Abstract: Quality in hospital administrative databases

Alberto Freitas[1, 2, *], Juliano Gaspar[1, 2], Nuno Rocha[1, 2], Goreti Marreiros[3], Altamiro da Costa-Pereira[1, 2]

[1] *Department of Health Information and Decision Sciences, Faculty of Medicine, University of Porto, Portugal*

[2] *CINTESIS - Center for Research in Health Technologies and Information Systems, Portugal*

[3] *GECAD – Knowledge Engineering and Decision Support Group, Institute of Engineering – Polytechnic of Porto, Portugal*

{alberto, jgaspar, nunorocha}@med.up.pt, mgt@isep.ipp.pt, altamiro@med.up.pt

Abstract

The clinical content of administrative databases includes, among others, patient demographic characteristics, codes for diagnoses and procedures. The data in these databases is standardized, clearly defined, readily available, less expensive than collected by other means, and normally covers hospitalizations in entire geographic areas. Although with some limitations, this data is often used to evaluate the quality of healthcare. Under these circumstances, the quality of the data, for instance, errors, or it completeness, is of central importance and should never be ignored. Both the minimization of data quality problems and a deep knowledge about this data (e.g., how to select a patient group) are important for users to trust and to correctly interpret results. In this paper we present, discuss and give some recommendations for some problems found in these administrative databases. We also present a simple tool that can be used to screen the quality of data through the use of domain specific data quality indicators. These indicators can significantly contribute to better data, to give steps towards a continuous increase of data quality and, certainly, to better informed decision-making.

Acknowledgement

The authors wish to thank ACSS, for providing access to the data, and the support given by the research project HR-QoD – Quality of data (outliers, inconsistencies and errors) in hospital inpatient databases: methods and implications for data modeling, cleansing and analysis (project PTDC/SAU – ESA /75660/ 2006). This work is partially supported by FEDER Funds through the "Programa Operacional Factores de Competitividade - COMPETE" program and by National Funds through FCT "Fundação para a Ciência e a Tecnologia" under the project: FCOMP-01-0124-FEDER-PEst-OE/EEI/UI0760/2011.