Researchers rely on data. For the life sciences industry, “real world” data (which are data collected outside of the highly controlled clinical trial setting) present an opportunity: to increase the efficiency of the process of developing new medicines; to improve the way healthcare is delivered; and to monitor the safety of existing treatments on a broader scale, as well as better track and monitor the spread of disease. Opening access to real world data could foster better health and a more effective and efficient healthcare system.

It isn’t easy (or obvious) how to best use real world data to achieve their promise to improve healthcare. Many datasets are incomplete, not connected to important information, or are otherwise not accessible for analysis. There is no generally accepted gold standard for using real world data the way there is with a clinical trial. Privacy concerns are real and valid, particularly when datasets include highly detailed information. Furthermore, the same analysis conducted on two different real world datasets can yield different answers. In a world where we are all living longer, developing new treatments is getting harder and more expensive, but the promise is too great to not rise to the challenge. There is a new future ahead, and we are just starting to see what is possible with this information.

For example, as of the early 2000s, there were no treatments for Hypereosinophilic Syndrome (HES), a rare white blood cell disease that can lead to fatal organ damage. Physicians attempted to treat HES with Gleevec® (imatinib mesylate), a medicine...
approved for chronic myeloid leukemia and recognized for its ability to prevent abnormal cell growth. By collecting and sharing healthcare data that demonstrated a masked improvement in patient outcomes, physicians and the innovator company were able to build the evidence base that ultimately led to Food and Drug Administration (FDA) approval of Gleevec® for HES. This treatment was thus approved and available to reduce patients’ suffering much faster than if it had gone through the standard clinical trial process for this indication. This is but one example of how the dissemination of healthcare data permits the discovery and approval of new therapies that may address an unmet patient need.

Federal and state entities produce some of the richest and most robust real world data sets. Valuable sources of publicly developed data for research include:

- Centers for Medicare & Medicaid Services (CMS) Medicare databases;
- Federally, regionally, or state-funded health information exchanges (HIEs), or all-payer claims databases (APCD);
- Veterans Administration datasets; and
- In the future, data captured from the plans run through the public health insurance exchanges.

Research based on these real world patient-level datasets can produce evidence that is critical to optimizing healthcare delivery and health outcomes. In turn, this evidence can ensure the adoption of best practices in care delivery, improve patient outcomes, spur the innovation of new technologies and treatments, facilitate medication safety monitoring, and promote medication adherence.

Real world data could help in the design of healthcare benefits so they are more value based. We have increasingly seen health plans manage costs by limiting access to specialty medicines. For patients with some of the highest burden of disease, this poses a serious problem because there are often no alternative treatments for their conditions. By using real world healthcare data, health benefits could be tailored to provide access to the patients mostly likely to benefit from the treatment. This could result in better allocation of healthcare resources through the development of sophisticated benefit designs that would allow for more targeted access to treatments for patients who are most likely to benefit. This ultimately leads to more cost-effective and efficient use of our country’s significant investment in healthcare.

For patient-level healthcare-related data to reach their full potential, the following issues all need to be considered and appropriately addressed.

**Provide timely access to personally funded data for all stakeholders with a valid research question.**

To accelerate research and care improvement in the United States, all stakeholders with a valid research question should be able to access and use publicly funded data. To ensure that the data are used appropriately, investigators interested in accessing the data should be required to provide a well-defined research proposal and explicitly identify planned uses of the data through an open, transparent, and well-defined process. Given the value that all researchers can provide in advancing the quality of care through greater access to healthcare data, access should not be limited based on the “commercial interests” of an entity. Access could be rendered meaningless, however, if the process for gaining access is not expeditious and efficient.

**Advance privacy standards that balance risk and benefit of real world data research.**

When providing access to healthcare data to researchers and other healthcare professionals, clear standards around appropriate mechanisms for the de-identification, encryption, and transmission of patient-level data should be in place. In addition, all research proposals to use healthcare data should clearly identify how patient identities will be protected in compliance with established regulations.

**Enhance data connection across data sources.**

Most patients receive care from a variety of healthcare providers and in a variety of care settings. Moreover, as patients undergo employment changes or move residences, they may change health plans or providers. To better understand how patients experience and respond to healthcare, it is important to be able to track patients across time and across these different locations or settings of care. As such, improved connection, or “interoperability,” would allow researchers to better assess the patient care journey, which will promote a better understanding of how to intervene to improve patient outcomes.

By using real world healthcare data, health benefits could be tailored to provide access to the patients mostly likely to benefit from the treatment.
Promote transparency around the sources and methods used to collect data.

