

Descrição da Proveniência de Dados para Extração de Conhecimento em Sistemas de Informação de Hemoterapia

Fernanda Nascimento Almeida

TESE APRESENTADA AO PROGRAMA
INTERUNIDADES EM BIOINFORMÁTICA DA
UNIVERSIDADE DE SÃO PAULO PARA
OBTENÇÃO DO TÍTULO DE DOUTOR EM
CIÊNCIAS

Programa: Interunidades em Bioinformática
Orientadora: Profa. Dra. Ester Cerdeira Sabino
Co-Orientador: Prof. Dr. João Eduardo Ferreira

Durante o desenvolvimento deste trabalho a autora recebeu auxílio financeiro da
CAPES.

São Paulo, maio de 2012.

Descrição da Proveniência de Dados para Extração de Conhecimento em Sistemas de Informação de Hemoterapia

Esta versão definitiva da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 23/05/2012.

Comissão Julgadora:

- Profª. Dra. Ester Cerdeira Sabino (orientadora) – Faculdade de Medicina da USP
- Profª. Dra. Anna Bárbara de Freitas Carneiro Proietti – Fundação Hemominas
- Profª. Dra. Sandra Fátima Menosi Gualandro – Faculdade de Medicina da USP
- Prof. Dr. Luciano Antonio Digiampietri – EACH – USP Leste
- Prof. Dr. Luciano Vieira de Araújo – EACH – USP Leste

Dedico aos meus amados pais, Marcio e Tereza, pelo esforço, empenho e dedicação que sempre depositaram em mim.

Acreditar e amar faz todo e qualquer ideal se tornar possível.

Agradecimentos

Agradeço aos meus orientadores, os professores João Eduardo Ferreira e Ester Cerdeira Sabino, por terem compartilhado um pouco do seu conhecimento comigo e por esta oportunidade.

Agradeço a professora Gisela Tunes da Silva (IME-USP) que supervisionou toda a parte de análise descritiva dos dados e classificação e, ajudou a fazer a análise de sobrevivência sugerida nesta tese que foi usada como resultado adicional. Agradeço também aos colaboradores deste trabalho, os professores Alfredo Mendrone Júnior (Fundação Pró-Sangue) e George Schreiber (Fundação Pró-Sangue), pelas contribuições e sugestões feitas ao longo do desenvolvimento desta tese.

Agradeço aos professores Anna Bárbara Proietti (Hemoninas), Sandra Gualandr (Faculdade de Medicina-USP), Luciano Digiampietri (EACH-USP) e Luciano Araújo (EACH-USP), pelos inúmeros comentários, sugestões, correções e ideias dadas que auxiliaram a tornar esta versão final da tese mais detalhada.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado, a Fundação Pró-Sangue pelo fornecimento dos dados e ao programa de Pós-Graduação Interunidades em Bioinformática da USP pela oportunidade.

Agradeço a Patrícia Martorelli, pela presteza, profissionalismo, dedicação e carinho sempre oferecidos.

Agradeço aos amigos Pedro Paulo da Silva e Kelly Braghetto pela ajuda, importantes críticas e incentivo dados durante o período de desenvolvimento desta tese. Deixo meus agradecimentos aos amigos, Eloiza Rezende, Daniel Cordeiro, Pedro Losko e aos colegas do Laboratório de Banco de Dados do IME pelos momentos de descontração e apoio.

Não poderia esquecer de agradecer a minha irmã, Caroline N. Almeida, o meu sobrinho Victor Hugo Almeida, por vibrarem a cada mínima conquista e, ao Saul Leite, por estar sempre ao meu lado nos momentos em que mais precisei. Por fim, gostaria além de dedicar agradecer aos meus pais, por sempre estarem ao meu lado apoiando todas as decisões que tomei até chegar aqui.

Resumo

O Hemocentro São Paulo é responsável por manter um banco de dados com informações sobre cada doação ou tentativa de doação de sangue. No entanto, os dados desse banco de dados não possuem a qualidade requerida pelas ferramentas/técnicas de análise. Por essa razão, fica difícil utilizar tais dados para estabelecer relações sistemáticas entre as variáveis armazenadas. A principal contribuição desta tese é a descrição da proveniência para atributos selecionados usando critérios de classificação definidos por especialistas. Este trabalho mostra que é possível fazer investigações detalhadas usando a descrição dos dados sem a necessidade de alterar a estrutura do banco de dados.

Durante o período de 1996 a 2006, 1.469.505 doadores foram responsáveis por mais de 2.8 milhões de doações. Após a descrição da proveniência, foram obtidos 252.301 doadores do sexo masculino e 133.056 doadores do sexo feminino e que atenderam aos critérios de inclusão usados nesta tese. Dos 385.357 doadores incluídos na análise, 21.954 (5,7%) tiveram suas doações adiadas devido a seus baixos níveis de hematócrito, 3.850 (1,5%) eram do sexo masculino e 18.104 (13,6%) do sexo feminino. Os resultados obtidos demonstram que, embora os intervalos de espera entre as doações de sangue sejam grandes entre os doadores do sexo feminino e masculino, as mulheres são recusadas mais cedo, por risco de desenvolver anemia, do que os homens.

Aproximadamente 12,84% das mulheres e 1,21% dos homens desenvolveriam hematócrito baixo antes da sétima doação. Os dados sugerem que indivíduos com baixo nível de hematócrito devem esperar mais tempo antes de executarem a próxima doação. Portanto, é importante compreender se existe uma ligação entre a doação de sangue e a diminuição no nível de hematócrito, a fim de evitar resultados indesejáveis para os doadores de sangue.

O modelo de proveniência apresentado nesta tese não foi definido de acordo com os modelos de proveniência genéricos já implementados. Esta tese apresenta um modelo de proveniência que foi capaz de acrescentar informações semânticas para adquirir conhecimento de um experimento *in silico*. Um dos principais objetivos foi desenvolver uma abordagem baseada em declarações, tentando responder a importantes questionamentos biológicos. O modelo descrito combina ricas informações em cada processo usando declarações, e se baseia no conhecimento de especialistas. Esta tese também utilizou estatística descritiva e Análise de Sobrevivência. Finalmente, com a validação do modelo em um domínio conhecido, é pretendido expandir esse método para outros sistemas de informação voltados para hemoterapia.

Palavras-chave: proveniência de dados, experimentos *in silico*, anemia, doadores de sangue, doação de sangue total.

Abstract

The São Paulo Blood Center is responsible to maintain a database with information on each donation. However, this database does not have the quality required by techniques of analysis. For this reason, it is difficult to use it directly to establish systematic relationships between the variables. The main contribution of this paper is a provenance description of attributes selected using classification criteria defined by specialists. We show that it is possible to make detailed investigations using the data description without the need to change the structure of the database.

During 1996 – 2006, 1,469,505 donors were responsible for more than 2.8 million of donation. After the provenance description, we obtained 252,301 male and 133,056 female that met our inclusion criteria. Of the 385,357 donors included in the analysis, 21,954(5.7%) were deferred due to low hematocrit, 3,850(1.5%) were males and 18,104(13.6%) were females. Our results show that, although the intervals between donations for female and male donors are wider, women presented anemia earlier than men. Approximately 12,84% of the females and 1,21% of the males would develop low hematocrit before the 7th donation. Our data suggest that individuals with low hematocrit level should wait longer before the next donation. Therefore, it is important to understand if there is a connection between blood donation and decrease in hematocrit level in order to prevent undesirable outcomes to blood donors.

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

