View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

The positive impacts of Real-World Data on the challenges facing the evolution of biopharma

John Wise[1], Angeli Möller[2], David Christie[3], Dipak Kalra[4], Elia Brodsky[5], Evelina Georgieva[6], Greg Jones[7], Ian Smith[8], Lars Greiffenberg[9], Marie McCarthy[10], Michael Arend[2], Olivier Luttringer[11], Sebastian Kloss[2], and Steve Arlington[1]

[1]Pistoia Alliance
[2]Bayer
[3]Amgen
[4]University of Gent, UCL & EuroRec Institute
[5]Pine Biotech
[6]Pryv
[7]Oracle
[8]MSD
[9]Abbvie
[10]ICON
[11]Novartis
Corresponding author: Wise J. (john.wise@pistoiaalliance.org)
Keywords: Real-World Data (RWD); Real-World Evidence (RWE); precompetitive collaboration.
Teaser: The biopharmaceutical industry must access, manage, and analyze Real-World Data (RWD) to generate Real-World Evidence (RWE), but does each company need to build their own in-house capability? Precompetitive collaboration should lead to better, more-cost-effective solutions.

Author biographies

John Wise
John Wise specializes in precompetitive collaboration in the life science R&D information ecosystem. He is a consultant to the Pistoia Alliance, a not-for-profit organization committed to lowering the barriers to innovation in life science R&D, and also serves as the programme coordinator for the PRISME Forum, a not-for-profit biopharma R&D IT/Informatics leadership group focused on the sharing of best practices. John has worked in life science R&D informatics in a variety of organizations, including academia, the pharmaceutical industry, a cancer research charity, as well as in the technology supply side of the industry. John graduated in physiology before obtaining a postgraduate certificate in education.

Dipak Kalra
Dipak Kalra is president of the European Institute for Innovation through Health Data and of the European Institute for Health Records. He has an international role in R&D of electronic health record systems and leads the development of ISO standards on EHR interoperability, personal health records, EHR requirements, and EHR security and confidentiality standards. Dipak was a co-lead on the IMI EHR4CR project and leads an EU Network of Excellence on semantic interoperability. Dipak is clinical professor of health informatics at University College London and a visiting professor at the University of Gent.

David Christie
David Christie is vice president of Research & Development Informatics at Amgen. David works closely with the head of R&D and the CIO to provide operational and strategic leadership in support of Amgen's worldwide initiatives in drug discovery and development. Previously, David was vice president of Global Commercial Operations and Corporate Functions IS. David has also led Amgen's International IS function based out of Zug, Switzerland, and before that, the IS group supporting Global Development functions. Before joining Amgen, David worked at Eli Lilly in various roles in Australia, New Zealand, and the USA. He holds a Bachelor of Business from the University of Technology, Sydney.

Demand for healthcare services is unprecedented. Society is struggling to afford the cost. Pricing of biopharmaceutical products is under scrutiny, especially by payers and Health Technology Assessment agencies. As we discuss here, rapidly advancing technologies, such Real-World Data (RWD), are being utilized to increase understanding of disease. RWD, when captured and analyzed, produces the Real-World Evidence (RWE) that underpins the economic case for innovative medicines. Furthermore, RWD can inform the understanding of disease, help identify new therapeutic

intervention points, and improve the efficiency of research and development (R&D), especially clinical trials. Pursuing precompetitive collaborations to define shared requirements for the use of RWD would equip service-providers with the specifications needed to implement cloud-based solutions for RWD acquisition, management and analysis. Only this approach would deliver cost-effective solutions to an industry-wide problem.

Introduction

According to The National Institute on Aging, 'The dramatic increase in average life expectancy during the 20th century ranks as one of society's greatest achievements' [1]. Vaccination is considered the most-successful and cost-effective medical intervention ever introduced [2]. Antibiotics (although emerging resistance trends are worrying) have made previously serious infections trivial and have made routine surgery safe. Biopharmaceutical industry products have been key to eliminating some diseases and are now treating many previously fatal diseases as chronic conditions. Yet, this success of the biopharmaceutical industry brings its own challenges. An increasing global population, an increasing life expectancy, and consequent percentage of older people, an increasing supply of advanced scientific and medical technologies, and an increased understanding of the causes and potential treatment of disease, are all creating an unpresented demand for treatments. In the 21st Century, society is struggling to afford these growing healthcare costs and the ever-expanding pharmacopeia on offer from the biopharmaceutical Industry.

All the key stakeholders in the healthcare system (i.e., patients, prescribers, providers, payers, developers, medicines regulators, health technology assessment agencies, governments, and the biopharmaceutical industry) need to recognize this. The biopharmaceutical industry must adapt rapidly to this existential truth. By 2030, the world population will be >8 billion. The population of India (~1.5 billion) will have overtaken that of China (~1.4 billion), while the population of Indonesia will have grown to ~300 million [3]. Urbanization will increase; approximately 80% of the population will live in cities with a concomitant increase in hypertensive, obese, noncompliant, diabetic asthmatics (HONDAs) as well as other diseases, such as schizophrenia. Furthermore, between 2015 and 2030, the number of people in the world aged 60 years or over is projected to grow by 56% to 1.4 billion [4], with a consequent increase in the incidence of the diseases of old age, such as the costly and devastating dementias.

In 2015, USA healthcare expenditure increased by 5.8% to reach US$3.2 trillion or 17.8% of its GDP [5] while total prescription drug spending increased by 9% to US$324 billion [6]. By 2020, USA healthcare expenditure is projected to reach 19.8% of GDP [7] and, according to a QuintilesIMS report, annual growth in the prescription drugs bill is expected to increase somewhere between 4% and 7% per annum up to 2021 [8]. Indeed, some projections have forecast healthcare consuming up to 25% of US GDP by 2030 [9]. By comparison, USA Government figures reveal that defence spending is anticipated to be around 3.8% of GDP by 2020 [10]. The projections for Europe are similar. For example, in 2014, according to World Bank figures, the EU average expenditure on healthcare was approximately 10% of GDP, with France and Germany each expending approximately 11% [11].

As a national example, in 2015, the National Health Service (NHS) in England cost approximately £116 billion [12], approximately 9% of the UK's GDP [13]. In 2010–2011, its drugs bill was £13 billion, and in 2014–2015 it was £15.5 billion [14], an increase of 19%. In 2011, the NHS drugs bill accounted for just 11% of the overall NHS budget; in 2016, that figure had risen to 15% and 'with a series of new and high-cost medicines due to become available in the next 12 months, senior figures have warned that urgent action is need to control the bill if other areas of care are not to suffer' [15].

It is self-evident that an inexorable increase in the costs of drugs would become at some point unsustainable. However, as a paper by researchers at Pfizer points out, 'excessive restrictions on the use of pharmaceuticals run counter to achieving a high performing health care system because they can result in the unintended consequence of increasing long-term costs and limiting investments in the research and development of future treatments and cures' [16]. New drugs might need to be accompanied by new diagnostics and be optimally delivered within redesigned care pathways to achieve their maximum benefit and to be demonstrably cost-effective. The increasing costs of the drugs bill can only be accommodated if the use of these drugs can be shown to reduce overall costs elsewhere (e.g., by decreasing the number of working-days lost both to patients and their carers, or decreasing patient-days spent in hospital) [17]. Society needs better tools to evaluate these downstream benefits: RWD and its derivative RWE provide the key to unlock this potential. The technical terms used in this review are defined in Table 1.

Key drug development concepts

The Duke-Margolis Center published a report in 2017 entitled *A Framework For Regulatory Use Of Real-World Evidence*'[18] that defines very carefully the terms RWD and RWE. RWD are 'data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources', including electronic Health Records (EHRs), administrative and claims data, captured directly in the course of an observational study, from sources of patient-generated information outside of clinical settings (e.g., in-home monitoring devices and wearable technologies), and in patient registries. It can also include contextual metrics not typically considered in routine clinical practice, such as environmental exposures and socioeconomic indicators. Importantly and interestingly, this definition does not preclude the incorporation of routinely collected data in traditional randomized clinical trials.

