ABSTRACT:
Stakeholders and decision makers are increasingly using real-world evidence (RWE) and technology to solve the problems of human health. Real World Data (RWD) and RWE will play a big role in health care decision making. Data sources can include claims data, electronic medical record data, genomics data, imaging data, sensors, wearables and many others. As big data gathered in real-world healthcare settings becomes more prevalent and robust, it is increasingly being used across the entire healthcare system for evidentiary purposes. This data has the potential to guide us create better study design and answer unanswered queries in trial set-up. The data from analytics can further form inputs to medical product development.

The utility of AI comes from its application to huge data arising out of RWE information bank. Natural Language Processing (NLP), an AI tool can help in charting unstructured data and provide a contextual meaning. Machine Learning (ML) is being utilized to search through volumes of data, looking for complex relationships using library of algorithms. ML will strengthen the way predictive analytics and prescriptive analytics are being transformed with data. Deep learning (DL) concepts will automate generation of predictive features and have its impact on analyzing data related to image processing, speech recognition and language translation.

Innovation in the form AI coupled with big data, real-world evidence, which is more dynamic, appropriate, illustrative, complete and cost-effective can be generated. This paper focus of areas of application of AI in ensuring fruitful RWE outcome.

INTRODUCTION
Real world data (RWD) and real world evidence (RWE) are playing an increasing role in health care decisions. FDA uses RWD and RWE to monitor post-market safety and adverse events and to make regulatory decisions. The health care community is using these data to support coverage decisions and to develop guidelines and decision support tools for use in clinical practice.

Medical product developers are using RWD and RWE to support clinical trial designs (e.g., large simple trials, pragmatic clinical trials) and observational studies to generate innovative, new treatment approaches.

While the definition of Real World Evidence is still evolving, most proponents associate RWE with data that is derived from medical practice among heterogeneous sets of patients in real life practice settings, such as insurance claims data and clinical data from electronic health records. The term is also stretched to encompass data that might not qualify strictly as outcomes data, such as genomic data, patient socioeconomic data and environmental data.

In clinical context, Big data includes data from EMR/EHR, Clinical, Claims, Labs, Drug data, Imaging, PROs, Consumer data and study data – Real world data too.

Beehive is one such complex structure which can give delicious honey if extracted in the right way using proper techniques.

Advent of AI has been a boon to transform the data available into a meaningful outcome like extracting honey from the complex structure of bee-hive.

FDA has been supporting initiatives leading to better outcomes and RWE, AI in clinical trials have been moving of sandbox environment.

FDA TALKS:
FDA uses RWD and RWE to monitor postmarket safety and adverse events and to make regulatory decisions.

The health care community is using these data to support coverage decisions and to develop guidelines and decision support tools for use in clinical practice.

On December 7, FDA published the much-anticipated “Framework for FDA’s Real-World Evidence Program” for drugs and biological products (the “Framework”). In a statement announcing the Framework, Commissioner Gottlieb recognized the opportunities and challenges of using real-world data (“RWD”) and real-world evidence (“RWE”) to enhance regulatory decision-making and noted that leveraging this information is “a top strategic priority for the
FDA.” FDA opened a docket for public comments on the Framework through February 5, 2019.

The Framework focuses on the use of RWE to support regulatory decisions about effectiveness. The agency outlines three considerations that will guide its overall RWE Program and inform the agency’s assessment of individual drug applications. The Framework also offers background on the agency’s previous use and current initiatives with respect to RWE and related topics, such as innovative clinical trial designs. This blog post provides an overview of FDA’s proposal and highlights a few initial takeaways noted by Covington’s Digital Health team.

The 21st Century Cures Act (“Cures Act”) required the agency to create a program for evaluating the use of real-world evidence (“RWE”) for two purposes: (1) to help support the approval of a new indication for an already-approved drug, and (2) to help support or satisfy post approval study requirements.

The Cures Act also mandated that the agency publish a framework for implementing the RWE Program that describes the sources of RWE; the gaps in data collection activities; the standards and methodologies for collecting and analyzing RWE; and the priority areas, remaining challenges, and potential pilot opportunities that the RWE Program will address. The new Framework covers drugs and biological products; FDA addressed the use of RWE in the context of medical devices separately in 2017.

