

IBM Research Report

Facilitating Clinical Data Analytics for Comparative Effectiveness Research

Xingzhi Sun¹, Linhao Xu², Yiqin Yu², Matthias Reumann¹

¹IBM Research Division
Level 5
204 Lygon Street
Carlton, VIC 3053
Australia

² IBM Research Division
China Research Laboratory
Building 19, Zhouguancun Software Park
8 Dongbeiwang West Road, Haidian District
Beijing, 100193
P.R.China



Research Division

Almaden - Austin - Beijing - Brazil - Cambridge - Dublin - Haifa - India - Melbourne - T. J. Watson -
Tokyo - Zurich

Facilitating Clinical Data Analytics for Comparative Effectiveness Research

Xingzhi Sun¹, Linhao Xu², Yiqin Yu², and Matthias Reumann¹
¹IBM Research – Australia, ²IBM Research – China

Introduction

Comparative effectiveness research (CER) is an important type of clinical research that compares the benefits and harms of alternative methods for a particular set of patients, to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. Compared with the classic Randomized Controlled Trial (RCT), which measures efficacy under ideal, well-controlled clinical conditions, CER evaluates the effectiveness of an intervention in a “real world” environment. Therefore, CER is critical for providing evidence and support for informed medical care and health policy decision making.

With the rapid development of clinical information systems (e.g., EMR), more and more patient/disease data are accumulated in daily clinical databases. Clinical data analytics over such databases for CER has many advantages, such as speeding up the research process, reducing the research cost and providing clinical evidence for real-life decision making.

To facilitate data analytics for CER, we propose a system called OCERA: ‘On-demand Comparative Effectiveness Research Accelerator’. OCERA can assist physicians to efficiently undertake high quality CER by addressing various types of complexities in the clinical research process. In this abstract, we will also report a case study that shows the effectiveness of our system.

Technology Brief

Epidemiology Support: In OCERA, epidemiological concepts and workflows are utilized to guide physicians to perform CER. This enhances the quality of CER studies by following the right methodology.

Hypothesis Generation: Hypotheses in CER are very broad. Given a clinical dataset with more than a hundred variables, it is a great challenge to identify all potentially valuable hypotheses for further analysis and evaluation. OCERA can help physicians quickly find the hypotheses that are appropriate to CER and are potentially valuable from the statistical viewpoint.

Statistical Method Recommendation: In clinical data analysis, the knowledge required to select the relevant statistical methods may be beyond the skills of some clinical researchers. OCERA has a comprehensive statistics library, and can smartly recommend the appropriate low-level statistical methods and settings according to the high-level task description and data characteristics.

Standard Clinical Insights Reporting: OCERA can automatically create research reports based on widely-accepted templates, and generate concluding remarks that are statistically significant.

A Case Study

In 2011, a clinical guideline is published to guide the practical insulin pump therapy in China. However, for different diabetes types, this guideline only provides general instructions for insulin dose when using insulin pump. To further investigate the effect of using insulin pump, the Endocrinology Department of Peking University People's Hospital (PKUPH) collaborated with IBM and leveraged our system in their research. The data in the study comes from a multi-centre insulin pump therapy research project with more than 20 hospitals involved, including thousands cases and more than 300 clinical factors for each case.

Cooperated with the physicians from PKUPH, several clinical factors such as diabetes mellitus type and insulin type are selected as exposure factors. For each exposure factor, the hypothesis generation component helps physicians identify a set of confounding factors (such as age, gender and BMI) and clinical outcomes (such as total cost, total daily dose of insulin per kg, and frequency of hypoglycaemic-like symptom). Then, for each hypothesis of interest, physicians can initialize a CER project that consists of a workflow of activities, where the involved statistical methods can be recommended for various types of analyses, such as baseline comparison and multivariate regression. Finally, the research reports are automatically generated. The research results and the positive feedback from the physicians indicate that the system is helpful to conduct the practical CER studies.

Conclusion

We have introduced a system for conducting Comparative Effectiveness Research (CER) via clinical data analytics. A case study shows that the system can hide the complexities of CER from the clinical researchers and improve their research productivity.

Keywords: Comparative Effectiveness Research, Clinical Data Analysis