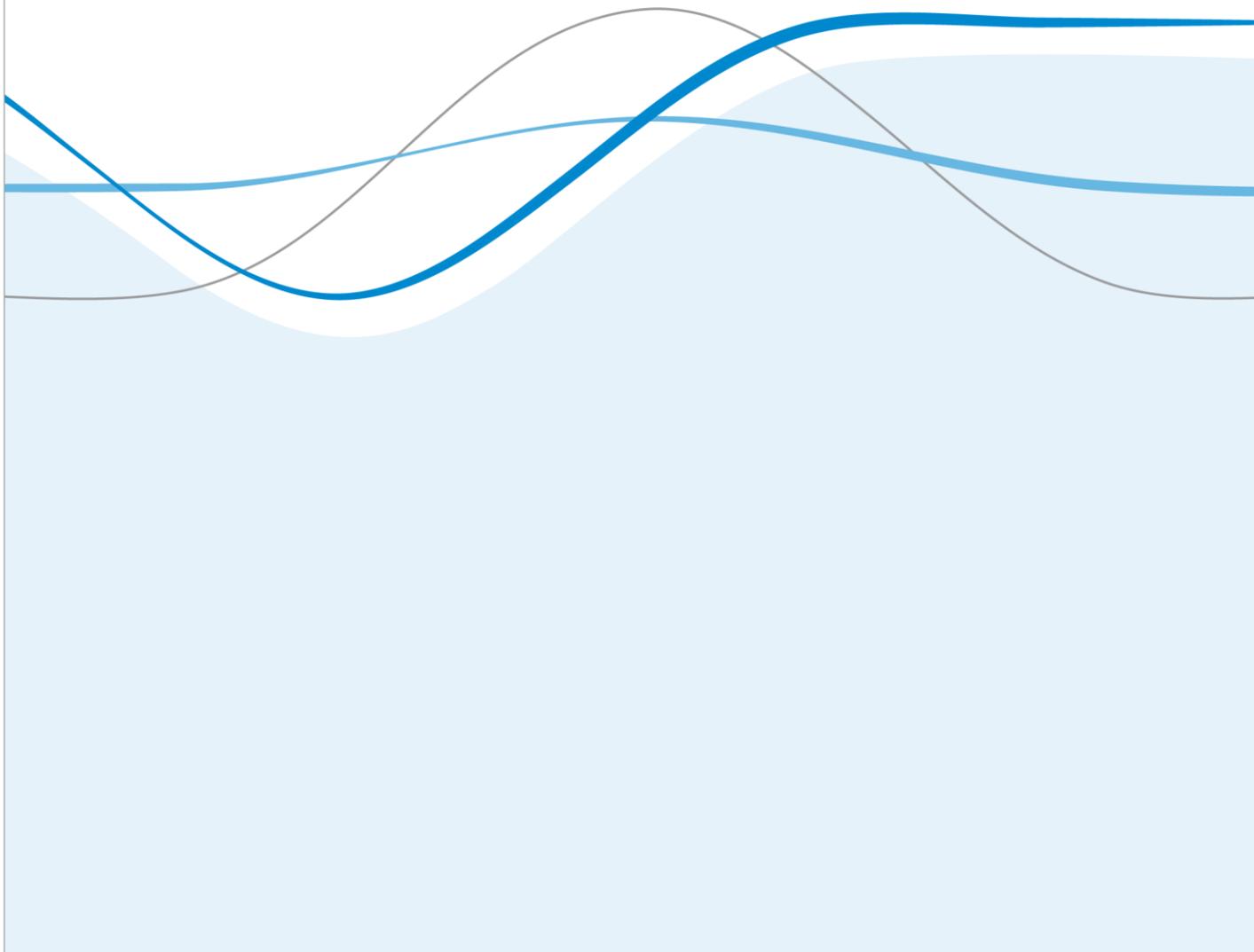


Legal Barriers to the Better Use of Health Data to Deliver Pharmaceutical Innovation

December 2018

Amanda Cole and Adrian Towse



Legal Barriers to the Better Use of Health Data to Deliver Pharmaceutical Innovation

Amanda Cole and Adrian Towse

December 2018

Please cite this report as:

Cole, A. and Towse, A., 2018. Legal Barriers to Better the Better Use of Health Data to Deliver Pharmaceutical Innovation. OHE Consulting Report, London: Office of Health Economics. Available at: <https://www.ohe.org/publications/legal-barriers-better-use-health-data-deliver-pharmaceutical-innovation>

OHE Consulting Ltd
(a registered company number 09853113)
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT

For further information please contact
Adrian Towse
Director
Tel: +44(0)207 747 8858
atowse@ohe.org

About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE as OHE Consulting Reports are subject to internal quality assurance and undergo a rapid review by a member of the OHE Editorial Panel. Any views expressed are those of the authors and do not necessarily reflect the views or approval of the OHE's Editorial Panel or Research and Policy Committee, or its sponsors.

Funding and Acknowledgements

This study was commissioned and funded by EFPIA. The research team are very grateful for the time and input of the interviewees consulted as part of this project, who included members of the pharmaceutical industry and representatives of public data protection authorities, reimbursement agencies, research authorities, and experts in health informatics and cyber security.

Table of Contents

Executive Summary.....	iii
1. Introduction	1
1.1. European needs for health data	1
1.2. The contribution of health data to pharmaceutical R&D	3
1.3. Project objectives and methods	3
1.4. Structure of the report.....	4
2. Data governance and protection of personal data in the EU	4
2.1. Balancing public and privacy interests	5
2.2. Protection of personal data.....	5
2.3. Consent, anonymisation, and authorisation according to intended use	6
2.4. GDPR	7
3. Framework: identifying the barriers to better use of data	9
3.1. Epidemiology and pharmacoepidemiology: identifying unmet need.....	12
3.2. Pharmacogenetics: targeting development	13
3.3. Interventional studies.....	13
3.4. Non-interventional studies	14
3.5. Pharmacovigilance	16
3.6. Managed entry agreements	17
3.7. Summary.....	18
4. Cross-cutting themes	18
4.1. Data subject rights.....	18
4.2. Anonymisation.....	19
4.3. Consent.....	20
4.4. Uncertainties around appropriate legal basis: GDPR	21
4.5. Compatibility of primary and secondary (re-)use of data	22
4.6. Heterogeneity.....	23
4.7. Issues relating to digital health.....	23
4.8. Engendering trust	24
5. Discussion and proposed next steps	24
References	28

EXECUTIVE SUMMARY

The sources of and potential uses for health data are growing rapidly in our ever more digitalised world of health care, as we increasingly seek to understand the potential or realised effectiveness of medical treatments in the “real world”. Yet, opportunities are necessarily bounded by the legal framework which governs how data can be collected, accessed and used.

The value of health data cannot be underestimated. Its value extends beyond fostering more effective and efficient pharmaceutical R&D and evaluation (although this is the main focus of this report), as its use is critical to the functioning and improvement of health care delivery. The main objective of our research was to build a consensus and prioritisation of the **main legal barriers to the better use of health data to deliver pharmaceutical innovation**, which was supported by interviews and a workshop with pharmaceutical industry members, as well as interviews with external experts in data protection, health research, informatics and cyber security. To identify tangible and specific legal barriers we investigated the legal issues arising in relation to six key activities along the pharmaceutical lifecycle. We then analysed the issues according to eight cross-cutting themes, which provided a platform for recommendations.

The introduction of the General Data Protection Regulation (GDPR) in 2018 has important implications for the issues discussed. A theme running across all uses of data considered in this report was establishing the appropriate legal basis for data processing, whether this be the consent of the individual or whether by meeting exemptions for scientific research, public interest, or provision of health or social care.

For data contributing to early activities around identifying unmet need, epidemiological and pharmacoepidemiologic studies often rely on aggregated datasets; legal barriers are generally low. Pharmacogenetics research usually relies on external datasets (e.g. cell line repositories and population genetics databases) for which the onus is on the data provider to meet the relevant legal standards. However, barriers remain including heterogeneity of data access models, how to maintain the possibility to re-use data to strengthen later research, and how “anonymisation” relates to genomic data.

The legal issues arising from the collection and analysis of health data in support of interventional (randomised controlled trials and pragmatic trials) and non-interventional (observational) studies can be distinguished according to primary data collection versus secondary use of already-collected data. For the former, maintaining the possibility to re-use data at a later point was highlighted, along with the challenges associated with obtaining re-consent at a later date. For secondary studies, of paramount concern is judging the compatibility of the purpose of data use with the original purpose of data collection, and the feasibility of anonymising all types of observational data.

Heterogeneity of data governance requirements between countries, and restrictive data access for pharmaceutical companies are perceived to be relevant barriers to efficacy and effectiveness research.

Ambiguity around pharmacovigilance reporting obligations for pharmaceutical companies was highlighted by industry. For managed entry agreements, the main barrier articulated was understanding the legitimate basis for processing data in the absence of consent. An understanding of this single issue has important ramifications across all activities studied; clarity is required to create an improved environment for the better use of health data. Where interpretation or implementation of these legal bases diverge across

different national contexts, this propagates uncertainty for the global pharmaceutical industry and further hampers opportunities for research. **A shared understanding that crosses borders and bridges stakeholders is required.**

In order to take the issues forward and consider them at a policy level, we assess the legal barriers according to eight-cross cutting themes.

