

Mineração de dados e características da mortalidade infantil

Data mining and characteristics of infant mortality

Rossana Cristina Xavier Ferreira Vianna ^{1,2}
 Claudia Maria Cabral de Barra Moro ²
 Samuel Jorge Moysés ^{2,3}
 Deborah Carvalho ^{2,4}
 Julio Cesar Nievola ²

Abstract

This study aims to identify patterns in maternal and fetal characteristics in the prediction of infant mortality by incorporating innovative techniques like data mining, with proven relevance for public health. A database was developed with infant deaths from 2000 to 2004 analyzed by the Committees for the Prevention of Infant Mortality, based on integration of the Information System on Live Births (SINASC), Mortality Information System, and Investigation of Infant Mortality in the State of Paraná. The data mining software was WEKA (open source). The data mining conducts a database search and provides rules to be analyzed to transform the data into useful information. After mining, 4,230 rules were selected: teenage pregnancy plus birth weight < 2,500g, or post-term birth plus teenage mother with a previous child or intercurrent conditions increase the risk of neonatal death. The results highlight the need for greater attention to teenage mothers, newborns with birth weight < 2,500g, post-term neonates, and infants of mothers with intercurrent conditions, thus corroborating other studies.

Database; Infant Mortality; Information Systems; Artificial Intelligence; Epidemiologic Surveillance

Introdução

Devido à vulnerabilidade da faixa etária de menores de um ano, a mortalidade infantil tem sido considerada importante indicador para medir a situação de saúde e a qualidade das ações de saúde de uma população ^{1,2,3,4}. Nesse sentido, importa para este trabalho registrar três importantes iniciativas relacionadas ao monitoramento, controle e avaliação da mortalidade infantil. A implantação, no Brasil, em 1975, do Sistema de Informações sobre Mortalidade (SIM); em 1990, a definição do Sistema de Informações sobre Nascidos Vivos (SINASC) e, em 2000, a criação, no Estado do Paraná, do Sistema de Investigação da Mortalidade Infantil (SIMI) permitiram a geração de grande quantidade de dados, que geralmente são subutilizados ⁵. O SIMI é utilizado no Paraná, desde sua criação, para facilitar a análise dos dados de mortalidade infantil pelos Comitês de Prevenção da Mortalidade Infantil Estadual, regionais e municipais ⁵.

Com a criação dos Sistemas de Informação em Saúde (SIS), técnicas como a Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Database – KDD*) têm se mostrado efetivas na função de gestão de informações ⁶. O KDD é um processo composto pela seleção de dados, pré-processamento, transformação dos dados e estabelecimento de padrões úteis na extração de conhecimento, ou seja, tradução de dados brutos em informações relevantes. Uma

¹ Secretaria do Estado da Saúde do Paraná, Curitiba, Brasil.

² Programa de Pós-graduação de Tecnologia em Saúde, Pontifícia Universidade Católica do Paraná, Curitiba, Brasil.

³ Secretaria Municipal de Saúde de Curitiba, Curitiba, Brasil.

⁴ Instituto Paranaense de Desenvolvimento Econômico e Social, Universidade Tuiuti, Curitiba, Brasil.

Correspondência

R. C. X. F. Vianna
 Secretária do Estado da Saúde do Paraná.
 Rua Piquiri 170, Curitiba, PR 80230-140, Brasil.
 rossanacxfv@yahoo.com.br
 rossanavianna@sesa.pr.gov.br

das etapas do processo de KDD é o *Data Mining* (DM), mineração de dados propriamente dita, que consiste da interpretação de padrões e da geração de conhecimento após a análise dos resultados obtidos⁶. Permite, por exemplo, examinar as relações de similaridade entre as informações⁷. O DM é uma área emergente dentro da inteligência computacional usada na análise de grandes bancos de dados⁸, com a geração de padrões e a extração de informações dessas bases⁹.

Por referência a grandes bancos de dados, é importante assinalar que, em média, no Paraná, são registrados anualmente, no SINASC, 150 mil nascidos vivos, com mais ou menos 2 mil óbitos infantis registrados no SIM ao ano e 1.500 óbitos infantis analisados por ano no SIMI.

