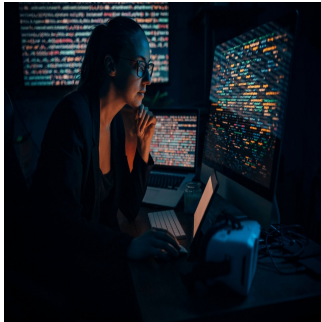

Enhancing Clinical Trial Insights with Real-World Data: Proxy Endpoint Models



Randomised Controlled Trials (RCTs) are the gold standard for assessing drug efficacy, but they come with limitations, such as small sample sizes and highly controlled environments. On the other hand, real-world data (RWD) offer extensive datasets drawn from everyday medical practice, providing insights into larger populations. However, RWD lacks specific clinical endpoints typically recorded in RCTs, making it challenging to evaluate treatment effects. To bridge this gap, recent advancements in machine learning, particularly the use of proxy endpoint models, offer a promising solution. These models allow researchers to infer missing disease endpoints in RWD using data from RCTs, improving the capacity to analyse and apply real-world evidence in clinical settings.

Using Proxy Endpoints to Connect RCTs and RWD

Proxy endpoint models are a powerful tool for overcoming the challenges of integrating RWD with RCT insights. RCTs offer high-quality, detailed data about patient responses to treatments, but they often involve fewer patients due to their controlled and expensive nature. In contrast, RWD encompasses vast patient populations but lacks the precision of RCTs, particularly when it comes to endpoints like disease severity scores. Proxy models enable researchers to use available RWD features to predict these endpoints, thereby unlocking the potential of larger datasets for clinical analysis.

One method to develop these models is by leveraging machine learning algorithms that identify key features in RCTs and map them to data points in RWD. For example, in studies of rheumatoid arthritis (RA) and atopic dermatitis (AD), researchers have successfully employed predictive models like Explainable Boosting Machines (EBMs). These models allow for the incorporation of non-linear relationships and interpretability, which is essential for gaining actionable clinical insights. By using selected features from RCT data that are also present in RWD, the proxy models can predict missing disease severity scores in broader real-world populations.

Critical Methodologies for Developing Proxy Endpoint Models

The process of creating reliable proxy endpoint models requires several critical steps. First, the integration of RCT and RWD datasets must be done carefully to ensure the features in both datasets are aligned. Researchers typically apply multi-stage feature selection methods that filter out irrelevant or redundant data. For instance, lab tests or demographic factors not available in RWD might be excluded, while significant biomarkers common to both RCT and RWD datasets are retained.

After feature selection, models are trained on the RCT data. EBMs, due to their interpretability and accuracy, are widely used in these scenarios. These models not only provide high predictive power but also offer insights into which features have the most significant impact on the prediction. For example, in predicting RA severity using the DAS28-CRP score, features like C-reactive protein (CRP) levels and erythrocyte sedimentation rate (ESR) are critical. In AD, variables such as eosinophil counts and lymphocyte levels play a prominent role in predicting the EASI score, a standard measure of disease severity.

Once the models are trained and validated using RCT data, they can be deployed in RWD environments. However, one of the challenges of deployment is that the more features a model requires, the fewer patients in the RWD dataset will have all the necessary data points. Therefore, models must balance complexity (number of features) and generalizability (applicability to a large RWD cohort). By refining feature sets to include only the most widely available and clinically relevant factors, proxy models can be applied to larger populations, greatly enhancing the potential for real-world analysis.

Applications and Benefits of Proxy Endpoint Models

Proxy endpoint models offer multiple benefits across different stages of drug development and clinical analysis. First, these models can dramatically increase the number of patients available for treatment effect studies by filling in missing endpoint data in RWD. For example, in the study of RA, proxy models that predict DAS28-CRP scores can more than double the size of the patient cohort available for analysis. This larger dataset can then be used to assess the effectiveness of treatments across more diverse populations, improving the generalizability of the results.

Additionally, these models can be used for constructing external control arms in clinical trials, offering a cost-effective way to compare treatment outcomes without enrolling new patients. Proxy endpoint models also enable the simulation of clinical trials using real-world data, allowing researchers to refine trial designs and explore different treatment strategies before conducting expensive and time-consuming RCTs.

Another significant application is in drug repurposing. By analysing predicted disease outcomes in RWD, researchers can identify potential new uses for existing drugs, as demonstrated in studies on diseases like Alzheimer's and coronary artery disease. These models can also facilitate personalised medicine by predicting how individual patients might respond to treatments based on their real-world medical data, helping clinicians make better-informed decisions.

Conclusion

Proxy endpoint models represent a significant advancement in integrating RWD with insights from RCTs. By leveraging machine learning techniques to predict disease severity scores, these models unlock the vast potential of real-world data for clinical research, drug development, and treatment analysis. In diseases like rheumatoid arthritis and atopic dermatitis, the successful application of proxy models has already demonstrated the ability to extend findings from small, controlled clinical trials to broader, real-world populations. As the methodologies and technologies continue to evolve, the use of proxy endpoints will likely become a cornerstone of real-world evidence, driving innovation in healthcare and improving patient outcomes.

While challenges remain—particularly around the availability of features in RWD datasets and the complexity of certain models—the benefits of proxy endpoint modelling are undeniable. These models bridge the controlled world of RCTs and real-world clinical data's vast, untapped potential, offering a path forward for more comprehensive and impactful medical research.

Source Credit: [Journal of Biomedical Informatics](#)

Image Credit: [iStock](#)

Published on : Tue, 24 Sep 2024