Data subject rights have been further enshrined by the GDPR, among them the right to erasure (“to be forgotten”). However, these rights are not absolute and will usually be more limited within the context of health research, due to the scientific or public interest merits of the data use, along with the high level of safeguards in place to protect data. Data ceases to be personal once **anonymisation** has been achieved, in which case it does not fall under data protection legislation. Introduction by the GDPR of the term “pseudonymisation” captures the concept that anonymisation is not absolute; re-identification risk must be considered along with the safeguards and precautions in place to protect data. Where anonymisation is not possible, processors must have a legal basis for processing data. One such basis is (informed) **consent** from the data subject. The GDPR heightens requirements for very clear and explicit statements of consent, raising the challenge of constructing consent that is broad enough to permit later research, but specific enough to meet legal standards. We argue that these requirements are not usually compatible with medical research, and whilst consent is critical for other reasons (clinical trial regulations, confidentiality obligations, ethical considerations, etc), it need (and should) *not* usually be the legal basis for data processing; choosing consent to be the legal basis puts companies at greater risk of non-compliance, as well as implying that no other legitimate bases apply. That said, **uncertainty around the appropriate legal basis** for processing data under the GDPR – in the absence of explicit consent – is currently a major issue for industry. Under the GDPR, pharmaceutical companies should be considered to have “legitimate interests” in processing data (Article 6), which due to its sensitive nature must and does undergo rigorous safeguarding activities, and must meet one of the *additional* legal bases for processing special category data (which includes health) outlined in Article 9: (h) provision of health or social care, (i) public interest in the area of public health, or (j) scientific research. We speculate which legal bases may be most appropriately applied to the six pharmaceutical activities studied, but clear guidance and consensus on this is required.

There is a need for a shared and consistent understanding of the **compatibility of primary and secondary (re-)use of data**, which addresses the **heterogeneity** both within and between countries that arises from divergent interpretations. Heterogeneity hampers cross-border research; developing common standards would support health care innovation and in doing so benefit industry, patients and society. Whilst the GDPR intends to harmonise practice across Europe, there is some concern that it could do the opposite. In addition, there is a need for clear guidance or minimum standards for industry in the emerging area of **digital health**. Promoting confidence and **engendering trust** is fundamental and is achieved through being transparent and sharing good practice. The public must be convinced of the benefits of data processing for pharmaceutical research, and of the high safeguarding standards employed in handling their sensitive information.

Many of the perceived legal “barriers” to better use of data are in fact uncertainties. There is a strong case for **industry to deal proactively with the uncertainties, sharing good practice and engendering trust by co-creating a code of conduct**, outlining the principles of responsible use.

1. INTRODUCTION

The term 'health data' encompasses a broad spectrum of information about health and health care, which is increasingly diverse in nature. Potential sources of health data are also evolving rapidly. Alongside traditional avenues of data collection on a treatment's effect through clinical trials, data reflecting use and outcomes in the "real world" are progressively being used to inform decision-making across the entire development pathway of a medicine. The list of stakeholders that have a direct interest in what and how health data are used in the development and evaluation of health treatments is extensive, and the transformation of this information into evidence and learning has benefits across all of society. It is therefore in the shared interest of all stakeholders that the legal framework is appropriate, both in terms of the constraints it places on data access and use to protect privacy concerns, and in the benefits that can accrue to society from the optimal use of health data. Whilst in this report we structure our discussion of these issues according to key activities performed across the pharmaceutical lifecycle, it should be noted that the ramifications of each are of course felt beyond the pharmaceutical industry alone, and extend to patients and the general public at large.

Despite the evolving opportunities for using new sources of health data, these opportunities are bounded by the legal framework within which data generation, access, and use must be conducted. In this report we describe research on the legal barriers to better use of health data across the pharmaceutical lifecycle.

It should be noted that the term 'real world data' (RWD) is used in this report to denote data that are collected outside of an experimental clinical trial setting¹, whilst 'real world evidence' (RWE) refers to the knowledge that is created when RWD have been used for a specific purpose: to generate insight.

1.1. European needs for health data

Technological developments are such that the supply of health data is growing, supported by policy efforts to enable and enhance the supply of, and interoperability of, health data across Europe. On the demand-side, initiatives that seek to better align regulatory and reimbursement decisions with outcomes in real-practice require good quality, readily-available, health data. Thus, personal data markets are becoming more important, and this raises privacy challenges. Whilst there is increasing demand and increasing supply of health data, it can be argued that the two do not meet in a functional way for many of the purposes that health data could fulfil.

The Digital Single Market strategy was adopted as one of 10 political priorities by the European Commission in 2015, and consists of: improving access to digital goods and services, fostering an environment where digital networks and services can prosper, and promoting digital as a driver for growth (European Commission, 2017a). Focussing on health in particular, a report on the State of Health in the EU concluded that patient-centred health data are still under-developed across the EU (European Commission, 2017b). The role of health data is highlighted as a key enabler for the "digital transformation" vision of the European Commission, as summarised in a communication

¹ RWD can also be defined as data collected in a routine care setting, which may be part of a pragmatic clinical trial.

on enabling the digital transformation of health and care in the Digital Single Market (European Commission, 2018), developed as part of the mid-term review of the Commission's Digital Single Market strategy. The report highlights that market fragmentation and lack of interoperability across health systems hinder an integrated approach to disease prevention and care. Succeeding depends on *"the availability of vast amounts of high quality data and appropriate regulatory frameworks that will safeguard the rights of the individual and society as well as stimulating innovation"* (European Commission, 2018, p.2).

Recent EU initiatives – which have stimulated the demand for further collection and use of health data – span regulatory, reimbursement and pharmacovigilance activities. The efficacy-effectiveness gap is well known: how a drug works under "ideal conditions" (efficacy) can differ substantially from how the drug works in a "real-world" health care environment that offers routine care (effectiveness). The value of effectiveness evidence is increasingly well recognised, though health technology assessment (HTA) agencies across Europe differ in the extent to which they incorporate RWD into decision-making (Makady et al., 2018). In parallel, there is a gradual shift towards earlier drug approvals. Notably, models of adaptive licensing, early access schemes, and managed entry agreements (MEAs) which may include performance-based payments all employ a system of provisional approval which is subject to later or continual re-assessment, drawing on data collected alongside product use. Eichler et al. (2015) describe a flexible life-span approach to bringing new drugs to a patient, whereby the trade-off between timely access and the need for more evidence is balanced within a model of progressive development and access expansion, coordinated by regulators and payers. Pilots were run by the European Medicines Agency (EMA), but it is not yet clear how the concept will be rolled out.

EU level initiatives to increase the utility of RWE include:

- Patient registries (such as PARENT Joint Action, ENCR - European Network of Cancer Registries, Eurocourse and the EMA Initiative on Patient Registries);
- Electronic health records (such as EH4CR, EMIF, EU-ADR Alliance, RD-Connect, epSOS, EuroRec);
- Initiatives aimed at establishing methods and platforms to enable and facilitate data access, analysis and collaboration (such as IMI GetReal, IMI PROTECT, IMI ADAPT SMART, IMI ADVANCE, the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance - ENCePP);
- Initiatives on HTA (e.g. EUnetHTA JA3 aims to conduct pilots on post-launch evidence generation and tools to support collaboration on post-launch evidence generation);
- Approaches aimed at the exploitation of social media (IMI WebRADR).

All of these initiatives, however, must work within the technical, ethical and legal frameworks that govern data collection, and address the challenges around data sharing (Auffray et al., 2016). A critical development in the creation of a Digital Single Market for Europe is the harmonization of the legal framework for data protection. The General Data Protection Regulation, approved in April 2016 and implemented in May 2018, is a pan-European data protection law which has a significant impact on perceived and realised legal barriers to better use of data, which are assessed in this report.

1.2. The contribution of health data to pharmaceutical R&D

The contribution of health data to pharmaceutical R&D and innovation includes: opportunities to explore new research and innovation areas, for example by identifying unmet need or by allowing evidence collection where RCTs are impractical or unethical; operational and cost-efficiencies in R&D, in particular reducing the costs of evidence generation; and improved research quality in relation to the evidence generated to support an innovation (Marjanovic et al., 2017). The potential sources of health data to support these benefits are vast, especially when we consider data that are collected outside the setting of a clinical trial, often referred to as RWD. While the range of data sources and technologies relevant to pharmaceutical R&D continues to expand, actual practice is constrained by the legal environment, which must catch-up with these opportunities.

In order to identify tangible and specific legal barriers to the better use of data, we investigate the following six specific activities:

1. Epidemiology and pharmacoepidemiology: Identifying unmet need
2. Pharmacogenetics: targeting development
3. Interventional studies
4. Non-interventional studies
5. Pharmacovigilance
6. Managed entry agreements

1.3. Project objectives and methods

The objective of this research is to identify and analyse the main legal barriers to the better use of health data. In short, we discuss: **To what extent is using health data (for a number of specified activities, as listed above) either possible or problematic given the legal framework for data collection, sharing and analysis?** Our research involved various stages of work:

- 1) *Generating a framework* to map the relevant data sources and activities, in consultation with the literature.
- 2) *Interviews* with industry experts across the six activities to identify the legal barriers. 12 interviews were conducted between January and March 2018. Candidates were selected based on their diverse expertise across the six activities, or their oversight of RWE activities or data protection and privacy. 10 different companies / organisations were represented among the 12 interviewees. Interviews were conducted over the phone using teleconference facilities. One or two members of the OHE research team were present and summarised the project, presented the “framework” and opened a discussion around the legal barriers for the (usually one or two) specific activities relating to the interviewee’s area(s) of expertise. Discussions were summarised and collated, and shared with workshop participants (see below).
- 3) *Workshop*. A workshop was hosted by EFPIA on the 16th April 2018, which brought together around 15 industry and policy experts. In advance of the workshop, attendees were sent a *pre-workshop exercise*: a summary of the interview data was provided in the form of a “long-list” of legal barriers or issues against each activity. For each, attendees were asked to select whether or not the issue was relevant, and to rank their “top three” along with their views on the

main policy priorities that could deal with that barrier. During the workshop, the research team presented the framework and a summary of the barriers identified, as well as the results of the pre-workshop prioritisation exercise. As a group, participants talked about the activities, barriers, and opportunities for better use of data, and discussed potential recommendations.