Nesse cenário, este trabalho tem por objetivo relatar a experiência da integração das bases de dados de três diferentes sistemas de informação (SINASC/SIM e SIMI) no período de 2000 a 2004; demonstrar a viabilidade e a possibilidade de replicação em saúde pública da Mineração de Dados (MD), retirando, dessa forma, a subjetividade/viés do analisador, e verificar se é possível utilizar de modo produtivo o resultado obtido com as regras geradas pela ferramenta utilizada¹⁰.

Métodos

A população de estudo na presente pesquisa totalizou 7.256 óbitos infantis, equivalendo a 77% do total de óbitos analisados pelo Comitê Estadual de Prevenção da Mortalidade Infantil (CEPMI) do Paraná (incluindo Comitês municipais e regionais), no período de 2000 a 2004; este, por sua vez, analisou 9.372 óbitos, que correspondem a 66% do total de óbitos ocorridos no estado no período estudado. Os 23% que foram excluídos são dados de crianças que nasceram antes de 2000 e, por isso, suas informações estão na base do SINASC e/ou SIM de 1999, ano que não foi considerado para a integração das bases no presente estudo. Também não foram consideradas as crianças que nasceram em 2004, mas que tiveram o óbito analisado pelos Comitês em 2005. Além disso, existiam algumas informações incompletas, tais como: dados repetidos (0,1%) e falta de dados em uma das bases dos sistemas (1%).

O período de tempo foi escolhido a partir da criação do SIMI, em 2000, quando os registros das análises de óbitos infantis começaram a ser coletados sistematicamente, até 2004, período em que os dados coletados e analisados já estavam previamente consolidados, durante a realização deste trabalho, nos três SIS escolhidos.

As informações constantes no SIMI foram catalogadas após a investigação dos óbitos infantis com informações constantes no SINASC, no SIM, nas entrevistas domiciliares e nas análises realizadas pelos Comitês de Prevenção da Mortalidade Infantil. As informações deste sistema, portanto, foram recuperadas e/ou modificadas após todo o processo de investigação dos óbitos. Por exemplo, se a variável peso ao nascer não existia no SIMI, mas existia no SINASC, era completada na nova base com o valor referente ao SINASC.

O processo de KDD envolve o uso dos dados escolhidos, neste caso, nas bases dos SIS do Governo, SINASC, SIM e SIMI, selecionados de janeiro de 2000 a dezembro de 2004. Essas bases foram pré-processadas e transformadas em base única para o DM. Nos casos de dados duvidosos, os óbitos faziam parte dos critérios de exclusão, relatados anteriormente.

A próxima fase do KDD utilizada foi o pré-processamento, quando foram categorizadas, isto é, definidas classes para determinados atributos ou variáveis em algumas informações, como a idade e o apgar. A idade do óbito, por exemplo, foi dividida em quatro categorias: perinatal (óbitos de 22 semanas de gestação até 6 dias após o nascimento), neonatal precoce (de 1 a 6 dias após o nascimento), neonatal tardio (de 7 a 27 dias após o nascimento) e tardio ou pós-neonatal (entre 28 e 360 dias de nascido), utilizando os dados do SIM. Essa preparação foi estabelecida conforme a necessidade apontada por três especialistas em mortalidade infantil das secretarias estadual e do Município de Curitiba, os quais fazem parte do CEPMI.

Durante a fase de pré-processamento, também foi realizada a limpeza dos dados para o DM, onde foram corrigidas, por exemplo, datas de óbitos que ocorreram antes do nascimento, com a realização da troca destas para utilização posterior desses dados¹¹. Também foi criado um novo atributo indicando a diferença das causas dos óbitos segundo a Classificação Internacional de Doenças, em sua décima revisão (CID-10). Observou-se o que foi alterado em capítulos (tendo 21 capítulos na CID-10) e categorias, que variam em número segundo o capítulo e subcategorias ou classes¹².

Na etapa seguinte do processo, que consistia na mineração de dados propriamente dita, foram extraídas regras que permitem relacionar as diversas variáveis sob estudo. Para sua realização, optou-se pelo uso do algoritmo J48 do WEKA (*Waikato Environment for Knowledge Analysis*; <http://www.cs.waikato.ac.nz/ml/weka>), que permite a descoberta de padrões na forma de um classificador representado como árvore de decisão, constituindo uma versão do algoritmo

C4.5, que representa as relações entre as variáveis no formato de uma árvore de decisão, a partir da qual podem ser lidas as regras. Uma árvore de decisão é um padrão tanto na área acadêmica, quanto comercial, e consiste em um conjunto hierarquicamente ordenado de conceitos, que iniciam em um nó raiz, sendo cada conceito dividido recursivamente em subconceitos no próximo nível mais inferior da árvore. Cada elemento terminal de uma árvore (conhecido como nó folha) está associado a um conceito objetivo.

