

1 Introdução

Esta tese propõe uma metodologia, baseada em procedimentos quantitativos, para estimação do risco de evento adverso (recaída ou morte) em crianças portadoras de Leucemia Linfoblástica Aguda (LLA). A metodologia proposta (Macrini *et al*, 2003a, 2003b) foi implementada em um grupo de crianças diagnosticadas no Setor de Hematologia do Instituto de Puericultura e Pediatria Martagão Gesteira (IPPMG) da UFRJ e no Serviço de Hematologia Hospital Universitário Pedro Ernesto (HUPE) da UERJ.

A estimação do risco de evento adverso foi realizada através de um modelo de Redes Neurais após uma seqüência de procedimentos de pré-tratamento de variáveis. Este tratamento é fundamental uma vez que o número reduzido de amostras é uma característica intrínseca deste problema, dada a raridade da doença, de aproximadamente 1 caso por 100 mil habitantes por ano.

1.1. Tratamento Direcionado ao Risco de Recidiva na Leucemia Linfoblástica Aguda na Infância

Sob o termo Leucemia Linfoblástica Aguda (LLA) há um conjunto de doenças neoplásicas que correspondem a 80% das leucemias da infância e a cerca de 30% dos casos de câncer até os 15 anos de idade. A LLA é um tipo de câncer proveniente das células precursoras de linfócitos, que infiltram a medula óssea, levando a uma diminuição da produção das células sanguíneas normais. É a neoplasia mais freqüente da infância. Nos Estados Unidos da América, a taxa de incidência anual é maior nos primeiros cinco anos de vida, com um pico em torno de quatro anos de idade, onde se encontra uma taxa de 4,4 casos/100.000 crianças. (Kersey, 1997; Rivera *et al*, 1993; Greaves, 1993; Azevedo, 2003). A Leucemia Linfoblástica Aguda pode ocorrer tanto em crianças como em adultos, mas o comportamento da doença e o tratamento são diferentes para cada um dos grupos.

Por esta razão, esta tese se ocupará de casos de LLA apenas em crianças abaixo de 12 anos.

Os passos decisivos no desenvolvimento do tratamento da LLA são a estratificação e a orientação da intensidade terapêutica de acordo com o risco de recaída correspondente (Riehm, 1995). Os estudos iniciados nos anos 70 e consolidados nos anos 80 contribuíram para a estratificação dos pacientes de acordo com um conjunto de fatores prognósticos. Esta estratégia permite fornecer esquemas quimioterápicos adequados às variáveis clínicas e laboratoriais de cada paciente, evitando toxicidade excessiva para os casos de bom prognóstico e o tratamento insuficiente para os casos de mau prognóstico. A avaliação do risco de recidiva é a etapa mais importante na estratégia do tratamento após a realização do diagnóstico, já que a intensidade do tratamento será estabelecida essencialmente a partir do risco de recidiva. (Rivera *et al*, 1993).

Existem diversas formas utilizadas atualmente para estratificar o risco de recidiva. Em geral, os protocolos de tratamento da LLA da infância estratificam seus pacientes em dois ou três grupos de risco. Pacientes com baixo risco de recidiva podem receber tratamento com menor toxicidade do que pacientes com risco alto ou intermediário. Alguns dos pacientes classificados como alto risco para recidiva podem ainda ser considerados para transplante de medula óssea em primeira remissão completa, por não responderem a esquemas quimioterápicos intensivos (Riehm, 1995).

Na proposta do Workshop de 1993 (Smith, 1996) promovido por The Cancer Therapy Evaluation Program (CTEP) do National Cancer Institute (NCI) dos EUA, houve consenso entre os grupos participantes (Children's Cancer Group (CCG), o Pediatric Oncology Group (POG), o Dana Farber Cancer Group e o St. Jude Children's Research Hospital) de que a contagem de leucócitos maior do que $50.000/\text{mm}^3$ no momento do diagnóstico e a idade menor do que 1 ano ou maior do que 10 anos conferem um mau prognóstico, esquematizados no quadro 01.

Quadro 01 Classificação Prognostica do National Cancer Institute para Leucemia Linfoblástica Aguda

Fatores Prognósticos de acordo com NCI	Sobrevida Livre de Eventos
Idade >1 ano ou <10 anos e ou Contagem de leucócitos <50000/ μ l	80%
Idade <1 ano ou >10 anos e ou Contagem de leucócitos >50000/ μ l	65%

Apesar da proposta ser interessante pela facilidade com que se pode dispor destes dois parâmetros, as taxas de recidiva não iriam provavelmente baixar dos patamares atuais. Isto porque o poder de predição da combinação das duas variáveis poderia ser dito insuficiente, já que a proposta não incorpora várias informações fundamentais do ponto de vista do prognóstico (Costa, 2003).

O sistema de estratificação de risco proposto em 1995 pelo grupo alemão para tratamento de LLA na infância - BFM95- é mais complexo, utilizando outras informações além de leucometria global e idade, além de estratificar os pacientes em 3 grupos de risco, como descrito no quadro 02 a seguir (Riehem, 1995). Para cada grupo de risco definido por estes critérios há um esquema terapêutico distinto.