RWE is 'evidence derived from RWD through the application of research methods. For regulatory applications, RWE can further be defined as clinical evidence regarding the use and potential benefits or risks of a medical product derived from analysis of RWD. RWE is not simply 'anecdotes' based on RWD – it involves data curation, validation, and standardization to ensure that the data themselves are adequately fit-for-purpose. It requires thoughtful study

designs to assess the effects of the treatments on the outcomes of interest, and an understanding of the context in which the treatments are used.'

Key concepts relevant to the interpretation of clinical evidence are 'efficacy', 'effectiveness', and 'outcomes'. These terms have distinct and very different meanings in the contemporary biopharmaceutical industry. Efficacy refers to the extent to which a drug does more good than harm in randomized, controlled trials where the patients are carefully selected and monitored. Effectiveness refers to the extent to which a drug does more good than harm in real life where patients are not so narrowly selected, are treated in every-day practice, might have comorbidities, likely consume concomitant medication, and often are not closely monitored. Outcomes are 'measures of the end result of what happens to patients as a consequence of their encounter(s) with the healthcare system' [19] and provide an instrument for the analysis and publication of the results of delivered care, and provide the opportunity to compare and contrast the effectiveness of different therapeutic approaches (e.g., drugs of the same class but from competing pharmaceutical companies).

Understanding these differences, especially the importance of patient outcomes, is central to addressing the current challenges facing the industry, not least of which is improving the productivity of its R&D. 'Pharma is developing drugs that bring incremental benefits, but at a premium price. This has given rise to the debate between the providers and payers—what is the value of the extra benefit?' [20].

Key challenges

The challenge for society is twofold: to understand and acknowledge the contribution that the biopharmaceutical industry makes to population well-being and to encourage a business environment that supports the investment in, and the adoption of, innovative therapeutics. By contrast, the challenge for the biopharmaceutical industry is quickly and efficiently to discover and develop therapeutics that provide value; that is, the therapeutics must meet critical unmet patient needs at a cost that society can afford. Furthermore, the biopharmaceutical industry needs to communicate the unique, market-differentiating value of its therapeutics in a clear and compelling manner such that they will be used and reimbursed.

For biopharma to overcome this challenge, it needs to have the appropriate data, methodologies, and expertise to exploit that data, such that it can assess and communicate the value of its therapeutics. Biopharma must find ways to reduce the silos between its own internal disciplines that are involved in the generation of value evidence. Management must create an alignment of interests and structures to support collaboration between these disciplines such as drug discovery, clinical development, pharmacometrics, biostatistics, health economics and outcomes research, epidemiology, and IT.

Another challenge for biopharma is that the term 'value' might mean different things to different stakeholders. For example, 'value' perceived as important by the regulatory agency for a therapeutic for a disease in a child, might not be the value that is being sought by the patient's parent carer. Outcomes and endpoints might be defined differently by different stakeholders for different clinical scenarios. For example, the parents of a child with advanced cancer might assess drug safety risks differently from regulators; wearable sensors might identify novel measures of Parkinson's disease control that favor different drugs over those assessed using classical measures. A richer evidence base and more-open dialog are needed if society is to become more patient-centered in its authorization of innovative therapies. This phenomenon underlines the need for biopharma to grasp the challenge of the relevant framing of medical needs, and to get access to the relevant data, analytics, and the cross-functional skills to convert these data into knowledge for decision-making at the right point in time during development.

The challenge for the regulatory agencies is to enhance the current regulatory paradigm for demonstrating that a drug is safe and effective. This uses the well-established methodology [21] of the different phases of clinical trials carried out sequentially and taking many years (say ~10). However, this approach is too slow, too costly, and is no longer fit for purpose. As Peter Huber wrote in his seminal paper *The Digital Future of Molecular Medicine: Rethinking FDA Regulation* [22], the current regulatory agency approach 'reflects obsolete policies and rules put into place to regulate ignorance, not knowledge…'.

Regulatory agencies are charged with not only protecting public health, but also advancing public health [23]. They are by nature conservative, but if they are to advance public health, to allow patients timely, cost-effective access to promising therapies to address unmet medical needs, then they will need to adapt. The Medicines and Healthcare products Regulatory Agency (MHRA)'s early access to medicines scheme (EAMS) appears such a positive evolution [24] and the European Medicines Agency's 'adaptive pathways… to improve timely access for patients to new medicines' [25] could be the 'preferred approach in the future' according to Hans-Georg Eichler of the EMA [26]. In the USA, the 21st Century Cures Act (2016) established new expedited product development programs for Regenerative Medicine Advanced Therapy and Breakthrough Devices and, in October 2017, the US Food and Drug Administration (FDA) published its *Draft Guidance for Industry on the Breakthrough Devices Program* [27]. Furthermore, the Commissioner of the FDA, Scott Gottleib, commented on the importance of 'Model-Informed Drug Development' and how the FDA has been working collaboratively to build the 'expertise and infrastructure necessary to advance the development of these state-of-the-art modeling and simulation technologies' [28] for improved regulatory decision making.

The challenge for the healthcare system is choice, for healthcare systems will continue to be cost-constrained as advances in science, technology and medicine offer increasing opportunities for useful, but not necessarily affordable, therapeutic interventions in an increasingly aging and needy society. The healthcare system will have to ration its

resources, but it should do so in a transparent and demonstrably equitable way. As such, the demonstration of the value of a therapeutic in the real-world setting is important and the healthcare system needs to be able to collect and interpret the relevant RWD/RWE in an accurate and publicly accessible way.

A key objective for healthcare systems is to create a 'closed loop' system for the patients in their care. As written in the *21st Century Cures Discussion Document Summary* of January 27, 2015… 'The discovery, development, and delivery process is a cycle, meaning that data captured and analysed on the delivery side informs new discoveries and better, more targeted solutions for patients' [29]. As such, all healthcare interventions, such as doctor visits, prescriptions issued and dispensed, medical devices deployed (including unique device identifier data [30]), laboratory tests, hospital visits, and so on. should all be captured. Such information would enable health systems to create more-efficient services and to optimize health outcomes better. Furthermore, there is a rich clinical data set (e.g., 'patient reported outcomes, clinician reported outcomes, observer reported outcomes and performance outcomes' [31]) that can be used for assessing the clinical value of products.

Healthcare systems would then be able to connect and indeed enrich these data sources and be in a position to compare and contrast different therapies in specific populations, and to be able to calculate the downstream cost savings of early interventions with the concomitant contractual and pricing implications for the biopharmaceutical industry.

This task is at best challenging, and becomes even more challenging if drug-unrelated factors (e.g., patient location and consequent environmental factors) are contributing to the responses to the drug and the subsequent outcomes. As David Houle *et al.* comment [32] 'Phenotypic variation is produced through a complex web of interactions between genotype and environment… and our ability to characterize phenomes — the full set of phenotypes of an individual — lags behind our ability to characterize genomes. Phenomics should be recognized and pursued as an independent discipline to enable the development and adoption of high-throughput and high-dimensional phenotyping.

What is the impact of these challenges on biopharma?

To address these and other challenges, each biopharmaceutical company must invest, either internally or externally, individually or collaboratively, in the knowledge, tools, and expertise of capturing, accessing, managing, and analyzing RWD and also the essential disciplines of health economics and outcomes research, pharmacoeconomics, pharmacoepidemiology, pharmacometrics, and quantitative systems pharmacology, with its underlying systems of modeling and simulation to derive convincing, standards-based RWE.

Health technology assessment

To assist in the determination and selection of those therapies that have effectiveness, are safe, and provide value, Health Technology Assessment (HTA) agencies have been established in many countries and indeed regions. The WHO defines HTA as [33] '… the systematic evaluation of properties, effects and/or impacts of health technologies and interventions. It covers both the direct, intended consequences of technologies and interventions and their indirect, unintended consequences.'

The work of these HTAs can be illustrated by the National Institute for Health and Care Excellence (NICE; the HTA in England and Wales). NICE measures the merit of a medical intervention using a unit called the 'quality adjusted life year' (QALY) [34]. This metric is intended to provide a way of comparing the costs and benefits of sometimes very different technologies in different conditions. It does this by looking at the gains to the quality of life and life expectancy. This way of comparing different interventions provides insight into the impact of introducing a new treatment on the ability of the NHS to maintain the services it already provides. However, use of this tool is not without criticism [35]. One such criticism is that the healthy individuals who ascribe value to states of illness might ascribe a value that is very different from the patients' valuations of treatments (including health gains, adverse effects, or avoided health losses), which are the economic evaluations with which society should be concerned.

