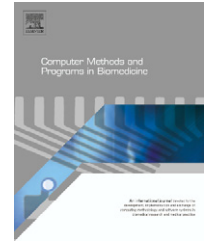




ELSEVIER

journal homepage: [www.intl.elsevierhealth.com/journals/cmpb](http://www.intl.elsevierhealth.com/journals/cmpb)

# Alternatives to relational database: Comparison of NoSQL and XML approaches for clinical data storage

## 1. Introduction

Clinical data is dynamic, sporadic, and heterogeneous in nature [1]. While they share some of the characteristics of the data managed by conventional data warehouse, special attention is required in the design of database schema because of the unique features they possess. Currently, storage of clinical data largely relies on relational database management systems. The relational database model is the most common and a proven approach to store and query data in various forms [2]. However, the major drawback is the need to pre-design the exact field structures of the data, which is required in

the process of database normalization to ensure data consistency [3]. In addition, the relational database model is not practical for certain forms of data that require a lot of fields to handle different types of data involved, where most of the data fields are indeed left unused due to the nature of the data. A relational database storing these kinds of data will contain many empty fields, resulting in inefficient storage and poor performance. Medical data, especially clinical notes, are such an example. To deal with these issues, we attempt to use a class of database known as NoSQL and extensible markup language (XML) to develop databases that can cope with the special features of clinical data more effectively.