

Ambiente para extração de informação epidemiológica a partir da mineração de dez anos de dados do Sistema Público de Saúde

FÁBIO ANTERO PIRES

Orientador: Prof. Dr. Marco Antônio Gutierrez

Programa de Cardiologia

RESUMO

PIRES FA. Ambiente para extração de informação epidemiológica a partir da mineração de dez anos de dados do sistema público de saúde [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2011. 186p.

A utilização de bases de dados para estudos epidemiológicos, avaliação da qualidade e quantidade dos serviços de saúde vem despertando a atenção dos pesquisadores no contexto da Saúde Pública. No Brasil, as bases de dados do Sistema Único de Saúde (SUS) são exemplos de repositórios importantes que reúnem informações fundamentais sobre a Saúde. Entretanto, apesar dos avanços em termos de coleta e de ferramentas públicas para a pesquisa nessas bases de dados, tais como o TABWIN e o TABNET, esses recursos ainda não fazem uso de técnicas mais avançadas para a produção de informação gerencial, como as disponíveis em ferramentas OLAP (On Line Analytical Processing) e de mineração de dados. A situação é extremamente agravada pelo fato dos dados da Saúde Pública, produzidos por vários sistemas isolados, não estarem integrados, impossibilitando pesquisas entre diferentes bases de dados. Conseqüentemente, a produção de informação gerencial torna-se uma tarefa extremamente difícil. Por outro lado, a integração dessas bases de dados pode constituir um recurso indispensável e fundamental para a manipulação do enorme volume de dados disponível nesses ambientes e, assim, possibilitar a produção de informação e conhecimento relevantes, que contribuam para a melhoria da gestão em Saúde Pública. Acompanhar o

seguimento de pacientes e comparar diferentes populações são outras importantes limitações das atuais bases de dados, uma vez que não há um identificador unívoco do paciente que possibilite executar tais tarefas. Esta Tese teve como objetivo a construção de um armazém de dados (data warehouse), a partir da análise de dez anos (período de 2000 a 2009) das principais bases de dados do SUS. Os métodos propostos para coleta, limpeza, padronização das estruturas dos bancos de dados, associação de registros ao paciente e integração dos sistemas de informação do SUS permitiram a identificação e o seguimento do paciente com sensibilidade de 99,68% e a especificidade de 97,94%.

Descritores: Relacionamento de registros, Mineração de dados, Armazém de dados, Sistema Público de Saúde, Estudos epidemiológicos.