Currently, NICE advises to support the use of new therapies within a worth of up to £20 000–30 000 per QALY [36]. However, in April 2017, this was supplemented by the 'Budget Impact Test' [37], which can significantly accelerate approvals for therapies with a cost per QALY of less than £10 000. 'The aim is to provide quicker access for patients to the most cost-effective new treatments'. For biopharma successfully to embrace this opportunity, it is clear that it must have a good understanding of, and ability to manage, the relevant RWD and its associated RWE.

The Fourth Hurdle

The biopharmaceutical industry needs not only to obtain regulatory approval for a new therapy, by demonstrating that it is safe, effective, and can be manufactured with appropriate quality (three well-recognized hurdles),but also to show that the potential new therapy provides more value for the same money or the same value for less money. Improvement is needed in the processes of this regulatory–value paradigm. According to an article published in 2015 in *Scientific American* [38], it already costs >US$2 billion to bring a drug to market (including the costs of those drugs that fail). The additional effort of overcoming the fourth hurdle [39] (of demonstrating the value of a therapeutic to the many different HTAs, with their many different assessment methodologies) will only add to this cost. Clearly, any additional costs and time delays will have a negative impact on investment in innovation and, as such, constrain the availability of new therapies for unmet patient needs. However, should be noted that, just because one HTA forms its opinion, it does not mean that another HTA will form the same opinion.

An interesting review was undertaken by the Heron Group [40], which compared the reasons for recommendation and rejection of drugs made by four different HTA systems [NICE, Scottish Medicines Consortium (SMC), Canadian

Agency for Drugs and Technologies in Health (CADTH) and Australian Pharmaceutical Benefits Advisory Committee (PBAC)]. Of 73 HTA submissions analyzed, 44 incidents of conflicting guidance between the four appraising bodies were identified. It was noted that the main area of conflict between the decisions of the different HTA agencies related to the economic analysis.

These many and different HTAs enforce an exacting exercise on the biopharmaceutical industry. The establishment of EUnetHTA [41] 'to create an effective and sustainable network for HTA across Europe… to help developing reliable, timely, transparent and transferable information to contribute to HTAs in European countries' is a voluntary attempt to address some of these issues. It is as yet unclear how countries will balance a European-level HTA decision-making process with national assessments of comparative value within their own health systems.

The new regulatory approval–reimbursement paradigm

The key to this new paradigm is RWD and its derivative, RWE. The data are typically derived from electronic systems used in healthcare delivery, patient registry information, genomics data, data contained within medical devices, and/or in tracking the patient experience during care, including in their own homes. Some key points about RWD include: (i) RWD are being generated with rapidly increasing velocity, volume, and variety (but are often noisy, unstructured, and of poor quality; RWD are needed by all healthcare stakeholders to identify gaps in healthcare systems and to understand drug value in target populations; (iii) the right RWD need to be identified, generated, and captured to provide relevant information about therapeutic performance; (iv) RWD accessibility, quality, linkages, and analytical methods need to be considered carefully to ensure study validity and to define pathways for data usage in product labeling and reimbursement decisions; (v) RWD need sophisticated infrastructure to support analysis to ensure the validity of the resultant RWE and to answer the most critical questions being asked by all stakeholders; and (vi) RWE evidentiary standards can vary between stakeholders.

All the stakeholders in the biopharmaceutical and healthcare industries are eager to mine RWD to create RWE that might provide insights that can enhance the scientific understanding of disease, drive down the cost of therapeutics, and enhance patient safety. Biopharmaceutical companies can exploit RWE to generate advantage across the value chain, because RWE informs the design of therapies (e.g., drugs with companion diagnostics) and the associated clinical trials to address those opportunities. RWE will be the cornerstone of value-based pricing that is redefining the basis of competition and market access and the defence of market position.

RWD and clinical trials

Randomized clinical trials

Randomized clinical trials (RCTs), or explanatory trials, generally measure efficacy; that is, the benefit a treatment produces under ideal conditions, often using carefully defined subjects in a research clinic and aims primarily to further scientific knowledge [42].

RWD can inform the design of RCTs both from a scientific perspective (e.g., what might be measured?) and a logistical perspective (where can we find the patients to participate in the trial?). For example, if one wanted to test a candidate therapy to address a specific medical, need one would need to identify potential subjects with that need, perhaps in a specific age group, with inclusion and/or exclusion criteria based on medical history and current medication, and with specific biomarkers within target ranges.

Pragmatic clinical trials

Pragmatic trials [43] measure effectiveness (i.e., the benefit the treatment produces in routine clinical practice). The design of a pragmatic trial reflects the variations between patients that occur in real clinical practice and aims to inform choices between treatments. To ensure generalizability, pragmatic trials should represent the patients to whom the treatment will be applied.

Many healthcare decision-makers are developing policies that allow the integration of evidence from pragmatic trials and RWD [44]. For example, many study arms would be required if using RCTs to assess the effectiveness of combinations of treatment. It would take 32 different study arms to examine all the possible combinations of just five treatments. Such an approach would make the costs prohibitive. 'Platform trials' allow multiple treatments to be evaluated simultaneously and 'offer flexible features such as dropping treatments for futility, declaring one or more treatments superior, or adding new treatments to be tested during the course of a trial' [45]. This is just one area where RWD can contribute [46].

The IMI project EH4CR [47] has been instrumental in demonstrating the value of the reuse of health data for research. Its sustainability model is in the European Institute for Innovation through Health Data (i~HD) [48–51].

RWD and pharmacoepidemiology

Pharmacoepidemiology is a well-established discipline and can be described as a bridge science spanning both pharmacology and epidemiology and can be defined as the application of epidemiological methods to pharmacological issues [52]. The discipline studies the utilization and effects of drugs in large numbers of patients. The increasing availability of RWD increases the utility of, and the focus on, this discipline. It can provide an estimate of the probability of beneficial effects, or indeed adverse effects, in populations. Other research questions relating to drug use could benefit from using epidemiological methodology.

Considerable research is being invested in pharmacoepidemiology. One such example is the Innovative Medicines Initiative (IMI)'s project on Pharmacoepidemiological Research on Outcomes of Therapeutics (IMI PROTECT) [53].

Pharmacovigilance can be considered as a subdiscipline of pharmacoepidemiology; the European Union-Adverse Drug Reaction project (EU-ADR) [54] explored the understanding of adverse drug reactions by integrative mining of clinical records and biomedical knowledge. This work led to the establishment of the EU-ADR Alliance, which is a 'collaboration Framework that has the goal of running studies and answering drug safety questions with the use of extracted data from multiple European privately and publicly owned Electronic Healthcare Records (eHR) databases [55].

RWD and disease taxonomy

A new and more-precise understanding of diseases, through molecular medicine, has a profound impact on how we relate therapeutic interventions to clinical endpoints. An important determinant of how RWE is produced and interpreted is the way in which we classify diseases.

Taxonomy is the practice and science of classification, typically considered in the context of biology (e.g., the Linnaean system for classifying living organisms). In medical practice, taxonomy often refers to the International Classification of Diseases (ICD), a system established more than 100 years ago that the WHO uses to track disease incidence, and physicians use as a basis for standardized diagnoses. Better classification of disease will lead to better, quicker, and more-precise diagnoses and treatments.

Improved understanding of the molecular make-up of diseases, associated with clinical data on individual patients, could drive the development of a 'new taxonomy' that would define diseases by their underlying molecular causes and other factors in addition to their traditional physical signs and symptoms [56].

A report commissioned by the US National Research Council [57] recommends the construction of a new data network that would improve biomedical research by enabling scientists to access patients' information during treatment while still protecting patients' rights. This would allow the marriage of molecular research and clinical data at the point of care, as opposed to research information continuing to reside primarily in academia. RWD, including medical histories, environmental exposures, and treatment outcomes of individual patients, enriched with metabolomics, genomics, and proteomics provide a rich source of information to inform disease taxonomies.

