Real-world data: Narasimha Kumar

How data from medical practice can inform care

The gap between randomised clinical trial outcomes and the actual safety and effectiveness of a drug in heterogeneous patient populations presents a major challenge for the global biopharmaceutical industry. To overcome this, regulatory agencies are increasingly mandating the use of real-world data (RWD), or data collected from clinical practice, at various stages of a drug's lifecycle. Traditionally, biopharma organisations have operated linearly on the basis of trial, regulatory approval, and post-marketing studies. RWD has only entered into the picture after a drug launch on the market. Integrating RWD further upstream – before regulatory approval – can help optimise clinical trial design, refine market access strategies, identify unmet medical needs, and continuously generate evidence about a drug after its launch.

While randomised clinical trials remain the gold standard for demonstrating efficacy and safety, they often capture only a fraction of the complexity seen in real clinical practice. Once therapies reach the market, physicians end up treating populations that are heterogeneous and under diverse healthcare constraints. RWD can therefore enable researchers to observe how diseases are managed across populations.

Standards of care have been traditionally identified based on usage guidelines, expert consensus, and published articles. However, due to challenges in data acquisition and analysis and the time it takes to submit this data to regulators, drug usage guidelines can lag behind clinical practice. This can be for several months, and in some cases, for years. As a result, in many therapeutic areas, the 'real' standard of care can differ from published clinical guidelines.

RWD has emerged as a way for biopharma teams to understand standard of care and treatment patterns, not as static guidelines, but as evolving clinical behaviours. This shift is transforming how companies design trials, prioritise indications, and demonstrate evidence of value to regulators and payers. Standard of care definitions based on clinical practice derived from RWD from heterogenous patient populations can be an outcome of observational studies. This can then lead to an innovation in drug targets that meet or exceed current care standards.

RWD can also reveal off-label use of drugs, regional and ethnic variations of use, treatment sequencing, therapy-switching patterns, delays in diagnoses, and gaps in a patient's adherence to guidelines. As a result, in many therapeutic areas, the 'real' standard of care can be shown to be substantially different from published clinical guidelines. This is why RWD analysis can be essential ahead of the design of comparator trials.

For example, in oncology, where patients may receive multiple unofficial sequences of a medication, designing comparator trials can be an exercise in hypothesis testing. With RWD these hypotheses can be validated. In rare diseases, where the standard of care is fragmented or inconsistent, RWD can be used to generate digital twins for simulation. In all of these cases of clinical trial design, the key advantage is the speed with which RWD can bring value to decision-making.

Integrating RWD into decision-making is not without constraints. These include the need to follow rules for privacy, data integrity or provenance, and data residency. Furthermore, RWD must overcome problems of interoperability or the conversion of new data from legacy data formats, and the need to adhere to data model standards. At the same time, RWD must address the multi-modal nature of the data itself. As a result, RWD projects can be expensive and time-consuming.

Rapid evolution of AI

A solution to these problems is emerging. The rapid evolution of artificial intelligence (AI) and the maturity of cloud computing services, which can handle big data at scale, have the potential to offset these shortcomings. In other words, the convergence of pharma decision-making, informed by RWD and AI, should make it possible to bring safer and more effective drugs to market while reducing the cost of drug discovery. Companies are therefore investing heavily in data harmonisation, analytics governance, and expertise across multiple disciplines.

At a conceptual level, the architecture of the future may be described as a convergence of multiple sources of data and modes of action into a secure location with agentic AI workflows enabling data ingestion, deidentification and harmonisation. Causal AI models then can create a knowledge/concept graph using medical ontologies, dictionaries, and hybrid data models that are a combination of clinical, observational and deep learning. The resulting, curated RWD can be made available for multiple use cases. These include aggregate-level analytics to validate protocol assumptions and synthetic control arms for trials or for a deeper understanding of the standard of care.

The promise of AI is changing the linear approach to trials, approvals, and post-marketing studies for industry and regulatory authorities. The new approach, for many, is to embed RWD use at every stage of the drug lifecycle. This use begins with trial design and leverages trial outcomes alongside RWD to support a drug's launch and post-market monitoring. The ongoing learnings from this continuous collection and analysis of RWD can then inform decisions about market access, comparator studies, and label expansion.

