

International Journal of Population Data Science



Journal Website: www.ijpds.org

A big data analytics platform to support simulation modeling for osteoarthritis care pathways

Vik, S¹, Sharif, B¹, Seidel, J², and Marshall, DA¹

¹University of Calgary

²Alberta Health Services

Introduction

Technical solutions have been used in industry settings for many years to facilitate efficient management and analyses of big data sources. An initiative to apply a business solution to support development of simulation models for health systems research using nearly two decades of provincial administrative health data is described.

Objectives and Approach

Administrative data including practitioner claims, hospitalizations and ambulatory care visits for patients with a diagnosis of osteoarthritis were obtained from Alberta Health for the period 1994/95 to 2012/13. These data were incorporated into a multidimensional data cube using Microsoft SQL Server Analysis Services. Initial steps required dimensional modeling to restructure the data into a star schema format. This involved appending several data sets and defining additional reference tables to contain stratification variables and denominator data for rate calculations. The modeling expert worked closely with the information technology team throughout the process and assessed validity of the output.

Results

Development and validation of the multidimensional cube occurred in iterations over approximately 12 months. The final solution resulted in an analytics platform that compiled data from approximately 400 million records obtained from four different administrative data sources. Ten dimension tables containing 102 variables provided enhanced flexibility to conduct ad hoc stratified analyses in a fraction of the time that would be required using conventional methods. For example, some analyses that previously required a day of analyst time could be performed in less than 15 minutes. The efficiencies in analytic time were achieved by the pre-aggregated measures and slice and dice capability of the data cube, which negated many intermediary steps for data extraction and time consuming iterative analyses required for development of the simulation models.

Conclusion/Implications

This project demonstrated how a technical solution applied in industry can be utilized to address challenges encountered by researchers related to managing and analyzing large administrative health data sets. The methods could be applied in many other research settings to facilitate access to and analyses of information using big data.