FDA underscores the distinction between real-world data and real-world evidence. The Cures Act defined “real world evidence” as “data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than randomized clinical trials.” In contrast, but similar to FDA’s approach for medical devices, the Framework differentiates between RWE and RWD as follows: RWD is defined as “data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources”; RWE, on the other hand, is “clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD” (emphasis in original). FDA explains that evaluating the use of RWE for regulatory decision-making involves assessing both the reliability and relevance of the underlying RWD and the methodologies used to generate RWE from the RWD.

Building on the definitions of RWD and RWE, FDA’s framework establishes a “three-part approach” for incorporating this information into regulatory decision-making for drugs and biological products:

- Whether the RWD are “fit for use” in regulatory decision-making;
- Whether the methodologies used to generate RWE can provide “adequate scientific evidence” to address the regulatory questions presented; and
- Whether the approach used in a case meets FDA’s regulatory requirements, such as established standards for data collection and study monitoring.

**Effectiveness Data.** The RWE Program will focus on the potential use of RWE to support changes to labeling about drug product effectiveness, including adding or modifying an indication, adding a new population, or adding comparative effectiveness or safety information.

**Concerns About Observational Studies and RWD.** The Framework acknowledges “observational studies may provide credible evidence,” but finds a stronger scientific justification for using randomized controlled trials as evidence of drug effectiveness. Indeed, the Framework indicates FDA will “consider reporting requirements for [observational studies] used to support effectiveness determinations.” These statements in the Framework signal that FDA remains cautious about the potential uses of RWE for regulatory purposes; FDA appears focused on an incremental approach, such as RWE to support a supplemental indication of an approved oncology drug or RWE from a “hybrid” clinical trial with both traditional RCT and RWE elements used to generate data. Bottom line, the burden will be on relevant stakeholders to demonstrate to FDA the ways that RWD and RWE can and should be used to support regulatory decisions.

**Data Standards for Submissions.** FDA recognizes the importance of developing data standards for submissions, to help ensure efficient review of RWD by the agency. FDA indicates it has “already been active in developing data standards for regulatory use and will continue to expand its work in this area.” This work will include identifying the relevant standards and methodologies to maximize the utility of RWD.

**Use of Electronic Source Data for RWE.** FDA touches on some of the regulatory considerations that arise from use of electronic source data, such as electronic health records and electronic data from clinical studies. The agency points out that it already has published some relevant information on these topics, including regulations that focus on the quality, authenticity, and reliability of electronic records (21 CFR Part 11) and a related guidance published in 2017. The agency highlights several key regulatory compliance issues, including informed consent, validation of electronic systems, audit trails for electronic records, and agency inspections. FDA is considering whether additional guidance on the use of electronic source data is needed.

FDA stresses the importance of continued engagement with all stakeholders in building out its framework. As FDA continues to develop policy in these areas, the agency will provide an opportunity for stakeholders to comment on specific regulatory issues and the agency’s proposed guidance documents. The current docket gives stakeholders an
opportunity to weigh in on FDA’s overarching strategic vision for implementing the RWE Program and the potential to offer important perspectives on how FDA can optimize the use of valuable real-world experiences in regulatory decision-making.

The Information Exchange and Data Transformation (INFORMED) is currently using RWD to examine the impact of a recent FDA labeling change for two approved products from weight-based dosing to flat-dosing of immune checkpoint inhibitors. This project aims to review how community practices are adopting the flat-dose label changes and the barriers or factors affecting this adoption.

Finally, to ensure that the agency can truly benefit from the volume, velocity and variety of real world data, the FDA will be looking to develop the capabilities of their staff with deep quantitative expertise.

The FDA aims to publish draft guidance for the use of RWE by October 2021, and consultation is already underway with healthcare sector stakeholders including the pharmaceutical industry. Both the FDA and the European Medicines Agency (EMA) have stated their wish to see increased use of RWE in supporting indications. In Asia, the growing maturity of real world data sources has led to the recent use of RWE in regulatory discussions, for example, in the decision in Japan on the use of raloxifene for the treatment of osteoporosis. Indian regulatory authorities are also looking to embed routinely collected electronic health records into their decision-making process.

