International Journal of Population Data Science

Journal Website: www.ijpds.org

Design and operation of a distributed health data network

Brown, J¹

¹Harvard Pilgrim Health Care Institute

Introduction

Several large health data networks such as FDA Sentinel, PCORnet, and the Canadian Network of Observational Drug Effect Studies (CNODES) facilitate multi-site research using real-world electronic health data such as administrative claims data, electronic health record data and registries. Experience in operation of multiple health data networks will be described.

Conclusion/Implications

Distributed health data networks offer the opportunity to use real-world data for public health surveillance and comparative safety and effectiveness research across large populations. The operational design, technical and analytic architecture, and governance models of networks drive their acceptance and success.

Objectives and Approach

Over the past 15 years substantial progress has been made in developing the optimal network operational design, governance, and technical architecture to facilitate the creation and operation of large-scale distributed health data networks. The design, architecture, and operation of a sustainable health data network requires balancing the needs of the network stakeholders such as funders, data sources, investigators, and regulatory bodies while enabling rapid and efficient use of data to support evidence generation and decision making. Important topics include protection of patient privacy, security, data autonomy, distributed analytics, data quality, and protection of confidential information.

Results

The design and architecture of existing distributed health data networks provides guidance regarding the potential operational model for new networks and identifies areas of research to improve network functionality and capabilities. Most health data networks adopt a common data model approach to facilitate multi-site querying and data quality assessment. This approach is coupled with distributed querying in which data partners maintain physical and operational control of their data. This design maximizes protection of confidential and proprietary information and minimizes the need to share patient-level data. Privacy-preserving distributed regression approaches and methods that obviate the need to share person-level data while generating robust results help to ensure network participation. Strong security and governance structures are also necessary for effective operation of a distributed network.