

ROSSANA CRISTINA XAVIER FERREIRA VIANNA

**IDENTIFICAÇÃO DE CARACTERÍSTICAS RELACIONADAS À MORTALIDADE
INFANTIL UTILIZANDO A DESCOBERTA DE CONHECIMENTO EM BASE DE
DADOS DE SAÚDE PÚBLICA**

CURITIBA

2007

ROSSANA CRISTINA XAVIER FERREIRA VIANNA

**IDENTIFICAÇÃO DE CARACTERÍSTICAS RELACIONADAS À MORTALIDADE
INFANTIL UTILIZANDO A DESCOBERTA DE CONHECIMENTO EM BASE DE
DADOS DE SAÚDE PÚBLICA**

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná como requisito parcial para a obtenção do título de Mestre em Tecnologia em Saúde.

Área de Concentração: Informática em Saúde.

Orientadora: Prof.^a Dr.^a Claudia M. C. Moro Barra.

Co-orientador: Prof. Dr. Samuel J. Moyses.

CURITIBA

2007

- Vianna, Rossana Cristina Xavier Ferreira.
- V617i Identificação de características relacionadas à mortalidade infantil
2007 utilizando a descoberta de conhecimento em base de dados de saúde pública. /
Rossana Cristina Xavier Ferreira Vianna ; orientador, Claudia M. C. Moro Barra
; co-orientador; Samuel J. Moyses. – 2007.
xii, 118 f. : il. ; 30 cm
- Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,
Curitiba, 2007
Bibliografia: f. 81-95
1. Saúde pública – Brasil. 2. Mortalidade infantil. 3. Sistemas de recuperação da informação. I. Barra, Claudia M. C. M. II. Moyses, Samuel Jorge. III. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação Em Tecnologia em Saúde. IV. Título.

CDD 20. ed. – 614.0981



Pontifícia Universidade Católica do Paraná
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Tecnologia em Saúde

**ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
DO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE DA
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ**

DEFESA DE DISSERTAÇÃO Nº048

Aos 23 dias do mês de março de 2007 realizou-se a sessão pública de defesa da dissertação **“Identificação de característica relacionadas à mortalidade infantil utilizando a descoberta de conhecimento em base de dados de Saúde Pública”**, apresentada por **Rossana Cristina Xavier Ferreira Vianna** como requisito parcial para a obtenção do título de Mestre em Tecnologia em Saúde – Área de Concentração – Informática em Saúde perante uma Banca Examinadora composta pelos seguintes membros:

Profª. Drª. Claudia Maria Cabral Moro Barra,
PUCPR (Orientadora)


assinatura

APROVADO
parecer (aprov / reprov.)

Prof. Dr. Samuel Jorge Moyses
PUCPR (Co-Orientador)



APROVADO

Prof. Dr. Aristides Schier da Cruz
(FEPAR)



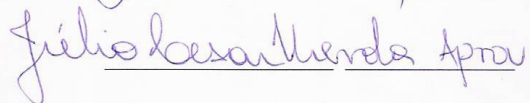
APROV.

Profª. Drª. Débora Ribeiro Carvalho
(UTP)



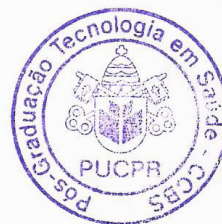
Aprov

Prof. Dr. Julio César Nievola
(PUCPR)


Aprov.

Conforme as normas regimentais do PPGTS e da PUCPR, o trabalho apresentado foi considerado **APROVADO** (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora registradas no Livro de Defesas do Programa.


Prof. Dr. Laudelino Cordeiro Bastos,
Diretor do PPGTS PUCPR



“Somos o que repetidamente fazemos. A excelência, portanto, não é um feito, mas um hábito.”

Aristóteles

Agradecimento especial

Ao meu esposo Eduardo, aos meus filhos, aos meus pais e a minha neta Julia, que suportaram a minha ausência durante a realização deste trabalho;

À Coordenação do Curso, professores e demais funcionários, pela dedicação e paciência com que mostraram os caminhos, apesar da minha impaciência e falta de colaboração em muitos momentos;

Os sinceros agradecimentos, à professora orientadora Dra. Claudia Moro, que tanto auxiliou e com suas palavras colocou em uma linha de pensamento que culminou nos frutos deste trabalho; **ao professor co-orientador Dr. Samuel Moyses e aos professores Dra. Deborah Carvalho e Dr. Julio C. Nievola**, que ajudaram também a desvendar os mistérios da pesquisa;

Aos colegas de turma, estagiários do PIBIC, em especial ao Thiago K. Shiono, pela amizade cultivada e pela troca de experiências que, com alegria e jovialidade, também lutaram pelos seus trabalhos;

À Secretaria de Estado da Saúde do Paraná, à Pontifícia Universidade Católica do Paraná e aos especialistas consultados, que permitiram a realização deste estudo, os mais sinceros agradecimentos.

PUBLICAÇÕES durante o Mestrado

-VIANNA, R. C. X. F.; MORO, C. M. C. **Sistema de Investigação da Mortalidade Infantil (SIMI)**. Anais do IV Congresso de Pediatria-Pediatria Social, Curitiba, 2005.

-VIANNA, R. C. X. F.; MORO, C. M. C. **Análise da Situação da Mortalidade Infantil no Estado do Paraná, Brazil, em 2004**. Anais do II Congresso Internacional de Especialidades Pediátricas, Criança 2005, Curitiba, 2005a.

-VIANNA, R. C. X. F.; MORO, C. M. C. **Aplicação da Descoberta de Conhecimento em Base de Dados para a identificação de características relacionadas à Mortalidade Infantil**. In Proceedings da Conferência IADIS (International Association for Development of the Information Society) Íbero-Americana, Lisboa, Portugal, 2005b.

-VIANNA, R. C. X. F.; SOARES, H. B. **A Mortalidade Materna e Infantil como articulador na Gestão e Atenção à Saúde**. Anais da 1ª Mostra Nacional de Vivências Inovadoras de Gestão no SUS. Brasília, 2006.

-VIANNA, R. C. X. F.; MORO, C. M. C. **Histórico da Mortalidade Infantil no Estado do Paraná, Brazil, antes e após o Sistema de Investigação da Mortalidade Infantil (SIMI)**. Anais do XIº Congresso Mundial de Saúde Coletiva. Rio de Janeiro, 2006a.

RESUMO

Os dados armazenados nos Sistemas de Informação de Saúde (SIS) no Brasil são muitas vezes subutilizados por várias razões, dentre as quais a dificuldade em analisar o seu grande volume. Neste trabalho buscou-se identificar características relacionadas à mortalidade infantil, a partir da utilização de uma base gerada com a integração dos dados de três SIS oficiais (de Nascidos Vivos - SINASC, de Mortalidade - SIM e de Investigação de Mortalidade Infantil - SIMI). O procedimento metodológico adotado foi a Descoberta de Conhecimento em Base de Dados (*KDD*) e, dentro deste, a Mineração de Dados (*DM*) foi a técnica utilizada para o estabelecimento de padrões de comportamento dos dados nesta nova base com os óbitos infantis analisados pelos Comitês de Prevenção da Mortalidade Infantil no Estado do Paraná, de 2000 a 2004. Para a integração dos três SIS, partiu-se de 9372 casos de óbitos analisados no SIMI, 14164 óbitos infantis do SIM e 801077 nascidos vivos do SINASC, gerando 7256 casos analisados na nova base. Utilizou-se o algoritmo classificador C4.5 da ferramenta *Waikato Environment for Knowledge Analysis (WEKA)*. Os atributos meta, a partir dos quais foram gerados padrões do que estava relacionado a esta característica escolhidos foram: a evitabilidade dos óbitos e a alteração da causa do óbito na declaração de óbito e após as análises dos Comitês. No *DM* quando o atributo meta foi a evitabilidade foram geradas 441 regras, sendo 78% classificadas corretamente, com precisão de 83% para identificação dos óbitos evitáveis e 62% para os inevitáveis. A sensibilidade alcançada foi de 80% para a identificação dos óbitos evitáveis. A especificidade foi de 62% para identificação dos inevitáveis. Com uma taxa de erro de 12% para os evitáveis e 44% para os inevitáveis. Quando o atributo meta foi a alteração da causa do óbito atribuída na declaração de óbito e após as análises dos Comitês, foram geradas 162 regras. Em 58% dos casos houve alteração da causa básica originalmente atribuída na declaração de óbito, sendo que 88% das regras foram classificadas corretamente, com precisão de 93% e sensibilidade de 88%, especificidade de 49% e taxa de erro igual a zero em 7 regras e 3% em 3 regras que foram avaliadas pelos especialistas em epidemiologia da mortalidade infantil. Três regras foram consideradas relevantes: se a mãe era adolescente e o peso ao nascer entre 2500 e 3500g, então óbito perinatal seria esperado. Se a mãe era adolescente e tinha escolaridade entre 4-7 anos, então desconforto respiratório e óbito eram esperados. Se o peso ao nascer era < 2500g com 41 semanas de gestação e presença de afecções maternas, então óbito < 30 dias era esperado. Tais resultados sugerem a necessidade dos serviços de saúde de estabelecerem medidas de maior atenção às adolescentes, às crianças com peso ao nascer < 2500g, as que nascem depois de 41 semanas e as com as mães com afecções maternas. A integração de bases de dados demonstrou ser importante para conhecer melhor a mortalidade infantil, podendo ser estendida a outros Sistemas de Informação em Saúde Pública. O *KDD* mostrou ser pertinente e adequado para utilização em Saúde Pública, visto que confirmou as características da mortalidade infantil com as existentes na literatura.

Palavras Chaves: 1. Saúde pública - Brasil. 2. Mortalidade Infantil. 3. Sistemas de recuperação da informação.

