PharmacoEconomics
https://doi.org/10.1007/s40273-020-00927-1

CURRENT OPINION

Future of Data Analytics in the Era of the General Data Protection Regulation in Europe

Katarzyna Kolasa[1,2] · W. Ken Redekop[3] · Alexander Berler[6] · Vladimir Zah[4] · Carl V. Asche[5]

Abstract

The development of evidence to demonstrate 'value for money' is regarded as an important step in facilitating the search for the optimal allocation of limited resources and has become an essential component in healthcare decision making. Real-world evidence collected from de-identified individuals throughout the continuum of healthcare represents the most valuable source in technology evaluation. However, in the European Union, the value assessment based on real-world data has become challenging as individuals have recently been given the right to have their personal data erased in the case of consent withdrawal or when the data are regarded as being no longer necessary. This act may limit the usefulness of data in the future as it may introduce information bias. Among healthcare stakeholders, this has become an important topic of discussion because it relates to the importance of data on one side and to the need for personal data protection on the other side, especially when it comes to "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveals information about his or her health status". At the forefront of these discussions are data protection issues as well as the population's trust in digital services. It seems that the new era has begun, where citizens and patients will have the ability to manage their personal or self-generated data. The European Commission has laid the groundwork for this paradigm shift that will steadily emerge in the coming years. To prepare for this change, we believe attention should be given to data security and other rules of data privacy. It has become increasingly important to ensure that individuals are properly introduced into complex environments with multiple sources of Big Data for clinical and behavioral purposes to provide an optimal balance between societal and individual benefits. In this article, a number of issues are considered and discussed, based upon the authors' experience, with the aim of helping the reader better understand the implications of the use of Big Data and the importance of data protection in the coming years.

1 Introduction

In Western European countries, the average life expectancy will increase by almost a year and the share of individuals aged 65 years and older in the total population will reach

Key Points for Decision Makers

Evaluation of new health technologies is becoming more difficult owing to a number of data protection issues related to how Big Data can be used to assess the personal preferences and behavior of individual customers.

The processes for limiting the use of data need to be systematic, transparent, and easy to handle.

Approaches to address data security and other principles of data privacy merit significant efforts to ensure that they are properly introduced.

Information exchange and interoperability have an important role in the secondary use of personal and clinical data because the medical information of any individual is and should be decentralized.

✉ Katarzyna Kolasa
kkolasa@kozminski.edu.pl

1 Health Economics and Healthcare Management Division, Kozminski University, 57/59 Jagiellonska St., 03-301 Warsaw, Poland

2 Global Market Access, Straub Medical, Switzerland Straub Medical AG, Wangs, Switzerland

3 Erasmus School of Health Policy and Management, Erasmus University, Rotterdam, The Netherlands

4 ZRx Outcomes Research Inc, Mississauga, ON, Canada

5 University of Illinois College of Medicine, Peoria, IL, USA

6 Gnomon Informatics SA, Thessaloniki, Greece

Published online: 08 June 2020

22% by 2022 [1]. As a consequence of aging and medical innovation, global healthcare spending is expected to increase at an annual rate of 5.4% until 2022. It is nearly twice as much as the rates observed during the period 2013–2017 [1]. In the search for the best approach towards the distribution of a limited healthcare budget, the "value for money" concept has been adopted most often [2, 3]. It refers to the amount of health outcomes achieved in a given budget. The value for money introduced a new dimension to evidence-based decision making that requires a greater reliance on different types of data in allocative choices in healthcare.

In 2017, The '*Economist*' declared data, and not oil, as the world's most valuable resource of the twenty-first century [4]. The growing "digital universe" explains this shift. Data production is estimated to increase from fewer 50 currently to 175 zettabytes in 2025 [5]. Professor Klaus Schwab proclaimed that we are witnessing the birth of the fourth industrial revolution, which is fueled by a staggering mountain of data. As he notes, it is "characterized by a range of new technologies that are fusing the physical, digital and biological worlds, impacting all disciplines, economies and industries, and even challenging ideas about what it means to be human" [6].

The amount of available data is growing in the healthcare sector as well. For example, in Sweden alone, there are 103 health registries [7]. In an Organization for Economic Co-operation and Development study in 2016, 23 out of 28 countries reported the implementation of electronic health records [8]. In the USA, while only 9% of non-federal acute care hospitals had a basic electronic health record in 2008, this percentage had risen to 96% in 2014 [9]. New health data sources are emerging as well. There are as many as 325,000 mobile health apps available and more are being launched regularly [10]. With the growing velocity, variety, volume, and veracity of data, we are entering the era of Big Data.