**RWD:**

What are RWD and where do they come from?

Real world data are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. RWD can come from several sources, for example:

- Electronic health records (EHRs)
- Claims and billing activities
- Product and disease registries
- Patient-generated data including in home-use settings
- Data gathered from other sources that can inform on health status, such as mobile devices

What is RWE?

Real world evidence is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective).

In December 2016, the 21st Century Cures Act became law in the USA, aiming to expedite approval for new medicines. Towards that aim, it included provision for RWE to be used in place of evidence from randomized controlled trials (RCTs), if judged appropriate by the FDA.

RWE is derived from the analysis of data collected from a healthcare setting, outside the context of prescriptive RCTs. One of the key objectives of RWE is to understand observations and events in patients in routine clinical practice. RWE complements RCTs, which are carefully controlled experiments to test specific hypotheses on the efficacy and safety of new drugs, and which by design do not reflect current clinical practice. Owing to the mechanism of data collection and experimental design, RWE studies generally may not yield definitive causal inference because of the many confounders of variability.

**RWE**

RWE is derived from data associated with outcomes from the care of heterogeneous patients as experienced in real world practice settings. Data relevant to RWE comes in multiple types and forms. For Example: • Claims Data derived from insurance reimbursements. • Clinical Trials Data derived from the outcomes of randomized clinical trials. • Clinical Setting Data derived from patient medical records and patient care. • Pharmacy Data derived from prescription orders and fulfillments. • Patient-powered Data derived directly from the patient experience.
APPLICATIONS:

There are numerous implications:

• Quality improvement: RWD, such as patient clinical data captured in electronic health records, can be utilized for health care quality improvement. Example - the Hospital Corporation of America (HCA) utilized routine clinical data to identify a best practice for infection control leading to significant reductions in MRSA infections. At the December 2014 NEHI roundtable on RWE Dr. Joe Selby of PCORI noted that many health care delivery systems regard quality improvement as the most practical use of patient data under their control.

• Regulatory approval for new products: RWE may augment RCT data on the safety and efficacy of new drugs and medical devices. Data on the use of more heterogeneous (real world) sets of patients may:
  • Create greater precision and clarity as to the safety and efficacy profile of new products, thus improving the labeling and approved indications of products.
  • Improve the amount and quality of information available to patients and physicians relative to informed decision-making about new products.
  • Demonstration of value: Examination of post-approval, real world use of products may accelerate the rate at which products prove their value to patients, providers and payers, including:
    • More precise identification of safety risks and risk/benefit trade-offs.
  • Identification of heterogeneous responses, including identification of sub-population effects of products, the value of products when used among complex and co-morbid patients and value derived when products are delivered in diverse practice settings.
  • Longitudinal study: RWE draws on data sources that are based on recurring events such as patient visits and insurance claim submissions, thus facilitating long-term study of patient outcomes and health care utilization that may generate new findings on the appropriate use and the value of innovations.
  • Hypothesis generation: Analysis of RWD sources is widely considered to be a valuable source for the generation of research hypotheses and research questions that can be tested in randomized trials of drugs, devices and procedures, including trials for new or expanded uses of existing products
  • Patient recruitment: RWD sources can be utilized to expedite identification and recruitment of patients for clinical research.

Real world data (RWD) involves data collected outside of clinical trials, and produces real world evidence (RWE), becoming actionable when powered by analytics, machine learning & artificial intelligence (AI). RWE provides insight beyond traditional clinical trial data, adding potential to link data from different sources; improve trial efficiency; identify new indications, and a real-world perspective of risks/benefits to make informed decisions beyond traditional clinical trials.

