests (such as blood pressure, weight, height, heart function) and bodily materials will be collected (such as blood, urine, DNA, hair). Feedback will be given on the results of these measurements and tests to the participant and his or her GP. The participants also receive an extensive questionnaire every one to two years, with questions about health, lifestyle and dietary habits, among other things.

The research focuses on the question of why one patient develops a chronic condition at a relatively young age while another remains healthy until old age. The researchers' starting point is that the onset of chronic conditions such as asthma, diabetes or kidney disorders is due to a complex interplay of factors. The influence of the different factors and the way in which they interact can only be understood by monitoring the health of a large group of people in different generations long term. The outcomes of LifeLines should result in quicker identification of illness, finding new treatments or even prevention of various chronic conditions (UMCG).

Further information about LifeLines: <https://www.lifelines.nl>

3.2 To prevent is better than to cure

Prevention is also an application area where the potential value of big data is evident. With the introduction and further development of smartphones, a world of opportunities has opened up for the gathering of information about behaviour, lifestyle and health. Smart applications turn mobile phones into pedometers, sleep monitors or medication reminders. Another development is the rise of wearables from companies like Fitbit, Jawbone and Withings, through which users can measure, save, share and compare their behaviour and certain health indicators with others (Nictiz, 2014).

While we currently use this information mainly to monitor our own progress and compare it with others who use the same wearable or application, this information can also be used in healthcare. People could share information about diet and exercise directly with the doctor treating them. If information is also shared for (scientific) research, researchers can describe the state of health in populations and the information can be used to identify problems and conditions before they manifest themselves.

An example of this is the American company Propeller Health, which focuses on disease management for asthma and COPD patients.

For this purpose, patients place a sensor on their inhaler, which registers where, when and how often the inhaler is used. By combining this data with around forty other sources of information, such as weather conditions, traffic, air quality or pollen in the air, the company can help patients and healthcare providers using the results of smart analysis of this data. By creating risk profiles and mapping risk areas, showing trends and predicting which patients are most at risk, for example. Patients can prevent an asthma attack as they have been warned in good time. The company recently entered into a partnership with the city of Louisville in Kentucky (United States) and its local pharmacies, to map risk factors in the city.

In a clinical setting too, analysis of the data produced by sensors from patients who are continuously monitored helps to recognise irregularities at an earlier stage and thus intervene earlier. For example, the University of Ontario's Institute of Technology, in cooperation with IBM, has launched the Artemis project. A platform that carries out real-time analysis of the data from monitoring premature babies in the ICU. Through smart analysis, hospital infections are recognised 24 hours before the symptoms appear in these babies, which gives extra time for adequate treatment (Soares, 2012).