Trabalhou-se com o programa WEKA, para a obtenção das regras, em virtude da sua adequação ao estudo e por três importantes razões: (a) é uma ferramenta amigável ao uso por profissionais de saúde, muitas vezes não afeitos à informática; (b) por ser *software* livre, pode ser utilizado sem custo e com grande facilidade nas secretarias de saúde; (c) é eficaz, rápido e efetivo. Foi utilizada a tarefa de classificação, que verifica se os fatores são causais e como eles influenciam o atributo-meta (variável escolhida), para observar a ocorrência de novas características relacionadas à mortalidade infantil.

Como o processo de DM é iterativo, foram realizados vários experimentos tendo como atributos-meta a evitabilidade dos óbitos e as diferenças entre a causa do óbito na declaração de óbito e após as análises dos Comitês. Durante esse processo, o estudo foi avaliado por um painel de especialistas, que processou as escolhas das melhores opções das variáveis submetidas à mineração. Esses atributos foram escolhidos baseados nos objetivos principais dos Comitês, que são a caracterização da evitabilidade dos óbitos e a especificação da causa do óbito, de maneira que se possa atuar em sua redução. Foi realizada a tarefa de mineração da classificação, a qual verifica se os fatores são causais e como eles influenciam o atributo-meta, para observar a ocorrência de novas características relacionadas à mortalidade infantil. A classificação utiliza dados do passado (processo designado como “treinamento”) para induzir regras para o futuro, isto é, regras para prever o valor do atributo-meta, objetivamente, predizem o futuro de um problema não determinístico. Na classificação, ocorre a previsão (prevê-se a possibilidade de o óbito ocorrer e com qual padrão de atributos está relacionado) de uma classe de valor, que utiliza o atributo-meta com outros atributos como preditores (como, por exemplo, o baixo peso ao nascer e a baixa escolaridade da mãe). A técnica da classificação indica os padrões de atributos que se encontram no caminho causal para o desfecho do óbito infantil no caso em estudo.

Após a tarefa de mineração de dados, 22 especialistas dos Comitês Regionais de Saúde do

estado realizaram a avaliação dos resultados, buscando relacionar os padrões encontrados no DM com a literatura existente e com suas próprias experiências em analisar óbitos infantis. Com essa interpretação, gerou-se a confirmação dos conhecimentos existentes na literatura sobre a mortalidade infantil.

O primeiro e o segundo atributos estudados foram os códigos da CID-10 na declaração de óbito. Após as análises dos Comitês, os atributos classificados em grupos por causa do grande número de possibilidade de diagnóstico, facilitando, assim, a posterior interpretação das regras geradas pelo DM. O terceiro atributo analisado e escolhido como atributo-meta foi se o óbito era evitável ou não. O quarto atributo foi a classificação da evitabilidade desse óbito, baseando-se na classificação já estabelecida e padronizada pelos Comitês para essa análise (sendo nove grupos baseados na sua redutibilidade, por exemplo: óbitos reduzíveis por imunoprevenção, por saneamento básico). E assim sucessivamente, conforme observado na Tabela 1.

Após a geração de 4.230 regras pelo DM, as 20 melhores foram escolhidas pelos três especialistas inicialmente selecionados e apresentadas aos 22 especialistas das 22 Regionais de Saúde do estado. Estes selecionaram as quatro regras de maior relevância e frequência e com menor taxa de erro selecionada pela ferramenta. A avaliação objetiva das regras foi realizada pelo WEKA. A própria ferramenta realiza a validação cruzada para estabelecer o aparecimento da regra e já faz o conjunto de treinamento e teste. São geradas algumas medidas de suporte, como a porcentagem de casos onde a regra apareceu, e medidas de confiança, como a porcentagem de acerto das regras, além da taxa de erro e a consistência destas. Após a avaliação, para cada regra descoberta, a partir de medidas estatísticas, o WEKA atribui valores, tais como: cobertura, taxa de acerto e erro, sensibilidade, especificidade e precisão.