ABSTRACT

Data has been stored in Brazilian Health Information Systems (SIS) and such information is, many times, subused due to difficulty in analyzing such great volume. In this work characteristics related to infant mortality were identified, starting from use of a basis generated with integration of data of three Systems (Born Alive - SINASC, Mortality - SIM and Infant Mortality Investigation - SIMI). Knowledge Discovery in Database (KDD) was used and, within it, Data Mining (DM) for determination of patterns of behavior in such new basis. The infantile deaths analyzed by Committees of Infant Mortality Prevention were used in State of Paraná from 2000 to 2004. The classifying algorithm C 4.5 of the tool Waikato Environment was used for the Knowledge Analysis. The meta attributes, for generated patterns related about these characteristics, were unavoidability of deaths and the alteration of death cause in the death declaration and after analyses done by Committees. Linking the data bases, there were 9372 cases of deaths analyzed in SIMI, 14164 infantile deaths from SIM and 801077 born alive from SINASC which, after linking, came to 7256 cases. In DM, when meta attribute was the unavoidability, 441 rules were generated, being 78% classified correctly, with accuracy of 83% for identification of avoidable deaths and 62% for the inevitable ones. 80% of sensitivity for identification of avoidable cases. 62% of specification for identification of inevitable ones. The error rate for avoidable ones was 12% and for unavoidable ones 44%. With the target attribute the alteration between the cause of death in the death declaration and after analyses by the Committees, 162 rules were generated, for cases in which there was alteration of basic cause (58%), with 88% successes, accuracy of 93% and sensitivity of 88%, specification 49% and error rate 0% in 7 rules and 3% in 3 rules that were appraised by consulted specialists. Among these, 3 rules were relevant: if mother was adolescent and weight, when being born, ranged between 2500 and 3500g, then it was verified to be a per birth death. If mother was adolescent and her level of education ranged among 4-7 years, then there was breathing discomfort followed by death. If weigh when being born was less than 2500g, 41 weeks of gestation and maternal diseases, then death within less than 30 days. It was suggested need for services in health to establish measures for larger attention to adolescents, to children that weight less than 2500g at birth, ones that are born after 41 weeks and mothers with maternal disease. Linking of data bases demonstrated that it is important to know infant mortality better, having possibility of extending it to other Information Systems on Public Health. KDD showed to be important and appropriate for use in Public Health because it confirmed characteristics of infant mortality which were described in relating literature.

Key words: 1. Public health. 2. Infant Mortality. 3. Recuperated Information Systems.

LISTA DE FIGURAS

Figura 1 - CMI no Paraná - Evolução Histórica 1979 a 2004.	11
Figura 2 - Fases do Processo de Descoberta do Conhecimento	13
Figura 3 - Processo de KDD no contexto de solução de um problema.	14
Figura 4 – Árvore de Decisão.	23
Figura 5 – Material e Métodos.	35
Figura 6 - Porcentagem da categorização da base segundo a CID após as análises.	55
Figura 7 - Análise da evitabilidade dos óbitos.	58
Figura 8 - Tela da Ficha de Análise do Óbito Infantil.	100
Figura 9 - Sistema de Vigilância dos Óbitos Infantis no Estado do Paraná-Brasil.....	106

LISTA DE TABELAS

Tabela 1 - Categorização da Mortalidade Infantil:.....	9
Tabela 2 - Características dos Critérios de Risco:	9
Tabela 3 - Categorização das Taxas de Mortalidade Infantil:	10
Tabela 4 - Seleção dos atributos da base para DM:	44
Tabela 5 - Questionário com as regras de mineração com o atributo meta de evitabilidade:	50
Tabela 6 - Questionário com as regras de mineração com o atributo meta alteração da CID:	52
Tabela 7 - Regras Meta Evitabilidade:	64
Tabela 8 - Escolha das regras, meta evitabilidade:	65
Tabela 9 - Escolha das regras, meta alteração entre as CID:.....	66
Tabela 10 - Critérios de Evitabilidade:	97
Tabela 11 - Medidas de prevenção e intervenção da mortalidade infantil:.....	99
Tabela 12 - Seleção dos atributos das três bases de dados SINASC/ SIM/SIMI:	109
Tabela 13 - Códigos das ocupações:.....	110
Tabela 14 - Categoria para idade da criança:	110
Tabela 15 - Distribuição da CID na DO:.....	110
Tabela 16 - Distribuição da CID após a investigação:.....	111
Tabela 17 - Categoria CID mais encontradas:	111
Tabela 18 - Categoria para idade da mãe:	112
Tabela 19 - Estado Civil da mãe:	112
Tabela 20 - Escolaridade da mãe:	112
Tabela 21 - Quantidade de filhos vivos:	113
Tabela 22 - Quantidade de filhos mortos:	113
Tabela 23 - Número de semanas de gestação:	113
Tabela 24 - Tipo de gravidez:	114
Tabela 25 - Tipo de Parto:	114
Tabela 26 - Número de consultas de pré-natal:	114
Tabela 27 - Média da classe Apgar no primeiro minuto:	114
Tabela 28 - Média da classe Apgar no quinto minuto:	115
Tabela 29 - Raça/Cor:	115
Tabela 30 - Categoria do peso ao nascer:	115
Tabela 31 - Categoria da idade no óbito:	115
Tabela 32 - Alteração entre as CID10 antes e após as análises dos Comitês:	115
Tabela 33 - Critérios de evitabilidade:.....	116

LISTA DE ABREVIATURAS E SIGLAS

APGAR -	Resposta clínica de viabilidade do RN, no primeiro e no quinto minuto de vida, baseada em critérios pré-estabelecidos de cor da pele, frequência cardíaca, esforço respiratório, tônus muscular, atividade reflexa.
BR -	Brasil.
CID10 -	Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde, Décima Revisão de 1989, utilizada sua nona edição revista em 2003.
CIDS -	Centro de Informação e Diagnóstico em Saúde da SESA-PR.
CEP -	Comitê de Ética em Pesquisa.
CMI -	Coeficiente de Mortalidade Infantil.
CPMI -	Comitê de Prevenção da Mortalidade Infantil.
DM -	<i>Data Mining</i> ou Mineração de Dados.
DN -	Declaração de Nascidos Vivos.
DO -	Declaração de Óbito.
IBGE -	Instituto Brasileiro de Geografia e Estatística.
KDD -	<i>Knowledge Discovery in Database</i> ou Descoberta de Conhecimentos em Base de Dados.
MS -	Ministério da Saúde.
NV -	Nascido Vivo.
OMS -	Organização Mundial de Saúde.
OPAS	Organização Pan-americana de Saúde.
PUCPR -	Pontifícia Universidade Católica do Paraná.
PR -	Estado do Paraná.
RN -	Recém-Nascido.
SESA-PR -	Secretaria do Estado da Saúde do Paraná.
SIH-SUS -	Sistema de Informação Hospitalar - SUS.
SIM -	Sistema de Informações de Mortalidade.
SIMI -	Sistema de Investigação de Mortalidade Infantil.
SINAN -	Sistema Nacional de Agravos de Notificação.
SINASC -	Sistema de Informações sobre Nascidos Vivos.
SIS -	Sistema de Informação em Saúde.
SUS -	Sistema Único de Saúde.
WEKA -	<i>Waikato Environment for Knowledge Analysis</i> .

SUMÁRIO

1. INTRODUÇÃO	1
1.1. JUSTIFICATIVA	2
1.2. OBJETIVOS	4
1.2.1. OBJETIVO PRINCIPAL	4
1.2.2. OBJETIVOS ESPECÍFICOS	5
2. REVISÃO DA LITERATURA	6
2.1. MORTALIDADE INFANTIL	6
2.1.1. DEFINIÇÕES	6
2.1.2. AVALIAÇÃO	10
2.1.3. PANORAMA ATUAL	10
2.2. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS	12
2.2.1. MINERAÇÃO DE DADOS	15
2.2.2. TAREFAS E ALGORITMOS DE MINERAÇÃO DE DADOS	18
2.2.3. TÉCNICAS E FERRAMENTA DE MINERAÇÃO DE DADOS	22
2.3. SISTEMAS DE INFORMAÇÃO EM SAÚDE	26
2.3.1. SISTEMA DE INFORMAÇÃO SOBRE NASCIDOS VIVOS	28
2.3.2. SISTEMA DE INFORMAÇÃO SOBRE MORTALIDADE	29
2.3.3. SISTEMA DE INVESTIGAÇÃO DA MORTALIDADE INFANTIL	31
3. MATERIAL E MÉTODOS	34
3.1. FASE DE PRÉ-PROCESSAMENTO	37
3.1.1. SELEÇÃO DOS DADOS	38
3.1.2. LIMPEZA DOS DADOS	40
3.1.3. ENRIQUECIMENTO E CONSTRUÇÃO DOS ATRIBUTOS	40
3.1.4. NORMALIZAÇÃO DOS DADOS	41
3.2. MINERAÇÃO DE DADOS	46
3.3. PÓS-PROCESSAMENTO e AVALIAÇÃO	49
4. RESULTADOS E ANÁLISES	54
4.1. RESULTADOS E ANÁLISES DA INTEGRAÇÃO DAS BASES	54
4.2. RESULTADOS E ANÁLISES DO DM	64
4.2.1. ATRIBUTO META EVITABILIDADE	64
4.2.2. ATRIBUTO META ALTERAÇÃO ENTRE AS CID	65
4.3. RESULTADOS E ANÁLISES DOS ESPECIALISTAS	66
4.3.1. ATRIBUTO META EVITABILIDADE	66
4.3.2. ATRIBUTO META ALTERAÇÃO ENTRE AS CID	69
5. DISCUSSÃO E CONCLUSÃO	72
6. REFERÊNCIAS BIBLIOGRÁFICAS	81
7. ANEXOS	95
Anexo 1: Declaração de Nascido Vivo	95
Anexo 2: Declaração de Óbito	96
Anexo 3: Tabela 10 - Critérios de Evitabilidade	97
Anexo 4: Tabela 11 – Medidas de prevenção e intervenção da MI	99
Anexo 5: Tela do SIMI	100
Anexo 6: Descrição da Ficha de Análise de Óbito Infantil	101
Anexo 7: Sistema de Vigilância dos Óbitos Infantis no Estado do Paraná	106
Anexo 8: Declaração de Consentimento Declarado para Liberação de Bancos de Dados	107
Anexo 9: Liberação do CEP para Realização da Pesquisa	108
8. APÊNDICES	109
Apêndice 1: Tabela 12 - Seleção dos atributos das três bases de dados SINASC/SIM/SIMI	109
Apêndice 2: Categorizações normalizadas	110
Apêndice 3 : Tabelas das análises da junção das bases	112

1. INTRODUÇÃO

A mortalidade infantil é um dos indicadores clássicos da saúde de uma população, principalmente para avaliar as condições da qualidade dos serviços prestados e as mudanças sociais e econômicas ocorridas ao longo do tempo (MINISTÉRIO DA SAÚDE, 2004). O estudo deste indicador neste trabalho é feito com a utilização de informações contidas nas bases de dados dos Sistemas de Informação em Saúde, com o propósito de transformação destas informações em conhecimento e capacidade de implementação de ações efetivas na melhoria da saúde da população infantil (CARROL et al., 2003). Os entendimentos de como estes óbitos ocorreram e a identificação dos que se podem prevenir ou evitar, (WIGGLESWORTH, 1980), pode auxiliar no entendimento das causas dos óbitos (LANSKI, FRANCA e LEAL, 2002) e, desta forma, poder atuar para sua redução. Para melhorar as investigações, análises e conhecimento das características relacionadas à mortalidade infantil foram criados Comitês multiprofissionais.