RWD and quantitative systems pharmacology

One of the important emerging methodologies is quantitative systems pharmacology (QSP). QSP is an umbrella term for modeling approaches that integrate a mathematical representation of the biological system with pharmacological information about a drug of interest to facilitate improved understanding of human drug responses. QSP is intended to help identify and validate targets, reveal possible biomarkers, support drug design, inform dose and regimen selection, and help proactively identify responders and nonresponders. QSP is a powerful translational science tool; although it is very promising, it needs quantitative patient data.

However, RWE, derived from RWD, is currently mainly providing contextual input, with qualitative insights into disease, epidemiological patterns, and effects of treatment. In the future, one can anticipate more-quantitative data sets coming directly from patients through electronic devices, mainly from wearables (e.g., the IMI RADAR-CNS project). Such data will increase significantly what can be done with RWD, because this type of data is quantitative and can be used for modeling and simulation in complex disease models. Both preclinical and clinical pharmacology require this kind of data for QSP modeling.

RWD and precision medicine

The practice of medicine is moving from classifying (and then treating) patients according to coarse-grained and traditionally defined diseases to a stratified medicine, or cohort-based medicine (or more commonly but somewhat misleadingly called 'personalized medicine') approach, which is based on identifying subgroups of patients with distinct mechanisms of disease, or particular responses to treatments [58]. Precision medicine, 'an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person' [59], is increasingly being viewed as the strategic direction for healthcare delivery. Different patients respond to the same drugs differently, in large part because of underlying differences in their genomes. RWD, along with its associated methodologies of QSP and modeling and simulation, can be exploited better to understand these important molecular biology differences between patients. If the biopharmaceutical industry can deploy stratified medicine and precision medicine approaches that demonstrate to the regulator and the payer a therapeutic value proposition that identifies those patients who should respond to a therapy (and, conversely, those patients who will not respond) then the effectiveness and the safety (i.e., the benefit:risk of the therapy) will be enhanced, its value improved, and the possibility of a listing in the pharmacopoeia enhanced.

The Royal Marsden Hospital in London is assessing the feasibility of using genetic sequencing to identify patients suitable for targeted drugs. Its FOrMAT [60] study is recruiting up to 200 patients with advanced gastrointestinal cancer, including stomach, pancreatic, bowel and esophageal cancers, to take part. In addition, this trial aims to establish the logistical and technical infrastructure that will routinely and rapidly enable the identification of patients in this way.

The 100,000 Genomes Project, being carried out by Genomics England [61], is an example of RWD/RWE being used to inform science and clinical practice. Genomes are analyzed from patients with cancers or with rare diseases. Correlation of phenotypic and genotypic information can lead to deeper scientific insights into the etiology of the disease and to more-precise diagnoses and, as such, more-precise therapeutic interventions.

Precision medicine accelerated by new technologies

There are some significant technology developments that are accelerating this drive towards precision medicine. Two strategies are currently being promoted to ensure that data stored in electronic health records (which by their nature can be very unstructured) can be effectively mined: (i) increasing the structure in such documents along with the use of controlled vocabularies; and (ii) applying Natural Language Processing (NLP) to unstructured or semistructured health data. However, advances in NLP in medicine and biology are dependent upon the size and quality of biomedical data sets used to train such algorithms.

Wearable smart devices (e.g., smart phones, smart watches, or Fitbits) can capture a wealth of patient-generated RWD without the patient needing to visit a clinical care environment. The capability to aggregate data from different wearables and track trends has already been demonstrated by the documented change in sleep patterns of Americans during the 2016 elections [62]. However, well-defined standards to allow the seamless interoperability and data exchange for such 'wearables' are still not within reach. The use of artificial intelligence (AI) to generate predicted outcomes based on diverse, heterogenous data sets is viewed as a necessary step to deal with the explosion of data generated by wearables and smartphone 'apps' transmitting patient-reported outcomes. Whether such algorithms will be validated and, as such, usable in a GxP environment for submission-relevant analyses, remains to be seen. Research and investment in this area are intense, with IDC reporting an expected increase in global AI-generated revenues from nearly US$8.0 billion in 2016 to more than US$47 billion in 2020 [63]. AI (including machine learning, deep learning, and cognitive computing) is now being exploited by many technology-focused companies, from small start-ups to large multinationals.

RWD and social listening

Social listening, the mining of social media, has the potential to provide valuable insights into a patient's real-life use of therapeutics. Routinely used by commercial organizations to gain brand insight, biopharmaceutical companies, regulators, and others are now actively investigating social listening as an innovative and additional method of pharmacovigilance. With the number of global social media users expected to reach approximately 2.95 billion by 2020 [64], social media sites, such as Facebook, Twitter, Redditt, and LinkedIn, have significant potential to harness the patient's opinion and, as such, become a valuable source of adverse event and postmarket, patient-surveillance data. Indeed, specialized health social networks. such as HealthUnlocked and PatientsLikeMe. could provide even greater value. The immediacy of these data, along with their widespread availability, could provide insights into adverse effects and noncompliance. Indeed, social listening can be used to track the most-common reasons for patients switching between therapies [65]. It has the potential to provide a large amount of information about 'real-world benefit' as discussed from the patient's perspective [66]. Both the FDA (MedWatcher Social) [67] and IMI (Web-RADR) [68] have initiated digital-listening programs that automatically collect and collate social media information about drug safety events and experiences with medical products.

Social listening is a powerful, low-cost, real-time data source; however, the use of the data is not without limitations. Many issues exist, including data validation, source verification, the complexity of managing large unstructured data sets, identifying signals from the noise, interpreting free text interspersed with 'Text Speak', misspellings and slang, and multiple languages. These issues are not insurmountable, but require powerful natural language processing, machine learning, and AI platforms.

Patient centricity and patient engagement are becoming increasingly important in healthcare delivery. Social listening provides unprecedented access to the issues that are important to patients and offers significant opportunities to supplement traditional postmarket data.

Nonetheless, for these real benefits of social media to be maximized, standards, methodologies, and agreed practices regarding privacy and regulatory guidance are needed.

RWD and regulatory guidance

There is considerable opportunity for RWD to enhance all areas of the drug lifecycle. For example, combining clinical data and RWD to enhance clinical trial design could reduce the number of patients required, thereby simultaneously delivering ethical, cost, and time benefits. RWD are nonstandardized and are often noisy, unstructured, and federated, and currently have no formal regulatory requirement. Substantial effort is required to find formalized ways of standardizing RWD. There is also the need to look at processes to enable end-to-end traceability and quality of the data to ensure trust for regulatory scrutiny. Such issues can be addressed in isolation, through clearly documented processes. However, regulatory guidance would be welcomed to promote industry-wide standardization, which would bring benefit to all, including, and perhaps especially, the regulators.

RWD and governance

If we are to scale up access and use of RWD, much effort will be required to overcome some of the challenging legal and ethical issues of data ownership and data privacy as well as the technical issues of performant infrastructure and security. Good practices in information governance, especially for research uses of health data, are emerging and have been documented elsewhere [69].

Given the profound opportunities of RWD, an industry-wide governance body, with membership from the biopharmaceutical industry, regulatory bodies, and subject-matter experts, could help drive some of the required change. This governance would help build the guardrails and lay the foundation to exploit true 'value' for both patient outcomes and truly transformational industry change to the benefit of all parties.

Cyber security [70] will also need to become more robust and resilient if the opportunities of RWD are fully to be realized. RWD security breaches, including attacks on stakeholders in healthcare systems, such as a ransomware attack on the NHS in the UK [71], are a threat to data availability, integrity, confidentiality, and interoperability and, as such, are also a threat to the development and advancement of precision medicine. Blockchain [72], an open-source technology that came into prominence as the technology underpinning the virtual currency Bitcoin [73], could enable patients to share personal data securely with healthcare stakeholders and could prove to be a useful tool to enhance RWD cyber security. For example, in Estonia, a country that is very dependent on its digital infrastructure, scalable blockchain technology has been developed and deployed to ensure the integrity of data stored in government repositories [74].