Many examples are already available that illustrate how much Big Data are reshaping decision-making processes in healthcare. For instance, the analysis of genetic records of 35,000 patients enabled the discovery of a genetic variant related to schizophrenia [11]. Another example is the study of 7700 brain images from 1171 people that led to the discovery of the first physiological sign of Alzheimer's disease resulting from decreased blood flow in the brain [12]. Beyond the clinical records of an individual's medical history, diagnoses, laboratory results, prescriptions, and healthcare service consumption, there is a growing amount of behavioral data to contend with. It is mainly the digital footprint left after the adoption of different devices that collect health-relevant information, such as dietary patterns, smoking habits, daily activities, and sleep–wake cycles. There is an ever-growing number of significant examples of how behavioral data can contribute to decision-making processes in healthcare. For example, a cross-sectional regression model based upon 826 million tweets collected between June 2009 and March 2010 in the USA predicted atherosclerotic heart disease mortality significantly better than a model that combined ten common demographic, socioeconomic, and health risk factors, including smoking, diabetes mellitus, hypertension, and obesity [13]. The analysis of behavioral data can ultimately help to change activities across the entire spectrum of disease development, including disease prevention, early diagnosis, and treatment monitoring.

With the growing amount of available data, the concept of "data-driven innovation" has been launched. It refers to a significant improvement of existing products and services and development of new products or services. The emergence of digital solutions and Big Data phenomena has also been acknowledged by the European Commission and its Horizon 2020 funding mechanism through the proposal of several activities linked to the Precommercial Procurement tool [14–17].

The European Union (EU) data economy represented 2% of the gross domestic product in 2016, and it was expected to reach 4% in 2020 [18]. In light of the growing importance of "data-driven innovation", the EU initiated the development of the digital single market in 2015 [19]. The underlying rationale is to maximize the growth potential of the European digital economy and to ensure the right conditions for digital networks and services to grow and thrive. One of the key prerequisites for the development of the digital single market is the data protection and trust in digital services amongst the population [20].

The General Data Protection Regulation (GDPR) [21] came into force in May 2018 to harmonize and unify the legal regulation across the EU. The key focus of the GDPR is to foster innovation while strengthening the privacy rights of individuals [22, 23]. It offers a new set of rules designed to give EU citizens more control over their personal data [23] and it provides rules for the protection and processing of personal data [24].

To a great extent, the GDPR has already been analyzed by many experts [24–26]. Nevertheless, to our knowledge, there has been limited discussion regarding its impact on the growing role of Big Data in decision making in healthcare. It is therefore important to consider how the rules of data privacy are being implemented in such a complex environment with multiple sources of both clinical and behavioral data. An ever-increasing number of comprehensive datasets are available that allow researchers not only to gain insights into the efficacy of treatments, but also to analyze genetic information and utilize self-generated data from mobile applications, wearables, and social media [11, 27].

In this article, we focus on two frequently mentioned issues with the implementation of the GDPR in the

healthcare sector. First, we address the question of how to ensure GDPR-compliant research in the era of "patient-centric" data ownership. Second, we discuss how to ensure free data transfer across multiple and diverse health data sources. Both issues are especially important for evidence generation as a basis for health technology evaluation in the new era of Big Data.

2 How to Ensure General Data Protection Regulation (GDPR) Compliant Research in the Era of 'Patient-Centric' Data Ownership?

The GDPR defines health-related data as "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status" [28].

2.1 Why is Consent a Key Element of the GDPR?

The provision of consent to the processing of personal data is the most obvious legal requirement when the data are directly collected from subjects. However, for consent to have a valid legal basis, it must be: (1) freely given; (2) specific; (3) informed; and (4) unambiguous. The first criterion means the consent must be a real choice of the subject and that the subject is in control of the provided information [28]. The second criterion of specificity is linked to the purpose limitation: the data subject's consent relates to the specific purpose of processing and cannot be used for alternative purposes. The third criterion for the consent to be informed means that adequate information must be provided to the data subject to enable him or her to make the informed choice. The information requirement is linked to the principle of transparency. The patient must at least have information on the identity of the "controller" (organizations that process personal data), the (type of) data collected and used, the existence of the right to withdraw consent, the potential use of the data for automated decision making (if relevant), as well as the possible risks involved with data transfers outside of the European Economic Area, if applicable [26]. The fourth and final criterion is unambiguity (or clarity), which means that the consent must be provided in the form of a precise statement that requires clear affirmative action (an opt-in; not necessarily in the form of an opt-in box, but also as a signature or oral confirmation). Whenever possible, consent should be gathered in a written form, as it will make it easier to demonstrate the consent's validity, if necessary [28].