Os Comitês de Prevenção da Mortalidade Infantil foram efetivados no Paraná a partir de 1999 e, desde então têm como objetivos principais a identificação da evitabilidade destes óbitos e a especificação da causa básica dos mesmos para melhorar as ações de redução dos óbitos infantis neste Estado. Uma das suas ações após as investigações é a correção da causa básica no sistema oficial de mortalidade, o que confere maior credibilidade ao mesmo (MANSANO et al, 2004). Os Comitês buscam informações sobre os óbitos infantis nos Sistemas de Informação em Saúde Pública brasileiros.

Desde 1975, com a implantação no Brasil do Sistema de Informação sobre Mortalidade (SIM); desde 1990, com a definição do Sistema de Informação sobre

Nascidos Vivos (SINASC) e desde 2000, com a criação, no Estado do Paraná, do Sistema de Investigação da Mortalidade Infantil (SIMI) tem sido gerada grande quantidade de dados, nem sempre utilizadas de maneira efetiva (VIANNA, 2004, 2004a, 2005a). Em média, no Paraná, são registrados anualmente 150.000 nascidos vivos no SINASC, com mais ou menos 2.000 óbitos infantis registrados no SIM ao ano e 1.500 óbitos infantis analisados ao ano no SIMI. Neste cenário, apresenta-se uma relativa dificuldade no gerenciamento destas informações de modo produtivo para a tomada de decisões em saúde e para a formulação / implementação / avaliação de políticas públicas baseadas em evidências extraídas de tais sistemas. Para auxiliar o gerenciamento destas informações a tecnologia em saúde pode-se mostrar de grande utilidade.

Métodos e técnicas computacionais, como o *Knowledge Discovery in Databases (KDD)* - Descoberta de Conhecimento em Base de Dados -, podem auxiliar os serviços de saúde no desenvolvimento de conhecimentos retirados de grandes bases de dados (BERRY e LINOFF, 1997). O *Data Mining (DM)* ou mineração de dados é uma parte dentro do processo de *KDD*, que consiste na aplicação de análises de dados e descoberta de algoritmos que produzem regras ou modelos sobre os dados colhidos (FAYYAD, WEIR e DJORGOVSKI, 1993). Estas regras podem ser relevantes para os especialistas da área de saúde na utilização dos conhecimentos gerados nas bases de dados coletadas (CIOS, PEDRYCZ e SWINIARSKI, 1998).

1.1. JUSTIFICATIVA

A mortalidade infantil é historicamente (ROSEN, 1958) um indicador que está diretamente relacionado às condições de vida de um país (KOONTZ,

KATHLEEN e RUDERMAN, 2004). É um dos fatores mais importantes na avaliação do risco de morrer do ser humano antes de completar um ano de vida (MARTIN, CORDONI e BASTOS, 2005) e nas relações com as condições de vida da saúde da mulher (MINISTÉRIO DA SAÚDE, 2004).

Foi escolhido este tema, devido ao interesse em conhecer melhor as características relacionadas à mortalidade infantil, e para atuar, em conjunto com os Comitês e serviços de saúde, de maneira condizente com as informações históricas contidas nas bases de dados e padronizadas pela ferramenta.

Vários estudos apontam a redução da mortalidade infantil desde as primeiras décadas do século XX, principalmente, por causa da melhoria do saneamento básico, da assistência pré-natal, das imunizações e do incentivo ao aleitamento materno (MELLO JORGE, 2001). No Brasil, a redução da mortalidade infantil pode estar relacionada com a ampliação destes serviços, além do aumento das coberturas vacinais e da queda da fecundidade (COSTA et al., 2003). Devido à grande quantidade de informações existentes nos serviços de saúde relacionada à mortalidade infantil, a informática vem demonstrando sua utilidade com a nova geração de teorias e ferramentas computacionais. Com a criação dos Sistemas de Informação em Saúde, as técnicas de gerenciamento de informações, como o *KDD*, têm se mostrado efetivas também nesta função (FAYYAD, SHAPIRO, SMYTH, 1996) e (FAYYAD et al., 1996a).

Neste contexto, o *KDD* é uma ferramenta de suporte para análise de dados que permite, por exemplo, examinar as relações de similaridade entre os mesmos (LINARES, 2003). O *DM* faz parte do *KDD* e constitui uma área emergente de inteligência computacional para análise de grandes bancos de dados (KUSIAK et

al., 2000), utilizando a busca de padrões e a extração de informações destas bases (McDONALD, BROSSETTE e MOSER, 1998).

O processo de *KDD* envolve o uso dos dados, neste caso, escolhidos nas bases dos Sistemas de Informação de Saúde do Governo, SINASC, SIM e SIMI, selecionados de janeiro de 2000 a dezembro de 2004. Estas bases foram pré-processadas e transformadas em uma base única para o *DM*. Os anos foram escolhidos com base na criação do SIMI, em 2000, a partir do qual 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, nos três Sistemas de Informação em Saúde escolhidos.

A justificativa maior para este trabalho reside no pressuposto de que, conhecendo com maior profundidade e nível de detalhamento o problema objeto desta pesquisa, podem-se estabelecer metas para melhoria das condições de vida da população infantil no Estado do Paraná (VIANNA e MORO, 2005b).

1.2. OBJETIVOS

1.2.1. OBJETIVO PRINCIPAL

Identificar características relevantes relacionadas à mortalidade infantil com a utilização do *KDD*, a partir da base de dados gerada com a integração dos SIS – SINASC, SIM e SIMI.

1.2.2. OBJETIVOS ESPECÍFICOS

- - Integrar três bases de dados de diferentes SIS públicos brasileiros.
- - Conhecer na nova base integrada as características relacionadas à mortalidade infantil, geradas pelo processo de *DM*.
- - Identificar as relações existentes entre os padrões encontrados no processo de *DM* e os conhecimentos estabelecidos na literatura.

2. REVISÃO DA LITERATURA

Devido à vulnerabilidade dos menores de um ano, os dados de mortalidade infantil podem servir como parâmetro para que sejam observadas falhas nos sistemas de prevenção, promoção e no desenvolvimento de ações de saúde voltadas à população (HOLCMAN, LATORRE e SANTOS, 2004). No Paraná (PR), os dados de Mortalidade Infantil são coletados no Centro de Informação e Diagnóstico em Saúde (CIDS) e analisados por 151 Comitês Municipais de Prevenção da Mortalidade Infantil (CMPMI), 22 Comitês Regionais de Prevenção da Mortalidade Infantil (CRPMI) e um Comitê Estadual de Prevenção da Mortalidade Infantil (CEPMI). Para auxiliar esta análise, propôs-se utilizar ferramentas não convencionais na Saúde Pública, no esforço de aclarar os conhecimentos disponíveis (COLLAZOS e BARRETO, 1999).

Uma destas ferramentas é o *DM*, constatada sua importância para compreender tendências, padrões e correlações nas grandes bases de dados (QUONIAN et al., 2001). Como os SIS abarcam grandes quantidades de dados, o *DM* pode constituir importante ajuda na aquisição de informações que poderão no futuro auxiliar na promoção de saúde da população (RICHARDS et al., 2001).

Com este conhecimento adquirido, pretende-se melhorar o foco das intervenções, com o estabelecimento dos pontos vulneráveis que causam doenças, injúrias ou incapacidade (CARROL et al., 2003).

2.1. MORTALIDADE INFANTIL

2.1.1. DEFINIÇÕES

A Mortalidade Infantil é o número de crianças que morrem com menos de um ano de idade, representado pelo coeficiente ou taxa de mortalidade infantil,

que é o número de óbitos de menores de um ano por mil nascidos vivos em um período e local. A mortalidade infantil divide-se em: Neonatal (Precoce e Tardia) e Tardia ou Pós-Neonatal.

A Mortalidade Infantil Neonatal é segundo a Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde em sua décima revisão - CID 10 (OMS, 2004, 2004a e 2004b), o número de óbitos de crianças nascidas vivas, ocorridos entre 0 a 27 dias de vida completos (27 dias, 23 horas e 59 minutos), por mil nascidos vivos, na população residente em determinado espaço geográfico, no período de tempo considerado. A mortalidade infantil neonatal determina o risco de um nascido vivo morrer nas primeiras quatro semanas de vida. Taxas elevadas deste coeficiente estão, em geral, relacionadas a condições insatisfatórias em relação ao nível socioeconômico da população, à saúde da mãe, bem como à inadequada assistência pré-natal, ao parto e ao recém-nascido (RN).

A Mortalidade Infantil Neonatal Precoce é o número de óbitos de crianças nascidas vivas ocorridos de 0 a 6 dias completos de vida (6 dias, 23 horas e 59 minutos), por mil nascidos vivos, na população residente em determinado espaço geográfico, no período de tempo considerado (OMS 2004, 2004a e 2004b). Pode ser subestimado pelo registro de óbitos declarados como natimortos, que na realidade, são óbitos de nascidos vivos ocorridos pouco após o parto, fato que também é uma das causas de subregistros (menor número de registros do que o real) de nascidos vivos.

A Mortalidade Infantil Neonatal Tardia é o número de óbitos de crianças nascidas vivas, ocorridos entre 7 e 27 dias de vida completos (27 dias, 23 horas e 59 minutos), por mil nascidos vivos, na população residente em determinado

espaço geográfico, no período de tempo considerado (OMS 2004, 2004a e 2004b).

A Mortalidade Infantil Pós-Neonatal é o número de óbitos de crianças nascidas vivas, ocorridos de 28 a 364 dias de vida completos (364 dias, 23 horas e 59 minutos), por mil nascidos vivos, na população residente em determinado espaço geográfico considerado, no período de tempo considerado (OMS 2004, 2004a e 2004b). Taxas elevadas de Mortalidade Infantil Pós-Neonatal, de maneira geral, refletem baixos níveis de saúde e desenvolvimento socioeconômico. Quando a taxa de mortalidade infantil é alta, frequentemente, a mortalidade infantil pós-neonatal é o componente mais elevado.

