

Statistical Insights

Going Synthetic with Real World Data

The benefits of synthetic data generation and application in clinical trial study design and analysis.

Going Synthetic with Real World Data

The benefits of synthetic data generation and application in clinical trial study design and analysis.

Real-world data has the potential to have a positive impact on the treatment of serious conditions, providing much-needed insights into the health and treatment of patients in an everyday setting.

However, the collection of this data can be met with resistance due to privacy concerns surrounding its use and anonymity. Aiden Flynn, CEO of clinical trial data and design experts, Exploristics, explores the power of synthetic data as a means of countering these concerns.

Over the past decade, there has been a surge in realworld data collection. With a multitude of emerging data sources ranging from electronic health records (EHRs), smartphone apps and medical insurance claims, real-world data offers useful information on patients and disease areas that can widen our understanding of the value of new medicines from early clinical development through to post-authorisation. Clinical data gathered in everyday settings has the potential to be used to complement the information garnered from the artificial environment of the randomised controlled trial (RCT), filling knowledge gaps on the use of medicines in the real-world or enabling the evaluation of unstudied factors influencing a patient's outcome.

However, despite such promise the collection of real-world data is not without its challenges. Drug developers not only face cost and time hurdles in exploring these burgeoning data resources but also the associated issues surrounding patient privacy.

Keeping it private

So, how is it possible to harness the possibilities that real-world data offers whilst overcoming justifiable concerns regarding patient privacy? To understand a patient's response to a given treatment or the evolution of their condition, access to their clinical data is required. However, any ability to identify who a patient is from this data compromises their privacy, breaching data privacy regulations and deterring further data sharing for the purposes of medical research.

Consequently, to counter these issues several methods are now employed postcollection to conceal patient identity while still gaining maximum benefit from the information in their data. Such methods include:



Anonymisation

where patients are assigned numbers rather than being identified by their names.



Pseudo-anonymisation

where patient information is mostly anonymised but certain identifiable data is retained, such as a person's home town.

Summarising

where rather than referring to an individual patient, their data is summarised within a patient group with meaningful information extracted from the group.

However, whilst these anonymisation methods are routinely used, an alternative approach is emerging which does not to compromise patient privacy and yet offers to unleash the power of real-world settings to inform clinical development; the generation of synthetic data.



Exploristics Ltd. 24 Linenhall Street Belfast, Northern Ireland BT2 8BG

The Benefits of Synthetic Data

Key reasons why synthetic data is being increasingly exploited.







Synthetic data offers a useful tool for statisticians as it can **replicate the main characteristics of real patient data, such as the range, distribution, averages and interrelationships.** It can be used to **increase the amount of available information**, either by supplementing real data sets or by being used in its absence. Importantly, synthetic data offers researchers **the benefit of circumventing all privacy concerns** as it can closely mimic real-world data sets but does not relate directly to real individuals.

Already, synthetic data sets are being used in diseasespecific research areas such as oncology to provide banks of information from which researchers can gain insights without the need for real patient data. Indeed, facilities like the Simulacrum, a bank of free-to-use synthetic patient-like cancer data developed by Health Data Insight and released in 2018, have been set up to support such research efforts. The Simulacrum offers a successful example of how a disease-specific synthetic data bank can transform research by providing an accurate representation of real data. Its constituent data is based on real data collected by the National Cancer Registration and Analysis Service (NCRAS) from individuals diagnosed with cancer in England. However, despite not being comprised of real patient data its synthetic data sets mimic the properties of the real NCRAS data, enabling researchers to still derive useful insights into cancer outcomes in England.

The need to be dynamic and evolving

Undoubtedly, the use of synthetic data offers a viable approach to replicate and augment the use of real-world data to inform clinical development. Yet, it also has notable drawbacks. Synthesised data is essentially static and based on a snapshot of the real-world at a given point in time. As such, it can very quickly become outdated, no longer reflecting the realities of a disease or patient population as environments and pathologies evolve. Therefore, to continue to be relevant, synthetic databanks must be regularly updated with newly synthesised data as the knowledge base grows. Indeed, it is the longitudinal characteristics of these data that will generate a greater understanding of the natural history of disease, the effectiveness of treatment interventions and the emerging unmet medical needs.

The future

With various strategies being employed to liberate the opportunities offered by real world data, synthetic data provides an emerging and useful tool for augmenting the still patchy clinical information available from real settings whilst eliminating patient privacy concerns. As such, its use looks set to grow in unlocking the potential of real-world information. As the field of real-world data matures, our increasing ability to harness the available data, both real and synthetic, to derive meaningful clinical insights from it will transform the clinical development process and the delivery of new medicines.



Exploristics Ltd. 24 Linenhall Street Belfast, Northern Ireland BT2 8BG find out more T +44 (0) 28 9600 1996 info@exploristics.com