It has to be emphasized that the consent to participate in a study or a trial must be distinguished from the consent to the processing of personal data [29–31]. Data protection legislation only governs the processing of personal data in the frame of research and has little bearing on the actual participation of individuals in research projects.

The GDPR recognizes that allowances must be made when the processing is carried out for scientific research purposes. Therefore, the terms "broad concept" and "dynamic consent" were introduced.

The broad concept refers to the circumstances when "it is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognized ethical standards for scientific research" [28]. An Independent European Advisory Body on data protection and privacy (Article 29) further specifies that broad consent may be an option when the purpose cannot be fully explained at the onset of the project [32].

The alternative to "broad concept" is "dynamic consent", which consists of using an IT solution, such as an app or a platform, to engage individuals and have active participants. This makes it possible to easily inform research participants and ask for re-consent or additional consent. "This approach is "dynamic" because it allows interactions over time; it enables participants to consent to new projects or to alter their consent choices in real time as their circumstances change and to have confidence that these changed choices will take effect" [33]. The advantages of dynamic consent appear to be numerous. In particular, it would be easier to seek re-consent from data subjects. It would also facilitate the re-use and further processing of the data and ease the possibility of withdrawal. However, dynamic consent also has some drawbacks. It might limit the possibility for the engagement in research only to those having access to the appropriate electronic devices and applications and would inevitably lead to additional implementation costs.

2.2 Is Consent the Only Option for Researchers to Access the Data According to the GDPR?

Given the importance assigned to consent, one may consider the feasibility of complying with such strict regulations when collecting data for research purposes. In the age of Big Data, however, research occupies a privileged position in the GDPR [24]. It adopts a "broad" definition of research. It encompasses the activities of public and private entities. Article 89 and related Recital 159 elaborate on "technological development and demonstration, fundamental research, applied research and privately funded research." When consent might pose significant challenges, the GDPR introduces five other legal bases described in Article 6 (1) that can be chosen to justify the collection and processing of personal data [21].

The most commonly used legal basis for research at public universities is the notion of "public interest". Researchers may meet that requirement by referring to the legal acts indicating that a given activity is fulfilling the purposes of the organization. Research should be dedicated to addressing societal challenges and providing long-term benefits to humanity. In the case of non-public research institutions, the "legitimate interests" may be a more appropriate lawful basis for processing personal data. This is a broad term that lacks a strict definition. Generally, it refers to the circumstances when the data processing takes place within an already established client-provider relationship. Hence, its usefulness for research may be limited [34].

2.3 May Public or Legitimate Interests be a Legal Basis Instead of Consent for Research According to the GDPR?

The GDPR has been the subject of intense debate, in particular concerning its potential effect on scientific research [35]. While the initial proposal of the Commission provided the option to carry out scientific research on a legal basis other than consent, the European Parliament Committee on Civil Liberties introduced a revision requiring that "consent should always form the correct basis for the processing of personal health data in a research context unless such research serves a purpose of 'exceptionally high public interest'". It also recommended that "where possible, health data was to be anonymized or at least pseudonymized to the highest possible technical standards." The European Parliament Committee on Civil Liberties justified its revision by arguing that processing sensitive data for scientific research was not as urgent or as compelling as public health or social protection; as a result, there was no need to provide an exception to the consent requirement [36]. At the time, there were serious concerns about whether this would hinder health research significantly. However, the Council of Ministers of the EU, the third actor of the EU legislative procedure, did not agree with the obligatory consent and the request for a fully anonymized approach favored by the Parliament; as a consequence, the text finally adopted in 2016 provides for derogations. Nevertheless, the literature still refers to the consent as the first possible legal ground, fostering the impression that consent is the principle, and the rest is an exception.

There is still one key distinction between the secondary processing of previously collected data for research purposes and the projects where research is the primary purpose. The consent provided for the primary processing of personal data may be withdrawn at any time, and this introduces a factor of uncertainty in a research project. If data subjects exercise their right to withdraw, the processing of their data must stop.

In the case of secondary processing of previously collected data, organizations may process personal data without consent, when "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject" [Article 6(1)(f)] [21]. Therefore, when a research project is carried out on a different legal basis such as legitimate interest or the public interest, data subjects may not withdraw consent. However, they may oppose the data processing by other means, such as exercising their right to object to the data processing, which may also be derogated (or partially suppressed) in certain circumstances.

Although the GDPR aims to encourage innovation by relaxing some regulations on further processing of personal data for research purposes, the "data minimization principle" of Article 5 requires personal data to be "limited to what is necessary" to complete the task successfully. In other words, it allows for the processing of personal data only to the extent needed to fulfill the research purpose.