Têm-se também o coeficiente de Mortalidade Perinatal, que é o número de óbitos fetais (a partir de 22 semanas completas, ou 154 dias de gestação, ou fetos com peso igual ou superior a 500 g ou estatura a partir de 25 cm) acrescidos dos óbitos neonatais precoces, por mil nascimentos totais (fetais mais nascidos vivos), da população residente em determinado espaço geográfico, no período de tempo considerado (OMS 2004, 2004a e 2004b). Esse indicador refere-se a óbitos ocorridos antes, durante ou logo após o parto sendo influenciado por fatores vinculados à gestação e ao parto, dentre os quais, o peso ao nascer e a qualidade da assistência prestada. Taxas elevadas estão geralmente associadas a condições insatisfatórias de assistência pré-natal, ao parto e ao RN, por isso, é considerado um bom indicador da assistência obstétrica e neonatal, além das condições socioeconômicas. A subnotificação de óbitos fetais é ainda muito importante. Hoje as informações sobre o peso ao nascer e a duração da gestação são ainda omitidas na Declaração de Óbito (DO). Esta categorização está detalhada na Tabela 1, baseada no Ministério da Saúde (MS), na Organização

Pan-americana de Saúde (OPAS) e na Organização Mundial de Saúde (OMS) e, especificada na CID 10 (OMS 2004b).

Tabela 1 - Categorização da Mortalidade Infantil:

PERINATAL	NEONATAL		PÓS-NEONATAL
De 22 semanas de gestação até 7 dias de vida.	Neonatal precoce	Neonatal tardia	De 28 a 364 dias de vida.
	Até 7 dias de vida.	De 7 a 28 dias de vida.	

Tabela 2 - Características dos Critérios de Risco:

RISCO BIOLÓGICO	RISCO SOCIAL
Peso baixo ao nascer (< 2500 g).	Renda familiar baixa (= ou $< \frac{1}{4}$ salário mínimo por pessoa).
Prematuridade (gestação < 37 semanas).	Mãe sem companheiro.
Doenças preexistentes (doenças genéticas, p.e.).	RN com 2 ou + irmãos < 4 anos.
Gestações anteriores (aborto anterior ou óbito infantil em outro filho).	Escolaridade da mãe (mãe analfabeta ou < 4 anos de estudo).
	Número de consultas de pré-natal (< 3 consultas).
	Desemprego.
	Idade da mãe (< 18 anos).

As principais características estabelecidas na literatura relacionadas ao óbito infantil são os critérios de risco (MISRA et al., 2004), (CESAR, 1990), (SILVA et al., 1999) e (ANDRADE et al., 2004) que estão apresentados na Tabela 2 e são os biológicos: peso baixo ao nascer (< 2.500 g), prematuridade (gestação < 37 semanas), doenças preexistentes (genéticas p.e.); gestações anteriores, aborto ou óbito infantil em outro filho (ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2004) e sociais (RUMEL et al., 1992): renda familiar baixa ($< ou = \frac{1}{4}$ salário mínimo por

pessoa), mãe sem companheiro, RN com dois ou mais irmãos menores de 4 anos, escolaridade da mãe (mãe analfabeta ou com menos de 4 anos de estudo), número de consultas de pré-natal (menos de 3 consultas), desemprego, idade da mãe (< 18 anos) (OLIVEIRA, 2001), (VICTORA et al., 1996) e (VICTORA et al., 2005).

2.1.2. AVALIAÇÃO

A evolução da mortalidade infantil no mundo passa por disparidades relacionadas com a evolução socioeconômica dos países. A categorização das taxas de mortalidade infantil, segundo o MS, pode ser observada na Tabela 3.

Tabela 3 - Categorização das Taxas de Mortalidade Infantil:

TAXAS DE MORTALIDADE INFANTIL	
altas	50 por mil nascidos vivos ou mais
médias	20-49 por mil nascidos vivos
baixas	menos de 20 por mil nascidos vivos

Esses parâmetros devem ser periodicamente reajustados às mudanças no perfil epidemiológico (evolução dos casos em uma comunidade) em cada local. Atualmente, vários países, porém poucos municípios brasileiros apresentam valores abaixo de 20/1000 nascidos vivos.

2.1.3. PANORAMA ATUAL

As maiores taxas de mortalidade infantil no mundo ocorrem em Papua Nova Guiné, onde, em 1990, eram 741/1000 e, em 2001, eram 701/1000. Na América Latina, o pior índice fica com a Bolívia, em 1990, com 85/1000 e, em 2000, com 35/1000 (Organização das Nações Unidas - ONU, 2004).

Segundo a ONU (2004), o menor índice de Mortalidade Infantil no mundo está na Suécia, com 6/1000 em 1990 e 3/1000 em 2000 e na América Latina, o menor índice foi no Chile, com 17/1000 em 1990 e 11/1000 em 2000 (Organização das Nações Unidas, 2004).

Em 1990, o Brasil registrou um índice de Mortalidade Infantil de 50/1000 e, em 2000, 28/1000 (MINISTÉRIO DA SAÚDE, 2005), ou seja, taxa classificada como média.

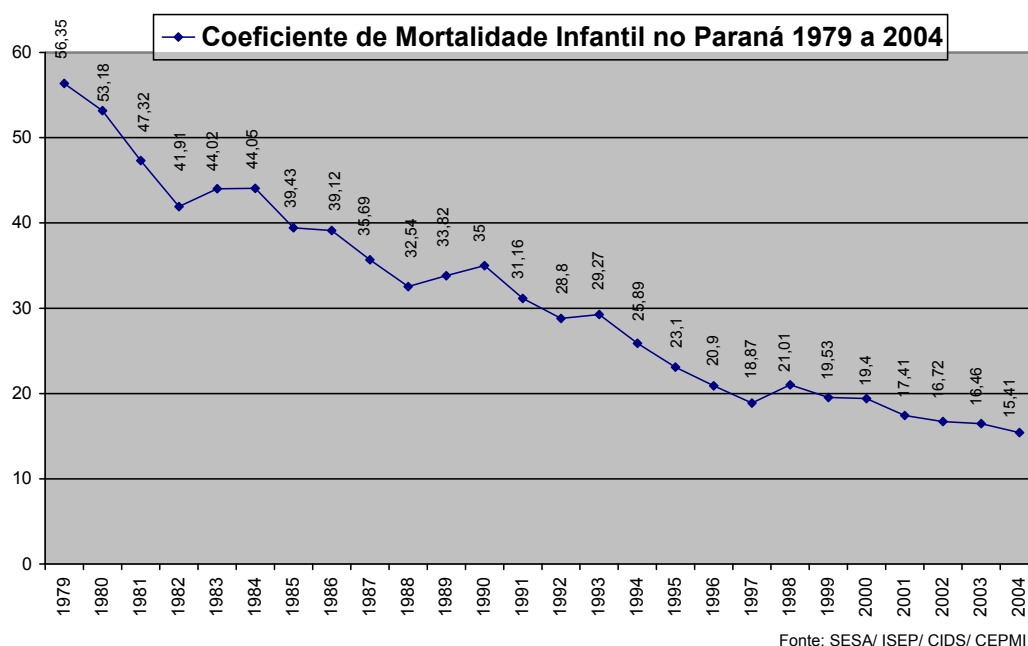


Figura 1 - CMI no Paraná - Evolução Histórica 1979 a 2004.

O Paraná apresentou tendência geral de decréscimo com 25,89/1000 em 1994 e 15,41/1000 em 2004 (MINISTÉRIO DA SAÚDE, 2005), taxa classificada como baixa, com redução de 45% destes óbitos. Estes dados são observados na Figura 1.

2.2. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

O *KDD* é um método iterativo¹, com algoritmos iterativos², que envolve numerosos passos e muitas decisões a serem tomadas pelos usuários. Como um campo de conhecimento multidisciplinar, evoluiu e continua evoluindo a partir da interseção de pesquisas em diversas áreas como: bases de dados (LIU, LU e YAO, 2001), aprendizado de máquinas (HAN e KAMBER, 2001), inteligência computacional (GOLDSCHIMIDT e PASSOS, 2005), reconhecimento de padrões (QUONIAN et al., 2001), estatística, inteligência artificial, aquisição de conhecimento para sistemas especialistas, visualização de dados, descoberta científica, recuperação de informação e computação de alto desempenho (HAND, MANNILA e SMYTH, 2001). Sistemas de software de *KDD* incorporam teorias, algoritmos (VAN BEMMEL e MUSEN, 1997) e métodos destes campos (SHORTLIFFE e PERREAULT, 2000).

O termo *KDD* foi formalizado em 1989, (PARSAYE, 1989), em referência ao processo de descoberta de conhecimento em dados. Este processo tem como objetivo extrair as relações implícitas e previamente desconhecidas entre os dados na busca de informações potencialmente úteis. Pode ser aplicado em várias áreas: análises de mercado, de crédito, de tendências, marketing, astronomia, medicina.