FEW CASE STUDIES:

- The Salford Lung Studies are described as the world’s first Phase 3 pragmatic randomized clinical trials of a novel drug therapy. The studies are located in the United Kingdom and are comparing the use of a once-per-day inhaled corticosteroid against normal course of care for both COPD and asthma patients. Patients are treated and monitored across encounters with physicians, physicians’ staff and community pharmacists, through use of electronic medical records. The drug trial will test both the clinical effectiveness of treatments and their impact on patients’ ability to adhere and realize long-standing benefit

- Hospital-acquired infections, including MRSA, are a growing threat to hospital patients and staff. There has been limited evidence to support adoption of competing infection control strategies for MRSA, leading at least nine states to simply mandate a screening and isolation strategy for hospitals. A pragmatic, cluster-randomized clinical trial conducted by Hospital Corporation of America (HCA) staff within the HCA system, and peer reviewed for publication in the New England
Journal of Medicine, found strong evidence for a single approach. The trial cost less than $3 million – substantially less than a classic randomized clinical trial, although an indication that “high-quality delivery science is not free.”

PROBABLE USE CASES IN DRUG LIFECYCLES:

Disease strategy and early discovery disease segmentation

RWE has the potential to be used early in drug discovery and development programs, facilitating product development by identifying diseases or indications that represent a significant burden in populations. Electronic health records to support differentiation of patients’ needs have been used within the National Institutes of Health (NIH), and the ability to characterize patient populations before conducting a trial has enabled the NIH to design trials that accelerate innovative interventions to testing phase in patient subgroups of need.

Phase 1–3 clinical study design

To ensure a clinical trial protocol has internal validity, trial design teams will often use a set of restrictive eligibility criteria that may remove from the trial large segments of a population with the disease of interest. The impact of these eligibility criteria is often not understood or in most cases is not tested until the question of generalizability is raised at the stage of regulatory or reimbursement submission. This has been recognized as a limitation of RCTs by many regulators, including the FDA, in response to many approved medicines being withdrawn owing to safety problems being identified once a therapy has been exposed to a broad patient population. RWE can aid clinical study design; for example, in assessing the population size of patients with different sets of inclusion or exclusion criteria.

Indication-seeking and label-expanding studies

In order to license a therapy in a new indication or to expand the label into a new population, it is mandatory to establish evidence to support the efficacy claim. Traditionally, explanatory trials determine whether the intervention produces the expected result under controlled circumstances, generated through careful design of RCTs. As the need for larger RCTs increases, owing to low-rate event endpoints, potentially differential efficacy throughout subpopulations of patients and the need to observe larger populations for rare adverse events after intervention, the cost of running the trials increases. The time to run these trials also impacts on the potential profitability of indication expansion. Therefore, new thinking is required on how and if explanatory trials can leverage some of the features of real-world trials to deliver accelerated efficacy studies.

Pragmatic clinical trials – a hybrid approach

The main features of an RCT are the randomization of patients, enrolment into a controlled trial setting and follow-up specified in a study protocol. Applying this concept while also using real-world data may provide a hybrid approach to running pragmatic clinical trials. The levels of pragmatism can be understood within the context of the PRagmatic Explanatory Continuum Indicator Summary (PRECIS) framework. In the regulatory context, a balanced approach of using real-world data to execute large-cohort phase 3 trials may generate enough of a reward to risk taking the step towards an innovative execution model. This hybrid approach to running studies has been taken in examples such as the Salford Lung Study.

Post-authorization studies

Regulators including the FDA, EMA and China Food and Drug Administration increasingly ask pharmaceutical companies to implement ‘post-marketing commitment’ studies as a condition of approval.
In some cases, these commitments are requested after a product launch, for example, in light of new safety concerns. The studies may cover safety, efficacy, effectiveness or optimal use. One specific type of study, a post-authorization safety study, is usual for product authorizations: a large group of patients receiving the new medicine is tracked, often for a longer time period than covered by the registrational trial. Pharmaceutical companies are also obliged to enforce systems for spontaneous safety reporting, capturing and assessing adverse event data received from prescribing physicians. These data are consolidated into reports for regulators and are typically used for pharmacovigilance rather than for public reimbursement by each country’s national and local bodies, based on its effectiveness and safety, value for money and affordability. These are the key questions covered by health technology assessments, answered by health economic models that use data from RCTs and RWE studies, plus financial estimates and calculations.

Physicians also need to know how best to use new treatments in the broad patient population, not just in the restricted clinical trial sample. To give prescribers, guideline committees and formularies confidence to offer the medicines to patients, companies and independent investigators run retrospective and prospective RWE studies, showing outcomes from treatments in their region.

