

Prezado Editor,

Muito interessante a abordagem do artigo “Vinculação Determinística de Bancos de Dados sobre Mortalidade por Aids” de Bronhara B et al. *Rev Bras Epidemiol* 2008; 11(4): 709-13.

Apesar dos objetivos dos autores terem sido avaliar o desempenho da vinculação determinística em bancos de dados e estimar a cobertura de cada banco, uma sutil informação esta contida no artigo que acredito ser fundamental para o futuro de trabalhos na área de epidemiologia brasileira, a qualificação da informação.

Podemos dividir a qualificação da informação em três partes:

- *Qualidade*: ato de preencher adequadamente cada atributo em um sistema de coleta de informação;
- *Vocabulário*: utilização de um vocabulário padronizado e único para retratar a informação em todos os registros em um sistema de coleta de informação. Como exemplos podemos citar, a codificação de um determinado diagnóstico, procedimentos realizados, forma de identificar o sexo, entre outros.
- *Identificação inequívoca do paciente*: devemos insistir e o mais rápido possível nesta identificação. Não me refiro aqui a deixar dados pessoais do indivíduo acessíveis a todos, mas sim uma identificação que permita acompanhar todo o tratamento dispensado a ele.

Tendo a qualificação da informação adequada teremos a possibilidade de tornar os estudos epidemiológicos retrospectivos simples e baratos, o que possivelmente poderão auxiliar na determinação de futuras políticas de saúde públicas.

Nosso grupo está trabalhando em um projeto de mineração de 10 anos nas bases de dados (AIH, SIA, SIM, SINASC) do DATASUS, e os grandes desafios estão sendo a limpeza e preparação destes dados para posterior aplicação dos algoritmos de mineração¹.

Um recurso adicional que aplicamos em nomes de pacientes é a pesquisa fonética e desta forma poder reconhecer nomes semelhantes foneticamente, mas diferentes na forma ortográfica. Por exemplo, JOSÉ DA SILVA e JOZÉ DA SILVA ou TOMÁS DOS SANTOS e THOMAZ DOS SANTOS.

Fabio Antero Pires
Instituto do Coração do Hospital
das Clínicas da FMUSP

Referencias

1. Santos RS, Gutierrez MA. MINERSUS - Ambiente Computacional para Extração de Informações para a Gestão da Saúde Pública através da Mineração de Dados do SUS. *Rev Bras Eng Biom* 2008; 24; 77-94.