- 4) *Interviews with independent external experts.* The research team conducted six further interviews with (non-industry) experts representing the following expertise: academic experts in digital health policy, cybersecurity, health informatics and information governance, leaders in European health data initiatives, payers, data protection authority (EU-level) and research authority. Experts had experience across various geographies (UK, Sweden, Poland) as well as pan-European institutions or initiatives. The purpose of the interviews was to corroborate findings and/or seek further inputs, to discuss potential policy solutions, and to resolve uncertainties.

The research was designed to obtain insights from industry members of the legal barriers experienced in using health data to support key activities along the pharmaceutical lifecycle, to build a consensus and prioritisation of those issues, and to discuss these with external experts in order to provide a thorough and well-rounded commentary on the issues and possible ways forward. Our analysis was supplemented by a review of the relevant literature, both for this and previous projects that have informed our background to the topic.

1.4. Structure of the report

In order to provide some background to the relevant issues, in Section 2 we describe some of the most important concepts and recent developments in data governance and data privacy. Section 3 contains the main findings from our interactions with industry on the key legal barriers to the better use of data, first presenting our framework, and then outlining the main legal barriers for each of the six activities explored. In Section 4 we reflect on the issues raised by grouping the barriers under several cross-cutting principles: data subject rights, anonymisation, consent, uncertainties around appropriate legal basis, compatibility of primary and secondary (re-)use of data, heterogeneity, issues relating to digital health, and engendering trust. These reflections are based on our collective interpretation of the issues as well as consultation with experts through our second stage of interviews, and further analysis of the relevant literature. In Section 5 we discuss the implications and potential ways forward.

2. DATA GOVERNANCE AND PROTECTION OF PERSONAL DATA IN THE EU

Whilst the demand for and potential benefits of expanding the collection and utilisation of health data are clear, these activities are necessarily constrained by legal and ethical considerations and regulations. In this section we provide a brief overview of the fundamental principles by which health data and its lawful processing are considered, governed, and defined. The concepts discussed represent the underpinning of the legal and regulatory contexts for the collection and management of health data, and therefore the key considerations for our framework (presented in Section 3) to identify the barriers to better use of data. The concepts will be revisited in more detail in Section 4.

2.1. Balancing public and privacy interests

The central aim of the legal framework for data governance is to balance public and privacy interests: of advancing society's understanding of medical treatments through evaluation and research, on the one hand, and protecting individuals' privacy, on the other. Protection of privacy is necessary in order to safeguard against improper use of personal information, thereby protecting patient identity and prohibiting discrimination. On the other hand, rich patient-level data can enable important effectiveness and cost-effectiveness research and service evaluation, optimise R&D and treatment targeting, and support sustainable health systems. Protection of privacy and the risks associated with disclosure of personal information must be set against the potential benefits to patients and the general public arising from making use of that information.

2.2. Protection of personal data

The right to data protection is underwritten by *Article 8* of the EU Charter of fundamental rights (Official Journal of the European Union, 2012)², according to which:

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for **specified purposes** and on the basis of the **consent** of the person concerned **or some other legitimate basis laid down by law**. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

Source: GDPR Art. 4

Note: words underlined are additions to the definition of personal data by the GDPR, in relation to the previous Directive (Art.2(a))

The interpretation and implementation of the fundamental rights summarized in this Article – the definition of 'specified purpose', 'consent' and 'legitimate basis' for processing data – are the foundation of data governance, and guide all subsequent discussions in this report.

There is a great deal of country variation in data governance requirements. Oderkirk, Ronchi and Klazinga (2013), in a paper that provides an overview of data privacy protection challenges internationally, explain that this variation is primarily driven by differences between countries in risk management of granting exemption to patient consent for use of their data (for sharing and/or data linkage activities). As described by the authors, this arises from different approaches by individual countries in weighing the trade-offs between data risks and data utilities. In a previous study conducted by OHE, we assessed the governance arrangements in place in eight countries, which demonstrated this variation, and developed an illustrative framework of a top-performing data governance model to support a favourable environment for the development and

² Another relevant article in the EU Charter is Article 13 on the Freedom of the arts and sciences: "The arts and scientific research shall be free of constraint. Academic freedom shall be respected".

use of RWD (Cole et al., 2015). In late 2016, the OECD Council adopted a recommendation which calls upon countries to “develop and implement health data governance frameworks that secure privacy while enabling health data uses that are in the public interest” (OECD, 2017). The recommendation was structured on 12 high-level principles:

1. Engagement and participation
2. Coordination within government and promotion of cooperation among organisations processing personal health data, whether in the public or private sectors
3. Review of the capacity of public sector health data systems used to process personal health data to serve and protect the public interest
4. Clear provision of information to individuals
5. Informed consent and appropriate alternatives
6. Review and approval procedures, as appropriate, for the use of personal health data for research and other health-related public interest purposes
7. Transparency, through public information mechanisms which do not compromise health data privacy and security protections or organisations’ commercial or other legitimate interests
8. Maximising the potential and promoting the development of technology
9. Monitoring and evaluation mechanisms
10. Establishment of appropriate training and skills development in privacy and security measures for those processing personal health data
11. Implementation of controls and safeguards
12. Require organisations processing personal health data to demonstrate that they meet national expectations for health data governance

Whilst there are encouraging examples internationally of the utility of health data in improving patients’ lives, several countries must go further in providing a facilitative framework for research whilst protecting privacy; the OECD recommendations could support this (Oderkirk and Ronchi, 2017).

2.3. Consent, anonymisation, and authorisation according to intended use

Data ceases to be ‘personal’ once it has been fully anonymised. However, for many of the uses which could offer value to the pharmaceutical R&D process, relevant data insight can often only be obtained through opportunities around data linkage and follow-up. Data protection legislation outlines the fair and lawful means by which personal data can be obtained and processed. There is a trade-off between degree of anonymisation and utility; where anonymisation is not possible, consideration of the purpose to which data is put is critical.

Where data are completely anonymous, the barriers to processing those data from a legal perspective are generally very low. However, it should be recognised that “anonymisation” is not usually considered an absolute term. “Pseudonymisation” refers to a process by which data are de-identified by replacing personally identifiable fields with artificial identifiers or “pseudonyms”. The level or degree of pseudonymisation should be assessed in relation to the amount of effort or technical resources that would be required to re-identify a patient, given the safeguards that have been put in place.

Personal, patient-level data are considered sensitive because of the personal-identifying information that are attached to the record, and the consequences that might arise for the individual from unrestricted dissemination. In its simplest form, data protection legislation permits that identifiable (i.e. personal) data may be processed if either (a) consent is obtained from the data subject, or (b) the law permits the processing of those data on some other legitimate basis; legitimate bases are defined by the intended use to which those data will be put.

Throughout this report we regularly use the terms “primary” data use, and “secondary” data use which involves some re-purposing of the data beyond the context for which it was originally collected (its primary use). This is very related to the notion of collecting consent from patients to process their data for a specific purpose, and then “further processing” data for a new purpose. Legal bases for the secondary use of data without explicit patient consent are generally determined by consideration of the intended use of the data, and the *compatibility* of this secondary use with the original *purpose* for initial data collection. For electronic health records, for example, this original purpose is to directly manage a patient’s care. Therefore, the later use of this data beyond direct care of a patient must be considered carefully; this includes the use of data for research purposes, which will not directly impact an individual’s care pathway but may inform and shape future health policy which is in the public’s interest. For this reason, health data are generally given a special status in data protection legislation, in recognition of its fundamental contribution to health care evaluation or audit, prevention practices, and medical research, as well as its highly sensitive nature. The use of health data for research poses some fundamental challenges, as there may be multiple purposes to which the data are put. In addition, research is enquiry-driven, and the pertinent research questions may evolve over time. For personalised medicine in particular, access to big data sets will become more and more necessary, which may require ongoing and long-term relationships with data subjects around the purpose of data. What constitutes acceptable secondary use of health data is a grey area and practice has varied substantially by country.

2.4. GDPR

In an effort to harmonise regulatory frameworks across Europe, the new General Data Protection Regulation (GDPR) came into force in May 2018, replacing the EU Data Protection Directive (Directive 95/46/EC). The previous *Directive* left considerable room for interpretation and resultant differences in local data protection laws and policies. The current *Regulation*, on the other hand, by its nature is self-executing, and immediately became law across all member states when it was enacted.

Rapid technological developments, new challenges to data protection, and lack of harmonisation were all cited as drivers in the need for change in the legislation (European Commission, 2010). The European Commission published a proposal for the GDPR in 2012 (European Commission, 2012), to which the European Parliament proposed several amendments in response, most notably eliminating the possibility for use of personal data for scientific research purposes without specific consent (European Parliament, 2012). Reacting to these proposed amendments, Di Iorio, Carinci and Oderkirk (2014) commented that (if the amendments were to stand as written) “*the right to privacy is likely to override the right to health and health care in Europe*”. Concern around these clauses led to significant debate across Europe, which in the end

successfully led to an agreement between the European Parliament, the Council and the Commission to include provisions in the GDPR to support scientific research (European Data in Health Research Alliance, 2015).

In considering our research question (can the legal conditions in Europe support expanding uses of health data?) it is worth noting the key passages of the GDPR that impact this question.

Chapter 2 of the GDPR text (Articles 5 to 11) outlines the 'Principles'. **Article 5 1(b)** states that personal data shall be "*collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')*". This is an important summary of the fact that, whilst the processing of data should not be incompatible with the original purpose for data collection, certain activities are deemed compatible, including scientific research. In addition, Article 5 specifies 'lawfulness, fairness and transparency', 'data minimisation', 'accuracy', 'storage limitation', 'integrity and confidentiality', and 'accountability' on behalf of the data controller.