RWD and the need for a catalog and a marketplace

RWD sources are many and large, and data can be captured from any sources that formalize information related to medicine adherence and experience. These can range from, for example, hospital-level diagnostic notes, data from Internet-of-things (IoT)-enabled medical equipment, smartphones, and wearables collecting vital and environmental information. Unless effectively managed, data are therefore federated and difficult to discover and use. RWD vendors have addressed this by compiling data sets that can be licensed at a cost, but there is no standardization to this approach and no easy way to compare content. Some data sets can only be accessed from the provider's own infrastructure (e.g., Genomics England or Farr Institutes). Some full data sets must be purchased even if only a small cohort is required and, given the current, unmodernized delivery methods (e.g., via CD-ROMs), data are often considered out-of-date when they arrives and, as such, of limited value. Therefore, more work is required to ensure RWD follows the FAIR data principles (findable, accessible, interoperable, and reusable) [75] and opportunity exists for a global catalog or marketplace for data to be implemented. An idealized, long-term vision would be for this service to drive standardization via a common data model, enable FAIR principles [as identified as important in, e.g., the IMI European Medical Information Framework (EMIF) project], and provide real-time access and interoperability.

RWD and the need for a common data model

RWD are often noisy and unstructured. Effort has been made to address these constraints via the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) [76]. However, it has not been as widely adopted as it might and is limited by data loss that often occurs doing conversion, because of limitations with data dictionaries. More work is required to build confidence in its use and further support is required by industry to encourage broader adoption if it is to be successful. With a standard data format, data interoperability and analytics across multiple data sets become more powerful, allowing new questions to be addressed. There is already significant interest in OMOP and CDM from the FDA and the EMA and expectations of enforcement by regulators might be one opportunity to develop its use.

RWD and technology

The biopharmaceutical industry must find ways to reduce the silos between its different disciplines and functions that are involved in the generation of RWE. Management must create synergies between them to enable better collaboration. There is an urgent need for a change in mindset. As PwC noted in their report *Revitalizing Pharmaceutical R&D* [77] '…The real constraints are governance considerations and the lack of determination by senior leadership to drive the required level of change'. There needs to be more exchange of RWD/RWE and their associated ideas. One approach is to embrace technology, specifically a technology platform, to serve as an enabler for the business process transformation that needs to occur. Such an approach would appear to be encouraged by the National Institutes of Health, because, according to GenomeWeb [78] 'it has awarded $9 million in fiscal 2017 to fund 12 projects kicking off the initial phase of an effort to build a cloud-based platform for storing, sharing, accessing, and computing biomedical data, and other digital objects such as analytical tools'.

What if all the different disciplines across the organization, but now coordinated, could access a single, trusted, secure enterprise technology platform? What if all these different disciplines stopped purchasing independently the same data sources from the same external vendors? This could be achieved if they had access to an enterprise technology platform and were supported by clearly defined, internal business processes that enabled them to buy the data sources once and reuse them across all the disciplines. What if this enterprise technology platform could provide the scalability to deliver searchable, real-time data to the end-users with high-performance access to hundreds of millions of patient claims, EHRs, 'omics', mobile device, sensors, and other related data sources?

To analyze these data sources, many organizations spend substantial time, effort, and money on 'data wrangling' (i.e., the process of cleaning, transforming, and aggregating the large-scale data sources with reference to a consistent enterprise taxonomy). Indeed, some organizations spend more time, effort, and money on 'data wrangling' than they do on the actual data analysis. What if this technology platform enabled low-cost, rapid 'data wrangling', such that dynamically formed, multidiscipline work groups from across the organization could spend most of their time doing analysis?

What if the technology platform leveraged industry and technology standards, such as: (i) Health Level Seven (HL7), involving Fast Healthcare Interoperability Resources (FHIR), Continuity of Care Document (CCD), and Clinical Document Architecture (CDA); (ii) Clinical Data Interchange Standards Consortium (CDISC), involving Operational Data Model (ODM), Study Data Tabulation Model (SDTM), and Shared Health And Research Electronic

library (SHARE); and (iii) Observational Health Data Sciences and Informatics (OHDSI), involving OMOP CDM, as low-cost enablers and key elements to drive the efficiency of data wrangling?

What if the team members in these workgroups could define secure work areas so that they themselves could perform individual analysis and then bring that analysis into the broader team's shared and secure work area so that analysis can be shared and compared within the team?

What if the best of those analysis results were combined for a successful, consistent, cross-discipline perspective for the RWE business question(s) that need to be answered? For example, what outcomes are achieved, or what is the value of the extra benefit (e.g., addressing the fourth hurdle), or how do I satisfy all the different HTAs with their differing assessment methodologies?

What if every result of the analysis, including reports, line listings, programs, additional derived data sets, and so on, were all fully traceable to satisfy regulatory requirements and engender both internal and external trust especially with organizations including the HTAs, regulators, payers, and so on?

If such a technology platform with all of these capabilities were available, it could be the catalyst required to initiate the transformative task of business process alignment across the disciplines and start to achieve many of the business objectives identified as desirable. Clearly, such a business transformation process is substantially more complex than implementing the underlying technology, but the deployment of such a technology platform would provide a highly efficient enabler of such a business process transformation.

Skills for RWD/RWE

However, research skills specific to the exploitation of RWD/RWE, such as domain knowledge as well as methodological and technical expertise, health informatics, and epidemiological capabilities, are not in such abundant supply within the pharmaceutical industry as those much longer-established skills, such as biostatistics required for the design, implementation, and analysis of the more-traditional RCTs.

Thus, the biopharmaceutical industry needs a multipronged approach to optimize the power of RWD collection and RWE generation including: (i) training in the strengths, weaknesses, and opportunities of RWD and RWE, with special emphasis on Mendelian randomization [79]; (ii) cross-organization cultural change to break down the silos, both technical and intellectual, between all the disciplines involved in RWE generation; (iii) access to the enabling data and technology infrastructure while cognizant of the philosophy 'nothing about me, without me' [80]; (iv) deploying the right talent in the right roles (whether it be health economics, epidemiology, QSP, data scientists, IT, etc.); and (v) empowering challenges to the *status quo* – 'speaking truth to power'.

Clearly, on-going investment in staff training, development, and education is required for the maintenance and continued development of these RWD/RWE skills for the future.

Concluding remarks

Personal communication with a senior leader of observational research in a major biopharmaceutical company revealed that its pharmacoeconomic function was originally established with a complement of six. In 2017, the complement was over 100. Clearly, this is a not unimportant discipline.

RWD and its derivative, RWE, should be used to inform and guide the biopharmaceutical industry from drug discovery onwards [81]. RWE can provide insights into the pathophysiology of disease, enable the construction of more-precise disease taxonomies, be used better to design and execute RCTs, and be used to stratify patient populations, to identify more clearly those patients who might benefit from a clinical trial, to provide earlier signals of safety issues, and, indeed, be used to avoid those safety issues. RWE needs to be more extensively used by the biopharmaceutical industry, regulators, and HTAs to establish endpoints and clinical outcomes that are more patient centered and establish the value of innovative therapies with greater sophistication than at present.

While adhering to good practices, not least in the protection of privacy, all stakeholders need greater access to fine-grained, high-quality, and interoperable health data, and need to develop an integrated, evidence-based development model that brings together RCTs and RWE to enable an adaptive licensing [82] approach throughout the life cycle of each drug.

RWE is nondiscretionary when demonstrating the value of a therapy to HTAs. Pharma needs to work constructively with the HTAs to develop a common language, supported by commonly agreed methods, that unequivocally exposes the value of a therapy, not only in the near term, but also in the longer term, by identifying the cost savings of avoided downstream morbidities (e.g., patient hospital days, lost working time for patients and indeed carers, and added family costs).

The burden is now on biopharma R&D organizations, and their corporate leadership, to work constructively with all the healthcare system stakeholders to overhaul the current drug discovery, development, marketing authorization, and market value models. There is also a burden on the regulatory agencies and the health technology agencies and the payers to see how they might work better together to facilitate for patients, timely and affordable access to new medicines.

However, some points emerge. For example, is it cost-effective for each biopharmaceutical company individually and separately to: (i) invest in the significant IT infrastructures required to capture, manage, and analyze RWD?; (ii) build up and retain a staff of engineers and data architects to manage the servers and the Hadoop data lakes?; (iii) retain a cadre of cyber-security specialists to protect its RWD infrastructure?; (iv) get expert legal counsel to advise on the legal and regulatory aspects of the use of RWD?; and (v) consider the ethical implications of big data linkage?

