Mohammed Aldhoayan - Abstract

Comprehensive Management and Analysis of Complex and Large Research Datasets: An Application in Communication Science and Disorders

Today, healthcare is awash in data. With the increasing number of data resources and advancements in information technology, large and complex research datasets are at hand to be used and converted into knowledge, which is one critical driver in our journey toward effective, efficient, and safe healthcare. However, when these datasets are mismanaged or corrupted, they could produce low-quality and even misleading results. Thus, it is critical to maintain high quality standards of data management and promote for the concepts of data lifecycle and data curation due to their direct impact on the quality of research datasets and the results of their analysis.

This dissertation research is aimed at demonstrating a data management and quality assurance process through the implementation of a full research dataset lifecycle, building a centralized and secured database with a user-friendly interface and assess its usability, and testing a theory-based model of sentence comprehension using structural equation modeling for four types of sentences.

Results of this study have shown that automating the process of extracting, verifying, transforming, and storing large research datasets from their source files to a structured and analysis-friendly database increases the quality of the data, reduces the burden and time waste of researchers, and facilitates the communication process between the research team members. The conducted usability study has shown that users spend less time and make fewer errors when retrieving datasets using properly designed interfaces than using Excel spreadsheets.

Results from the sentence comprehension analysis have shown that language processing, short-term memory, and working memory can be measured separately. The structural equation modeling analyses have also shown the importance of working memory and how it significantly predicts sentence comprehension of object cleft and garden path sentences.

The data processing techniques used in this dissertation have laid the cornerstone of a generic data processing and management system that gives clinical researchers, especially the ones with limited resources, the ability to manage and process their datasets by enabling them to define and execute different extraction rules. The results of the dissertation also provided clarification on the nature of the interactions between critical cognitive systems and their impact in sentence comprehension deficits.