Article 6 describes the lawful bases for processing data, for which one of the following must apply: (a) *consent* – for a specific purpose, (b) *contract*, (c) *legal obligation* – necessary to comply with the law, (d) *vital interests* – protecting life, (e) *public task* – performing a task in the public interest or (f) *legitimate interests* – "*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 which require protection of personal data*".

The GDPR also has separate provisions for the processing of "special categories" of personal data, which includes health data and is set out in **Article 9**. It states that the processing of these personal data shall be prohibited, unless the data subject gives explicit consent *or* one of ten conditions apply. The most relevant for pharmaceutical activities are: **2(g)** processing is necessary for reasons of "*substantial public interest*", **(h)** processing is necessary for the "*provision of health or social care or treatment or the management of health or social care systems*", **(i)** processing is necessary for "*reasons of public interest in the area of public health*" including "*ensuring high standard of quality and safety of health care and medicinal products*"³, and **(j)** processing is necessary for "*scientific or historical research purposes*", which like Article 5 references Article 89 of the GDPR, that outlines the conditions and safeguards that must be in place.

On 25th May 2018, coinciding with the implementation of the GDPR, the Article 29 Working Party ceased to exist; this was an advisory board formed of representatives from all EU Member State's data protection authorities as well as the European Data Protection Supervisor and the European Commission, and provided advice on the application of the data protection Directive. This has now been replaced by the European Data Protection Board, which has an enhanced status as an independent legal body of

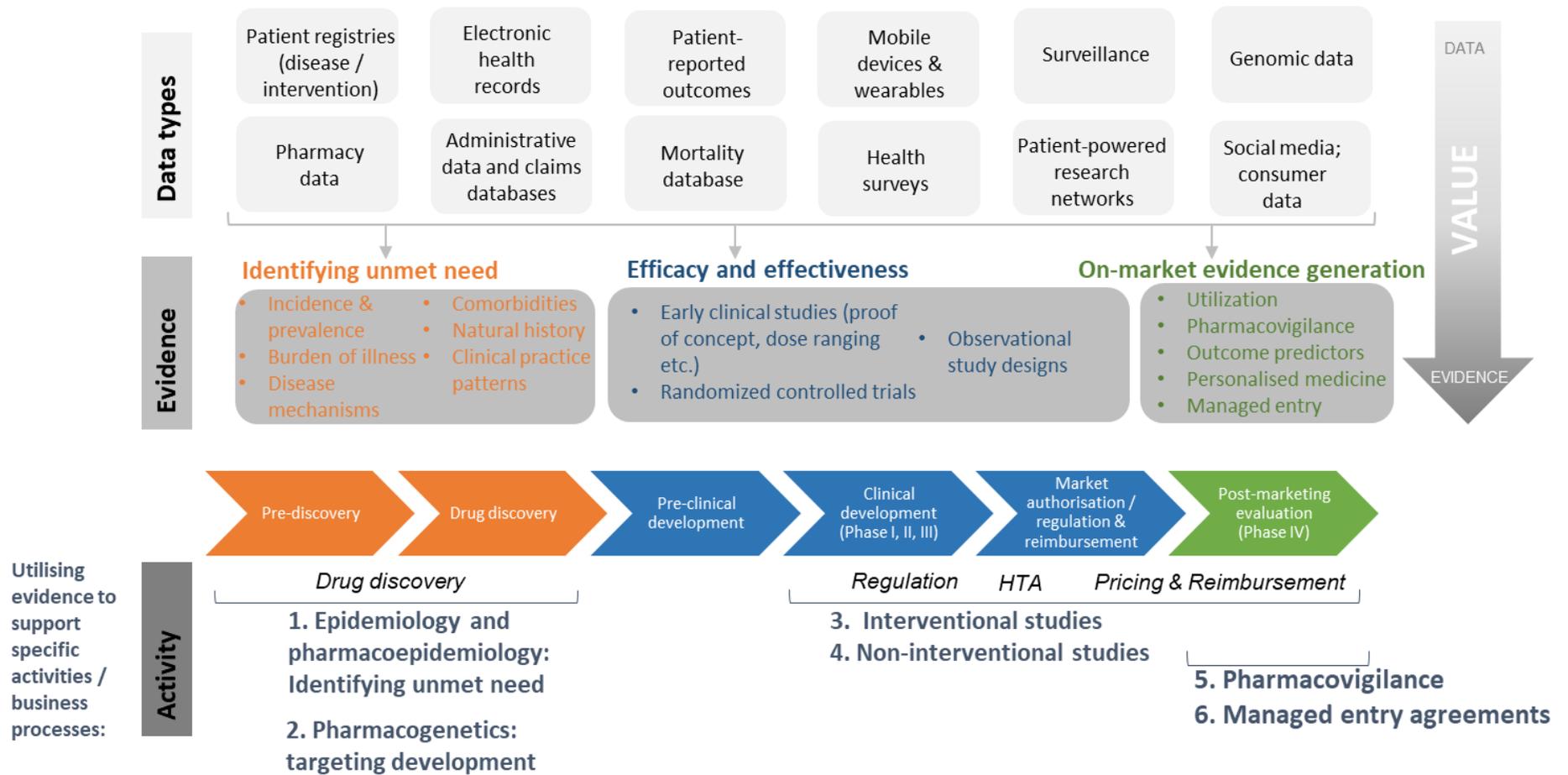
³ Recitals 53 and 54 provide more detail on processing sensitive data in health and social care sector (53), and in the public health sector (54)."

the EU; its primary aim is to contribute to the consistent application of the GDPR throughout the EU.

3. FRAMEWORK: IDENTIFYING THE BARRIERS TO BETTER USE OF DATA

In order to understand specific legal issues arising from the use of health data, we set out a framework (Figure 1) to visualise the six specific pharmaceutical activities of interest, set out in Section 1.2 above. The Figure depicts these activities in relation to the lifecycle of a medicine – linking these with the stages of evidence generation and data types to support them. The framework modifies and expands the work of others (e.g. Galson and Simon (2016)) to include evidence generation across the whole life of a medicine, and linking these with specific activities in the medicine’s discovery, regulation, pricing and reimbursement.

Figure 1. Framework for evidence requirements during the lifecycle of a medicine



Across the top of Figure 1 we detail examples of the various data types that can contribute to evidence generation activities, which cover: identifying unmet need, demonstrating efficacy and effectiveness, and on-market evidence generation. These correspond with the phases outlined across the bottom of early drug discovery, pre-clinical and clinical development, and evaluation for marketing and post-marketing purposes. The six activities were selected to represent a broad range of processes for which access to health data is fundamental, and – due to the spread of their positioning along the development process and reliance on different data types – give rise to a diverse set of legal issues.

The broad range of data types contribute to a much richer understanding of patient and population needs, treatment effect, and meaningful outcomes. Data types include traditional clinical data sets (registries, mortality databases etc.); databases that support the delivery of clinical care (e.g. health records) or administer claims and payments; real-time medical data that can be tracked through mobile devices and wearables; surveys and other data collection activities to capture patient outcomes or population needs; genomic data which opens up possibilities to stimulate innovation as well as guide development (Genomics England, 2017), and; data from other sectors entirely such as social media and consumer data, for which there may be opportunities to link with or inform our understanding of health. “Real-world” sources of data are now often being used to supplement clinical development and provide evidence to demonstrate the clinical and/or cost-effectiveness of new technologies. With this in mind, it is important to distinguish between ‘de-novo’ health data (for example a dataset created specifically for the purposes of a clinical trial or research project) and routinely-collected health data. Whereas for de-novo data collection the requirements around data governance and protection (e.g. ethical approval and consent) can be planned prospectively, the utilisation of routinely-collected data (e.g. electronic health records) for research or other purposes can be more problematic, as this involves the re-purposing of data for a ‘secondary use’ (i.e. beyond what the data were originally collected for: managing a patient’s care). In addition, where de-novo data could be usefully analysed at a later date beyond the scope for the original data collection/analysis activity, this also becomes a secondary use. Legal frameworks are playing catch-up in order to try to accommodate these new uses of data, which benefit the public and society but in a different way. This is explored further below, in relation to the six specific activities.

Whereas data represent the raw material, evidence provides the insight. Along the right of the diagram there is an arrow which depicts that there is a transformation of data to evidence, which in Cole et al. (2015) we describe as a “value chain”. It is the evidence generated from this process that is of value to stakeholders in health care. The steps that facilitate the transformation of raw data into analysis and results include: accessing the raw data, cleaning and managing the data, linkage and aggregation, and analysis. The rules around how and by whom these steps of the value chain can be conducted are dictated by the legal context, and practice varies by country.

For each of the six activities, we consider what are the relevant data sources, what the “purpose” of data use is, the opportunities, and the ensuing legal barriers. Our findings are summarised in the remainder of this section. In Section 4 we will group the findings

under eight cross-cutting themes, and in doing so discuss the issues in light of further research and propose potential solutions.

3.1. Epidemiology and pharmacoepidemiology: identifying unmet need

Overview

Sources of data relevant to this activity include patient registries, administrative data, patient-powered research networks, health surveys, genomic data and social media/consumer data.