Advances in AI and data platforms are accelerating this transition by enabling a dynamic analysis of treatment pathways and patient outcomes at scale. In this emerging model, RWD becomes not merely supportive evidence, but a core strategic capability.

Compliance with quality guidelines (GxP), especially in

areas such as data integrity and the change management of continuous learning AI models, is yet to be fully embedded into the AI solution lifecycle. Problems remain, including the cost of AI and data initiatives for medium-sized companies and academic institutions and the need for large companies to define and implement a unified AI and RWD strategy across their portfolios. It is therefore useful for companies taking this technology forward to ask themselves five questions. These are first, what RWD data are available; second, what patterns emerged from the use of this data; third, what changes in trial design should be made on the basis of this data; fourth, how can the value of this data be proven continuously; and finally, what lessons can be learned for the future?

The next generation

The next generation of pharmaceutical innovation will be defined not only by scientific discovery but by how well therapies align with real clinical behaviour. Companies that embed RWD into standard of care and treatment pattern analyses have the opportunity to reduce uncertainty across their development and commercialisation decisions. Rather than replacing randomised controlled trials, the full use of RWD could enhance them.

References:

1. Application of Real-World Data to External Control Groups in Oncology Clinical Trial Drug Development

https://pmc.ncbi.nlm.nih.gov/articles/PMC8771908/

2. Real-world evidence to support regulatory submissions: A landscape review and assessment of use cases

https://pmc.ncbi.nlm.nih.gov/articles/PMC11295294/

3. RWE – A Guide to RWE Analysis and Application

https://intuitionlabs.ai/articles/real-world-evidence-analysis

4. The future of RWD and RWE in healthcare decision-making: Applications of novel real-world data collection methods for healthcare decision-making

https://www.drugdiscoverytrends.com/real-world-evidence-frameworks-pharma-decisions/

5. Use of real-world evidence in regulatory decision making – EMA publishes review of its studies

https://www.ema.europa.eu/en/news/use-real-world-evidence-regulatory-decision-making-ema-publishes-review-its-studies

6. Utilization of real-world evidence in regulatory approvals for multiple myeloma therapies

https://pmc.ncbi.nlm.nih.gov/articles/PMC12660783/

7. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics

https://pmc.ncbi.nlm.nih.gov/articles/PMC7196856/

8. FDA use of Real-World Evidence in Regulatory Decision Making

https://www.fda.gov/science-research/real-world-evidence/fda-use-real-world-evidence-regulatory-decision-making

This article was written by Narasimha Kumar, global head of technology consulting, BC Platforms AG of Zurich, Switzerland.

Definitions

Term	Definition
Real-World Data (RWD)	Data that is collected in normal clinical practices and not in controlled clinical trial settings about patients' health status and delivery of health care, from sources such as electronic medical records, claims, patient-reported outcomes and disease registries.
Real World Evidence (RWE)	The clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.
Standard of Care (SoC) – Medical definition	The 'benchmark' of best practice based on scientific evidence, consensus and clinical guidelines. The level of and type of treatment that a competent health care professional would provide to a patient under the same or similar circumstances.
Standard of Care (SoC) – Legal definition	From a medico-legal perspective SoC is the minimum accepted level of care. This definition could change country by country or even within a country. Most common definition in the US is that level of care, skill and treatment which, in light of all relevant surrounding circumstances, is recognised as acceptable and appropriate by reasonably prudent similar health care providers.
Interoperability	The ability of different information systems, devices, and applications to securely access, exchange, and cooperatively use data. It enables seamless communication between EHRs, labs, and pharmacies to improve patient care, reduce errors, and streamline workflows.
Causal AI models	AI models that go beyond correlation-based predictions and look at cause-and-effect scenarios. By doing so, they enable 'what-if' analyses to help in decision making.
Hybrid data models	Data models that combine structured (tables and rows), with semi-structured (documents and file formats with sections and defined information units) and unstructured (narratives in clinical summaries) to provide context-aware insights for Gen AI applications.