Alternatively, can the biopharmaceutical industry come together with other key healthcare stakeholders and review the current state-of-the-art, identify best practices, and chart a standards-based way forward for the optimal exploitation or RWD/RWE? Can the biopharmaceutical industry also come together with some of its major technology suppliers and collectively define the requirements for commonly accessed, regulatory-compliant, trusted third-party RWD platforms that can provide the RWD services for individual biopharmaceutical companies and indeed other stakeholders in the healthcare system?

Such a precompetitive, collaborative, shared approach would create economies of scale. The cost savings enjoyed by each participating company would free up resources to invest in the strategic specialties of RWD/RWE, such as pharmacoepidemiology and quantified systems pharmacology, along with the data scientists and their deep domain knowledge and statistics and 'R' expertise.

We strongly advocate a precompetitive, collaborative approach to these challenges. Perhaps a not-for-profit, precompetitive collaboration organization, such as the Pistoia Alliance [83], might provide the infrastructure to create a Community of Interest (CoI) to undertake this work to create the specification for a shared, third-party RWD/RWE capability. By joining such a cross-industry precompetitive collaboration, individual companies (drawn from the healthcare system but with a strong presence from the biopharmaceutical industry) could share understanding, elucidate best practices, specify required functionality, identify appropriate standards, describe the skills required, and review the most-appropriate technological approaches.

Declaration of Interest

J.W. is a contractor to the Pistoia Alliance. As such the Pistoia Alliance funded the coordination of this multi-author paper. S.A. is the President of the Pistoia Alliance. Amgen, Bayer, Pine Biotech, Pryv, MSD, Abbive and Novartis are members of the Pistoia Alliance.

References

1 National Institute on Aging (2011) *Global Health and Aging*, National Institute of Health

2 Berkley, S. (2014) Improving access to vaccines through tiered pricing. *Lancet* 383, 2265–2267

3 Population Pyramids of the World from 1950 to 2100 - http://www.populationpyramid.net/china/2020/

4 UN (2015) *World Population Ageing*, United Nations

5 Centers for Medicare & Medicaid Services, (2016) - https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html

6 Centers for Medicare & Medicaid Services, (2017) - https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html

7 Health Affairs Blog, (2011) - http://healthaffairs.org/blog/2011/07/28/u-s-health-spending-projected-to-grow-5-8-percent-annually/

8 Medicines Use and Spending in the U.S. A Review of 2016 and Outlook to 2021 - http://www.imshealth.com/en/thought-leadership/quintilesims-institute/reports/medicines-use-and-spending-in-the-us-review-of-2016-outlook-to-2021

9 KHN Morning Briefing (2009) - http://khn.org/morning-breakout/dr00039364/

10 http://www.usgovernmentspending.com/defense_spending

11 https://data.worldbank.org/indicator/SH.XPD.TOTL.ZS

12 The NHS in England (2016) - http://www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx

13 The World Bank (2017) - http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS

14 Health and Social Care Information Centre (2015) - http://content.digital.nhs.uk/catalogue/PUB18973/hosp-pres-eng-201415-report.pdf

15 "The Times", Monday, 15th December 2016

16 Role of Pharmaceuticals in U.S. Health Care Spending (2015) - http://www.pfizer.com/files/about/Position-Role-of-Pharmaceuticals-in-US-Healthcare-Spending.pdf

17 Congressional Budget Office (2013) Offsetting Effects of Prescription Drug Use on Medicare's Spending for Medical Services - https://www.cbo.gov/sites/default/files/presentation/44364offsetsacademyhealth-buntin6-20130.pdf

18 https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf

19 Krousel-Wood, M.A. (1999) Practical considerations in the measurement of outcomes in healthcare. *Ochsner J.* 1, 187–194

20 Eichler, H-G. (2010) The Tale of Health Care Reform - DIA Global Forum December 2010 p20

21 Bhat, A. (2010) Article title. *Perspect. Clin. Res.* 1, 6–10

22 Huber, P. (2013) *The Digital Future of Molecular Medicine: Rethinking FDA Regulation*, FDA

23 Statement of FDA Mission - https://www.fda.gov/downloads/aboutfda/reportsmanualsforms/reports/budgetreports/ucm298331.pdf

24 https://www.gov.uk/guidance/apply-for-the-early-access-to-medicines-scheme-eams

25 European Medicines Agency, Adaptive Pathways - http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp

26 Eichler, H-G. *et al.* (2015) From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin. Pharmacol. Ther.* 97, 234–246

27 https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm581664.pdf

28 Remarks by Scott Gottlieb, M.D. Commissioner of Food and Drugs (2017) - https://www.fda.gov/NewsEvents/Speeches/ucm578381.htm

29 21ST CENTURY CURES DISCUSSION DOCUMENT SUMMARY – JANUARY 27, 2015 - https://energycommerce.house.gov/sites/republicans.energycommerce.house.gov/files/114/Analysis/Cures/20150127-Cures-Discussion-Document-Section-by-Section.pdf

30 McClellan, M. et al. (2015) Faster, More Efficient Innovation through Better Evidence on Real-World Safety and Effectiveness, The Brookings Institution

31 ISPOR TASK FORCE REPORT - https://www.ispor.org/clinical-outcomes-assessment-guidelines.pdf

32 Houle, D. *et al.* (2010) Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–886

33 WHO (2014) Health Technology Assessment, WHO

34 NICE (National Institute for Health and Care Excellence), Glossary - https://www.nice.org.uk/glossary?letter=q

35 Nord, E. *et al.* (2009) QALYs: some challenges. *Value Health* 12 (Suppl. 1), S10–S15

36 NICE, (2013) Judging whether public health interventions offer value for money - https://www.nice.orgRCTs.uk/advice/lgb10/chapter/judging-the-cost-effectiveness-of-public-health-activities

37 https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/budget-impact-test

38 Mullin, R. *et al.* (2014) Cost to Develop New Pharmaceutical Drug Now Exceeds $2.5B, Scientific American - https://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/

39 Paul, J.E. *et al.* (2001) 'Fourth hurdle reviews', NICE, and database applications. *Pharmacoepidemiol. Drug Saf.* 10, 429–438

40 Cann, K. *et al.* (2008) A comparison of reasons for recommendation and rejection in four HTA systems: NICE, SMC, CADTH and PBAC. Heron Evidence Development Ltd. ISPOR 11th Annual European Congress Scientific Awards Winners. - https://www.ispor.org/awards/11euro/HT3-Plested%20Podium.pdf

41 EUnetHTA - http://www.eunethta.eu/about-us

42 Kabisch, M. *et al.* (2011) Randomized controlled trials. *Dtsch. Arztebl. Int.* 108, 663–668

43 Roland, M. *et al.* (1998) What are pragmatic trials? *BMJ* 316, 285

44 FDA (2016) Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, FDA

45 Saville, B. *et al.* (2016) Efficiencies of platform clinical trials: a vision of the future. *J. Soc. Clin. Trials* 13, 358–366

46 Mahajan, R. (2015) Real world data: additional source for making clinical decisions. *Int. J. Appl. Basic Med. Res.* 5, 82

47 EH4CR (Electronic Records for Clinical Research) - http://www.ehr4cr.eu/

48 The European Institute for Innovation through Health Data - http://www.i-hd.eu

49 Dupont, D. *et al.* (2017) Business analysis for a sustainable, multi-stakeholder ecosystem for leveraging the Electronic Health Records for Clinical Research (EHR4CR) platform in Europe. *Int. J. Med. Inform.* 97, 341–352

50 Dupont, D. (2016) Assessing the financial impact of reusing electronic health records data for clinical research: results from the EHR4CR European Project. *J. Health. Med. Informat.* 7, 235

51 Beresniak, A. (2016) Cost-benefit assessment of using electronic health records data for clinical research versus current practices: contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project. *Contemp. Clin. Trials* 46, 85–91

52 International Society for Pharmacoepidemiology - https://www.pharmacoepi.org/about/about.cfm

53 http://www.imi-protect.eu/about.shtml

54 Coloma, P.M. (2011) Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol. Drug Saf.* 20. 1–11

55 http://eu-adr-alliance.net/#aboutus

56 Kola, I. *et al.* (2011) A call to reform the taxonomy of human disease. *Nat. Rev. Drug Discov.* 10, 641–642

57 National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease (2011) *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, National Research Council