It is important to distinguish between *primary* and *secondary* data in epidemiological studies to identify unmet need. Where there are no large-scale secondary sources to support the early investigation of unmet need for new treatments, then a primary study would be conducted, for which patient consent would be collected (e.g. laboratory data to characterise the presence of a pathogen in support of vaccine development). However, the main source of information at these very early stages of identifying unmet need would usually be secondary studies, i.e. relying on other already-existing sources of data. There are many potential sources of data that could support this activity, for example surveillance networks, cancer registration data, WHO's burden of disease, Biobanks, etc. Anonymised population-level data (large datasets containing no identifiable patient information) are regularly used, and usually meet the purposes of this business activity; therefore, barriers to access and ethical concerns are generally low. There are several models of data access for this purpose: (1) retrospective anonymised data, which can be hosted internally (as licensed data), (2) access through a 'safe haven' (access to a data platform to run analytics), or (3) buying the evidence (e.g. from academic groups or other data provider organisations). However, data linkage is sometimes required to gain further important insight, for which the industry is reliant on the access provisions made available by data hosts and/or the nature of the consent already collected.

Opportunities for better use of data

Interviewees suggested that a currently under-developed opportunity is the further linkage of data from diverse sources and across non-health sectors. For studies looking to characterise unmet need, of interest are not only those patients suffering from a condition, but also those that might in the future. For the second group, lifestyle factors can be very important, and linkage between health and other datasets could be extremely valuable (e.g. wearables, energy use, spending patterns, employment, etc.).

Perceived legal barriers or uncertainties identified

The following three issues were identified as the most important in relation to this activity:

1. Strong reliance on **anonymised data**; recognition that there are degrees of anonymisation, and lack of clarity around whether GDPR will alter the definition of this.
2. Heterogeneity in **data access** models and legal frameworks across countries.
3. Reliance on data providers to collect appropriate **consent** which can permit valuable linkage, and lack of guidance on legal framework to permit **data linkage** across wider (non-health) sectors e.g. energy use, lifestyle data etc.

In addition, the “conservatism” of the pharmaceutical industry in adopting global policy that adheres to the most stringent legal environments was noted to limit opportunities to utilise data in support of this activity. There was also a perceived regulatory ambiguity for primary data collection studies around what is considered ‘interventional’ and ‘non-interventional’, which has important implications for requirements around monitoring, good clinical practice, and consent; in turn, the different consent requirements have implications for the later utilisation of data.

3.2. Pharmacogenetics: targeting development

Overview

Sources of data relevant to this activity include patient registries, administrative data, health surveys, and genomic data. Pharmacogenetics is the study of inherited genetic differences in drug metabolic pathways and is conducted early in clinical development. Useful data sources are often external to the company (e.g. cell line repositories and population genetics research and databases). In these cases, the onus is on the data provider – such as the national repositories or academic institutions – to operate within the legal framework. This includes applying the appropriate level of anonymisation or collecting the appropriate consent, and then sharing these details with the company.

Opportunities for better use of data

There are expanding opportunities for the use of genomic and other high-dimensional data throughout the R&D lifecycle, and an expansion of advanced therapeutic medicinal products (ATMPs), which are typically more personalised treatments.

Perceived legal barriers or uncertainties identified

The following two issues were identified as the most important in relation to this activity:

1. Recognition that true “**anonymity**” of genomic data is unlikely, in which case consent or other legitimate bases should be relied upon.
2. How to ensure the company can **re-use data** at a later point in development (iteratively and across the lifecycle of a product: personalised medicines, safety, minimising risk, etc.).

Other issues raised included the fact that, under GDPR, genetic material is now explicitly included in the definition of ‘personal data’, which may impose a more restrictive framework. Linkage between genetic material and other health data was also highlighted, as was the question as to whether genetic material collected as part of clinical care versus that collected in research should be treated differently.

3.3. Interventional studies

Overview

Sources of data relevant to the assessment of efficacy and effectiveness in clinical studies could span the whole range of data types included in Figure 1. Both randomized controlled trials (RCTs) and pragmatic trials are primary, interventional studies, that are defined by their randomised treatment allocation. However whereas RCTs are explanatory trials designed to measure efficacy (the benefit of treatments under ideal conditions), pragmatic trials measure effectiveness, and are designed in such a way that maximises generalisability of results: eligibility, recruitment and setting are by design

more akin to 'real practice', and follow-up is generally more reliant on routinely collected sources of data such as electronic health records. Patients provide consent when they enter any interventional trial, which are associated with a high level of legal and ethical obligations. Whilst patients consent to the processing of their information for analysis of study data as specified within the trial protocol, at the point of initial consent it may not be possible to anticipate and specify how data could be used to optimally address future research questions that may subsequently arise.

Opportunities for better use of data

Digital tools to support treatment provision and monitoring are of increasing quality and relevance. There is also a growing utilisation of RWD in general to capture and assess outcomes in interventional studies. Further opportunities include linkage with data sources from other (non-health) sectors to predict outcomes or identify trial participants.

Perceived legal barriers or uncertainties identified

The following three issues were identified as the most important in relation to this activity:

1. Challenge: constructing **consent** in such a way that facilitates later use of data; envisaging future research questions can be difficult, and the need to be explicit may pose a barrier to future research.
2. Feasibility of **re-consent** (or "obtaining new consent"). For primary data, there are at least opportunities for re-contact with the patient, but this is difficult and there is likely to be high drop-out.
3. In making judgements on the compatibility of research applications (to **re-process** data) with the original trial protocol, ethics committees and their interpretations are highly variable.

Other issues identified were:

- For cross-country studies: incompatibility between sources of data, for which technical issues could derive from different legal environments.
- Compatibility between EU Clinical Trials Regulation and GDPR? (e.g. consent being required outside protocol: Article 28.2 versus GDPR Article 9.2.).
- Can consent collected by digital means ever be truly "informed"? (collection of consent not on the basis of a conversation). Are methods to obtain informed consent compatible with routine collection of RWD?
- Lack of information or guidance on the legal framework for linking with data across (non-health) sectors.
- Unique challenges for gene therapy products.

3.4. Non-interventional studies

Overview

Non-interventional or 'observational' studies involve no treatment allocation. The possible sources of data for observational studies are multiple and span all those listed in Figure 1. Prospective and retrospective observational studies should be distinguished. To the extent that some prospective observational studies also involve obtaining patient consent, some of the issues described above relating to 'pragmatic trials' also relate to observational studies, and are therefore duplicated below. While there are issues around

collecting consent from patients for prospective observational studies in such a way that covers all requirements, the legal environment becomes more complicated for retrospective studies which involve secondary uses of routinely collected data. For secondary use, we must consider the compatibility of the purpose of processing data, with the original premise under which those data were collected (e.g. electronic health records: to manage a patient's health care provision), *and* to what extent the 'insight' obtained is in the public's interest, supports scientific research, or supports the provision/management of health care (each of which represents a legal basis for processing sensitive data under the GDPR).

There are several models of data access for secondary data sources, including retrospective anonymised data which can be hosted internally (licensed data), access through a 'safe haven' (access to data platform to run analytics), and buying the evidence (e.g. from academic groups or other data provider organisations). However, for observational studies, data linkage is often key to obtaining valuable insight, which means that data cannot be fully anonymised.

Opportunities for better use of data

Digital tools to support treatment provision and monitoring are of increasing quality and relevance. As well as supporting research and insights into disease progression, they can also provide further information to health professionals on a patient's needs, and provide aggregated data for health providers (to support patients' management) as well as pharmaceutical companies (to support R&D). Utilisation of observational data can provide the means to change advice to patients based on the analysis of their individual data. Data linkage opportunities with wellness and activity data are also expanding, as are possibilities to re-examine bio-samples for the purposes of genetic profiling.

Perceived legal barriers or uncertainties identified

The following five issues were identified as the most important in relation to this activity:

1. **Divergent** data governance requirements in multi-country studies (e.g. granularity of data, levels of access, permitted use, etc.).
2. Access to data for pharmaceutical companies in particular can be **restrictive**, limiting the utility of data for analysis (models of access and requirements for partnerships vary).
3. It is difficult to make electronic health records (in the form of a free-text clinical record) **truly anonymous**, which may place limits on their use for research.
4. Poor standards of existing (external) data repositories which do not meet the legal / governance standards required; also relevant for partnering with small third-party tech firms. **GDPR** places new legal responsibilities on data processors, not just data controllers. There is therefore an increasing risk of legal exposure from association with non-compliant data sources.
5. Feasibility of **re-consent**. For primary data, there are at least opportunities for re-contact with the patient, but this is difficult and there is likely to be high drop-out. For secondary use data, the same challenges arise but it is even more challenging as there are no open lines of communication with the patient.

Other issues identified were:

- Re-purposing wellness and activity data for health care purposes (e.g. apps, fit bits etc.): understanding the complex legal landscape and multiple entities receiving those data. Given this complex landscape, is “informed consent” possible?
- It is difficult to anonymise certain types of data, e.g. relating small population such as rare diseases.
- The use of digital technologies to support drug delivery: blurs boundaries between “research” and “provision / management of health care”, for which the legal bases for data processing are distinct.
- Digital health: requirement to work through third party to address data quality / device functionality issues, which is complex and time consuming.
- Inadvertent re-identification of an individual: clearer guidance required; conservatism of the pharmaceutical industry currently poses a barrier to important analyses.
- Implementing the “right to be forgotten”: implications for already-processed data?
- Can consent collected by digital means ever be truly “informed”? (i.e. collection of consent not on the basis of a conversation).
- Is specific informed consent compatible with the collection of RWD or the process of conducting big data analyses?

3.5. Pharmacovigilance

Overview

Sources of data relevant to pharmacovigilance include patient registries, surveillance, pharmacy data and social media / consumer data. Data to support pharmacovigilance could be based on *primary data* (e.g. data collected by a company through clinical studies) or *secondary data* (individual case reporting, e.g. from medical records, insurance or reporting schemes).

Opportunities for better use of data

It has been suggested that the U.S. Sentinel initiative – which relies on a distributed data network with set algorithms – could serve as a model for Europe. In addition, social media could play a more prominent role in pharmacovigilance.