On that note, it must be mentioned that GDPR introduces a high degree of scrutiny regarding the processing of sensitive data. Article 9 introduces a special category of data that includes among others, genetic and biometric data related to physical, physiological, or behavioral characteristics of a natural person as well as non-health-related information concerning political opinions and religious or philosophical beliefs. Unless there is an explicit consent given, these data cannot be accessed, except in specifically defined circumstances: for example, threats to public health and preventive or occupational medicine [21]. It should be stressed that explicit consent further raises the standards of the "regular" consent. The consent must be clearly and explicitly expressed by the subject, thereby leaving no place for misinterpretation. Explicit consent can be expressed in a written or spoken format, but an electronic format is acceptable as well. A signed written statement is an appropriate means of demonstrating consent in the case of a scientific research project [37].

2.4 How to Ensure that the Conducted Retrospective Data Analysis is Compliant with the GDPR?

Article 6.4 of the GDPR [21] indicates that the purpose of further processing must be compatible with the purpose of the initial processing. In this case, it is not necessary to use a legal basis other than the basis used for the initial processing. This means that the research project may re-use an existing data set without relying on a new specific legal basis. When possible, data subjects should be informed and the research sponsor must ensure that all their obligations as controllers are respected; in particular, in the case of medical

research, the research sponsor must ensure that one of the exceptions of Article 9.2 [21] is applicable. In the case of further processing of scientific research, the controller can continue to rely on the legal basis of the initial processing provided the appropriate safeguards are implemented in compliance with Article 89.1 [21], which governs processing for scientific research purposes and triggers the scientific research regimen of the GDPR.

2.5 How May Pseudonymization Support Compliance with the GDPR in Scientific Research?

If a legal basis other than explicit consent has been implemented or re-use of personal data without consent is planned, Article 6 lists additional requirements for data processing [21]. These additional requirements include the existence of appropriate safeguards, which may include pseudonymization or anonymization [38]. Pseudonymization is really a new term introduced by the GDPR that has become a key gateway for researchers to ensure compliance with GDPR regulations. Article 3 defines pseudonymization as "The processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information" [21]. The identifiable parts of personal data are translated into unique artificial identifiers (pseudonyms). The GDPR requires the additional information needed to re-identify the person to be kept separately from the pseudonymized data. Different pseudonymization methods are available. The simplest form is through scrambling, which involves mixing, or masking (the obfuscation of letters), where an important unique part of the data is hidden with random characters or other data. Certain parts of the GDPR explicitly refer to the use of another type of pseudonymization, namely encryption, which involves using an algorithm to transform plain text information into a non-readable form (ciphertext). The encryption is a two-way function (encryption to encode the information and decryption to return it to its original form); this differs from hashing, which is a cryptographic one-way function without the possibility of decryption. Another type of pseudonymization is tokenization, which, unlike encryption, does not use a mathematical process and instead transforms the sensitive data into a token (a random string of characters). Tokenization uses a database, called a token vault, which stores the relationship between the sensitive data and the token. The real data in the vault are secured, often via encryption [39].

The alternative to pseudonymization is anonymization, where the identifiable information is fully masked. Article 4 and the related Recital 26 [21] define anonymized data as "data rendered anonymous in such a way that the data subject is not or no longer identifiable." Anonymization places the processing and storage of data outside the scope of the GDPR because the data are no longer personal.

It is important to note that pseudonymization, anonymization, and de-identification techniques can only reduce the privacy risk for uncovering sensitive patient data. As such, de-identified data still need extensive data protection measures and patient consent procedures. Several techniques, like those described above, are proposed and the most common of them are found in the "IHE Information Technology Infrastructure Handbook on De-Identification" created by the non-profit organization named "Integrating the Healthcare Enterprise" [40].

2.6 How does the GDPR Align with National Regulations Regarding Research?

The regimen described above is applicable at the European level, but national legislation should still be carefully considered. Some provisions in the GDPR provide the possibility to the EU Member States to further legislate on some specific points, such as the possible derogations to data subject rights in the frame of scientific research or processing of data concerning health, genetic, and biometric data (Article 89.2 of the GDPR) [21].

The Member States retain the possibility of introducing or maintaining further conditions, including limitations, concerning the processing of health data, genetic data, and biometric data (Article 9.4) [21]. In practice, this means that the conditions for use of data will not be the same from one Member State to the other. Obviously, this could negatively impact transnational European research projects. Still, the additional requirements should not limit the free flow of personal data within the EU with a special focus on cross-border activities (Recital 53) [21]. The possible variation in the applicable rules in different Member States may complicate the application of cross-border projects. As, however, the basis for the data processing will remain the same, such challenges should be resolvable.