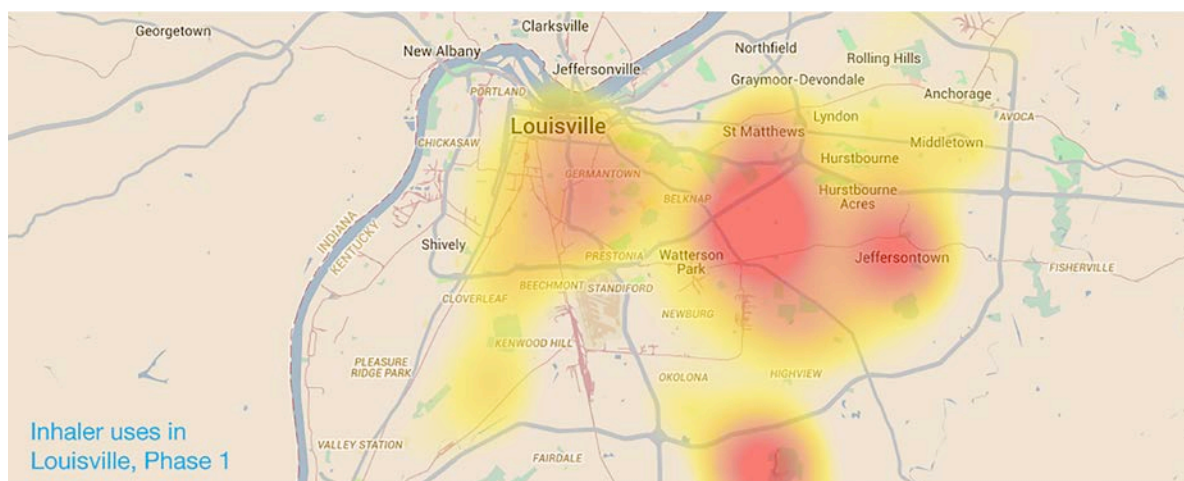


Figure 5. Use of asthma inhalers in the city of Louisville in the United States (AIR Louisville, 2015)

3.3 High-quality, safe and affordable healthcare

In addition to the potential for detecting problems and conditions before they reveal themselves as a disease, big data also offers opportunities to provide patients with better healthcare. This can manifest itself in better healthcare results or in increased affordability.

An interesting example that paints a clear picture of the potential of this can be found in the United States, at the healthcare organisation Aurora Health Care. This organisation consists of fifteen hospitals, 185 clinics and more than eighty pharmacies. The more than thirty-thousand employees within these healthcare institutions create a huge quantity of data every day. A number of years ago, the organisation resolved to use

this data to gain better insight into the quality provided, in order to subsequently improve this and reduce costs.

To achieve this, 'Smart Chart' was developed (Forbes, 2012), a system that gathers all the data collected within the organisation over the last ten years in a single data warehouse. This includes invoicing data, medical data, laboratory values, pharmacy information and data about appointments. Using all these data sources, insight can be gained into how the quality of the healthcare provided and its costs differ between all Aurora institutions. Aurora can also look at both individual patients and groups of patients with the same condition. This provides answers to question such as: how long do patients stay in the healthcare institution, what does the

Social media to support monitoring

More and more patients, family members and carers are using Facebook, Twitter or other social media to give feedback about the quality of a treatment, contact with the doctor or communication with a hospital. One of the largest rating websites is ZorgkaartNederland.nl, where patients can rate their dentist, doctor, GP or medical specialist.

A literature study by the Radboudumc showed that patients' reactions on social media about their treatment can be used to gain insight into the quality of healthcare (Van de Belt, 2014). The conclusion: there is a link between information on social media and traditional quality measurements, such as patient experiences and complications. The Dutch Health Care Inspectorate (IGZ) can use the information that emerges from this as an additional source of information for incident and risk-driven monitoring. So use of new data sources can contribute to improving the quality of healthcare.

Further information:

<http://www.zonmw.nl/nl/projecten/project-detail/effectiever-toezicht-met-sociale-media-toegevoegde-waarderandvoorwaarde-en-valkuilen/samenvatting/>

treatment process look like, what complications do they have, how many readmissions are there? This creates an image of the optimum outcomes for patients and the associated treatment. Using all the available data and the almost real-time data analysis, Aurora can even predict the outcomes of various treatments fairly accurately. Another advantage is the fact that this helps doctors and researchers prescribe the right medication for the right patients.

By comparing these results with national benchmarks, institutions and their departments and doctors also gain insight into their own performance. This contributes to the constant drive for improvement across the whole organisation. It has resulted in Aurora Health Care becoming one of the best-scoring healthcare providers in the United States. Since the introduction of the new working method using data as a tool, the number of readmissions has fallen by ten percent over recent years, resulting in a saving of millions of euros. Use of the knowledge hidden in the data therefore helps to continuously improve healthcare and reduce costs.

Big data in surgery

At the Radboudumc, the first steps in terms of big data are visible in the diagnosis-planning-treatment-evaluation cycle for surgery. For a complete and adequate diagnosis, different imaging techniques are combined. The objective is to build a virtual patient that is the best possible representation of the clinical and anatomical reality, with use of 3D photography among other techniques. By comparing patients' characteristics with an ever increasing group of previous patients, the most accurate possible diagnosis is made. On the basis of these insights, a treatment plan is drawn up, after which the patient is operated on virtually. A simulation of the postoperative situation is generated on the basis of the normal operation steps, which are now carried out virtually. An important advantage is the fact that the doctor can give patients an expected outcome on the basis of big data analysis, so that they are fully informed and have realistic expectations of the treatment. This planning step ensures better preparation and a clearer treatment plan prior to the operation, and as a result the operation takes less time and is more predictable.