Resultados e análises estatísticas da mineração de dados

Foram escolhidas quatro regras relacionadas aos óbitos evitáveis para os especialistas avaliarem e validarem, as quais não apresentaram qualquer taxa de erro e apareceram em pelo menos 29% dos casos, isto é, as mais frequentes. Também foram selecionadas seis regras relacionadas aos óbitos não evitáveis, as quais apareceram em, no máximo, 12% dos casos, isto é, as mais raras. Pensou-se em validar os extremos, porque, dessa forma, é possível estabelecer a validade da utilização dessa técnica não convencional de identificação

Tabela 1

Nome, descrição e origem dos atributos estudados.

Nome dos atributos	Descrição dos atributos	Origem dos atributos nos SIS
ClasseCodCID	Código da CID na declaração de óbito	Base do SIMI
ClasseCodCIDInv	Código da CID após investigação	Base do SIMI
CodTipoMorteEvitavel	Código do tipo de morte evitável ou não	Base do SIMI
ClasseCodEvitabilidade	Código de evitabilidade por grupos A até I, já categorizados no SIMI	Atributo-meta, base do SIMI
analiseobitoSEXO	Sexo: masculino, feminino, ignorado	Base do SIMI e confirmado na base do SIM
ClasselDADEMAE*	Categorização da idade da mãe (anos): < 16, 16-35 e > 35	Base do SIM e do SINASC
ESTCIVMAE	1. Solteiro, 2. Casado, 3. Viúvo, 4. Separado judicialmente, 5. União consensual, 9. Ignorado	Base do SINASC
ESCMAE	1. Nenhuma, 2. 1-3, 3. 4-7, 4. 8-11, 5. 12 e mais, 9. Ignorado	Base do SIM e do SINASC
QTDFILVIVO	Quantidade de filhos vivos	Base do SIM e do SINASC
QTDFILMORT	Quantidade de filhos mortos	Base do SIM e do SINASC
GESTACAO *	Categorizado em (semanas): 1. < 22, 2. 22-27, 3. 28-31, 4. 32-36, 5. 37-41, 6. 42 e mais, 9. Ignorado	Base do SIM
GRAVIDEZ	1. Única, 2. Dupla, 3. Tripla e mais, 9. Ignorada	Base do SIM e do SINASC
PARTO	1. Vaginal, 2. Cesário, 9. Ignorado	Base do SIM e do SINASC
CONSULTAS	Categorizado em: 1, 2-4, 5-6, +6, ignorado e sem preenchimento	Base do SINASC
ClasseAPGAR1 *	Nota de viabilidade dada no primeiro minuto de vida. Categorizado em: 0-5 e 6-10	Base do SINASC
ClasseAPGAR5 *	Nota de viabilidade dada no quinto minuto de vida. Categorizado em: 0-5 e 6-10	Base do SINASC
RACACOR	1. Branca, 2. Preta, 3. Amarela, 4. Parda, 5. Indígena	Base do SIM e do SINASC
ClassePESO *	Categorizado em gramas, ao nascer: < 2.500, 2.500-3.500, > 3.500 e sem preenchimento	Base do SINASC
IdadeClasse *	Categorização da idade (dias) no óbito: < 1, 1-6, 7-28 e 29-360	Base do SIM
Alteração da CID *	Capítulo: se alterou capítulo da CID-10; Grupo: se alterou o grupo da CID-10; Categoria: se alterou a categoria da CID-10; Subcategoria: se alterou a subcategoria da CID-10 **	Gerado para o trabalho e baseado no SIMI, escolhido também como atributo-meta

Nota: Os títulos dos nomes dos atributos estão segundo a padronização do SIMI.

CID: Classificação Internacional de Doenças; SIM: Sistema de Informações sobre Mortalidade; SIMI: Sistema de Investigação da Mortalidade Infantil; SINASC: Sistema de Informações sobre Nascidos Vivos; SIS: Sistemas de Informações em Saúde.

* Categorias estabelecidas para facilitar a mineração de dados;

** Fonte: Moro 12.

de características relacionadas à mortalidade infantil no máximo de acertos, validando também o SIMI e os erros, para poder estabelecer, como resultado deste trabalho, a melhoria das análises baseadas na utilização da ferramenta proposta.