3 How Can we Ensure Free Data Sharing Across Multiple and Diverse Health Data Sources?

3.1 How to Ensure that Data Sharing is Compliant with the GDPR?

The term "Big Data analytics" refers to the technical or analytical methods to extract information from multiple complex data sets [41]. Big Data analytics commonly rely on existing data sets collected for other purposes, and this is therefore considered as further processing.

Sharing personal data is a processing activity that is subject to the rules of the GDPR unless the data have been adequately anonymized. If Big Data analytics is used for scientific research, then the rules laid out earlier will apply; in particular, data subjects must be informed, if possible, and be given the possibility to exercise their rights. Recital 50 of the GDPR [21] clarifies that no legal basis separate from that which allowed the initial data collection is needed.

The fact that the further processing is for a different purpose does not necessarily mean that it is automatically incompatible; this needs to be assessed on a case-by-case basis, following the test of Article 6.4 of the GDPR [21]. If the new processing is not deemed incompatible with the initial processing, it may proceed, while still adhering to all the GDPR requirements.

If further processing is carried out by a different controller than the initial controller, the data must be obtained from the initial controller. This requires access by the controller to numerous data sets. Transfer of data is a processing operation; therefore, transfer within the EU must be compliant with Chapter II of the GDPR [21]. Data may be shared between controllers based on a data-sharing agreement organizing the responsibility of both controllers. Such an agreement could be a means for the initial controller to ensure the stewardship function. The data-sharing agreement may involve requirements as to the conditions and purposes of the processing, therefore, qualifying as a possible safeguard. It follows that data sharing must be envisaged from the beginning.

3.2 Will Patient Identifiers Support Interoperability?

The GDPR recognizes that data governance is an essential component of data sharing among institutions. That is why Recital 68 says, "Data controllers should be encouraged to develop interoperable formats that enable data portability" [21]. This necessity is formalized in the concept of data protection by design and by default.

The first concept of "privacy by design" focuses on integrating data protection into the product design process to ensure the inclusion of appropriate GDPR compliance measures. In contrast, the second requirement of "privacy by default" involves the restriction of processing of any personal data to the extent that is necessary for each specific business purpose.

The principles of data protection by design and by default bring together the legal and technical aspects of data protection. Privacy principles must be considered and implemented throughout the design cycle of processing. An example of where the principles must be carefully applied is the ability to easily identify the patient [42]. Efforts to implement a unique patient identifier across different databases can lead to errors in the patient healthcare continuum and increase the likelihood of privacy harm [43]. Health data are generated in multiple systems, and their integration may only be possible through a collaboration across different healthcare providers. This applies to all types of data used in healthcare, such as care plans, laboratory results, medical procedures, drug administration, community care, health records, and billing. With interoperable systems, data can be exchanged and stored automatically rather than re-typed into the system each time. Although presently data are still mainly collected through registries, healthcare systems are moving towards the holistic integrative analysis of multiple data sources, which will require specific expertise in data analytics. Data are not always available in a usable format, thus hindering the integration of data from various sources. As a solution to create a widely used and accepted data format, the integration profile process has been proposed as a way to enable end-to-end interoperability by sharing structured (and unstructured) data between the point-of-care systems [44].

3.3 New Challenges of Interoperability in the Era of the GDPR and Patient-Centric Data Sharing

Interoperability in healthcare is often focused on data exchange between business entities. In the era of big data and digitalization, the shift happens towards patient-centered interoperability, meaning that a patient has full access to the data provided and is able to edit or delete it at any point. Among other factors, it is owing to the emergence of mobile health applications. Together with automated data collection from connected medical and wellness devices, the patient, and especially the empowered patient, can now contribute actively and create part of the Big Data needed for research and public health [45]. For example, ChatBots allow new direct communications today, between clinicians and patients [46, 47], empowering both of them in the form of shared care plans, direct communications, and chatting and artificial intelligence algorithms that drive data-driven innovation to a new era. From a practical point of view, security and privacy by design and interoperability by design are prerequisites to master GDPR and data privacy compliance. Examples of system-embedded patient consent provision and revocation are now steadily emerging in the market, as well as proof of interoperability compliance and alignment to EU and other regulations [48]. All these new innovative tools are reshaping the medical software industry from a procedural approach to digitize the clinical process towards a workflow approach to digitize the interaction between patients and clinicians. As a result, software tools are now categorized as medical devices that fall under the regulation of the existing medical devices directive, soon to become the medical devices regulation [49, 50]. This is

a new compliance challenge that innovative solutions have to address from the design phase to data processing and reuse phases.