Even if the optimum treatment plan has been set up with good consultation, the ideal outcome can only be achieved if this plan can also be followed precisely by the surgeon in the operating theatre. Use of techniques such as navigation-assisted surgery, 3D-printed drill and saw guides, and augmented reality ensure smoother progress of the operation and a reduction in the number of complications and secondary corrections. Quantitative analysis and qualitative assessment by the doctor mutually reinforce each other in the evaluation process. This means that the next diagnosis, planning and operation with the same resources will go even better than with the last patient, precisely because the information obtained is added to the big-database. Thanks to this cycle, the database of virtual representations of patients is growing on a daily basis at the Radboudumc and in the future it will provide vital information for better and better big data support.

3.4 Personalised healthcare

Another opportunity for big data is the possibility of personalising healthcare, offering more customisation for the individual. Personalised healthcare is a promise that is gradually being realised. It's about the individual's unique characteristics becoming a determining factor in diagnoses, predicting the course of diseases and choosing treatments. For example, the presence of risk factors and the patient's specific genetic and metabolic characteristics. Using known information about the outcomes of treatment in patients with comparable characteristics, the treatments that offer the best outcomes for the patient can be chosen. This way, diseases can be treated in a more effective way than previously.

Centre for Personalised Cancer Treatment

Until recently, the treatment of cancer has mainly been based on 'the average cancer patient'. In practice, this means that the treatment of lung cancer or breast cancer, for example, is successful for some patients and unsuccessful for others. All patients have to deal with the side effects of the treatment, however.

Each (type of) cancer has different characteristics in each patient. In order to treat cancer as successfully as possible, it therefore seems more logical to take each individual case as the basis rather than the 'average cancer patient', which does not actually exist. No general treatment, then, but a personalised approach. The Centre for Personalised Cancer Treatment (CPCT) is trying to achieve this by analysing the genetic material of the cancer cells, the DNA, at the beginning of the treatment process and creating a profile. By linking this information to other data about treatment results and effects of specific medication, the results of a specific treatment for a patient can be predicted. So in the future, patients will receive customised treatment and if possible will no longer be exposed to ineffective treatments.

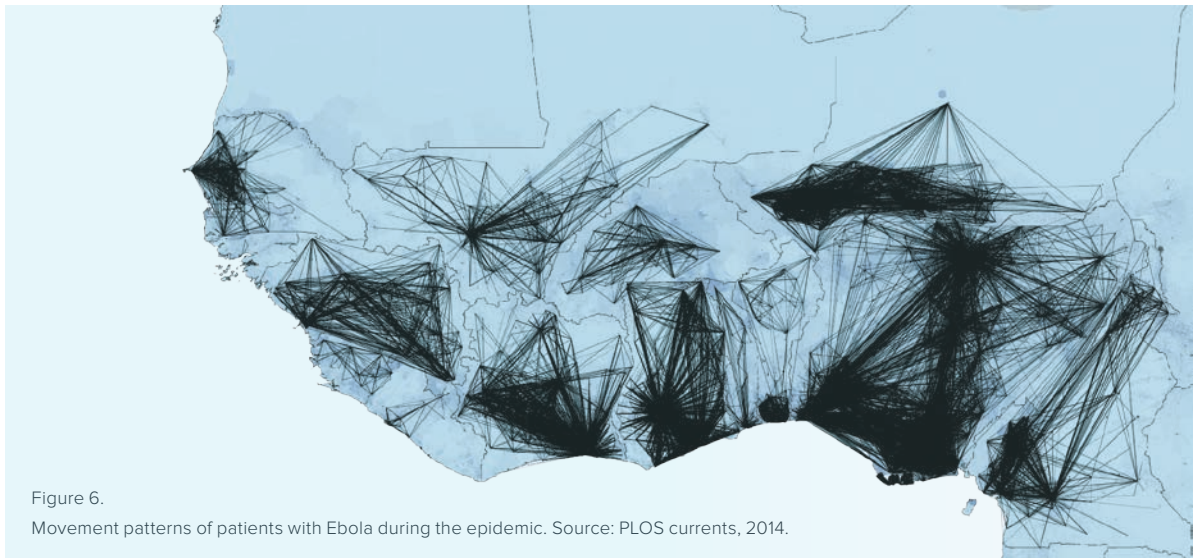
Further information: <http://www.cpct.nl>