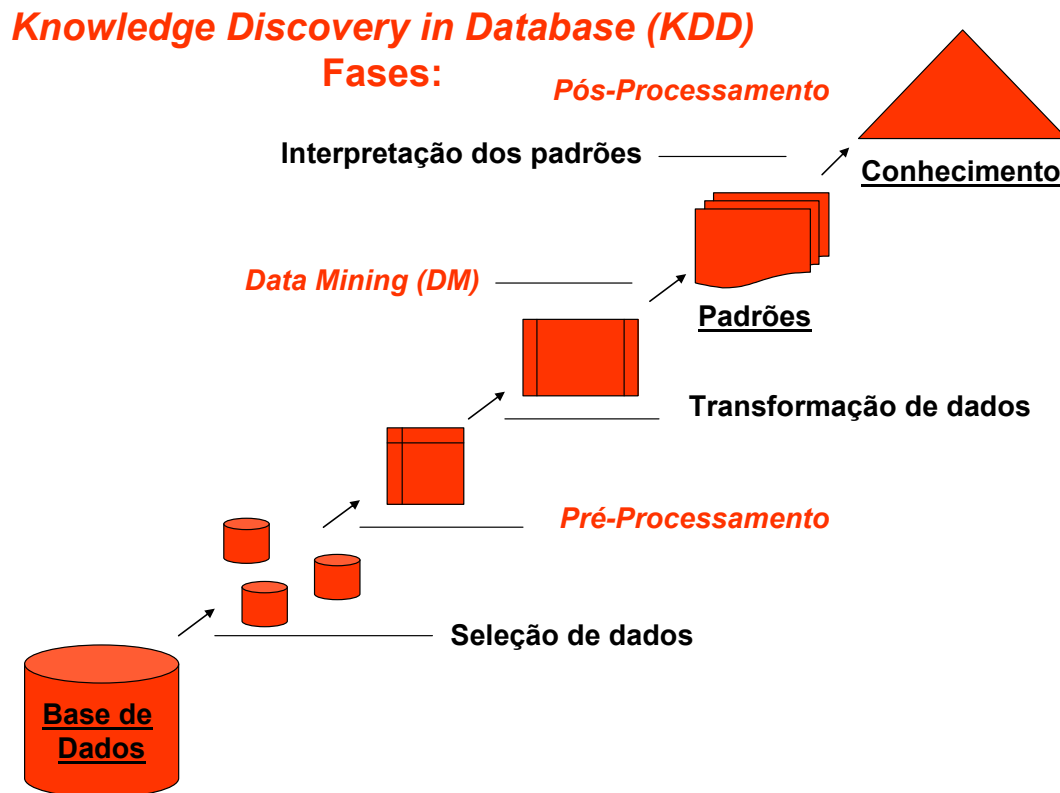
Na medicina pode ser útil, por exemplo, para analisar o histórico do paciente, aperfeiçoar o tratamento e auxiliar na obtenção de conhecimento implícito nas

¹ Interativo: ação exercida por mais de um agente.

² Iterativo: ato repetitivo.

bases de dados para sugerir mudanças e ações que visam a redução de doenças.

O *KDD* é composto por três etapas principais, mostradas na figura 2 adaptada de FAYYAD, SHAPIRO e SMYTH, 1996a, e que são: o pré-processamento, o *DM* e o pós-processamento (QUINLAN, 1986).



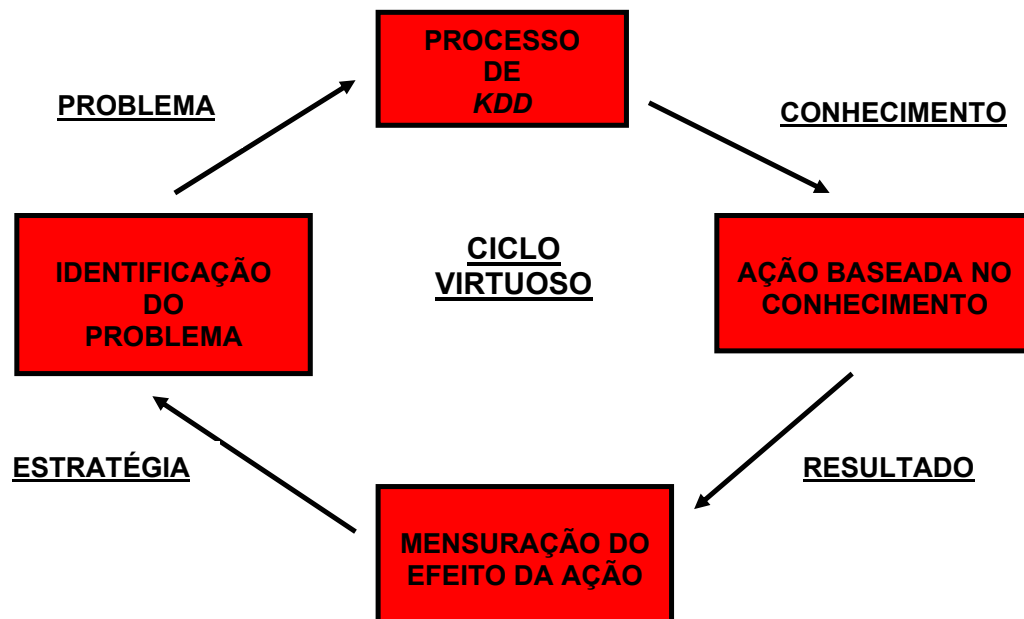
Adaptado de Fayyad, 1996

Figura 2 - Fases do Processo de Descoberta do Conhecimento

Em sua fase de pré-processamento (WEISS e INDURKHYA, 1998), o *KDD* subdivide-se em: definição do objetivo, seleção, depuração (HARRICSON, 1998), enriquecimento, codificação (estrutura dos dados, imputação de dados). Em sua segunda fase ocorre a extração de padrões ou *DM* (HAN e KAMBER, 2001), propriamente dito, com tarefas de associação ou classificação ou clusterização. E

sua terceira fase têm-se o pós-processamento, quando se faz a análise dos dados obtidos pelo *DM* (WITTEN e FRANK, 2000).

O *KDD* é uma das principais tecnologias que relacionam Bases de Dados, Estatística, Técnicas de Agrupamento, Técnicas Gráficas, Inteligência Artificial e Sistemas Especialistas (FAYYAD, SHAPIRO e SMYTH, 1996b). Considerando que o *KDD* é um processo contínuo de busca de conhecimentos e resolução de problemas, na figura 3 pode-se observar graficamente este ciclo de identificação do problema e ação baseada no conhecimento gerado, formando o “Ciclo Virtuoso” de Silver (1998), que auxilia no entendimento do *KDD* para a solução de um problema e para explicar a importância da iteratividade do mesmo.



Fonte: SILVER, 1998.

Figura 3 - Processo de KDD no contexto de solução de um problema.

Na figura 3 mostra-se primeiro a existência do problema e sua identificação, dando início ao processo de geração de conhecimento. A partir daí, geram-se resultados que podem ser mensurados com uma estratégia para identificação de

novos problemas que podem ser processados. Organizando desta forma, um ciclo contínuo de geração de conhecimentos e o estabelecimento de ações baseadas no conhecimento gerado e as estratégias para verificação do impacto dessas ações, ciclicamente. Deste modo, facilita o trabalho em questão, porque melhora a qualidade de atuação a partir do momento que se pode retornar ao problema de maneira iterativa e encontrar outras soluções com novas estratégias de intervenção, tendo sempre uma ação baseada em conhecimento. A partir do momento que um resultado é gerado pode-se mensurar o efeito da ação e, se necessário propor novas estratégias de intervenção, identificação do problema novamente, novo processo de *KDD* com novas ações baseadas em conhecimento e assim sucessivamente, até encontrar a melhor solução baseada nos padrões gerados pelo *DM* e que foram transformados em conhecimento. Portanto, o *KDD* compreende todo o processo de descoberta de conhecimento em bases de dados, com vários passos que envolvem a preparação dos dados, a busca por padrões, a avaliação do conhecimento e o refinamento, envolvendo iteração e modificação. Já o *DM*, descrito a seguir, refere-se a aplicação de algoritmos para a extração de padrões destas bases selecionadas para a mineração.

2.2.1. MINERAÇÃO DE DADOS

O *DM* é uma das fases do processo de *KDD*. É uma técnica que faz a extração de padrões utilizando os dados disponíveis (FAYYAD, SHAPIRO e SMYTH, 1996). Pode-se entender melhor o problema em estudo, com a aplicação de algoritmos de descoberta de padrões nos dados (IMIELINSKI e MANNILA, 1996), com certas limitações de eficiência computacional aceitável (FAYYAD et

al., 1996a). O *DM* tem foco na escolha dos dados (BERRY e LINOFF, 2000), na busca e representação, no desenvolvimento iterativo e na manutenção com reaprendizado (PELLEGRINI e COLLAZOS, 1999). O *DM* automatiza a descoberta de informação (GROTH, 1998) e o processo de levantamento de tendências e padrões (COLLAZOS e BARRETO, 2000). Pode-se inferir, portanto, que o *DM* é a aplicação de técnicas de indução sobre bases de dados a fim de obter uma generalização das informações representadas (ÁVILA, 1998). A base de dados é percorrida em busca de regularidades, ou seja, combinações de valores para certos atributos compartilhados por fatos na mesma base. Pode-se pensar que desta forma, uma regularidade é um resumo de informação na base de dados. E, cada regularidade como uma regra, pode prever o valor de um atributo com base em outros atributos (ÁVILA, 1998). O *DM* portanto, pode ser definido como a aplicação de um algoritmo sobre um conjunto de exemplos que possuem atributos e pertencem a uma determinada classe. Esses elementos necessitam de classificação, por exemplo, segundo os valores de seus atributos e a classe a que pertence, obedecendo aos padrões encontrados automaticamente pelo atributo (ÁVILA, 1998).

Existem fatores críticos para construir um sistema de *DM*: automação e clareza, por exemplo. A automação, pois a técnica utilizada deve ser fácil de usar, o mais transparente possível, tão automatizada quanto for possível e a clareza, pois as respostas devem ser compreensíveis e fazer sentido (BERSON e SMITH, 1998).

No trabalho de Ian e Frank (2005) foi feita uma revisão histórica do aparecimento do *DM* e relatada a sua importância e evolução para usos diversos até chegar as grandes bases de dados da área da saúde. Relatou-se nesta

revisão histórica a primeira geração de *DM*, em meados dos anos 1980, com ferramentas de análise voltadas a uma tarefa, geralmente classificadores com Indução de Regras (por exemplo o C4.5) ou Redes Neurais (por exemplo o *Back Propagation*); descoberta de agrupamentos (por exemplo o *K-Means*) ou a visualização de dados (QUINLAN, 1986). A segunda geração dos algoritmos de *DM* apareceu em meados dos anos 1990, com múltiplos tipos de análise (por exemplo, o SPPS, o *Clementine*, o *Intelligent Miner*, o *SAS Enterprise Miner*) e com tarefas de classificação, agrupamentos e visualização (PIATETSKY-SHAPIRO, 1991). A terceira geração, no final dos anos 1990, apresentava interfaces orientadas ao usuário (PIATETSKY-SHAPIRO, 1999). Na quarta geração de utilização do *DM*, a interação entre o homem e a máquina é maior na condução do *KDD*. Um dos primeiros livros sobre *DM* apresenta uma coleção de trabalhos apresentados em um Workshop sobre *KDD* do final dos anos 80 (PIATETSKY-SHAPIRO, 1991). Os estudos até estas datas foram realizados para pesquisas de negócios e marketing, com técnicas para mercados, vendas e compradores, utilizando também *data warehousing* e *DM* (BERRY e LINOFF, 1997). Depois, popularizando a revisão de muitas tecnologias envolvidas com perspectivas de negócios e *DM* (DHAR e STEIN, 1997). No ano seguinte, utilizando várias aplicações de *DM* para vários laboratórios de informática (CABENA, STADLER e ZANASI, 1998). Com uma variedade de técnicas estatísticas para fazer previsões sobre o *bug* do milênio, têm-se um guia prático de *DM* (WEISS e INDURKHYA, 1998). Organizou-se a partir desta data trabalhos mais específicos com *DM* em outras áreas que não somente a de mercados, *business* e *marketing*. Utilizando informações prospectivas, com foco na descoberta de conhecimento em bases de dados de grandes corporações,

escreveu-se sobre conceitos e técnicas de *DM* (HAN e KAMBER, 2001). Neste mesmo ano um grupo internacional elaborou um livro interdisciplinar sobre *DM* (HAND, MANNILA e SMYTH, 2001).