3.4 How Can Blockchain Protect Data Sharing in the Era of the GDPR?

There is a growing interest in blockchain technology that could potentially address the two challenges of a patient's identity and interoperability [51]. By definition, it allows a data owner to control his or her own medical records. Consequently, it is up to that individual to share his or her data without any traditional intermediary. The blockchain allows the data owner to assign access rules (like smart contracts) for other data users. In other words, the blockchain creates an information-sharing marketplace. In principle, it offers a centralized and shared mechanism for the management of authentication and authorization rules surrounding data. In practice, a patient creates his or her own medical history by connecting to a particular healthcare provider interface (like a patient portal). Then, he or she grants that institution the access key, along with permission to securely transmit data (or metadata) to the blockchain. If it is done across multiple healthcare providers, medical data can be aggregated to create a database.

One of the key components of the blockchain mechanism is the hashing function, which ensures both interoperability and security. Thanks to cryptographic algorithms, the patient data can be stored under unique identifiers while being safe and tamper proof. A ledger of hashes could be compared to the original data to prove it was not altered. Hence, the blockchain would allow for secure sharing of electronic medical information such as genomic data, clinical trial data, hospital, outpatient clinic, visiting nurse and immunization records, imaging and laboratory results, as well as pharmacy records, health data from mobile devices, wearables, and the "Internet of Things" at the click of a button. The potential issue of "blockchain technology" is the massive patenting spree, with over 114 pending patents granted (last updated 5 March, 2019) [52]. For the interoperability to function, a blockchain needs to be a "universally accepted" open source, such as The Linux Foundation initiative Hyperledger Frameworks (Fabric, Indy, Iroha) that makes software code open to the public. In addition, recent publications suggest that blockchain technology may apply to patient consent handling without putting the medical information itself into the blockchain and instead only the consent/revocation information from a procedural and legal perspective [53].

4 Conclusions

Big Data provide a multitude of opportunities to further develop the concept of value in the healthcare sector. Big Data facilitate the fight for better health by enabling faster identification of people at risk and better understanding of disease consequences [23]. Big Data analysis can help to combat rising health inequalities and improve the assessment of the effectiveness of costly treatments [23]. As Article 89 and the related Recital 157 in the GDPR explicitly state: "Registries provide solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services".

Personal data play a critical role in the development of data-driven healthcare. Still, it would be a great loss if we did not utilize the full potential of Big Data because of a lack of understanding of data privacy regulations. Therefore, a sensitive balance needs to be struck between protecting privacy and making the best use of health data. This is especially true with the growing availability of behavioral and digital data. Wearable and embedded devices (such as pacemakers, glucometers, and activity trackers) paired with remote monitoring and telemedicine services will ensure on-time care and patient monitoring with minimal disruption of day-to-day activities.

With the stunning increase in the variety of different data sources available to demonstrate the 'value for money' in the healthcare sector, it is more crucial than ever to preserve any information that reveals a patient's health status. Only when giving sufficient care to data security and other data protection principles, will we succeed to ensure that Big Data works to benefit all of the healthcare system stakeholders and, more importantly, the patients. The future ahead clearly shows that people, devices, and the software will seamlessly interact to provide better care to the citizens. This, of course, empowers but also challenges health policy makers to develop appropriate safeguards for the data privacy in the EU and on a global scale. Difficult challenges will undoubtedly emerge and these will only be resolved properly if we clarify and agree on the trade-off between the health maximization and limits of the use of personal data.

On a positive note, there are already some encouraging solutions within GDPR regulations that may actually encourage the growing use of data while keeping data privacy. For instance, the pseudonymization will inevitably help the Big Data industry to develop methods to grant

access to the personal health data for third parties. In due course, it will hopefully enable Big Data-driven health innovation and advance interoperability frameworks while taking into account ethical and security risks in a new digital era. The GDPR provides the legal platform to incorporate data privacy by design and security by design as well. This will hopefully help the developers to adopt appropriate safety measures of data management during the clinical development while limiting the risk of data privacy breaches. The GDPR should be seen as an enabler, not a barrier, to improve access to innovative digital health solutions in Europe. It has also increased the trend of renovation of laws and regulations beyond the EU, which will hopefully allow safer re-use of patient-driven data for research and public health globally.

In conclusion, GDPR has introduced a new era of lawful data processing, where the real owner of the data is the individual citizen. As we have already started our journey towards patient-centric healthcare systems, it is interesting to the observer what is the next destination in front of us. One can wonder whether the current clinical or healthcare provider-driven point-of-care system will be replaced with personalized reimbursement models based on health outcomes being reported on the individual level. Surely, the era of Big data welcomes us to the bright future. The data protection regulation is a traffic light system that ensures a very safe journey.