Quando o atributo-meta escolhido foi a evitabilidade dos óbitos, foram geradas 441 regras, com 78% classificadas corretamente, com precisão de 83% para identificação dos evitáveis e 62% para os não evitáveis; com uma sensibilidade de

80% para identificação dos evitáveis e uma especificidade de 62% para identificação dos não evitáveis. A taxa de erro para os evitáveis foi de 12% e, para os não evitáveis, de 44%.

Com o atributo-meta comparando a CID-10 existente na declaração de óbito e após as investigações, as instâncias (ou número de casos utilizados no DM) foram apenas as que sofreram alteração da CID-10 após a investigação dos Comitês, ou seja, 4.230 (58%) casos dos 7.256. Assim,

foram geradas 169 regras com 88% de acertos, precisão de 93%, sensibilidade de 88% e especificidade de 49%, sendo consideradas taxas de erro de 0% em sete regras. Nas outras duas regras, foram consideradas taxas de erro de até 3% nos casos de maior frequência de aparecimento, um valor menor que a taxa de erro gerada na matriz de confusão e que não ultrapassou 27%.

Discussão

Após a integração dos três SIS, confirmaram-se características observadas e relatadas anteriormente pelos Comitês de Prevenção da Mortalidade Infantil^{5,10}. A integração e relacionamento (*linkage*) de bases de dados diferentes auxiliam a complementar, rever e recuperar informações¹³. Apresentando-se como uma estratégia adequada para aprimorar a qualidade destas, permite uma aproximação da situação epidemiológica real existente na população estudada^{14,15,16}.

Observou-se a importância de dados coletados corretamente, justificados e criticados para que se possa trabalhar com as bases de dados sem a necessidade de dispensar muito tempo na preparação destes, para uma boa mineração de dados^{11,17} e para uma boa atuação dos Comitês de Prevenção de Mortalidade Materna e Infantil¹⁸. Essa técnica é importante ferramenta para utilização em grandes bancos de dados de saúde pública, visto já ser utilizada em igualmente grandes bancos de dados da Internet¹⁹.

Com a análise das regras geradas pelo DM, obteve-se o perfil da mortalidade infantil no Estado do Paraná, no período de 2000 a 2004. Desse modo, após a aplicação do KDD, destaca-se que 55% dos óbitos poderiam ser evitáveis se houvesse uma adequada atenção à gestação, parto e ao recém-nascido. O sexo predominante para óbitos foi o masculino (58%) e a idade prevalente da mãe ficou no intervalo de 16-35 anos (87%), sendo esta casada (39%), com escolaridade de 4-7 anos (40%), sem nenhum filho vivo (41%) e nenhum filho morto (84%), com gestação a termo (47%), gravidez única (91%), parto vaginal (63%) e com mais de seis consultas de pré-natal (39%). Para as mortes infantis, prevaleceu um apgar no 1º minuto > 5 (53%) e apgar no 5º minuto > 5 (71%), sendo a criança de raça/cor branca (91%), com peso ao nascer < 2.500g (56%) e idade do óbito < 1 mês (54%).

Essas características coincidem com as observadas pelas coortes históricas de 1982, 1993 e 2004, para a mortalidade neonatal²⁰. Coincidem também com o uso de *linkage* de sistemas de informação relacionados a estudos de coorte sobre mortalidade neonatal¹⁴, com o estudo de

características relacionadas à mortalidade perinatal e a avaliação da assistência ao parto²¹. No estudo da investigação dos óbitos perinatais e infantis para o planejamento de políticas públicas de saúde, aparecem características semelhantes²².

As características geradas pelo DM são importantes ainda para a construção de mapas de risco. Estes são utilizados como estratégia de ação para a redução dos óbitos, com a consolidação da integralidade da atenção à saúde e com o fortalecimento de ações intra e intersectoriais no território²³.

Os estudos sobre fatores de risco associados à mortalidade fetal e infantil no Brasil, em áreas da Região Metropolitana de São Paulo, bem como em outras localidades, com suas propostas de instrumentos preditivos e recomendações sobre assistência materno-infantil já relatavam características semelhantes aos encontrados pelo DM^{4,24,25,26,27,28,29,30}.