**RWE – CHALLENGES:**

Legal and Regulatory Compliance

• Need to promote open, transparent and replicable real world analysis

• Impact on public health and the efficiency of the health care system

• Respecting patient privacy and patient consent

• Rules for public access to the data

• Data holder – who is the owner/holder? And should access be granted to third-party?

• The business/sustainability model for maintenance of the data source?

• Data Sharing – Data sharing regulations?

**AI:**

Natural language processing (NLP) is an AI tool that can be described as the ability of a computer program to understand the human language and automatically extract contextual meaning. NLP can be particularly useful in processing and evaluating large amounts of unstructured data. In healthcare, a common application is to evaluate physician notes in patient medical records, and find relevant information. By applying NLP, a system can more easily and rapidly extract and analyze data that would otherwise be too burdensome for researchers. NLP replaces the highly cumbersome act of medical chart extraction using teams of researchers.

NLP Techniques range from simple word-based models for text classification to rich, structured models for syntactic parsing, collocation finding, word sense disambiguation, and machine translation.

Machine Learning (ML) is a library of algorithms that scour over large volumes of data to accurately and efficiently learn relationships found in recorded examples. Over the last 15-20 years, ML has gradually been replacing traditional statistical inference as the tool of choice for learning complex relationships in data. The key advantage of ML is the capability to operate on large numbers of engineered predictive features in
datasets including outliers, noise and collinearities, without the stability and reliability concerns of traditional statistical modeling.

Deep learning is a newer generation of learning algorithms rooted in an older concept called neural networks. Neural networks use an array of nodes to perform computations or decisions rapidly. Deep learning can be thought of as stacking many neural networks.

AI IN RWE:

From a real-world evidence perspective, one of the main advantages of Big Data infrastructure is the ability to maintain very large, heterogeneous and linked data sets that are highly available, where they can be queried and statistically processed rapidly and can be used in visualizations on a near real-time basis. For example, not only can data be updated and added to existing visualizations such as for tracking the opioid epidemic on a real-time basis across the entire U.S., but even extremely large custom cohort studies, such as answering questions on lipid lowering therapy or type 2 diabetes, representing millions of people and billions of data points, can be accomplished in hours or days rather than months or years (if data needs to be collected). In addition, important AI implementations have been made easier thanks to increased integration of AI algorithms with Big Data software.

The NLP demand for real-world evidence is highly driven by the tremendous increase in textual unstructured clinical data. The practice patterns used by physicians for the documentation of clinical notes as well as patient discharge summaries have generated an enormous amount of unstructured data. Such voluminous data needs to be structured and analyzed effectively for enhanced reporting and analytics. NLP, combined with machine learning and deep learning as described below, is rapidly becoming accurate enough to automate or replace abstraction. This drives significant efficiencies in generating information from text for real world evidence purposes.

For example, NLP can be applied to find information on treatment outcomes, adverse events, symptom presentation and referral patterns. Consider the following physician notes examples:

“He states the symptoms are controlled. Less than 1% BSA currently affected.”

“Stopped [Drug X] d/c ‘increased depression.’ On Paxil but ‘feels not helping.’ No psoriasis flares.”

“She has psoriasis on the back of her legs, torso, scalp. She uses a dermatologist. She was off [Drug X] for a URI and flared up.”

The underlined information are just examples of what is not captured in either billing or structured EMR data. The ‘old way’ would be to use nurse abstractors to chart review a small sample of patients. With advanced NLP, data on such things as reasons for discontinuation of a medication can now be captured at scale across tens of thousands of patients for less than the cost of a traditional chart review.

Machine Learning (ML) - One of our key applications of this capability has been in identifying patients with undiagnosed or underdiagnosed conditions. For example, the current approach is to use coded billing information or prescriptions to identify patients. Using ML, we are able to see much more complex patterns and interactions that are similar between patients with and without a particular diagnosis and able to confirm that the diagnosis is present but either unlabeled (such as in dementia) or unrecognized (such as in early presentation of rare diseases like muscular dystrophy). This technology has the promise of improving diagnosis in the clinic as well as in research studies.

