Mining some medical information in a fee database

Pablo E. Verde [*†], Christian Ohmann [‡], Christof Kugler [§] and Max Geraedts [¶]

February 10, 2004

Abstract

Billing collection systems of medical insurance institutions generate a massive amount of data, which are used for money transaction and for further administrative purposes. Recently, public health researchers highlighted the possibility that these databases could contain useful information for health-care evaluation. In this work, the R system is used to analyze a snapshot of 5 years of information of a fee database containing the clinical history of 38,153 heart illness patients. Two main problems were investigated. The first one involves learning about the dynamic of the course of patients clinical interventions. The second problem concerns assessing the effect of the first clinic of intervention. This analysis illustrates statistical data mining with R. The course of patients interventions is reduced to a multivariate contingency table with transition frequencies in the cells. Log-linear modelling is applied to analyze a Markov dependency structure and to assess baseline effects. This analysis shows that stationary Markov models, commonly used in health-care evaluation could be an over-simplified modelling approach.

KEYWORDS: data mining, health-care evaluation, Markov chains, log-linear models, over-dispersion.

[*]Coordination Center for Clinical Trials, Heinrich-Heine University Duesseldorf
[†]E-mail: pabloemilio.verde@uni-duesseldorf.de
[‡]Coordination Center for Clinical Trials, Heinrich-Heine University Duesseldorf
[§]Institute for Integrative Care in Medicine - University of Giessen
[¶]Public Health Program, Heinrich-Heine University Duesseldorf