Conclusão

Este trabalho demonstra que é possível utilizar o DM em Saúde Pública, obtendo-se conclusões relevantes. A finalização do DM permitiu inferir que foram geradas importantes regras relativas ao desfecho em estudo e, entre estas, 11 obtiveram mérito de 75% ou mais. Destas, três chamam a atenção pelo que representam:

- regra relacionada à mãe adolescente (< 16 anos) – mesmo quando a criança apresenta bom peso ao nascer (2.500 a 3.500g), a mãe surge como alto critério de risco para o óbito perinatal;
- regra também relacionada à mãe adolescente (< 16 anos) – com escolaridade razoável (4-7 anos), sem nenhum filho morto, mas já com outro filho, apresenta risco para desconforto respiratório do recém-nascido;
- regra que revela que o baixo peso ao nascer (< 2.500g), com pós-datismo (41 semanas de gestação ao nascer) e afetado por afecções maternas, está relacionado ao óbito em 30 dias de vida.

Ainda, uma característica gerada e que foi considerada surpreendente foi a de malformações e sexo feminino relacionadas ao óbito infantil inevitável.

Dessa forma, conclui-se que, para melhorar a atenção aos menores de um ano de idade no Paraná, valendo-se das informações fornecidas pelas bases de dados dos cinco anos selecionados e após o estabelecimento dos padrões gerados pelo DM, devem ocorrer ações voltadas:

- às adolescentes, principalmente as que já têm outro filho;
- às mães com problemas na gestação;

- às mães com filhos de baixo peso ao nascer e com pós-datismo;

Após a avaliação e análise desses resultados, confirmou-se que os instrumentos de análise (SIMI) utilizados pelos Comitês de Prevenção da Mortalidade Infantil no Estado do Paraná são eficazes e podem ser confirmados com outros estudos e com mais essa ferramenta computacional utilizada nesta investigação. Confirma também a necessidade do trabalho conjunto de pediatras, gineco-obstetras e gestores de saúde para, baseados em mais este estudo, subsidiar melhor a assistência prestada às crianças no estado e, futuramente, utilizar outras ferramentas computacionais como sistemas de alertas para os usuários de prestadores de serviços de saúde à população infantil³¹.

Como contribuições deste trabalho, destacam-se:

- a integração de bases reais;
- a importância dessa integração de maneira efetiva, pensando a construção dos SIS, que deve

ser realizada de maneira a fornecer características importantes para utilização no presente e alinhar estudos futuros com a população abrangida no sistema;

- a análise de situações reais, existentes na nova base gerada a partir dos três SIS;
- a análise das regras geradas para mudar o comportamento da mortalidade infantil, melhorando, dessa forma, as características de evitabilidade, já implementadas por este estudo;
- o estímulo à melhora das informações prestadas nas declarações de óbito;
- a percepção de que, para melhorar a aceitação do DM no meio médico³², é necessário o estabelecimento de procedimentos para reduzir os problemas de definição e resultados, apesar de sua boa acurácia preditiva (de 78% a 90%)^{33,34};
- a implantação de um “data warehouse” (armazém de dados) e da Sala de Situação para facilitar a aquisição de dados com base nas características estabelecidas e geradas por este estudo e já em utilização nos serviços.

Resumo

O estudo busca identificar padrões de características materno-fetais na predição da mortalidade infantil, por meio da incorporação de técnicas inovadoras, como a Mineração de Dados, que se mostram relevantes em Saúde Pública. Foi elaborada uma base de dados, com óbitos infantis analisados pelos Comitês de Prevenção da Mortalidade Infantil de 2000 a 2004, a partir da integração dos Sistemas de Informações de Nascidos Vivos, da Mortalidade e da Investigação da Mortalidade Infantil no Estado do Paraná. O programa da mineração foi o WEKA, de uso livre. A mineração faz busca em banco de dados e fornece regras que devem ser analisadas para transformação em infor-

mação útil. Após a mineração, selecionaram-se 4.230 regras, por exemplo: mãe adolescente e peso ao nascer < 2.500g, ou parto pós-termo e mãe adolescente com outro filho, ou com afecções maternas, aumentam o risco para óbito neonatal. Vê-se a necessidade de estabelecer maior atenção às adolescentes, às crianças com peso ao nascer < 2.500g, pós-termo e filhas de mães com afecções maternas, confirmando resultados de outros estudos.