Deep learning has introduced the capability to effectively automate the generation of predictive features in various types of inference problems and thereby achieve breakthrough performance in applications such as image processing, speech recognition and language translation. In healthcare, some of the key applications of deep learning that are being pursued are for reading radiology exams or pathology slides.
One of the most intriguing and potentially game changing examples of machine learning is in the area of predictive and prescriptive analytics. With traditional research approaches, evidence development focuses on evaluating and tracking what has already happened. But, how do we move from understanding what happened to being able to predict what will happen 6 months, a year, 5 years out?

Using different mathematical techniques and modeling, predictive analytics use existing data to find trends and patterns and tell us what might happen. They help to identify who is most at risk and what outcomes can be expected.

Traditionally, risk analytics have been performed using standard statistical techniques such as stepwise logistic regression. In these approaches, characteristics or risks are identified and added into models to determine their impact on the model performance. While predictive analytics can be generated using traditional statistical approaches, ML enables models to be generated to include thousands of variables and millions of data points. The result is usually more highly performant models as well as the ability to uncover more data relationships of importance that might not have been so prior to the analysis.

For example, we recently presented a machine learning based model for predicting heart failure readmissions that outperformed existing models (LACE Risk Score) by 10 points[1] and relied on another machine learning based variable that measures the aggregate disease burden of a patient (OM1 Medical Burden Index (OMBI™) and which is the strongest single predictor of many outcomes (heart failure admission and readmission, resource utilization).

Prescriptive analytics are an advanced form of predictive analytics. The goal of prescriptive analytics is to make the information presented actionable to a decision maker. Prescriptive analytics tell us what to do about the information that the predictive models generated and help us to know which ones matter most and what actions to take. For example, a clinician might use predictive analytics to understand who is most at risk for a cardiac event, whereas prescriptive analytics might tell the provider which patients have alterable factors, such as weight loss or smoking status, and which ones will have the greatest impact on outcomes.

The healthcare and real-world evidence applications of these AI driven capabilities are potentially enormous. Clinicians are already using these capabilities to identify which patients are most likely to have poor clinical or financial outcomes and to proactively take actions to minimize that risk. For example, avoiding a cardiac readmission can save a health care payer or at-risk provider $14,000-$18,000 on average per event. The implications are similarly large for manufacturers. Predictive analytics are now being applied to identify patients most likely to benefit from certain treatments, those likely to be adherent to therapy, or even those likely to suffer an adverse event.

PROBABLE USE CASES:

The reality is somewhere between 30-50 percent (30-50%) of a patient’s EHR data is structured in nature. Per above comments about the correct tabular data strategy, the recommendation is to use RDBMS technologies to manage this portion of the patient’s EHR data.

The remainder of the patient’s EHR data is unstructured in nature. That data is trapped in documents like clinician notes, patient discharge summaries, radiology reports, lab reports, etc. Oracle recommends using NLP algorithms to extract the relevant data from these documents and combine it with the structured portion of the patient’s EHR data to yield a richer patient EHR data source. After mastering this data source with NLP, researchers can use this same approach against other unstructured data sources like scientific literature (e.g. Pubmed Research Articles) and/or social media (e.g. PatientsLikeMe, CancerLinq, Twitter, Facebook, etc.).
CONCLUSION

Artificial intelligence and big data are transforming real-world evidence from a largely retrospective viewpoint to a more concurrent and forward-looking set of capabilities. This paradigm shift also will drive RWE to the forefront of strategy for both healthcare and life sciences organizations. While there are many different components of AI that offer new approaches and methods to evaluating and generating real-world evidence, one common thread throughout is the importance of big data and the interdependency on having access to enormous amounts of data.

By embracing the innovation in AI (and the availability of big data), researchers can generate real-world evidence that is more dynamic, timely, representative, comprehensive and cost-effective. This next generation of real-world evidence will also can be used to measure, predict and personalize care in a way previously not possible. In the end, all healthcare stakeholders’ benefit when medical products and services are focused on and delivered to those who will benefit the most.

REFERENCES

Pubmed scholar articles

ACKNOWLEDGMENTS

Acknowledging the support from Ephicacy management team.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Charan Kumar
Ephicacy Lifescience Analytics Pvt. Ltd.
charan.kumar@ephicacy.com
charan.kumar.kj@gmail.com