Na área da saúde, estudos como o de PELLEGRINI e COLLAZOS (1999) descrevem a extração de conhecimentos a partir dos Sistemas de Informação, comparando as principais diferenças entre utilizar apenas os sistemas de informação em saúde como armazenadores de dados e o *DM*, que estabelece padrões de comportamento e pode facilitar o entendimento das grandes bases de dados geradas nestes sistemas. Neste estudo foi testada o *DM* em uma base de dados de recém nascidos com malformações visando estabelecer fatores de risco e foi comprovada a necessidade de melhorar a informação nos sistemas pesquisados. Dois anos depois outro trabalho refere-se à importância da integração da estatística, do *DM* e da psiquiatria para entender melhor as grandes interações clínicas, genéticas e medicamentosas nesta área da saúde (PEREIRA, 2001). Este último estudo demonstra a importância da maior interação entre o pesquisador psiquiatra e as ferramentas computacionais para melhorar o conhecimento adquirido, o que nem sempre ocorria com o estatístico que era apenas o analisador final dos dados fornecidos pelo pesquisador. É necessário, neste novo paradigma, um maior conhecimento de várias áreas a todo pesquisador voltado à confirmação nos padrões estabelecidos pelo *DM* da sua hipótese inicial.

2.2.2. TAREFAS E ALGORITMOS DE MINERAÇÃO DE DADOS

As tarefas de *DM* (MENESES e GRINSTEIN, 1998), são os fundamentos computacionais ligados à Inteligência Artificial (QUONIAM et al., 2001), que

propiciam a construção de algoritmos para possibilitar a busca de padrões escondidos nos dados (FAYYAD, SHAPIRO e SMYTH, 1996). O *DM* tem o foco na escolha dos dados, na busca e na representação dos mesmos. As principais são: Regras de Associação, Agrupamento, Classificação (BRAUNER, 2003).

A Associação é a identificação de associações entre dados relacionados ou não (AGRAWAL, IMIELINSKI e SWAMI, 1993). A associação encontra elementos que implicam a presença de outros em uma mesma transação (SRIKANT e AGRAWAL, 1995), ajudando, por exemplo, a revelar afinidades entre a clínica de vários pacientes com a mesma doença para descobrir afinidades escondidas entre esses sinais/sintomas. A associação, por exemplo, contabiliza ocorrências associadas às combinações possíveis, geralmente, em forma de texto e de maneira descritiva (REZENDE, 2005).

O Agrupamento procura a similaridade no grupo para transformação do heterogêneo em homogêneo, por exemplo, agrupar pacientes. É um método em que dados parecidos são agrupados juntos, sendo uma técnica de aprendizado não supervisionado.

A Classificação examina características do grupo e enquadra os dados em conjuntos pré-definidos. A classificação estabelece a variável que caracteriza o objetivo como sendo o atributo meta. A partir da criação de um modelo que é o “conjunto de treinamento”, com generalizações e especializações de dados, distingue classes, predizendo dados ou classes de registros não classificados automaticamente. A classificação consiste na predição de um valor categórico (REZENDE, 2005), por exemplo, predizer se o peso ao nascer é adequado para a idade gestacional ou não. Um dos objetivos da classificação é tentar encontrar uma relação entre as variáveis de previsão, ou seja, os atributos das bases de

dados, e uma determinada classe, por exemplo, o óbito infantil. Resumindo, um dos objetivos da classificação é categorizar dados de arquivos.

A tarefa de classificação foi escolhida para esse trabalho devido à sua adequação ao estudo, entendendo esta tarefa como a busca por uma função que permita associar corretamente cada registro de um banco de dados a um único rótulo categórico, denominado classe (GOLDSCHIMIDT e PASSOS, 2005). A classificação é o aprendizado de uma função que mapeia um dado em uma cadeia de várias classes conhecidas. A classificação prediz um valor categórico, e, no caso deste estudo, pode prever se os atributos meta escolhidos são realmente um bom ou mau critério relacionado ao risco de morrer, por exemplo.

Alguns dos principais exemplos de algoritmos de classificação usando árvore de decisão e indução de regras são: CART, CHAID, ID 3, C4.5, SLIQ.

O CART (*Classification and Regression Trees*) é um algoritmo de exploração e previsão de dados que usa métricas de entropia para escolher ramificações ótimas (BERSON e SMITH, 1998).

O CHAID (*Chi Square Automatic Interaction Detection*) é similar ao CART no fato de construir uma árvore de decisão, mas difere no modo como escolhe suas ramificações: pelo teste do “chi quadrado” em tabelas de contingência (BERSON e SMITH, 1998).

O ID3 é um algoritmo que classifica os objetos testando suas propriedades (BERSON e SMITH, 1998). Foi proposto por J ROSS QUINLAN em 1978 e citado em outros trabalhos (QUINLAN, 1986).

O C4. 5 surgiu como uma evolução do ID3, funcionando mesmo na ausência de valores, realizando a poda e a derivação de regras, (BERSON e SMITH, 1998), (QUINLAN, 1986). O algoritmo C4. 5 *rules* constrói regras de produção a

partir da árvore de decisão. O processo de geração de regras de produção funciona em dois estágios: regras são induzidas inicialmente de uma árvore e posteriormente refinadas. As regras com coberturas mais altas após o refinamento são apresentadas para o usuário, e as demais descartadas.

O algoritmo C4. 5 divide inicialmente iterativamente um conjunto de treinamento, até que cada subconjunto contenha casos de uma única classe. Existem diversas medidas de diversidade envolvendo as densidades de distribuição dos dados em cada classe, geradas pelo teste. O ganho de informação de cada atributo tem prevalecido como fator de escolha do atributo (GOLDSCHIMIDT e PASSOS, 2005). Os passos do algoritmo C4. 5 são:

1. Selecionar um atributo como sendo o nodo raiz.
2. Criar galhos para todos os diferentes valores do atributo selecionado no item 1.
3. Testar se todos os exemplos de treinamento sobre uma folha pertencerem a uma mesma classe. Esta folha recebe o nome de classe. Se todas as folhas possuem uma classe, o algoritmo termina.
4. Senão, o nodo é determinado com um atributo que não ocorra no trajeto da raiz, e galhos são criados pra todos os valores. O algoritmo retorna ao passo 3 (QUINLAN, 1986).

O SLIQ (*Supervised Learning In Quest*) foi desenvolvido pela equipe do projeto *Quest*, da *International Business Machines* (IBM), em 1995. Ele usa pré-classificação na fase de crescimento da árvore, calculando valores de entropia de todos os nós. Posteriormente, em 1996, a equipe criadora desenvolveu um novo algoritmo chamado *SPRINT* (*Scalable PaRallelizable Induction of Decision Trees*),

que resolveu as restrições de memória existentes no SLIQ (BERSON e SMITH, 1998).

Os algoritmos ID3, C4. 5 e SLIQ tiveram origem na inteligência artificial, enquanto o CHAID originou-se na estatística. O CART é um híbrido entre IA e estatística (BERSON e SMITH, 1998).

2.2.3. TÉCNICAS E FERRAMENTA DE MINERAÇÃO DE DADOS

As formas de representação dos resultados da mineração de dados podem ser agrupadas, por exemplo, em: árvores de decisão, indução de regras e outros (IAN e FRANK, 2005). Serão descritas apenas as utilizadas neste estudo.

Uma árvore de decisão é um modelo de uma função, no qual é determinado o valor de uma variável que pode ser discreta (codificadas em um conjunto finito de símbolos) ou contínua (conjunto de valores infinito em um intervalo). E, com base nesse valor, alguma ação é executada, para prever padrões de comportamento dos dados da base selecionada.

A árvore de decisão classifica exemplos de uma base de dados em classes, conforme observado na Figura 4. Com a árvore de decisão é possível fazer a comparação dos valores dos atributos para decidir por uma categoria.

A árvore de decisão oferece uma visão gráfica da tomada de decisão necessária, conforme observado na Figura 4. Algumas especificam que variáveis são testadas, que ações devem ser executadas e a ordem em que a tomada de decisão deverá ser executada (MARTIN e MCLURE, 1991).

Os nós das árvores são representados pelos nomes dos atributos, as ligações por possíveis valores para o atributo e as folhas com as diferentes

classes. O registro é classificado seguindo o caminho da raiz da árvore até a folha segundo as ligações ou valores dos atributos que são satisfeitos (ÁVILA, 1998).