Perceived legal barriers or uncertainties identified

The following three issues were identified as the most important in relation to this activity:

1. Obligations on individual case reporting from **social media** are unclear.
2. For primary data collection studies: **ambiguity around what is considered ‘interventional’ and ‘non-interventional’**; important implications for requirements around monitoring, GCP, consent model etc.
3. **Ambiguity in the revised EMA Guidance** on Good Pharmacovigilance Practice (GVP) Module VI on non-serious adverse events in post-authorisation non-interventional studies: obligations to be explicit in what you are not going to capture.

The above three issues represent uncertainties, which are hoped to be clarified with further guidance from the EMA. Other issues arising were:

- Appropriateness of EudraVigilance requirements to share source data for 'inspection' purposes.
- Using routine data for pharmacovigilance: necessary identifiability of the individual in order to populate a returnable adverse event. Consent requirements?
- Patient support programmes: obliged to collect data on adverse events, but quality and governance arrangements in low cost international centres are problematic.

3.6. Managed entry agreements

Overview

The potential sources of data to support managed entry agreements (MEAs) could be broad, but are usually on the basis of clinical registries, administrative data, or ongoing trials. MEAs refer to any scheme between a payer and a pharmaceutical company to explicitly match payment with some performance indicator. For example, where the payer observes some risk that clinical trial performance will not be replicated in real clinical practice, the company could offer to pay the health service back in full or in part for any patient for whom the medicine did not achieve a certain pre-agreed outcome.

MEAs require data collection alongside clinical practice, and for these data to be processed in order to calculate payments or rebates. This is often operated through a third party, although data could be processed by the health care provider (with the company [contractually] reserving the right to audit data through a third party). The company is usually provided with aggregated data only, to process payments. Data used for this purpose will sometimes be data already collected (e.g. routine claims or resource use data); specific consent for processing data as part of an MEA is generally not collected.

Opportunities for better use of data

Current implementation of MEAs is relatively limited in most countries; there is therefore an opportunity for increased use of MEAs to improve value-for-money for health systems. A positive externality could be the improved collection of outcomes data by service providers.

Perceived legal barriers or uncertainties identified

The following was identified as the most important in relation to this activity:

1. Ambiguity: to what extent does processing data for the purpose of MEAs fall under one of the **legitimate bases** for processing data in the absence of consent: Public interest? Provision / management of health care? This must be supported by member state law.

Other issues raised were:

- Would it be ethical to collect explicit patient consent for processing data for this purpose?
- Third party auditor must be able to verify data processing system to ensure that the algorithms are working; the health system must be open to permit the processing of pseudonymous data in this way.

3.7. Summary

We set out to identify legal barriers to the better use of health data, by investigating the main issues in relation to specific activities along the pharmaceutical lifecycle.

For activities undertaken very early on in the pharmaceutical R&D process to identify unmet need – epidemiology, pharmacoepidemiology and pharmacogenetics research – the main issues identified relate to ensuring the company can re-use data at a later point in development, to enable important data linkages or to strengthen later research. In addition, heterogeneity of data access models and legal frameworks across countries was noted to be a significant barrier, as well as the degree of anonymity possible for certain types of data (e.g. genomic).

For demonstrating evidence of efficacy and effectiveness, there was a differentiation between legal barriers for primary studies (involving de novo data collection) versus secondary studies (involving repurposing of previously collected / routinely collected data). For primary studies (usually interventional, but sometimes observational) there were issues around envisaging future research questions and the related feasibility of obtaining re-consent later for new uses of the data collected. For secondary studies, the main issue related to judging the compatibility of original data collection with proposed “new uses” of data, and the feasibility of anonymising all types of observational data. For both primary and secondary studies, there were issues arising over the heterogeneity of data governance requirements in multi-country studies, and restrictive data access for pharmaceutical companies.

The major theme for on-market evidence generation was understanding the legitimate legal bases for processing data in the absence of consent; this theme was relevant across all six activities analysed.

By investigating the legal issues arising for six activities across the pharmaceutical lifecycle, we have articulated the specific issues being faced on the ground by individuals working for the better use of health data. To take the issues forward and consider them at a policy level, it is necessary to consider the legal issues at a more aggregated level. In the next section we offer an assessment of the legal barriers according to eight cross-cutting themes, which encapsulate all the specific barriers raised: data subject rights, anonymisation, consent, uncertainties around appropriate legal basis, compatibility of primary and secondary (re-)use of data, heterogeneity, issues relating to digital health, and engendering trust.

4. CROSS-CUTTING THEMES

In order to reflect on the issues arising and consider their implications more broadly, in this section we consider the barriers identified under several cross-cutting themes, which all relate to the legal basis for data collection and use. Within each theme, we speculate on the potential ways forward or clarifications required.

4.1. Data subject rights

The GDPR further enshrines data subject rights, including for example requirements around: transparent communication, time limits for complying with rights of the data subject, right to access free of charge, right of rectification, right to erasure (“to be forgotten”), and right to restrict or object to processing. Data subject rights arose as a

perceived legal barrier in terms of implementing the right to erasure, and concern over the implications for already-processed data.

It should be noted that the data subject rights listed above are not absolute rights. With regards to the right to erasure specifically, described in Article 17, this right applies if data are no longer necessary for the original purpose, or if a subject withdraws consent “*and where there is no other legal ground for processing*”. In addition, the right does not apply if the data processing is necessary for reasons for public interest in the area of public health. Whilst consent to participate in a clinical study can be withdrawn, data processors should be clear from the outset about the implications for use of the data, for example no further collection but retention of data already collected, and/or retention of data for use in the project but not in future research (HRA, 2018a). This must be specified upfront.

In summary, data subject rights are likely to be more limited where the processing of data is for health research, and the implementation of appropriate safeguards permits these research exemptions to be applicable. Nevertheless, guidance and an industry-wide position of how these rights apply to pharmaceutical research would be of great benefit.

4.2. Anonymisation

There is a critical policy distinction between personal and non-personal data. Where data are anonymous, they are considered non-personal data and therefore not subject to the rules and restrictions outlined in data protection legislation. However, the concept of anonymisation is not absolute; there are different levels of re-identifiability. Through Article 4(5) the GDPR introduced into legislation the term “pseudonymisation”, which is the separation of data from direct identifiers so that association with a personal identity is not possible without additional information that is held separately (and subject to “technical and organisation measures” to ensure the data are not attributed to a person). This preserves the ability to link data but reduces the risk of re-identification. Re-identification risk must be considered along with the safeguards and precautions in place to protect the data, and the technical resources that would be required to re-identify a person. In consideration of whether data are personal or not, in legal terms the focus should be on the *effect* of the measures in place to protect identity (is the person identifiable?) rather than the means. Whilst this effect-based definition is used in the GDPR, there is concern among stakeholders around the lack of clarity on how to meet the status of anonymity.

Several issues arose in relation to anonymisation for the pharmaceutical R&D or evaluation activities analysed, mainly in relation to the degree of anonymisation that would be required for specific uses. It was noted that for certain types of data – genomic, free-text patient record data and data relating to small populations – achieving full anonymisation would be impossible. Therefore, it is necessary to take a probabilistic approach to minimising risk, taking into consideration the safeguards in place to protect data.

By introducing the term pseudonymisation there is a perceived risk that the GDPR expands the scope of what is considered to be personal data, and thereby restricts research using pseudonymised datasets. In consideration of the GDPR’s implications for administrative data research in the UK, Mourby et al. (2018) argue that the GDPR will

not expand the category of personal data, and therefore will *not* bring more types of health research within the remit of the data protection regulations. Their interpretation is based on the text of the GDPR which states that whether data are considered personal should be determined by assessing whether there “is a means reasonably likely to be used to identify individuals”, rather than merely a theoretical possibility of identification. In this regard, anonymisation should be considered according to the context within which the data will be processed.

It would be useful to have a process guide around what level of de-identification is appropriate under which circumstances. Where complete anonymisation is not possible or appropriate (which is the case for most pharmaceutical activities we have discussed in this report), then processors must have a legal basis for processing the data: either consent or some other legal basis provided for by the legislation.

4.3. Consent

By obtaining explicit consent, the data controller can legitimately process personal data. The challenge is constructing consent in such a way that is *broad* enough to permit later use, but *specific* enough to comply with the legal standards. The GDPR heightens requirements for very clear and specific statements of consent. This can pose a challenge for medical research, as potential future research questions can be hard to envisage at the outset. Obtaining explicit consent that meets the requirements of the GDPR could therefore act as a barrier to future research. A related challenge is the feasibility of re-consent where the data controller wishes to process data for a new purpose, or to bring ongoing studies in line with the requirements of the GDPR. For primary data, there are at least opportunities for interaction with the patient, though re-consenting participants is difficult and associated with a high drop-out rate. For secondary data, there is reliance on data providers to collect the appropriate consent which can permit the processing activities required. A fluid model of re-consent that could be envisaged is dynamic consent. Dynamic consent could give patients the power to approve use of their data on a case-by-case basis, for example using digital models of iterative opt-in / opt-out. However, the feasibility of this approach and burden it would place on participants and researchers alike may preclude certain research activities.

The legal barriers and challenges outlined fall away if consent is not used as the lawful basis for processing data. Consent for participation in research is not the same as relying on consent as the legal basis for processing data under the data protection legislation. In the UK, for example, research authorities and councils have explicitly said that consent should not normally be the legal basis for processing data in health research (HRA, 2018b; MRC, 2018). Consent is an integral part of research studies, particularly those of an interventional nature, but this is for reasons other than data processing (for example meeting clinical trial regulations, confidentiality obligations, and ethical considerations). Rather, other legal bases should be used for processing data, which can be justified by the “legitimate interest” of the organisation (Article 6 of the GDPR), in combination with the safeguards and controls that are in place to protect the data⁴. The

⁴ As health data is within the “special category” outlined in the GDPR (Article 9), additional justifications apply: provision of health or social care, public interest, or scientific research. More on this in Section 4.4.

onerous consent requirements outlined by the GDPR, which are not very compatible with medical research, would therefore not apply.

