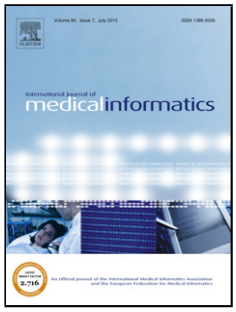

Title: Privacy-preserving Data Cube for Electronic Medical Records: An Experimental Evaluation

Author: Soohyung Kim Hyukki Lee Yon Dohn Chung

PII: S1386-5056(16)30201-5
DOI: http://dx.doi.org/doi:10.1016/j.ijmedinf.2016.09.008
Reference: IJB 3396

To appear in: *International Journal of Medical Informatics*

Received date: 23-6-2016
Revised date: 5-9-2016
Accepted date: 20-9-2016

Please cite this article as: Soohyung Kim, Hyukki Lee, Yon Dohn Chung, Privacy-preserving Data Cube for Electronic Medical Records: An Experimental Evaluation, *<![CDATA[International Journal of Medical Informatics]]>* (2016), http://dx.doi.org/10.1016/j.ijmedinf.2016.09.008

Privacy-preserving Data Cube for Electronic Medical Records: An Experimental Evaluation

Soohyung Kim[a], Hyukki Lee[b], Yon Dohn Chung[b,*]

[a]*Department of IT Convergence, Korea University, Seoul, Korea*
[b]*Department of Computer Science and Engineering, Korea University, Seoul, Korea*

Abstract

Introduction: The aim of this study is to evaluate the effectiveness and efficiency of privacy-preserving data cubes of electronic medical records (EMRs). An EMR data cube is a complex of EMR statistics that are summarized or aggregated by all possible combinations of attributes. Data cubes are widely utilized for efficient big data analysis and also have great potential for EMR analysis. For safe data analysis without privacy breaches, we must consider the privacy preservation characteristics of the EMR data cube. In this paper, we introduce a design for a privacy-preserving EMR data cube and the anonymization methods needed to achieve data privacy. We further focus on changes in efficiency and effectiveness that are caused by the anonymization process for privacy preservation. Thus, we experimentally evaluate various types of privacy-preserving EMR data cubes using several practical metrics and discuss the applicability of each anonymization method with consideration for the EMR analysis environment.

Methods: We construct privacy-preserving EMR data cubes from anonymized EMR datasets. A real EMR dataset and demographic dataset are used for the evaluation. There are a large number of anonymization methods to preserve EMR privacy, and the methods are classified into three categories (i.e., global

*Corresponding author. Address: 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. Tel.: +82 2 3290 3580

Email addresses: soohyung@korea.ac.kr (Soohyung Kim), hyukki@korea.ac.kr (Hyukki Lee), ydchung@korea.ac.kr (Yon Dohn Chung)