Commentary: Jennifer Cubino

Meeting the challenges of real-world data access

The advent of precision medicine has unlocked significant new potential for industry to provide more targeted therapies for patients. Here, the use of real-world data (RWD), or different types of data relating to patient health status and/or the delivery of healthcare, has become especially useful, expanding beyond its main original application in assessing product safety to providing population-based evidence about disease epidemiology and information about outcomes following medical interventions.

This can also involve real-world evidence (RWE), or the analysis of this data. Access to real-world data from diverse populations, for example, can enable the development of new therapies, arising from the discovery of druggable targets from population-based research with real-world data outcomes. Real-world data is also being used in drug development and increasingly, in the evaluation of new medicines at the European Medicines Agency and the US Food and Drug Administration.

Although real-world data plays an important role in developing effective new medicines and evaluating patient health outcomes per treatment pathway, this resource also faces challenges.

The challenges are both technical and ethical. Technical issues that also have ethical implications include ensuring data provenance, or being able to trace data back to a specific moment in time, and navigating the different legal systems across the world governing data privacy. There is also a need to understand disease-specific treatment outcomes for patients, as well as different standards of care across countries. Fortunately, there are a number of approaches for solving these problems that are underway at our company BC Platforms and elsewhere, and which are reviewed in this article.

The use of real-world data has already had a positive impact on how decisions are made during drug development. This applies to drug discovery, preclinical development and clinical development. What should not be overlooked however is that data provided to drug developers has provenance which means that it can be attributed to a particular individual, whilst ensuring that person's privacy. Effectively we are looking at that person's health journey. Data privacy is an important consideration.

The rules governing data privacy differ depending on the geographic region. In the US, the use of healthcare data is governed by the Health Insurance Portability and Accountability Act (HIPAA) of 1996 which protects sensitive patient information from being disclosed without a patient's knowledge or consent. However once de-identified, this data can still be sold to third parties. It can be linked with other information, enabling companies to use data aggregators to perform predictive analytics – a source of revenue for these companies. The information in question can include prescription, insurance and laboratory test data. The HIPAA does not require companies to ask patients to opt in or opt out of these sharing arrangements. Unfortunately, de-identified HIPAA-compliant data, especially where data linking is allowed, has been traced back to specific individuals[1-3].

By comparison, the European Union's General Data Protection Regulation (GDPR) requires healthcare organisations to conduct real-world data analyses using anonymised patient data, a higher standard than de-identification. Also, the legislation does not allow patient data to be aggregated with other data. EU patients who participate in real-world data research can see how their data is being used and can withdraw their consent at any time. Because patient consent is required, researchers can get a better understanding of the individual's medical condition and the outcomes of treatment. We believe that patient consent should be a starting point for real-world data research, and have built this into our business model.

How is this policy implemented at the local level? To start with, local regulatory compliance managers need to make sure that data is consented and/or anonymised and/or de-identified for use. Individual data, as required under GPDR, must follow specificity of use. For example, the project should operate within a GDPR-compliant framework and be based on ISO certified software. This should include standardisation and harmonisation to real-world data ready for analysis and adhering to institutional, local and national approval processes, alongside highly controlled data processing.

Where global or institutional rules require, additional data security and privacy can be achieved through federated analytics. Federated analytics is the use of information from different datasets without actually transferring the data to a central location.

The patient experience

Drug development is a complex process involving many actors. In the era of personalised medicine, the amount of data required to make sound decisions for individual patients is considerable. Companies like BC Platforms play a role in this process by collecting and analysing data and making it available to users such as researchers and drug developers. The process is not linear; it is a loop. This means that developers need data not just on a patient's disease but also on the outcomes of specific treatments.

For conditions of high unmet medical need, patient outcomes can be very disease-specific. For example if a patient has multiple sclerosis, is that person capable of ambulating, or providing for his or her own care? If the disease is asthma, how often does a patient need to use a rescue inhaler? How often does the patient miss work or school because of the condition? How has the condition changed over time?