It is likely that pharmaceutical companies, in wanting to avoid the risk of non-compliance with the GDPR, look to their process of obtaining consent as a way to explicitly set out and gain “permission” for all data processing activities; we argue that this may actually put companies in a more risky position, as achieving the high standards set in the data protection legislation would be difficult in a research context (whether commercial or public), and also implies no other legitimate basis for using the data is available.

Whilst there is some flexibility provided for in the regulation around ways to demonstrate compliance, data processors must be clear on which bases they are using. The Article 29 Working Party issued guidelines relating to consent under the GDPR, which has been “endorsed” by the European Data Protection Board that replaces it (Article 29 Working Party, 2017; EDPB, 2018). The guideline states that, in relation to Article 6 regarding legitimate bases, the controller cannot “swap” from consent to other lawful bases, which must be determined and stated at the time of collection of personal data.

4.4. Uncertainties around appropriate legal basis: GDPR

A major source of uncertainty for industry is the appropriate legal basis for processing data in the absence of explicit consent, and understanding what activities reasonably fall under the various exemptions provided by the GDPR. There is a *prima facie* case for assuming that these data processing activities are likely to be in the interests of society. For example, understanding the effectiveness and cost-effectiveness of a medicine is critical in ensuring patients are being offered the best available options. Similarly, ensuring payers obtain value for money through managed entry agreements is also in the interest of payers, as well as patients and industry if access would otherwise be denied. Yet, defining these bases legally is difficult. The Regulation’s most relevant text is described in section 2.4 of this report. In summary, Article 6 sets out the alternatives to consent, the most relevant for the activities discussed in this report being “legitimate interests”. For health data, an *additional* legal basis to processing this ‘special category’ personal data is required, as set out in Article 9. For our purposes, these are: (h) provision of health or social care, (i) public interest in the area of public health, or (j) scientific research. For research studies, the legal basis must be stated in the privacy notice.

Before speculating on the appropriate legal bases for data processing, it is pertinent to reflect again on the differentiation as between primary data collection and secondary re-use (or re-processing) of data. As described throughout this report, consent is generally already collected as part of primary data collection activities (though, as described above, this may not be the most appropriate basis for data processing). In Table 1, below, we speculate which legal bases from Article 9 (*other than consent*) could conceivably apply to the six activities examined in this report.

Table 1. Potential legal bases for processing data (*authors' speculation*)

		Relevance of additional legal bases for processing data in the absence of consent (GDPR Article 9) – Authors' speculation		
		(h) provision of health or social care	(i) public interest	(j) scientific research
Six key activities considered	Epidemiology and pharmacoepidemiology: Identifying unmet need		Pursuing new development opportunities to target unmet need	Researching health needs and possible solutions
	Pharmacogenetics: targeting development		Finding new targets	Early clinical development
	Interventional studies			Clinical development: efficacy studies
	Non-interventional studies	In addition to providing data for research, digital health tools can support clinical care and decision-making		Clinical development: Effectiveness studies
	Pharmacovigilance	Part of routine care and vigilance systems	Ensuring high standard of quality and safety of medicinal products	
	Managed entry agreements	Resources allocated to health care is part of routine management of health care service	In the interest of society to ensure health systems are able to provide access to cost-effective treatments	

All the potential legal bases described in Table 1 are contingent upon the safeguards that must be put in place to protect data, as well as requirements that the processing is fair, lawful, transparent, and accords with data minimisation standards and individual rights. The suggested justifications are conjecture: there is a need for clear guidance and consensus on how the exemptions to consent apply to specific pharmaceutical activities, in order to support pharmaceutical companies in being explicit in which legal basis they are using.

4.5. Compatibility of primary and secondary (re-)use of data

Having a good understanding of the 'purpose' of data collection, and the purpose of its subsequent use, is critical in understanding the legal requirements around how those data are managed and protected. This issue relates to some of the themes already outlined, but is worth pulling out separately given its importance for research using

secondary data, which includes most activities considered in this report outside of interventional studies, or where there is value in processing data from clinical trials at a later stage for a different purpose.

A particular barrier that was raised in relation to this theme was the variable judgements of ethics committees in considering the compatibility of research applications (to re-process data) with the original trial protocols. Individual interpretation appears to play an important role within ethics committees, leading to variable and unpredictable outcomes. Also of relevance is concern around the reliance on data providers, and the poor standards of existing (external) data repositories which do not meet the legal / governance standards required.

In understanding the legitimacy of re-processing data for research, the considerations outlined above in relation to legal exemptions stand.

4.6. Heterogeneity

Heterogeneity in data access models, data linkage, governance arrangements and legal frameworks across countries hampers valuable cross-border research, and leads to duplicated effort for multi-national organisations. Developing common standards and approaches to health care data access across Europe would support health care innovation and benefit industry, patients and society. Guidance on how best to navigate member state requirements would also be helpful.

The GDPR was introduced with the aim of harmonising the legal framework for data processing across Europe. However, there is some concern that the GDPR may serve to fragment rather than harmonise practice across countries. As well as leaving significant room for interpretation, there are many "opening clauses" within the GDPR which permit member state modifications, as well as specific reference to permitted divergence in member state law in recital 53 (processing of sensitive data in health and social care sector): *"Union or Member State law should provide for specific and suitable measures so as to protect the fundamental rights and the personal data of natural persons. Member States should be allowed to maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health."*

4.7. Issues relating to digital health

Several legal issues were raised relating to digital health specifically, an area of expanding interest and use. Among issues raised was the legality of re-purposing wellness and activity data for health care purposes (would these be included in special category personal data? Given the complex landscape and multiple organisations involved, would truly informed consent be possible?) In addition, digital health to support drug delivery could blur the boundary between "research" and "provision / management of health care", for which the legal bases for data processing are distinct. The costly requirement to work through a third party to address data quality and device functionality issues was also raised.

There is a need for some clear guidance or sets of minimum standards for industry in the emerging area of digital health. The benefits of sharing and linking data must be understood and shared with the wider community, to promote confidence and trust.

4.8. Engendering trust

Data protection is all about protecting people. Trust is essential, for which transparency is key. Patients and society must trust in the system that governs how personal data should be protected and used for good. They must also trust the organisations entrusted with their data to work within the legal and ethical standards set by that system. Losing trust is easy; isolated “bad examples” can cause considerable damage to a whole industry. Establishing trustworthiness is much more difficult, but is necessary and must be achieved through being transparent and sharing good practice. Alongside efforts to communicate the good to the public that is generated from pharmaceutical R&D activities, including the benefits of data processing for research, the public must also be convinced of the high safeguarding standards employed in their handling of sensitive information.

In a report by the International Risk Governance Center (IRGC) on governance of trust in precision medicine, the authors outline the central concept of trust across the three stages of the precision medicine value chain. The concepts can be applied across all pharmaceutical industry outputs. First, data collection must be fair and responsible: there must be trust that those who collect the data will protect privacy, and that there is a benefit returning to citizens and patients. Second, data analysis and governance must be transparent, reliable, and accountable; governance must establish and maintain trust in the data system. Finally, there must be trust in the whole system and its ability to improve public as well as individual health (EPFL IRGC, 2017).

To foster trust, the pharmaceutical industry needs to simultaneously expand the space for use of data without consent (by demonstrating other legitimate, worthwhile and legal bases) whilst also extending the concept of a sustained relationship with individual patients, based in part on consent, but also on mutual understanding and trust.

A perceived lack of trust in the pharmaceutical industry manifests in restrictive data access for companies, limiting the utility of data for analysis, or requiring partnerships with academic institutions. There is considerable variation between countries, which reflects the significant cultural differences. A report funded by the Wellcome Trust investigated public attitudes to commercial access to health data, finding that (in the UK) most are in favour as long as there is a clear public benefit and appropriate safeguards in place (Ipsos MORI, 2016). The key, therefore, is communicating this effectively with the public. Demonstrating the cybersecurity tools in place could help to build confidence, and sharing details with patients on the chain of custody of their data could increase transparency.

5. DISCUSSION AND PROPOSED NEXT STEPS

The opportunities to make better use of health data across the pharmaceutical lifecycle are high. However, the legal issues are significant, particularly as we lean more heavily on RWE to understand the potential for or realised impact of medicines in a real-world setting. In Table 2, below, we highlight (according to the main cross-cutting themes) some of the major challenges discussed in this report, alongside some possible solutions. Whilst some suggested solutions are specific to the pharmaceutical industry, most require the consideration or joint action of a broader set of stakeholders.