58 MRC - Stratified Medicine - https://www.mrc.ac.uk/research/initiatives/stratified-medicine/

59 https://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm455305.htm

60 The Royal Marsden - FOrMAT genetic sequencing trial - https://www.royalmarsden.org/format-genetic-sequencing-trial

61 Genomics England - https://www.genomicsengland.co.uk/#

62 Did Americans Lose Sleep Over the Election? (2016) - https://health.nokia.com/blog/2016/11/21/did-americans-lose-sleep-over-the-election/

63 Worldwide Cognitive Systems and Artificial Intelligence Revenues Forecast to Surge Past $47 Billion in 2020 - https://www.idc.com/getdoc.jsp?containerId=prUS41878616

64 https://www.statista.com/topics/1164/social-networks/

65 Risson, V. *et al.* (2016) Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. *J. Med. Internet Res.* 18, e62

66 Powell, G.E. *et al.* (2015) In their own words: social listening for "real-world benefits" from prescription and OTC products. *Value Health* 18, A86

67 FDA - Mining Social Media for Adverse Event Surveillance (2016) - https://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm455305.htm

68 https://www.imi.europa.eu/content/web-radr

69 Kalra D. *et al.* (2016) The European Institute for Innovation through Health Data. *Learn. Health Syst.* 1, 1–8

70 https://www.cisco.com/c/dam/global/th_th/assets/docs/seminar/cio_Cisco_Security_Overview.pdf

71 http://www.wired.co.uk/article/nhs-cyberattack-ransomware-security

72 The Telegraph (2014) - http://www.telegraph.co.uk/technology/news/10881213/The-coming-digital-anarchy.html

73 Bitcoin - https://bitcoin.org/en/

74 https://e-estonia.com/solutions/security-and-safety/ksi-blockchain/

75 Wilkinson, M. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018

76 Observational Medical Outcomes Partnership – Common Data Model - http://omop.org/CDM

77 PwC Revitalizing pharmaceutical R&D - The value of real world evidence https://www.strategyand.pwc.com/media/file/Revitalizing-pharmaceutical-RD.pdf

78 NIH Awards $9M in Funding to Support Data Commons Pilot Phase - https://www.genomeweb.com/informatics/nih-awards-9m-funding-support-data-commons-pilot-phase

79 Davey Smith, G. *et al.* (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98

80 Coulter, A. et al. (2011) *Making Shared Decision-Making A Reality - No Decision about Me, without Me*. The Kings Fund

81 Galson, S. et al. (2016) *Real-World Evidence to Guide the Approval and Use of New Treatments*, National Academy of Medicine

82 Eichler, H-G. *et al.* (2015) From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin. Pharmacol. Ther.* 97, 234–246

83 The Pistoia Alliance http://www.pistoiaalliance.org/

84 IMI - The Innovative Medicines Initiative - https://www.imi.europa.eu/

85 IMI - All Projects - https://www.imi.europa.eu/content/ongoing-projects

86 IMI – EMIF - http://www.emif.eu/

87 IMI – GetReaL - http://www.imi-getreal.eu/

88 IMI 2 - https://www.imi.europa.eu/content/imi-2

89 IMI 2 - Big Data for Better Outcomes Programme (BD4BO) - https://intranet.imim.es/files/news/Topic_Feb2016_BD4BO_DataDN.pdf

90 IMI 2 - European Health Data Network (EHDN) - https://www.imi.europa.eu/sites/default/files/uploads/documents/Future_Topics/IndicativeTopic_EHDN.pdf

Box 1. XXXX

RWD and IMI projects

The IMI is Europe's largest public–private initiative aiming to speed up the development of better and safer medicines for patients [84]. Launched in 2008 with a €2 billion budget, IMI had the goal of 'significantly improving the efficiency and effectiveness of the drug development process with the long-term aim that the pharmaceutical sector produce more effective and safer innovative medicines'. There are many (~85) IMI projects [85], some of which are of particular interest to the RWD community.

IMI EHR4CR

The EHR4CR project has developed a robust and scalable platform that can utilize deidentified data from hospital EHR systems, in full compliance with the ethical, regulatory, and data protection policies and requirements of each participating country. The EHR4CR created the 'The Champion Programme', which sets out to prove the value of RWD for clinical research and has also established i~HD.

IMI ADAPT-SMART

Inter alia, the IMI ADAPT-SMART project introduces the concept of 'Medicines adaptive pathways to patients' (MAPPs). This is a concept that seeks to provide patients with timely access to beneficial medicines, starting from a small group of well-identified patients (often those with little or no alternative treatments). As RWD/RWE accumulates on the benefits and risks of a medicine, access can be extended to other groups of patients. MAPPs relate to a flexible pathway covering the entire life cycle of a medicine, from development, through licensing, to patient access (pricing/reimbursement and healthcare delivery). The project sets out to embrace important healthcare stakeholders, such as regulatory agencies, HTA bodies, companies, payers, patients, and healthcare professionals.

IMI EMIF

The EMIF [86] was launched in January 2013 as a 5-year project to help improve access to patient-level data. A common Information Framework (EMIF-Platform) links up and facilitates access to diverse medical and research data sources. Two primary research areas are being addressed: the onset of Alzheimer's disease (EMIF-AD) and metabolic complications of obesity (EMIF-Metabolic).

IMI GetReal

The IMI GetReal [87] project was launched in 2013 as a 3-year activity that aimed to show how robust new methods of RWE collection and synthesis could be adopted earlier in the biopharmaceutical R&D and healthcare decision-making process. The project brought together key stakeholder groups (industry, academia, regulatory agencies, reimbursement agencies, healthcare budget holders, and patient groups) to share their insights and know-how. GetReal was challenged to develop new approaches for incorporating real-life data into drug development, and pave the way for greater consensus on this issue.

IMI 2

IMI 2 [88] was launched in 2010 and will run for 10 years with a budget of €3.276 billion. The goal of the IMI 2 program is to develop next-generation vaccines, medicines, and treatments, such as new antibiotics. One of the umbrella projects of IMI 2 is called 'Big Data for Better Outcomes' (BD4BO) [89].

Big Data for Better Outcomes

BD4BO 'aims to catalyse and support the evolution towards value-based and more outcomes-focused sustainable and therefore better-quality healthcare systems in Europe'. It aims to do this by maximizing the potential of large amounts of data from variable, quickly developing digital and nondigital sources, which will be referred to as 'big data' in the context of this initiative.

European Health Data Network

The European Health Data Network (EHDN) [90] is a BD4BO-enabling project with the aim of delivering the vision for large-scale medical outcomes research. It is anticipated that EHDN should take learnings from several initiatives, such as : (i) European Union FP7 project EU-ADR, TRANSFoRm; (ii) IMI projects EH4CR and EMIF; and (iii) the US-based OHDSI project.

The EHDN project has three goals: (i) to use the projects listed above to develop a standard methodology; (ii) to mature both the supply side and the demand side of the 'health data eco-system' in compliance with robust privacy and ethics governance; and (iii) to stimulate development of new and augmented health services through available and expanded technologies, in the interest of health outcomes.

In addition, the EHDN project has three primary domains of focus: (i) research; (ii) health services efficiency; and (iii) individual patient care. The project will apply two important standards: (i) tOMOP CDM; and (ii) International Consortium for Health Outcomes Measurement (ICHOM) standards.

IMI RADAR-CNS

The RADAR-CNS project aims to develop new ways of monitoring major depressive disorder, epilepsy, and multiple sclerosis using wearable devices and smartphone technology.