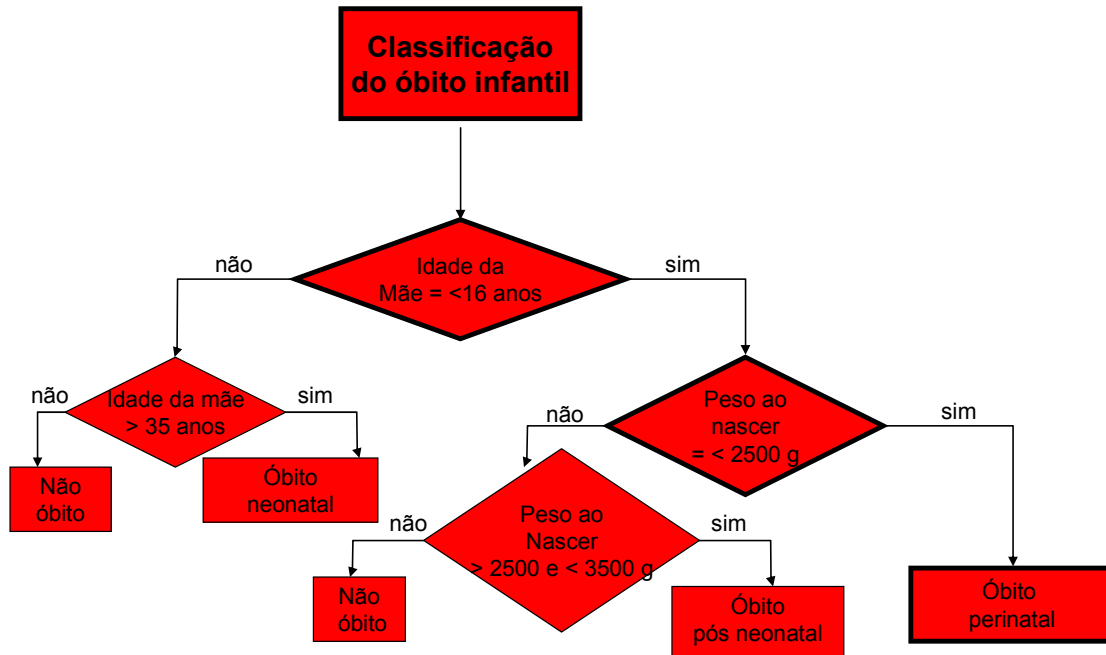


Figura 4 – Árvore de Decisão.

Árvores de decisão são modelos de previsão (classificação) que podem ser vistos exatamente como uma árvore, conforme Figura 4. Cada ramo é uma questão classificatória, e as folhas são partições do conjunto de dados com suas classificações (BERSON e SMITH, 1998).

O maior benefício desse enfoque é sua compreensibilidade, entretanto, para modelar dados com sucesso usando árvores de decisão, diversas eliminações de ramos de árvores ou podas são necessárias. A partir de uma árvore de decisão, podem ser obtidas regras caracterizadas por serem mutuamente exclusivas e coletivamente exaustivas (SHORTLIFFE e PERREAULT, 2000). A árvore de decisão é um bom método para classificação de dados ou predição de saídas

(prediz classes de amostras, levando em consideração os atributos mais relevantes), para categorizar dados de arquivos ou para gerar regras facilmente entendida, explicadas e traduzidas para linguagem natural (COLLAZOS e BARRETO, 2000). Outras facilidades na escolha da árvore de decisão é que ao escolher e apresentar os atributos em ordem de importância, permite aos usuários conhecer quais fatores mais influenciam seus trabalhos; elas também não assumem qualquer distribuição particular para os dados e os atributos podem ser categóricos (qualitativos) ou numéricos (quantitativos).

As regras de classificação são uma alternativa para a apresentação dos resultados, além das árvores de decisão. Lê-se um conjunto de regras diretamente de uma árvore de decisão. Uma regra é gerada para cada folha. O antecedente da regra inclui uma condição para cada nó na rota (vem geralmente concatenado com e ou *and*), desde a raiz até a folha, e o conseqüente da regra (vem geralmente concatenado com o ou *or*) é a classe assinalada pela folha. São utilizadas juntas como um conjunto. O que pode ser observado no exemplo a seguir:

Ex.: **Se** peso = 1500g

e semanas de gestação > 37,

então pequeno para a idade gestacional.

Regras de classificação são do tipo proposicional. Envolvem testar o valor de um atributo contra uma constante. As regras de classificação podem ter exceções. Modificações incrementais podem ser feitas a um conjunto de regras expressando exceções, ao invés de reconstruir o conjunto todo. Conforme o exemplificado a seguir:

Ex.: **Se** peso > 1500g **e** peso < 2500g,

Se gestação > 37 semanas

Então pequeno para idade gestacional

Senão prematuro.

Em muitas tarefas de classificação, regras proposicionais são suficientemente expressivas para descrições de conceitos com concisão e precisão. Entretanto, há situações onde uma forma mais expressiva de regra poderia prover uma descrição de conceitos mais intuitiva e concisa. São situações que envolvem relacionamentos entre exemplos, como o que segue:

Ex.: **Se** peso maior que altura,

então obeso.

Senão normal ou longilíneo.

Os valores dos atributos peso e altura não são importantes, apenas o resultado da comparação dos dois. Regras, desta forma são chamadas relacionais, porque expressam relacionamentos entre atributos, ao invés de proposicionais, que simbolizam um fato sobre apenas um atributo. A classificação faz a derivação de regras que classificam amostras similares em grupos, e a aprendizagem de uma função que mapeia um dado em uma de várias classes conhecidas. A classificação prevê um mapeamento prévio a partir de atributos para grupamentos especificados (GROTH, 1998). É também denominado aprendizado supervisionado, porque a entrada e a saída desejadas são fornecidas previamente por um supervisor externo (FAUSETT, 1994). Ex.: grupamentos de pessoas com dois anos ou menos podem ser mapeados para a categoria bebê. As outras classificações de pessoas podem ser como crianças, adultos, idosos. Resumindo, as regras do passado geram possibilidades para entender o futuro, sendo condicionadas à base (são probabilidades condicionais).

A ferramenta *WEKA* tem alguns pacotes, utilizando o algoritmo de classificação (com a implementação de algoritmos de aprendizagem supervisionada). Esta ferramenta é um método determinístico que dispõe de métodos de meta aprendizagem, no qual são construídos classificadores com sucessivos e independentes conjuntos de amostras de dados. Ao induzir o primeiro classificador, todas as instâncias são equiprováveis, isto é, tem o mesmo peso. Após ter induzido o primeiro classificador, os pesos das instâncias de treinamento classificadas incorretamente são alterados baseados nos classificadores que foram construídos anteriormente (WITTEN e FRANK, 2000). A ferramenta tem como características: acesso a fontes de dados heterogêneas, facilidade para inclusão de novos métodos e de novas operações, recursos para planejamento de ações, estruturas para armazenamento de modelos de conhecimento e estruturas para armazenamento de histórico de ações (GOLDSCHIMIDT e PASSOS, 2005).

Desta forma, a descoberta do conhecimento em bases de dados, na qual uma parte desta descoberta é a mineração de dados, é uma tecnologia que deve ser considerada pelos profissionais de todas as áreas para facilitar o trabalho de atuação frente à grande quantidade de informações contidas nas bases de dados, principalmente, na área de saúde, visto seu poder de transformação de dados em conhecimento útil e utilizável.

2.3. SISTEMAS DE INFORMAÇÃO EM SAÚDE

Sistema de Informação (SI) consiste nas técnicas de aquisição, organização, armazenamento e recuperação de informações (BARRETO, PELEGRINI e COLLAZOS, 2000) e (BARRETO, FALQUATO e BORGES, 2000a). Os SI são

constituídos por um conjunto de dados com atributos relevantes e disponíveis (VAN BEMMEL e MUSEN, 1997). Estes dados vêm de uma ou várias fontes e são armazenados em uma memória permanente, utilizando regras de combinação de informações sumarizadas e com visões diferentes sobre os dados manipulados previamente (SHORTLLIFE e PERREAULT, 2000).

Os Sistemas de Informação em Saúde podem ser classificados em administrativos, clínicos (prontuário eletrônico do paciente, sistemas de apoio à decisão, sistemas de recuperação de imagem) e epidemiológicos. Também podem ser classificados em públicos e privados.

O conhecimento atualizado das condições de saúde da população decorre da realização de estudos e análises das informações disponíveis, especialmente às referentes aos indicadores básicos de saúde selecionados para acompanhamento periódico nos Sistemas de Informação em Saúde Pública (MEDRONHO, 2003). A precisão desse conhecimento depende, em grande parte, da qualidade dos dados gerados nos sistemas de informação de saúde, fato que pode ser influenciado por fatores técnicos, administrativos e operacionais.

Um bom sistema de informações depende da periodicidade do fluxo de fornecimento dos dados e do criterioso preenchimento dos instrumentos de coleta (fichas de notificação e investigação, declaração de óbito, declaração de nascido vivo, boletins de atendimento, prontuários médicos, relatórios, entrevistas). A transformação desses dados (valor quantitativo obtido para caracterizar um fato ou circunstância) em informações (análise descritiva dos dados) pode ser feita em todos os níveis do sistema de saúde. Para isso, faz-se necessário organizá-los em tabelas e gráficos, que, conforme a complexidade pode ser realizada por todos os profissionais de saúde.

Entre os principais SI em Saúde Pública, alguns dos mais utilizados são o SINASC e o SIM.

2.3.1. SISTEMA DE INFORMAÇÃO SOBRE NASCIDOS VIVOS

Até 1990, os métodos para quantificar os nascidos vivos eram manuais, dificultando a execução e comprometendo a qualidade. Não havia um padrão operacional uniforme, coordenado e integrado com os diversos órgãos envolvidos, como hospitais, cartórios e instituições que elaboravam as estatísticas de saúde (MINISTÉRIO DA SAÚDE, 2005). As informações utilizadas apoiavam-se em levantamentos aleatórios e bases comparativas, (LAURENTI et al., 1987).

A partir de então, o MS começou a implantar o SINASC (MINISTÉRIO DA SAÚDE, 1999). O objetivo deste sistema era registrar os nascidos vivos, a partir de um documento básico: a declaração de nascido vivo, gerada nos hospitais, em outras instituições de saúde que realizam partos e nos cartórios do Registro Civil, para os partos ocorridos no domicílio (MELLO JORGE et al., 1993). Consideravam-se essas instituições como os locais que poderiam prover as mais completas e corretas informações relativas aos nascimentos, visto que cerca de 80% dos partos no Brasil são hospitalares (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 1999).

Assim, desde 1990, o SINASC mantém, em âmbito nacional, um banco de dados importante sobre os recém-nascidos. Criado através da coleta de dados sobre o nascimento e o parto e da geração de dados populacionais, caracteriza epidemiologicamente os nascimentos e disponibiliza essas informações a todos os níveis dos sistemas de saúde, segundo o (MINISTÉRIO DA SAÚDE, 1995) e (MINISTÉRIO DA SAÚDE, 1999).

A implantação do SINASC, de forma lenta e gradual, disponibilizou o primeiro Sistema de Informações de Estatísticas Vitais desenvolvido para microcomputadores desde sua criação, em 1990 (FUNDAÇÃO NACIONAL DE SAÚDE, 2001a).