Table 2. Summary of challenges and potential solutions

	Challenge	Solution
Data subject rights	<ul style="list-style-type: none"> - GDPR further enshrines data subject rights. Of particular concern: right to erasure (“to be forgotten”) 	<ul style="list-style-type: none"> - Guidance and an industry-wide position of how these rights apply to health research - Clear specification by industry upfront of retention periods and how data will be handled if consent is withdrawn
“Anonymisation”	<ul style="list-style-type: none"> - Limited utility of truly anonymous data - Uncertainty around degree of anonymisation required for different uses - How to address re-identification risk in small populations 	<ul style="list-style-type: none"> - Clarify scope / concept of anonymisation and pseudonymisation - A probabilistic approach should be taken, which includes consideration of the safeguards in place to protect data, and the context of its intended use - Generation of a process guide around what level of de-identification is appropriate under which circumstances
Consent	<ul style="list-style-type: none"> - Constructing consent to be specific enough to comply with the legal standards, but broad enough to permit re-use across the value chain - Questionable viability of re-consent 	<ul style="list-style-type: none"> - Alignment on the alternative legal bases for data processing, which may be more appropriate than consent in the context of medical research (GDPR: scientific research, public interest, provision of health care)
Uncertainties around appropriate legal bases	<ul style="list-style-type: none"> - Reliance on consent may be inappropriate or unfeasible for some pharmaceutical data processing activities 	<ul style="list-style-type: none"> - As above: there is a need for clear guidance and consensus on how the exemptions to consent apply to specific pharmaceutical activities
Compatibility of primary and secondary (re-) use of data	<ul style="list-style-type: none"> - Variable judgements of ethics committees 	<ul style="list-style-type: none"> - More consistent interpretations required
Heterogeneity	<ul style="list-style-type: none"> - Heterogeneity in data access models, data linkage, governance arrangements and legal frameworks across countries hampers valuable cross-border research 	<ul style="list-style-type: none"> - Developing common standards and approaches to health care data access across Europe would support health care innovation and benefit industry, patients and society - Given the number of opening clauses and provisions for member state divergence in GDPR implementation, a cross-border initiative and shared understanding would greatly benefit industry and researchers; this must include national authorities
Digital health	<ul style="list-style-type: none"> - Enabling the re-purposing of wellness data for research 	<ul style="list-style-type: none"> - Clear guidance or sets of minimum standards for industry in the emerging area of digital health. The benefits of sharing and linking data must be understood and shared with the wider community, to promote confidence and trust.

	- Blurred boundary between research and provision of health care, for which the legal bases for processing data are distinct	
Trust	<ul style="list-style-type: none"> - How to address ethical/legal obligations to patients and citizens in a big data environment - Preventing bad news from defining policy - How to address pharma-only data access restrictions 	<ul style="list-style-type: none"> - There is a case for industry leadership, and collaboration across stakeholders in dealing proactively with the uncertainties, sharing good practice, and promoting trust. Consider code of conduct based on: <ul style="list-style-type: none"> Co-development of principles for responsible use Agreement and communication around: <ul style="list-style-type: none"> - what level of de-identification is acceptable for what use - data 'chain of custody'; - cybersecurity and safeguards; - legal bases for data processing (a menu of channels for which GDPR exemptions may apply);

Given its broad implications, its Europe-wide scope, and its very recent introduction, it is perhaps unsurprising that the GDPR featured highly in discussions around the legal barriers to better use of data. It may take some time to understand the GDPR's real impact on the pharmaceutical industry as well as health care delivery more broadly, as countries and courts live with and test the Regulation. We conclude that the GDPR does not create new legal barriers. It sets out the various legal bases for processing data, and it is for those working in or with the health care sector to understand how these apply to them. It is important to remember that the GDPR covers all sectors, including the technology sector for which sensitivities around data sharing and commercial interests are extremely high; the Regulation was not designed to hamper important scientific research.

In assessing the legal barriers to better use of data, many of the issues we identified were *uncertainties* rather than barriers per se. Data protection authorities, whilst best placed to provide guidance, are hugely stretched. There is therefore a strong case for industry to deal proactively with the uncertainties, sharing good practice and engendering trust by co-creating a code of conduct, outlining principles of responsible use. This could include agreement and communication around a data 'chain of custody'; cybersecurity and safeguards; what level of de-identification is acceptable for what use; and clear alignment of GDPR Article 9 exemptions, including a menu of channels for which GDPR exemptions may apply and how these relate to specific pharmaceutical R&D and evaluation activities. There should also be a platform for engagement with patients, which will be critical in encouraging a shared understanding of the value to society of pharmaceutical research.

The penalties for non-compliance with the GDPR are clear and extremely high: up to 4% of annual global turnover or €20 million (whichever is greater). Yet, as described in this report, there is currently little clarity or confidence in how to ensure that the pharmaceutical industry remains compliant. This challenge is further pronounced by the ever-expanding opportunities for better use of data across the pharmaceutical lifecycle. National data protection authorities must find a way of working with industry, in a way that enables research and reduces the legal risk of important data processing activities. An industry code of conduct could significantly clarify the issues and promote understanding. All stakeholders must be on board, as all stakeholders stand to benefit from the better use of health data.

REFERENCES

Article 29 Working Party, 2017. *Guidelines on consent under Regulation 2016/679. WP259 rev.01.*

Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C. & Conesa, A., 2016. Making sense of big data in health research: towards an EU action plan. *Genome medicine*, 8 (1), 71.

Cole, A., Garrison, L., Mestre-Ferrandiz, J. & Towse, A., 2015. *Data Governance Arrangements for Real-World Evidence*. 2015. London: Office of Health Economics. Available: <https://www.ohe.org/publications/data-governance-arrangements-real-world-evidence>.

Di Iorio, C., Carinci, F. & Oderkirk, J., 2014. Health research and systems' governance are at risk: should the right to data protection override health? *Journal of medical ethics*, 40 (7), 488-492.

EDPB, 2018. *THE EUROPEAN DATA PROTECTION BOARD. Endorsement 1/2018*. Available: https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents_en_0.pdf.

Eichler, H. G., Baird, L., Barker, R., Bloechl-Daum, B., Børlum-Kristensen, F., Brown, J., Chua, R., Del Signore, S., Dugan, U. & Ferguson, J., 2015. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clinical Pharmacology & Therapeutics*, 97 (3), 234-246.

EPFL IRGC, 2017. *Governance of trust in precision medicine* Lausanne: Ecole Polytechnique Federale de Lausanne. EPFL International Risk Governance Center. Workshop Report, Available: <https://irgc.epfl.ch/wp-content/uploads/2018/10/IRGC-2018.-Governance-of-Trust-in-Precision-Medicine.-Workshop-report.pdf>.

European Commission, 2010. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS: A comprehensive approach on personal data protection in the European Union. Brussels, 4.11.2010. COM(2010) 609 final.

European Commission, 2012. *Proposal for regulation of the European Parliament and of the Council on protection of individuals with regard to the processing of personal data and on the free movements of such data*. Brussels. COM(2012) 11 final.

European Commission, 2017a. Shaping the Digital Single Market. Available: <https://ec.europa.eu/digital-single-market/en/policies/shaping-digital-single-market> [Accessed 31/07/2018].

European Commission, 2017b. *State of Health in the EU: Companion Report 2017*. Luxembourg: Publications Office of the European Union. Available: https://ec.europa.eu/health/sites/health/files/state/docs/2017_companion_en.pdf.

European Commission, 2018. *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society*. 25.4.2018. Brussels. COM(2018) 233 final, Available: file:///C:/Users/acole/Downloads/CommunicationeHealth.pdf.

European Data in Health Research Alliance, 2015. Good news for health research from trilogue talks on the Data Protection Regulation. 17/12/2015. Available: <http://www.datasaveslives.eu/news-resources/good-news-for-health-research-from-trilogue-talks-on-the-data-protection-regulation/> [Accessed 12/01/2016].

European Parliament, 2012. *Draft Report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011-C7-0025/2012-2012/0011(COD))*. PR\922387EN.doc. Committee on Civil Liberties, Justice and Home Affairs.

Galson, S. & Simon, G., 2016. *Real-World Evidence to Guide the Approval and Use of New Treatments*. National Academy of Medicine, Available: <https://nam.edu/real-world-evidence-to-guide-the-approval-and-use-of-new-treatments/>.

Genomics England, 2017. Genomics. NHS England. Available: <https://www.england.nhs.uk/healthcare-science/personalisedmedicine/genomics/> [Accessed 13/09/2017].

HRA, 2018a. Data subject rights. NHS Health Research Authority. Available: <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/data-subject-rights/> [Accessed 01/08/2018].

HRA, 2018b. Legal basis for processing data - GDPR technical guidance. NHS Health Research Authority. Available: <https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-detailed-guidance/legal-basis-processing-data/> [Accessed 02/08/2018].

Ipsos MORI, 2016. *The One-Way Mirror: Public attitudes to commercial access to health data*. Trust, W. Ipsos MORI's Social Research Institute, Available: <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>.

Makady, A., van Veelen, A., Jonsson, P., Moseley, O., D'Andon, A., de Boer, A., Hillege, H., Klungel, O. & Goettsch, W., 2018. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. *PharmacoEconomics*, 36 (3), 359-368.

Marjanovic, M., Ghiga, I., Yang, M. Z. & Knack, A., 2017. *Understanding value in health data ecosystems: a review of current evidence and ways forward*. Cambridge. RAND Europe.

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., Smith, H., Aidinlis, S. & Kaye, J., 2018. Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. *Computer Law & Security Review*, 34 (2), 222-233.

MRC, 2018. GDPR: What researchers need to know. Medical Research Council. Available: <https://www.insight.mrc.ac.uk/2018/04/16/gdpr-research-changes/> [Accessed 02/08/2018].

Oderkirk, J. & Ronchi, E., 2017. Governing data for better health and healthcare. *OECD Observer*, 309.

Oderkirk, J., Ronchi, E. & Klazinga, N., 2013. International comparisons of health system performance among OECD countries: Opportunities and data privacy protection challenges. *Health Policy*, 112 (1), 9-18.

OECD, 2017. *Recommendation of the OECD Council on Health Data Governance*. THE NEXT GENERATION of HEALTH REFORMS: OECD Health Ministerial Meeting, Available: <http://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>.

Official Journal of the European Union, 2012. CHARTER OF FUNDAMENTAL RIGHTS OF THE EUROPEAN UNION. 2012/C 326/02. Available: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT> [Accessed 17/01/2018].