Real-world data is not always readily available to drug developers, including hospital-based patient level treatment

data. Furthermore, the administration needed to maintain the diversity in data is overwhelming for drug developers to resource at scale. In cases where follow-up data is required, our company expects that patients be asked to give their consent to be re-contacted for more information. In these instances, the patient is asked to fill out a questionnaire providing supplemental information on healthcare status, treatments and nuanced patient outcome measures. The time advantage of this approach – where real-world data in a medical record is updated directly rather than through an observational study – can be considerable. One need only compare this approach with the cost and time of a formal study to appreciate its impact on drug development.

For several years, the regulatory authorities have been adjusting their procedures to take account of real-world data. In 2016, the US Food and Drug Administration launched Sentinel, a system for discovering and analysing data on drug safety. These analyses are conducted on products that are already on the market and complements the FDA's existing tools for monitoring drug safety. Two years later, in 2018, the agency introduced a policy for accepting real-world data in addition to clinical trial data in new drug applications – as long as the real-world data had provenance. This means that it has been possible for drug developers to significantly reduce the number of patients needed to conduct interventional clinical research.

The challenges

As discussed earlier, gaining patient consent and achieving data provenance are essential to the use of real-world data in drug development, especially for personalised medicines. There is a third challenge as well. This is knowing what the standard for care is for specific therapies. A standard of care is a treatment accepted by medical professionals as appropriate for a specific disease. These standards can differ across geographical regions, healthcare systems and even across individual hospitals. A diagnostic or treatment for a patient may vary depending on the standard in that patient's locality and this can result in different outcomes. Scores used for clinical assessment, for example joint mobility assessment scores in rheumatoid arthritis, can be different resulting in a different granularity of data. Such differences not only influence patient care, but affect the resulting real-world data. Comparing, contrasting and aggregating real-world data between patients, especially outcomes data, is often less than straightforward. Providing robust outcomes data, bearing in mind that standards of care differ, is fundamental to success. It is more complicated for complex diseases.

The outlook

Access to real-world data on the part of drug developers is accelerating the development of new medicines and shifting the industry towards a more agile and data-centric approach. Several steps are being taken to overcome current challenges, especially those which prioritise patient transparency and consent when collecting and using data.

Longstanding issues remain, in terms of differences in standards of care, as well as healthcare justice and global access to medicines. Access to adequate healthcare and treatments should be a fundamental focus for all healthcare systems.

The more real-world data that drug developers have at their disposal, the faster they can discover new medications to treat unmet medical needs. When combining omics, outcome data and longitudinal clinical data, drug developers can learn what makes some patients recover from an illness while others do not. The insights gained from real-world data could take years off the R&D process, generating cost savings that could improve global patient access to new drugs. The key to achieving this goal is to strive for, and to maintain, high data provenance with real-world data. This is why healthcare systems and global legislative bodies need to think about patient engagement and consent differently. It has great potential to open the door to important research and findings that could benefit everyone.

References:

1. Yoo J, Thaler A, Sweeney L, Zang J. Risks to Patient Privacy: A Re-identification of Patients in Maine and Vermont Statewide Hospital Data. *Technology Science*. 2018100901. October 08, 2018.

2. Na L, Yang C, Lo C, Zhao F, Fukuoka Y, Aswani A. Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. *JAMA Netw Open*. 2018;1(8):e186040.

3. Slabodkin G. Data Re-Identification Remains Risk Despite HIPAA Safeguards. Health Data Management.

Terms used in this article

Real-world data: Data about a patient's health and/or the delivery of healthcare.

Real-world evidence: The analysis of real-world data.

Data provenance: Data attributable to an individual.

Standard of care: A treatment accepted by medical professionals as appropriate for a specific disease.

De-identified data: A term used in the HIPAA legislation related to removing HIPAA defined identifiers from protected health data. De-identified data has removed: patient name, address, medical record number, among other personal identifiers. A de-identified data set is expected to not allow re-identification of an individual.

Anonymised data: A term used in the GDPR legislation related to the permanent removal of any identifiers that may identify an individual. The term is much broader than the de-identified definition under the HIPAA legislation. Anonymised data is prohibited from being used in such a way that may cause re-identification of an individual.

The author, Jennifer Cubino, is the Chief Operating Officer of BC Platforms AG. She holds a BSc from Cornell University's College of Agriculture and Life Sciences and a MA in Clinical Research from Boston University's College of Medicine, US.