3.5 Population management and public health

Public health institutions and organisations in the social field could also benefit from combining and analysing the data from the various sources that they manage. This way, different risk groups in populations could be distinguished. By linking this information to data about healthcare use, the healthcare prevalence for a certain area can be predicted. An example of the application of such software is what is known as the Adjusted Clinical Group (ACG) Case-Mix system, which was developed by John Hopkins University in Baltimore (United States). It is used to classify a population according to gender, age, geographic location and other characteristics, in order to establish the differences in health status and healthcare use. The software can calculate how much healthcare a population group will need, for example. The knowledge obtained from these kinds of predictive analyses can be used to organise the right healthcare offer in the region and to offer prevention programmes in the right place. Health insurers also use such predictions as the basis for their healthcare procurement.

Organisations that are responsible for public health may also benefit from the use of big data. Through monitoring of the spread of infectious diseases and analysis of disease patterns in order to improve public health and increase reaction speed in the event of incidents, for example. In 2014, during the Ebola outbreak in West Africa, data from telecommunications providers was used to map how groups of people travelled around the region. This helped to determine where aid could best be provided and the most effective way to set travel bans to prevent further spread.

This could be taken a step further if data from other sources (health records, social media, observations by aid workers on location, data about medication deliveries and use) could be combined in order to determine whether travel bans, education programmes and treatments actually work. Or even to predict which areas run the highest risk of rapid spread of infectious diseases.



3.6 Fraud detection

Big data also brings high expectations with it in terms of improving the detection of financial irregularities and fraud. Research by the Dutch Healthcare Authority (NZa, 2014) showed that medical specialists, GPs, nursing institutions and other healthcare providers submitted eight million euros of unusual claims in 2012. They charged notably more hours than colleagues, for example, or systematically allowed treatments to continue to the point when they could receive a higher fee. This is an estimation of irregularities, some of which may have been caused by fraud. Estimations about the financial damage of fraud within healthcare in the Netherlands vary from tens of millions to billions of euros per year (PWC, 2013).

With the help of smart analyses, irregularities and fraud could be recognised quicker and better in

the enormous quantity of data that healthcare organisations provide annually to health insurers. In the United States, the Centers for Medicare and Medicaid Services (CMS), a federal insurance institution, have done this to some extent. Until recently, patient declarations were compared digitally and manually with documentation submitted by healthcare professionals. Now the 'Fraud Prevention System' compares declaration patterns with fraudster profiles, healthcare providers who declare more hours than fit into a day are spotted and social-networking data is used to take previous fraudsters' cooperation partners into account. In this way, hundreds of millions of dollars' worth of fraud are detected every year, significantly more than in the years before this new system was introduced. The costs of the system are currently being earned back at a ratio of one in ten.

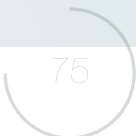
Health effects of air pollution

The Dutch National Institute for Public Health and the Environment (RIVM) together with Utrecht University, investigated the effect of air pollution on health. Using new research methods, Statistics Netherlands (CBS) registered who died among a study cohort of seven million Dutch inhabitants over the course of seven years. This data was anonymously linked to other Statistics Netherlands data, such as age, gender, marital status and home address. On the basis of the (encrypted) home addresses, data about airborne particulate matter (fine dust) and nitrogen dioxide was then linked to the home addresses of the seven million participants. The results show that the amount of both particulate matter and nitrogen dioxide has a direct impact on the total mortality and the mortality due to respiratory conditions and lung cancer in the Netherlands. In the future, both organisations aim to implement the method used to research the effects of other environmental factors on health.