Table 1. Definitions and relevant websites of terms used in this review

Term	Definition	Website
Adaptive pathways	A scientific concept for medicine development and data generation that allows for early and progressive patient access to a medicine	www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp
App	An application, especially as downloaded by a user to a mobile device	https://en.oxforddictionaries.com/definition/app
Artificial Intelligence (AI)	A branch of computer science dealing with the simulation of intelligent behavior in computers.	www.merriam-webster.com/dictionary/artificial%20intelligence
Biostatistics	Statistical processes and methods applied to the collection, analysis, and interpretation of biological data and especially data relating to human biology, health, and medicine	www.merriam-webster.com/dictionary/biostatistics
CDISC	A global, nonprofit organization that develops data standards to streamline clinical research and enable connections to healthcare, empowering the valuable information offered by patients participating in research studies around the world	www.cdisc.org/system/files/all/CDISC_Brochure.pdf

Clinical development	A collective term encapsulating the myriad details and extensive time required to test a potential therapy in humans. Typically considered as three sequential phases: Phase 1 to test safety and dosage and requiring several months and 20–100 volunteers; Phase 2 to test efficacy and adverse effects: requiring several months to 2 years and up to up to several hundred patients with the disease; and Phase 3 to test efficacy and monitor adverse reactions and requiring 1–4 years and up to several thousand patients with the disease	www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm
Companion diagnostic	A medical device, often an *in vitro* device, which provides information that is essential for the safe and effective use of a corresponding drug or biological product	www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm407297.htm
Cyber security	Measures taken to protect a computer or computer system (as on the Internet) against unauthorized access or attack	www.merriam-webster.com/dictionary/cybersecurity
Disease model	An animal or cells displaying all or some of the pathological processes that are observed in the actual human or animal disease. Studying disease models aids understanding of how the disease develops and testing potential treatment approaches	www.nature.com/subjects/disease-model
Drug discovery	The process through which potential new medicines are identified. It involves a range of scientific disciplines, including biology, chemistry, and pharmacology	www.nature.com/subjects/drug-discovery
Epidemiology	The study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems	www.who.int/topics/epidemiology/en/
European Medicines Agency	A decentralized agency of the European Union (EU) responsible for the scientific evaluation, supervision, and safety monitoring of medicines in the EU	www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000112.jsp&mid=WC0b01ac0580028a43
Farr Institute	A UK-wide research collaboration involving 21 academic institutions and health partners in England, Scotland and Wales. Publicly funded by a consortium of ten organizations led by the Medical Research Council, the Institute is committed to delivering high-quality, cutting-edge research using 'big data' to advance the health and care of patients and the public	www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000112.jsp&mid=WC0b01ac0580028a43
Fitbit	A device that contains a 3D motion sensor that accurately tracks your calories burned. steps taken, distance traveled, and sleep quality	www.collinsdictionary.com/submission/10039/Fitbit
Food and Drug Administration (FDA)	A federal agency in the US Department of Health and Human Services established to regulate the release of new foods and health-related products	www.thefreedictionary.com/FDA
Genomic data	Data derived from 'a branch of biotechnology concerned with applying the techniques of genetics and molecular biology to the genetic mapping and DNA sequencing of sets of genes or the complete genomes of selected organisms, with organizing the results in databases, and with applications of the data (as in medicine or biology)	www.merriam-webster.com/dictionary/genomics
Genomics	The large-scale study of groups of genes	www.nature.com/scitable/definition/genomics-126
Genomics England	A company set up and owned by the UK Department of Health to run the 100,000 Genomes Project, which aims to sequence 100,000 genomes from NHS patients with a rare disease and their families, and patients with cancer	https://en.wikipedia.org/wiki/Genomics_England
GxP	A general term for Good Practice quality guidelines and regulations. The titles of these good practice guidelines usually begin with 'Good' and end in 'Practice', with the specific practice descriptor in between. Three most commonly-used GxPs in the pharmaceutical industry are: Good Laboratory Practices (GLP), Good Manufacturing Practices (GMP), and Good Clinical Practices (GCP)	https://regulatory.usc.edu/consulting/resources/drugs/gxps/
Health economics	An applied field of study that allows for the systematic and rigorous examination of problems faced in promoting health	www.jhsph.edu/departments/international-health/global-health-masters-degrees/master-of-health-science-in-health-economics/what-is-health-economics.html
Health informatics	The interdisciplinary study of the design, development, adoption, and application of IT-based innovations in healthcare services delivery, management, and planning	www.himss.org/health-informatics-defined
Innovative Medicines Initiative (IMI)	Europe's largest public-private initiative aiming to speed up the development of better and safer medicines for patients	www.imi.europa.eu/content/adapt-smart
Mendelian randomization	A statistical technique used by genetic epidemiologists to determine causal effects, for instance whether a biomarker actually influences disease risk, or whether it is just a statistical association	https://biology.stackexchange.com/questions/3171/what-is-mendelian-randomization-and-how-is-it-used-to-infer-causality-in-epidem/3184
Metabolomics	Refers to the systematic identification and quantification of the small-molecule metabolic products (the metabolome) of a	www.nature.com/subjects/metabolomics

	biological system (cell, tissue, organ, biological fluid, or organism) at a specific point in time	
National Health Service (NHS)	The UK's healthcare service available to all, regardless of wealth. With the exception of some charges, such as prescriptions, optical services, and dental services, the NHS in England remains free at the point of use for all UK residents	www.nhs.uk/NHSEngland/thenhs/about/Pages/overview.aspx
National Institute for Health and Care Excellence (NICE)	The UK's Health Technology Agency, whose is to improve outcomes for people using the NHS and other public health and social care services	www.nice.org.uk/about/what-we-do
Natural language processing	The automatic (or semiautomatic) processing of human language	www.cl.cam.ac.uk/teaching/2002/NatLangProc/revised.pdf
Outcomes research	Seeks to understand the end results of particular healthcare practices and interventions	www.hopkinsmedicine.org/gim/research/method/outcomes.html
Pathophysiology	The physiology of abnormal states; specifically: the functional changes that accompany a particular syndrome or disease	www.merriam-webster.com/dictionary/pathophysiology
Patient Registry	An organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves a predetermined scientific, clinical, or policy purpose(s)	www.ncbi.nlm.nih.gov/books/NBK164514/
Personalized medicine	An ambiguous term that could be misinterpreted to imply that treatments and preventions are being developed uniquely for each individual	https://ghr.nlm.nih.gov/primer/precisionmedicine/precisionvspersonalized
Pharmacoeconomics	The branch of economics that uses cost-benefit, cost-effectiveness, cost-minimization, cost-of-illness, and cost-utility analyses to compare pharmaceutical products and treatment strategies	www.ncbi.nlm.nih.gov/pubmed/16120204
Pharmacoepidemiology	The study in real conditions and on large populations, of the use, effectiveness and risk of drugs	www.ncbi.nlm.nih.gov/pubmed/27476308
Pharmacometrics	The science that quantifies drug, disease, and trial information to aid efficient drug development and/or regulatory decisions	www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm167032.htm
Pharmacovigilance	The science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem	www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/
Phenomics	The use of large-scale approaches to study how genetic instructions from a single gene or the whole genome translate into the full set of phenotypic traits of an organism	www.nsf.gov/bio/pubs/reports/phenomics_workshop_report.pdf
Precision medicine	Identifying which approaches will be effective for which patients based on genetic, environmental, and lifestyle factors	https://ghr.nlm.nih.gov/primer/precisionmedicine/definition
Proteomics	The study of the structure and function of proteins, including the way they work and interact with each other inside cells	www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=306524
Quantitative systems pharmacology	An approach to translational medicine that combines computational and experimental methods to elucidate, validate, and apply new pharmacological concepts to the development and use of small-molecule and biologic drugs	www.ema.europa.eu/docs/en_GB/document_library/Presentation/2012/04/WC500126724.pdf
R	A language and environment for statistical computing and graphics	www.r-project.org/about.html
RCTs	A study in which several similar people are randomly assigned to two (or more) groups to test a specific drug, treatment, or other intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, a dummy intervention (placebo) or no intervention at all	www.nice.org.uk/glossary?letter=r
RWD	Data relating to patient health status and/or the delivery of healthcare routinely collected from a variety of sources	https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf
RWE	Evidence derived from RWD through the application of research methods. For regulatory applications, RWE can further be defined as clinical evidence regarding the use and potential benefits or risks of a medical product derived from analysis of RWD	https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf
Social listening	The analysis of comments and images that people put online and into the public domain (e.g., on Twitter or Facebook or on specialist blogs or forums)	www.aqr.org.uk/glossary/social-listening
Statistics	The practice or science of collecting and analyzing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample	https://en.oxforddictionaries.com/definition/statistics
Stratified medicine	Identifying subgroups of patients with distinct mechanisms of disease or particular responses to treatments	www.mrc.ac.uk/research/initiatives/stratified-medicine/