For data to be useful to researchers, healthcare professionals, and other stakeholders, they will need to be able to ensure that the data are of a high quality, complete, reliable, and were collected using standardized procedures. Establishing clear and uniform standards to report what has and has not been collected will enable stakeholders to perform analyses and make recommendations confidently based on their research.

Develop clear methodological standards for analysis of the data.

All stakeholders performing assessments using the data should do so in a manner that is robust, valid, and reliable. Given that evidence produced using healthcare datasets will inform delivery and payment decisions, it is important to prevent the development and use of poor, low-quality evidence, which would ultimately put patients’ health at risk. This means ensuring that appropriate and clear methodological standards for performing research using healthcare data are established, communicated, and broadly used. The research standards established by the Patient-Centered Outcomes Research Institute (PCORI) could be a good foundation. Moreover, existing research organizations (e.g., the Agency for Healthcare Research and Quality, the National Institutes of Health, PCORI) could assist with identifying and disseminating these best practices for research. In addition, pre-registration of studies that are intended as key evidence to inform health policy decisions in programs that receive significant federal or state funding should be required. The study design and analytic plan should be posted on a publicly available website prior to beginning the research. If a study is not pre-registered, it should be considered hypothesis generating, and findings should be corroborated in a study that is pre-registered. This standard should apply to everyone.

Establish appropriate communication and dissemination practices.

Analyses based on healthcare data must be broadly and appropriately communicated. Wide distribution of research can lead to greater awareness of novel healthcare research ideas and promote further research and data sharing. To ensure optimal use of research, it will be important for communication of evidence to be unbiased and clearly evidence-based. Government funded agencies, such as PCORI, may be good organizations to help outline best practices for clear, consistent, and reliable dissemination of healthcare data.

Expand access to government sponsored health claims data for research.

The elderly are the costliest and highest users of healthcare services; however, large Medicare datasets are out of reach to many qualified researchers. Entities should not be arbitrarily excluded from requesting access to these data because they are commercial in nature, provided such entities can prove that they are qualified to perform the research, have a clearly defined research proposal, and have agreed to abide by requirements for using the data. This would provide commercial entities (including pharmaceutical companies, healthcare data companies, device manufacturers and other healthcare providers) better tools for developing treatments.

Similarly, as the Affordable Care Act’s health insurance exchange plans become more established, it will be important to collect their data and provide access to qualified researchers. By expanding permitted uses of health exchange data, researchers would be able to track and measure care delivered through these new plans and ensure that patients in diverse locations have access to high-quality care in a cost-effective manner.

By expanding permitted uses of health exchange data, researchers would be able to track and measure care delivered through these new plans and ensure that patients in diverse locations have access to high-quality care in a cost-effective manner.

Kirsten Axelsen is Vice President of Worldwide Healthcare Research and Quality, the National Institutes of Health, PCORI could assist with identifying and disseminating these best practices for research. In addition, pre-registration of studies that are intended as key evidence to inform health policy decisions in programs that receive significant federal or state funding should be required. The study design and analytic plan should be posted on a publicly available website prior to beginning the research. If a study is not pre-registered, it should be considered hypothesis generating, and findings should be corroborated in a study that is pre-registered. This standard should apply to everyone.

Establish appropriate communication and dissemination practices.

Analyses based on healthcare data must be broadly and appropriately communicated. Wide distribution of research can lead to greater awareness of novel healthcare research ideas and promote further research and data sharing. To ensure optimal use of research, it will be important for communication of evidence to be unbiased and clearly evidence-based. Government funded agencies, such as PCORI, may be good organizations to help outline best practices for clear, consistent, and reliable dissemination of healthcare data.

Expand access to government sponsored health claims data for research.

The elderly are the costliest and highest users of healthcare services; however, large Medicare datasets are out of reach to many qualified researchers. Entities should not be arbitrarily excluded from requesting access to these data because they are commercial in nature, provided such entities can prove that they are qualified to perform the research, have a clearly defined research proposal, and have agreed to abide by requirements for using the data. This would provide commercial entities (including pharmaceutical companies, healthcare data companies, device manufacturers and other healthcare providers) better tools for developing treatments.

Similarly, as the Affordable Care Act’s health insurance exchange plans become more established, it will be important to collect their data and provide access to qualified researchers. By expanding permitted uses of health exchange data, researchers would be able to track and measure care delivered through these new plans and ensure that patients in diverse locations have access to high-quality care in a cost-effective manner.

Given that health insurance exchanges are still under development, the ability for stakeholders to be engaged around their formation may be particularly timely.

The challenge is there, but given what is at stake, isn’t the only choice to rise up to meet it?