Acknowledgements We give special thanks to the contributions made to this article by Mahault Piéchaud Boura from the Timelex law firm based in Brussels, which specializes in information and technology law. Portions of this work were presented during an Issue Panel at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 21st European Congress in Barcelona, Spain in November 2018.

Author Contributions KK, CA, KR, AB and VZ conceived of the presented idea. KK drafted initial manuscript with input from all authors. CA, KR, AB and VZ aided in interpreting the idea and worked on the manuscript. KK, CA, KR, AB and VZ discussed and equally contributed to the final version of the manuscript.

Funding No funding was received for the preparation of this article.

Compliance with Ethical Standards

Conflict of interest Katarzyna Kolasa, W. Ken Redekop, Alexander Berler, Vladimir Zah, and Carl V. Asche have no conflicts of interest that are directly relevant to the content of this article.

References

1. Deloitte. Global health care outlook: shaping the future. 2019. Available from: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf. Accessed 9 Mar 2019.
2. Okoli C, Ezenduka C, Uzochukwu B, Okoronkwo I, Onwujekwe O. Achieving value for money in healthcare: principles, methods and empirical applications. Afr J Health Econ. 2014;2.
3. Caro JJ, Brazier JE, Karnon J, Kolominsky-Rabas P, McGuire AJ, Nord E, et al. Determining value in health technology assessment: stay the course or tack away? Pharmacoeconomics. 2019;37(3):293–9.
4. The Economist. The world's most valuable resource is no longer oil, but data. 2017. Available from: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. Accessed 9 Mar 2019.
5. Data Age 2025. The digitization of the world from edge to core. An IDC White Paper. November 2018. Available from: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. Accessed 9 Mar 2019.
6. Schwab K. The fourth industrial revolution: what it means, how to respond. 2016. Available from: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond. Accessed 21 May 2020.
7. Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish healthcare quality registries. J Intern Med. 2015;277(1):94–136.
8. Oderkirk J. Findings of the 2016 OECD HCQI study of electronic health record system development and data use. 2016. Available from: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DELSA/HEA/WD/HWP(2017)9&docLanguage=En. Accessed 10 Jan 2020.
9. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008–2015. ONC data brief, No. 35. Washington, DC: Office of the National Coordinator for Health Information Technology; 2016. Available from: https://www.healthit.gov/sites/default/files/briefs/2015_hospital_adoption_db_v17.pdf. Accessed 1 Jan 2020.
10. Research to Guidance. 84,000 health app publishers in 2017: newcomers differ in their go-to-market approach. 2017. Available from: https://research2guidance.com/84000-health-app-publishers-in-2017/. Accessed 9 Mar 2019.
11. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature. 2009;460(7256):744–7.
12. Iturria-Medina Y, Sotero RC, Toussaint PJ, Mateos-Perez JM, Evans AC. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. Nat Commun. 2016;7:11934.
13. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on twitter predicts county-level heart disease mortality. Psychol Sci. 2015;26(2):159–69.
14. European Commission. European assistance for innovation procurement: eafip. 2019. Available from: https://eafip.eu/about/. Accessed 26 Apr 2019.
15. European Commission. Recovering life wellbeing through pain self-management techniques involving ICTs. 2017. Available from: https://cordis.europa.eu/project/id/689476. Accessed 26 Apr 2019.
16. European Commission. The DECIPHER Project (Distributed European Community Individual Patient Healthcare Electronic Record). 2017. Available from: https://cordis.europa.eu/project/id/288028. Accessed 26 Apr 2019.
17. European Commission. Procuring innovative ICT for patient empowerment and self-management for type 2 diabetes mellitus. 2019. Available from: https://cordis.europa.eu/project/id/727409. Accessed 26 Apr 2019.
18. European Commission. Building a European data economy. 2019. Available from: https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy. Accessed 10 Jan 2020.

