

Método de Extração de Coortes em Bases de Dados Assistenciais Para Estudos da Doença Cardiovascular

MARIA TEREZA FERNANDES ABRAHÃO

Orientador: Prof. Dr. Marco Antônio Gutierrez
Programa de Cardiologia

RESUMO

Abrahão MT. *Método de extração de coortes em bases de dados assistenciais para estudos da doença cardiovascular [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2016.*

A informação coletada de prontuários manuais ou eletrônicos, quando usada para propósitos não diretamente relacionados ao atendimento do paciente, é chamado de uso secundário de dados. A adoção de um sistema de registro eletrônico em saúde (RES) pode facilitar a coleta de dados para uso secundário em pesquisa, aproveitando as melhorias na estruturação e recuperação da informação do paciente, recursos não disponíveis nos tradicionais prontuários em papel. Estudos observacionais baseados no uso secundário de dados têm o potencial de prover evidências para a construção de políticas em saúde. No entanto, a pesquisa através desses dados apresenta problemas característicos a essa fonte de dados. Ao longo do tempo, os sistemas e seus métodos de armazenar dados se tornam obsoletos ou são reestruturados, existem questões de privacidade para o compartilhamento dos dados dos indivíduos e questões relacionadas ao uso desses dados em um contexto diferente do seu propósito original. É necessária uma abordagem sistemática para contornar esses problemas, onde o processamento dos dados é efetuado antes do seu compartilhamento. O objetivo desta Tese é propor um método de extração de coortes de pacientes para estudos observacionais contemplando quatro etapas: (1) mapeamento: a reorganização de dados a partir de um esquema lógico existente em um esquema externo comum sobre o qual é aplicado o método; (2) limpeza: preparação dos dados, levantamento do

perfil da base de dados e cálculo dos indicadores de qualidade; (3) seleção da coorte: aplicação dos parâmetros do estudo para seleção de dados longitudinais dos pacientes para a formação da coorte; (4) transformação: derivação de variáveis de estudo que não estão presentes nos dados originais e transformação dos dados longitudinais em dados anonimizados prontos para análise estatística e compartilhamento. O mapeamento é uma etapa específica para cada RES e não é objeto desse trabalho, mas foi realizada para a aplicação do método. As etapas de limpeza, seleção de coorte e transformação são comuns para qualquer RES. A utilização de um esquema externo possibilita o uso parâmetros que facilitam a extração de diferentes coortes para diferentes estudos sem a necessidade de alterações nos algoritmos e garante que a extração seja efetuada sem perda de informações por um processo idempotente. A geração de indicadores e a análise estatística fazem parte do processo e permitem descrever o perfil e qualidade da base de dados e os resultados do estudo. Os algoritmos computacionais e os dados são disponibilizados em um repositório versionado e podem ser usados a qualquer momento para reproduzir os resultados, permitindo a verificação, alterações e correções de erros. Este método foi aplicado no RES utilizado no Instituto do Coração – HC FMUSP, considerando uma base de dados de 1.116.848 pacientes cadastrados no período de 1999 até 2013, resultando em 312.469 registros de pacientes após o processo de limpeza. Para efetuar uma análise da doença cardiovascular em relação ao uso de estatinas na prevenção secundária de eventos evolutivos, foi constituída uma coorte de 27.915 pacientes, segundo os seguintes critérios: período de 2003 a 2013, pacientes do gênero masculino e feminino, maiores de 18 anos, com um diagnóstico no padrão CID-10 (códigos I20 a I25, I64 a I70 e G45) e com registro de no mínimo duas consultas ambulatoriais. Como resultados, cerca de 80% dos pacientes tiveram registro de estatinas, sendo que, 30% tiveram registro de estatinas por mais de 5 anos, 42% não tiveram registro de nenhum evento evolutivo e 9,7% tiveram registro de dois ou mais eventos. O tempo médio de sobrevida calculado pelo método Kaplan-Meier foi de 115 meses (intervalo de confiança 95% 114-116) e os pacientes sem registro de estatinas apresentaram uma maior probabilidade de óbito pelo teste log-rank $p < 0,001$. Conclui-se que a adoção de métodos sistematizados para a

extração de coortes de pacientes a partir do RES pode ser uma abordagem viável para a condução de estudos epidemiológicos.

Descritores: 1.Informática Médica 2.Sistemas de Informação Hospitalar 3.Registros Eletrônicos de Saúde 4.Estudos de Coortes 5.Estudos Retrospectivos 6.Mineração de Dados