Base de Dados; Mortalidade Infantil; Sistemas de Informação; Inteligência Artificial; Vigilância Epidemiológica

Colaboradores

R. C. X. F. Vianna participou de todas as etapas do estudo e da elaboração do artigo, sendo responsável pela versão final. S. J. Moysés participou do planejamento do estudo, orientação, discussão dos resultados e revisão do artigo. D. Carvalho participou do planejamento do estudo, construção e processamento do banco de dados, análise dos resultados e revisão do artigo. J. C. Nievola participou do planejamento do estudo, discussão dos resultados e revisão do artigo. C. M. C. B. Moro foi responsável pela orientação geral, participando de todas as etapas do trabalho.

Agradecimentos

A Tiago K. Shiono pela construção e processamento do banco de dados.

Referências

1. Ministério da Saúde. Saúde Brasil 2004. Uma análise da Situação de Saúde. Brasília: Ministério da Saúde; 2004.
2. Andrade SM, Soares DA, Matsuo T, Souza RKT, Mathias TAF, Iwakura MLH, et al. Condições de vida e mortalidade infantil no Estado do Paraná, Brasil, 1997/2001. *Cad Saúde Pública* 2006; 22:181-9.
3. Mello-Jorge MHP, Gotlieb SLD, Soboll MLMS, Almeida ME, Latorre MRDO. Registro dos eventos vitais: sua importância em Saúde Pública. *Rev Saúde Pública* 1983; 17:148-51.
4. Costa MCN, Mota ELA, Paim JS, Silva LMV, Teixeira MG, Mendes CMC. Mortalidade no Brasil em períodos recentes de crise econômica. *Rev Saúde Pública* 2003; 37:699-709.
5. Vianna RCXF. Análise da situação da mortalidade infantil no Estado do Paraná, Brasil em 2004. In: *Proceedings of International Congress Pediatrician Especiality, Children* [CD-ROM]. Curitiba: Instituto Pequeno Príncipe; 2005.
6. Fayyad UM, Shapiro GP, Smyth P. From data mining to knowledge discovery: an overview. In: Fayyad UM, Shapiro GP, Smyth P, Uthursamy R, editors. *Advances in knowledge discovery and data mining*. Cambridge: The MIT Press/London: AAAI Press; 1996. p. 1-34.
7. Linares KSC. Aspectos teóricos do data mining: descoberta do conhecimento em medicina [Tese de Doutorado]. Florianópolis: Universidade Federal de Santa Catarina; 2003.
8. Kusiak A, Kern JA, Kernstine KH, Tseng BTL. Autonomous decision making a data mining approach. *IEEE Trans Inf Technol Biomed* 2000; 4: 274-83.
9. McDonald JM, Brossette S, Moser SA. Pathology information systems, data mining leads to knowledge discovery. *Arch Pathol Lab Med* 1998; 122:409-11.
10. Vianna RCXF, Moro CMC. Aplicação da descoberta de conhecimento em base de dados para a identificação de características relacionadas à mortalidade infantil. In: *Proceedings of International Association for Development of the Information Society*. Lisboa: IADIS Press; 2005. p. 678-80.
11. Seth P, Guatam NT, Balint R. Preparing and mining data with Microsoft SQL Server 2000 and analysis services. <http://www.microsoft.com/learning> (acessado em 18/Out/2004).
12. Moro CMC. Medidas de similaridade entre conceitos da saúde [Tese de Doutorado]. São Paulo: Escola Politécnica, Universidade de São Paulo; 2001.