Further information: http://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Nieuwsberichten/2015/Big_Data_geven_meer_kennis_over_gezondheidseffecten_luchtverontreiniging

4.

Big Data: many opportunities, many challenges



4.

Big Data: many opportunities, many challenges

As demonstrated by the examples discussed previously, better use of the quantities of data in and around healthcare could have great significance. What is needed in order speed up this development? We briefly discuss a number of challenges we must address in order to make better use of big data.

4.1 Technology

For the collection, storage and analysis of big data sets, new technologies, beyond those that healthcare providers generally have access to, are required. The 'traditional' architecture is not sufficient to make the developments as described in this paper possible. Building the right technical foundation required for advanced big data analysis is a complicated process. The American National Institute of Standards and Technology lists the necessary changes in architecture for the storage of data, the processing of data, its analysis and finally the visualisation of the outcomes as the most important elements of this foundation. This requires a different perspective on information provision, for example by hospitals.

In the examples that are discussed in this publication, data is exchanged beyond the virtual boundaries of the organisation more than once. More integration of data sources therefore regularly demands new collaborations between organisations at different levels. What does this collaboration look like and what agreements are required? What data is necessary and how will it be integrated, who will analyse it and how,

and who will eventually use the results? Are technologies used that the organisation has available in-house, open source or not, stored locally or in the cloud? Many questions and challenges that organisations within healthcare will address in the coming years.

4.2 Standardisation

The technology is increasingly advanced and more and more capable of processing unstructured information. However, the fact remains that analysis of structured data is more reliable. Structuring data by standardising the way in which it is recorded, requested and exchanged is therefore an important challenge.

Although many standards are applied in healthcare, interoperability between the information silos is a rare occurrence (Nictiz, 2015). Healthcare providers register and collect information in different ways and in many different systems. Exchange between these systems is a difficult process. The same applies to the increasing quantity of data generated and stored by patients. Standardising the way in which data is recorded in source systems, makes the exchange and integration needed for big data analysis easier. It is important that public and private organisations jointly develop and implement standards. Not only at national level, but certainly also internationally.

We may have different thoughts about this in ten or twenty years' time, if better algorithms can extract all the necessary information from natural (unstructured) language in an instant with the same reliability as from structured information. However, standardisation is important in order to make use of the potential value in all the data today and tomorrow.

4.3 Access to data

The fragmentation of data between hospitals, patients, research institutions, policy organisations, industry and other parties is a major barrier to use of the potential of big data. Each of these organisations' data warehouses functions as a silo for potential information that is difficult to open. Meanwhile, big data analysis derives its strength precisely from the integration of the data from these various silos. Aside from the technical challenge posed by the capability to integrate different data sources, another challenge is to get owners of the data to the point that they actually want to share it. In healthcare, 'this is my data' is the culture that still prevails. New cooperations are necessary in order to change

this pattern and to demonstrate how interests can complement each other. The fact that patients will be managing more and more data about themselves has added a new dimension to this discussion (Jacobs, 2015).

4.4 Privacy

Another discussion around the use of big data in healthcare is about the question of who owns certain healthcare information, who can use the information and for what purposes. There are laws and regulations, at both national and European level, about the protection of healthcare data. In addition, professional groups and institutions have codes of conduct that impose additional requirements on the way in which sensitive private information is handled.

Absolute protection of privacy, in the sense of having complete control over personal information, seems impossible. For instance, because in many cases, it is not at all clear who owns the information and who can use it. People are often not aware that the data they collect is also used by commercial parties, for example. In addition, data users can be based in the Netherlands and abroad, which means

that laws and regulations in other countries are also relevant.

This is also one of the reasons that data collected by citizens and healthcare professionals is still rarely exchanged. In a future scenario in which people shape the care for their own health according to their own wishes and requirements as much as possible, widely supported, safe and reliable data processing will be even more crucial than it already is, according to RZV (2015). A model is needed that guarantees the privacy of patients, without compromising the necessary freedom of movement of data between different parties in healthcare. New regulations from the European Commission will also play a role in this in the Netherlands. Clarification of the laws and regulations is necessary, so that users of data know what the practical consequences of these laws and regulations are. The RVZ warns that if this does not happen, the standards, frameworks, facilities and procedures will be determined by the various commercial parties and their earning models. This could severely limit the opportunities to take advantage of for instance big data applications for citizens and society.