19. European Commission. EU leaders' meeting in Sofia: completing a trusted digital single market for the benefit of all. 2018. Available from: https://ec.europa.eu/commission/presscorner/detail/en/IP_18_3740. Accessed 9 Mar 2019.
20. Voss W. First the GDPR, now the proposed ePrivacy regulation. J Internet Law. 2017;21:3–11.
21. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed.
22. European Commission. A new era for data protection in the EU: what changes after May 2018? 2018. Available from: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf. Accessed 9 Mar 2019.
23. European Public Health Alliance. Health and care in the digital single market: reflection paper. September 2017. Available from: https://epha.org/wp-content/uploads/2018/02/Health-and-care-in-digital-single-market-position-paper.pdf. Accessed 9 Mar 2017.
24. Cornock M. General data protection regulation (GDPR) and implications for research. Maturitas. 2018;111:A1–2.
25. Mourby M, Mackey E, Elliot M, Gowans H, Wallace SE, Bell J, et al. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. Comput Law Secur Rev. 2018;34(2):222–33.
26. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience. 2017;11:709.
27. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med. 2019;2:45.
28. Voigt P, von dem Bussche A. The EU general data protection regulation (GDPR): a practical guide. Springer, Berlin 2017.
29. European Commission Directorate-General for Health and Food Safety. Question and answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation. 2019. Available from: https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf. Accessed 2 Jan 2020.
30. World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–4.
31. Gefenas E, Cekanauskaite A, Lekstutiene J, Lukaseviciene V. Application challenges of the new EU clinical trials regulation. Eu J Clin Pharmacol. 2017;73(7):795–8.
32. European Commission. Article 29 working party. 2016. Available from: https://ec.europa.eu/justice/article-29/documentation/index_en.htm. [Accessed 24 Feb 2020].
33. Farrell AM, Devereux J, Karpin I, Weller P. Health law: frameworks and context. : Cambridge University Press; 2017;262
34. European Commission. What does 'grounds of legitimate interest' mean? 2019. Available from: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/what-does-grounds-legitimate-interest-mean_en. Accessed 24 Feb 2020.
35. World Health Organization. Who owns our genes? 1999. Available from: https://www.who.int/genomics/elsi/regulatory_data/region/international/073/en/. Accessed 24 Feb 2020.
36. European Parliament Research Service. Rules for EU institutions' processing of personal data. 2018. Available from: https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/608754/EPRS_BRI(2017)608754_EN.pdf. Accessed 29 Apr 2019.
37. Agarwal R, Sands DZ, Schneider JD. Quantifying the economic impact of communication inefficiencies in U.S. hospitals. J Healthc Manag. 2010;55(4):265–81.
38. Elliot M, Mackey E, O'Hara K, Tudor C. The anonymisation decision-making framework. 2016. Available from: https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf. Accessed 10 Jan 2020.
39. Benschop T, Machingauta C, Welch M. Statistical disclosure control for microdata: a practice guide for sdcMicro. 2019. Available from: https://sdcpractice.readthedocs.io/en/latest/. Accessed 10 Jan 2020.
40. IHE IT Infrastructure Technical Committee. Handbook on de-identification. 2014. Available from: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Handbook_De-Identification_Rev1.1_2014-06-06.pdf. Accessed 27 Apr 2019.
41. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inform Manag. 2015;35(2):137–44.
42. Tucker K, Branson J, Dilleen M, Hollis S, Loughlin P, Nixon MJ, et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol. 2016;16(1):77.
43. Gliklich RE, Dreyer NA, Leavy MB (eds) Managing patient identity across data sources: registries for evaluating patient outcomes. A user's guide. 3rd ed. Rockville: Agency for Healthcare Research and Quality; 2014.
44. Hoerbst A, Ammenwerth E. Quality and certification of electronic health records: an overview of current approaches from the US and Europe. Appl Clin Inform. 2010;1(2):149–64.
45. Continua Design. Guidelines. 2017. Available from: https://www.pchalliance.org/continua-design-guidelines. Accessed 29 Apr 2019.
46. Futurist TM. The top 12 health chatbots. 2018. Available from: https://medicalfuturist.com/top-12-health-chatbots/. Accessed 29 Apr 2019.
47. Pereira J, Diaz O. Using health ChatBots for behavior change: a mapping study. J Med Syst. 2019;43(5):135.
48. National Multiple Sclerosis Society. 9-hole peg test (9-HPT). 2015. Available from: https://www.nationalmssociety.org/For-Professionals/Researchers/Resources-for-Researchers/Clinical-Study-Measures/9-Hole-Peg-Test-(9-HPT). Accessed 29 Apr 2019.
49. The Council of the European Communities. Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. 1993.
50. The European Parliament and The Council of the European Union. Regulation (EU) 2017/745 of The European Parliament and of The Council of April 2017 on Medical Devices, amending Directive 2001/83/EC, Regulation (EC) No. 178/2002 and Regulation (EC) No. 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.
51. Gordon WJ, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput Struct Biotechnol J. 2018;16:224–30.
52. USPTO Patent Full-Text and Image Database. [Search in Term 1 for "blockchain technology"]. 2019. Available from: https://patft.uspto.gov/netahtml/PTO/search-bool.html. Accessed 9 Mar 2019.
53. Houlding D. Eight opportunities to advance AI in healthcare using blockchain. 2018. Available from: https://www.linkedin.com/pulse/8-opportunities-advance-ai-healthcare-using-houlding-cissp-cipp/. Accessed 29 April 2019.