13. Cavalcante MS, Ramos Jr. AN, Pontes LRSK. Linkage of health information systems: a strategy to optimize the surveillance of pregnant women infected by HIV. *Epidemiol Serv Saúde* 2005; 14: 127-33.
14. Almeida MF, Jorge MHPM. O uso da técnica de "linkage" de sistemas de informação em estudos de coorte sobre mortalidade neonatal. *Rev Saúde Pública* 1996; 30:141-7.
15. Moraes Neto OL, Barros MBA. Fatores de risco para mortalidade neonatal e pós-neonatal na Região Centro-Oeste do Brasil: linkage entre bancos de dados de nascidos vivos e óbitos infantis. *Cad Saúde Pública* 2000; 16:477-85.
16. Camargo Jr. KR, Coeli CM. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. *Cad Saúde Pública* 2000; 16:439-47.
17. Pyle D. Data preparation for data mining. San Francisco: Morgan Kaufman Publishers; 1999.
18. Souza PTA, Tosetto VG. Mudança na estratégia de coleta de informações com alteração do perfil epidemiológico de mortalidade. In: *Anais da 4ª Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças*. Brasília: Ministério da Saúde; 2005. p. 100-2.
19. Cañizares SMS, Muñoz MAA, Ocerin JMC. Software de mineração de dados: análise de características. In: *Proceedings of International Association for Development of the Information Society*. Lisboa: IADIS Press; 2005. p. 380-5.
20. Menezes AMB, Victora CG, Barros FC, Albernaz E, Menezes FS, Jannke HA, et al. Mortalidade infantil em duas coortes de base populacional no Sul do Brasil: tendências e diferenciais. *Cad Saúde Pública* 1996; 12 Suppl 1:79-86.
21. Lansky S, França E, César CC, Monteiro Neto LC, Leal MC. Mortes perinatais e avaliação da assistência ao parto em maternidades do Sistema Único de Saúde em Belo Horizonte, Minas Gerais, Brasil, 1999. *Cad Saúde Pública* 2006; 22:117-30.
22. Aerts DRGC. Investigação dos óbitos perinatais e infantis: seu uso no planejamento de políticas públicas de saúde. *J Pediatr (Rio J.)* 1997; 73:364-6.
23. Lyra TM, Guimarães MJB, Melo Filho DA, Nascimento Junior JA, Lima AAF, Montenegro D, et al. Utilização de mapas de risco para vigilância em saúde: a experiência do Recife. In: *Anais da 4ª Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças*. Brasília: Ministério da Saúde; 2005. p. 26-7.
24. Caldeira AP, França E, Perpetuo IHO, Goulart EMA. Evolução da mortalidade infantil por causas evitáveis, Belo Horizonte, 1984-1998. *Rev Saúde Pública* 2005; 39:67-74.
25. Cesar CLG. Fatores de risco associados à mortalidade infantil em Cotia e Vargem Grande Paulista, SP, 1984-1985: uma proposta de instrumentos preditivos [Tese de Doutorado]. São Paulo: Faculdade de Saúde Pública, Universidade de São Paulo; 1988.
26. Koontz AM, Buckley KA, Ruderman M. The evolution of fetal and infant mortality review as a public health strategy. *Matern Child Health J* 2004; 8: 195-203.
27. Misra DP, Grason H, Liao M, Strobino DM, McDonnell KA, Allston AA. The nationwide evaluation of fetal and infant mortality review (FIMR) programs: development and implementation of recommendations and conduct of essential maternal and child health services by FIMR programs. *Matern Child Health J* 2004; 8:217-29.
28. Oliveira IC. Aplicação de data mining na busca de um modelo de prevenção da mortalidade infantil [Dissertação de Mestrado]. Florianópolis: Universidade Federal de Santa Catarina; 2001.
29. Rumel D, Costa Filho DC, Aiello S, Koyama SM, Rosenbaum W. Acurácia dos indicadores de risco do programa de defesa da vida dos lactentes em regiões do estado de São Paulo, Brasil. *Rev Saúde Pública* 1992; 26:6-11.
30. Silva AA, Bettiol H, Barbieri MA, Ribeiro VS, Aragão VMF, Brito LGO, et al. Infant mortality and low birth weight in cities of Northeastern and South-eastern Brazil. *Rev Saúde Pública* 2003; 37:693-8.
31. Ministério da Saúde/Conselho Nacional de Saúde. O Conselho Nacional de Saúde e a construção do SUS: referências estratégicas para a melhora do modelo de atenção à saúde. Brasília: Ministério da Saúde; 2003.
32. Collazos K, Barreto JM, Roisemberg M. Dificuldades na aplicação de KDD na medicina. In: *Anais do II Workshop de Informática Aplicada à Saúde [CD-ROM]*. Itajaí: Editora Ciências Brasileiras da Computação; 2002.
33. Shearer RC. The CRISP-DM model the new blueprint for data mining. *Journal of Data Warehousing* 2000; 5:13-22.
34. Bostwick DG, Burke HB. Prediction of individual patient outcome in cancer. *Cancer* 2001; 91:1643-6.

Recebido em 11/Ago/2009

Versão final reapresentada em 08/Dez/2009

Aprovado em 18/Dez/2009