Registration at the Source

'Registration at the Source' is a multi-year programme led by the Dutch Federation of University Medical Centres (NFU) and Nictiz. The programme envisages that, in the future, patients and healthcare professionals will have access to the necessary unambiguous healthcare information at any time and from any place. The 'registration at the source' principle means that healthcare professionals (and citizens) record healthcare information once and unambiguously using the healthcare information building blocks developed for this purpose, according to (inter)national standards. The recorded data can then be used for various purposes. In addition to provision of healthcare, for transfer between departments and institutions, quality registers, scientific research and administrative processes, for example.

The building blocks are developed by healthcare professionals (doctors, nurses and paramedics) themselves. Use is made of pre-existing (international) standards wherever possible. At present, around 90 healthcare information building blocks have been developed. These include wills, medication use, allergies, body weight and family history.

Further information:

<https://www.nictiz.nl/projecten/specialistische-zorg/zorginformatiebouwstenen>

5. Conclusion

This publication discusses just a few of the many inspiring examples that demonstrate how big data can play a positive role in healthcare; and this is only the tip of the iceberg. The quantities of data continue to increase, as do the possibilities for making use of the data. However, practical application is illusive and we have still not exited the gold-rush phase. Hopefully, over the coming years, we will succeed in overcoming the challenges step by step and mining the actual gold: preventing people from getting sick and helping patients as successfully as possible if they do get sick.

References

- Alavi & Leidner (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues.
- Beyer & Laney (2012). The Importance of 'Big Data': A Definition. Gartner, Connecticut.
- Bruijn, De (2013). Naar een fraudebeeld Nederland. PwC, Amsterdam.
- Forbes (2012). Using Big Data to Help A Hospital Meet The Financial Future. Consulted on 11 November 2015. <http://www.forbes.com/sites/tomgroenfeldt/2012/04/20/aurora-health-uses-big-data-to-reduce-risk-in-outcomes-based-pay/>
- Grimes (2008). Unstructured Data and the 80 Percent Rule. Consulted on 11 November 2015. <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>
- Groves, Kayyali, Knott & Van Keuken (2013). The Big Data Revolution in Healthcare. McKinsey&Company.
- Hengst, Van Pelt, Postema, Van der Sluis, De Jong, Van Dijk, Ekker & Pattel (2014). Zelfmetingen en de Nederlandse Gezondheidszorg. Nictiz, The Hague.
- Huysman (2006). De zin en onzin van kennismanagement. Vrije Universiteit, Amsterdam.
- Institute for Health Technology Transformation (2013). Transforming Health Care Through Big Data.
- Jacobs (2015). Gebruik persoonlijke gezondheidsdata: nieuwe wetenschappelijke omgangsvormen. Consulted on 11 November 2015. <http://www.smarthealth.nl/trendition/2015/09/30/gebruik-persoonlijke-gezondheidsdata-nieuwe-wetenschappelijke-omgangsvormen/>
- Krijgsman & Ottenheijm (2015). Advies ter verbetering van informatie-uitwisseling in de zorg. Nictiz, The Hague.
- Nederlandse Zorgautoriteit (2014). Onderzoek zorgfraude.
- Nationale Denktank (2014). Big Data in zicht.
- Raad voor de Volksgezondheid en Zorg (2015). Consumenten e-health. RVZ, The Hague.
- Rathenau Instituut (2015). Big data en slimme algoritmen. Consulted on 11 November 2015. <https://www.rathenau.nl/nl/page/big-data-en-slimme-algoritmen>
- Soares (2012). A Framework that Focuses on the Data in Big Data Governance. Consulted on 11 November 2015. <http://www.ibmbigdatahub.com/blog/framework-focuses-data-big-data-governance>

Betere gezondheid
door betere informatie