Quadro 02 Classificação Prognostica do Protocolo BFM95 para Leucemia Linfoblástica Aguda (LLA)

Fator Prognóstico	Idade	Contagem global de leucócitos ao diagnóstico	Imunofenotipo T	Resposta ao corticóide no 8º dia de tratamento	Citogenética t(9;22) ou t(4;11)	Remissão completa no 33º dia de tratamento
Grupo de risco						
Risco Padrão	>1 e < 6 anos	<20000/ μ l	Ausente	< 1000	Ausente	SIM
Risco Médio	<1 ou >6 anos	>20000/ μ l	Presente	< 1000	Ausente	SIM
Alto Risco	independe	independe	independe	> 1000	Presente	NÃO

Em 2002, o grupo alemão para o tratamento de LLA propôs um novo estudo terapêutico chamado ALL IC-BFM2002, incluindo além de modificações no tratamento propriamente dito, modificações na estratificação do risco, utilizando

uma avaliação da medula óssea no 15º dia de tratamento, além dos mesmos parâmetros do protocolo BFM95 citados acima (Schrappe *et al*, 2003). Entretanto, os resultados deste protocolo ainda não foram publicados, dado o curto tempo de sua implementação. Os dados utilizados nesta tese provém de pacientes tratados com o protocolo BFM95.

Apesar dos avanços na classificação de risco dos portadores de LLA na infância, cerca de 70% dos pacientes são atualmente classificados como risco intermediário de recidiva. Neste grupo de risco intermediário, cerca de 35% dos pacientes irá recidivar. Portanto, dentro do chamado grupo intermediário de recidiva há um conjunto heterogêneo de pacientes no que diz respeito ao risco real de recaída (Riehm, 1995). Isto indica a necessidade de novas estratégias de classificação de risco de recidiva, sobretudo que discriminem melhor os pacientes a que atualmente atribuí-se um risco intermediário de recidiva.

1.2. Objetivo

O objetivo principal desta tese é propor, implementar e analisar uma metodologia, baseada em um criterioso processo de seleção de variáveis e em um modelo não linear, com a finalidade de estimar o risco de evento adverso em crianças portadoras de leucemia linfoblástica aguda.

1.3. Contribuições da Tese

Principais:

- Resolver um problema complexo de modelagem com dados reais. Este problema é particularmente difícil por causa: 1) Do número reduzido de casos (intrínseco a este tipo de problema); 2) Do número elevado de possíveis variáveis de entrada, e 3) Da sua característica fortemente interdisciplinar;
- Propor uma nova metodologia para estimação de risco de evento adverso em pacientes com leucemia linfoblástica aguda. Esta metodologia envolve um pré-tratamento específico dos dados e uma modelagem através de Redes Neurais. Esta metodologia se mostrou:

- 1) Eficiente do ponto de vista de performance e do seu poder de discriminação (há pouquíssimos resultados na região de fronteira);
- 2) Mais adequada quando comparada a métodos lineares;
- 3) Capaz de gerar resultados interpretáveis e coerentes do ponto de vista médico.

Esta metodologia pode possivelmente ser estendida, obviamente com adaptações, a outros problemas com dados médicos, especialmente (mas não exclusivamente) àqueles que tenham muitas variáveis e pouca quantidade de dados disponível.

Específicas:

- Propor uma associação específica de várias técnicas quantitativas que geram uma solução adequada para o problema proposto;
- Propor refinamentos do modelo projetado especificamente para o problema desta tese.
- Implementar a metodologia proposta utilizando, devido ao número reduzido de amostras, o método “*leave-one-out*” para obter resultados fora-da-amostra.
- Realizar experimentos numéricos testando diversas situações da modelagem e do tratamento de dados propostos para o banco de dados utilizado na tese.
- Analisar e interpretar os resultados obtidos.

1.4. Organização da Tese

Esta tese tem um caráter fortemente multidisciplinar. Na tentativa de contribuir para uma melhor solução de um problema na área médica, lança-se mão de diversas técnicas e conceitos quantitativos, advindo das áreas de estatística e inteligência computacional (*machine learning*), muito destes bem estabelecidos na literatura. O texto da tese será, possivelmente, do interesse de leitores com níveis diversos de familiaridade com estas técnicas. Por esta razão, o desenvolvimento técnico vem exposto em diversos apêndices, ao invés de estar no corpo principal da tese. Embora importante para que o texto seja razoavelmente autocontido, a leitura destas partes mais técnicas fica assim facultada, já que seria possivelmente

árdua a um grupo de leitores e desnecessária a outros. Desta forma, o corpo principal da tese dirige o leitor de forma direta através das reais contribuições da mesma.

Além desta introdução, este documento encontra-se dividido em mais 3 capítulos. No capítulo 2 são apresentados o banco de dados e a metodologia proposta para estimação do risco de evento adverso. No capítulo 3 são apresentados os resultados e análises. Já no capítulo 4 encontram-se as conclusões. Esta tese inclui ainda além das referências bibliográficas, cinco anexos: No anexo A, são apresentados os aspectos técnicos do algoritmo de seleção de variáveis utilizado. O anexo B é dedicado às Redes Neurais. Nos anexos C e D faz-se uma descrição do modelo de Análise Discriminante e de Regressão Logística, respectivamente. Por fim, no anexo E encontra-se a relação das variáveis que compõem o banco de dados.