O sistema informatizado e de fácil instalação teve sua operação simplificada pelas rotinas auto-explicáveis, possibilitou a elaboração de relatórios de frequência e de cruzamentos entre variáveis e permitiu o manuseio dos bancos de dados em instância estadual. No Paraná, o SINASC foi implantado gradualmente, de maneira regionalizada e descentralizada a partir de 1994. Em 1998, foi desenhada uma nova versão da Declaração de Nascido Vivo (DN), com novo aplicativo informatizado e, em 2005, implementada uma nova versão para 2006 (FUNDAÇÃO NACIONAL DE SAÚDE, 2001b).

A DN é composta por sete blocos, contendo quarenta e uma variáveis: Bloco I - Cartório. Bloco II - Local de Ocorrência. Bloco III - Mãe. Bloco IV - Gestação e parto. Bloco V - Recém-nascido. Bloco VI - Identificação. Bloco VII - Responsável pelo preenchimento, conforme indicado no Anexo 1.

Outro SI em Saúde Pública importante no Brasil é o SIM.

2.3.2. SISTEMA DE INFORMAÇÃO SOBRE MORTALIDADE

O SIM foi criado pelo MS em 1975, para obter dados sistematizados sobre a mortalidade que, de modo íntegro e abrangente pudessem embasar os diversos níveis gerenciais em suas ações de saúde. Foi o primeiro sistema em informações de estatísticas vitais desenvolvido em microcomputadores no país (MINISTÉRIO DA SAÚDE, 1999). Implantado devido à necessidade de conhecimento real sobre os óbitos no Brasil, enfrenta dificuldades de

preenchimento das informações até os dias atuais (LAURENTI, HARAKI e GOTILIEB, 2005). No Paraná, passa a ser informatizado desde 1979, quando ocorre também a descentralização e regionalização do processo em todo o estado. Atualmente, uma nova versão está em testes e implantação (FUNDAÇÃO NACIONAL DE SAÚDE, 2003).

Esse sistema visa produzir estatísticas de mortalidade e construir os principais indicadores de saúde, permitindo, assim, a realização de estudos não apenas sob o ponto de vista estatístico epidemiológico, mas também sócio-demográfico (MINISTÉRIO DA SAÚDE, 1999). O documento padrão do SIM é a Declaração de Óbito - DO, padronizada em todo o território nacional, conforme indicado no Anexo 2.

As causas das mortes são preenchidas pelo médico e, posteriormente, recebem um código segundo a Classificação Estatística Internacional de Doenças e problemas relacionados à saúde - CID, no "Modelo Internacional de Certificado Médico da Causa de Morte", utilizado em todos os países e recomendado pela Assembléia Mundial de Saúde, desde 1948 (MINISTÉRIO DA SAÚDE, 1999). De 1979 a 1995, utilizava-se a CID-9 e, a partir de 1996, passou-se a utilizar a CID-10 em todo o território nacional.

A Lei dos Registros Públicos do Brasil de 1976 determina no seu Artigo 77 que nenhum sepultamento será feito sem Certidão Oficial de Registro expedida no lugar do falecimento e, quando se tratar do óbito de crianças com menos de um ano, o oficial verificará se ela foi devidamente registrada, em caso contrário, o documento de registro de nascido vivo será previamente feito.

O número da Declaração de Óbito é previamente atribuído e grafado na primeira linha (FUNDAÇÃO NACIONAL DE SAÚDE, 2001). A DO é composta por

nove blocos, com um total de sessenta e duas variáveis: Bloco I - Cartório. Bloco II - Identificação. Bloco III - Residência. Bloco IV - Ocorrência. Bloco V - Óbito Fetal ou Menor de um Ano. Bloco VI - Condições e Causas do Óbito. Bloco VII - Médico. Bloco VIII - Causas Externas. Bloco IX - Localidades sem médicos, conforme indicado no anexo 2.

Outro SI em Saúde Pública importante, e que veio facilitar o conhecimento sobre os óbitos infantis criado no Estado do Paraná, foi o SIMI.

2.3.3. SISTEMA DE INVESTIGAÇÃO DA MORTALIDADE INFANTIL

Este sistema foi criado e implantado na Secretaria de Estado da Saúde do Paraná (SESA-PR) em 2000 para facilitar a coleta e análise dos dados elaborados pelos Comitês de Prevenção da Mortalidade Infantil Municipais, Regionais e Estadual que atuam no estado desde 1997. Foram elaboradas fichas de análise de óbitos pelo Comitê Estadual de Prevenção da Mortalidade Infantil (CEPMI). Estas auxiliam a coleta inicial dos dados nos municípios, os quais repassam as informações às Regionais de Saúde do Estado, que transmitem estes dados via intranet para a SESA-PR. A investigação da mortalidade infantil é facilitada pelo SIMI, que vêm auxiliando no recebimento dos dados de maneira mais ágil, irrestrita e visual no estado (VIANNA, 2005).

Os critérios estabelecidos inicialmente para seleção dos óbitos a serem investigados foram: nascidos vivos com mais de 1500 gramas que morreram no primeiro ano de vida, com exceções das anomalias congênitas. A criação de Comitês de Prevenção de Mortalidade Materna e Infantil (MANSANO et al, 2004) é considerada importante para auxiliar a especificação das causas básicas de óbitos nestas populações, analisando e propondo ações para melhoria do

preenchimento dos atestados de óbito, das declarações de nascidos vivos e de ações para redução desses óbitos, propondo também a mudança na estratégia de coleta de informações com alteração do perfil epidemiológico de mortalidade (SOUZA e TOSETTO, 2005).

Os Comitês Municipais procedem às investigações dos óbitos infantis a partir da Vigilância Epidemiológica, utilizando como metodologia a busca ativa desses óbitos. Os dados são levantados por meio dos prontuários hospitalares, onde ocorreu o nascimento e o óbito nos casos neonatais, e ambulatoriais, quando houver; visita domiciliar; declaração de óbito; declaração de nascido vivo; carteira de pré-natal da gestação e de imunização da criança e da gestante; informações de agentes comunitários e em entrevistas domiciliares e em cemitérios clandestinos.

Os documentos necessários para análise final do óbito no Comitê Regional são reunidos, anexados e conferidos de acordo com um *check list*. O CEPMI elaborou uma ficha para padronização da análise do óbito, de forma que o processo pudesse ser informatizado. A ficha informatizada pode ser visualizada na figura 7, no Anexo 5.

Nesta ficha constam os seguintes dados: Comitê de Análise do Óbito, Data de Encerramento da Análise, Regional de Saúde e Município de residência, Número da Declaração de Óbito (DO), Número da Declaração de Nascido Vivo (DN), Sexo, Análise do Óbito, Causa básica do óbito na DO, Causa básica do óbito após a investigação, Morte Evitável, Critérios de evitabilidade, Responsabilidade, Medidas de prevenção, Membros presentes e profissão. Os dados estão descritos no Anexo 6. No Sistema de Investigação da Mortalidade Infantil (SIMI), ao todo, são vinte e dois relatórios e seis gráficos.

Em resumo, os Comitês analisam os óbitos infantis e buscam determinar sua evitabilidade, baseando-se em critérios pré-estabelecidos segundo o Manual de Criação dos Comitês de Prevenção de Óbito Infantil e Fetal (MINISTÉRIO DA SAÚDE, 2004a). Na figura 8, no Anexo 6, observa-se o fluxograma resumido da metodologia do sistema de vigilância dos óbitos infantis no Estado do Paraná.

O CEPMI empenha-se no cumprimento de dois objetivos principais:

- Identificar e investigar as causas de óbito no primeiro ano de vida com sua evitabilidade (hoje redutibilidade);
- Traçar e propor medidas preventivas, de promoção à vida, baseadas na maior especificidade das causas básicas dos óbitos e sua prevenção.

Desta forma, nos últimos anos o SIMI vem colaborando com os comitês para execução de seus objetivos. Os dados destas análises e as facilidades do SIMI encontram-se em diversos trabalhos publicados (VIANNA, 2004), (VIANNA, 2004a), (VIANNA, 2005) e (VIANNA e MORO, 2006).

3. MATERIAL E MÉTODOS

Os dados utilizados neste estudo compreendem o período de 2000, quando houve a implantação do Sistema de Investigação da Mortalidade Infantil (SIMI), até 2004, quando os dados já estavam consolidados antes da realização deste estudo.

Para identificar as características relacionadas à Mortalidade Infantil, utilizando o *KDD*, realizou-se a integração de três bases de dados: do SINASC, do SIM e do SIMI, completando-se informações existentes isoladamente em cada um dos SI. Esta integração corresponde ao primeiro objetivo específico.

O próximo passo foi o pré-processamento ou a preparação da base de dados foi realizada com a seleção dos dados considerados importantes para a Mortalidade Infantil. Além da categorização, que consiste na definição de classes para um atributo, como a idade e o apgar, por exemplo. A idade do óbito foi dividida em 4 categorias: perinatal (óbito antes de um dia após o nascimento), neonatal precoce (óbito entre 1 e 6 dias após a partir do nascimento), neonatal tardio (óbito entre 7 e 28 dias após a partir do nascimento), e tardio (entre 29 e 360 dias). Esta preparação foi estabelecida conforme a necessidade relacionada pelos três especialistas (considerados como grupo 1 de especialistas) em mortalidade infantil das secretarias estaduais e do município de Curitiba, para esclarecer melhor os atributos. Os outros especialistas que também foram consultados para avaliar as principais regras selecionadas são os que trabalham analisando os óbitos infantis nos Comitês de Prevenção da Mortalidade Infantil, sendo vinte e dois no total, um de cada Regional de Saúde do Estado (considerados como grupo 2 de especialistas). Nesta fase foram consultados os do grupo 1 para auxiliarem na escolha dos melhores atributos, e, na fase final do

processo de *KDD* foram consultados os do grupo 1 e do grupo 2 de especialistas, para auxiliarem na validação da ferramenta. Estes especialistas foram escolhidos por estarem habituados a analisarem os óbitos infantis e, desta forma, poder inferir se as escolhas dos padrões gerados pela ferramenta eram válidas e significativas.

Também foi categorizado um novo atributo para a alteração das CID. Observou-se o que foi alterado em capítulo (tendo 21 capítulos na CID-10) e categorias que variam em número segundo o capítulo e subcategorias ou classes propostos por MORO (2003).

No pré-processamento incluiu-se também a limpeza dos dados para o *DM*, conforme ilustrado na Figura 5. Foram corrigidas, por exemplo, datas de óbitos que ocorreram antes do nascimento, com a realização da troca das mesmas para utilização posterior desses dados (SETH, GUATAM e BALINT, 2002).

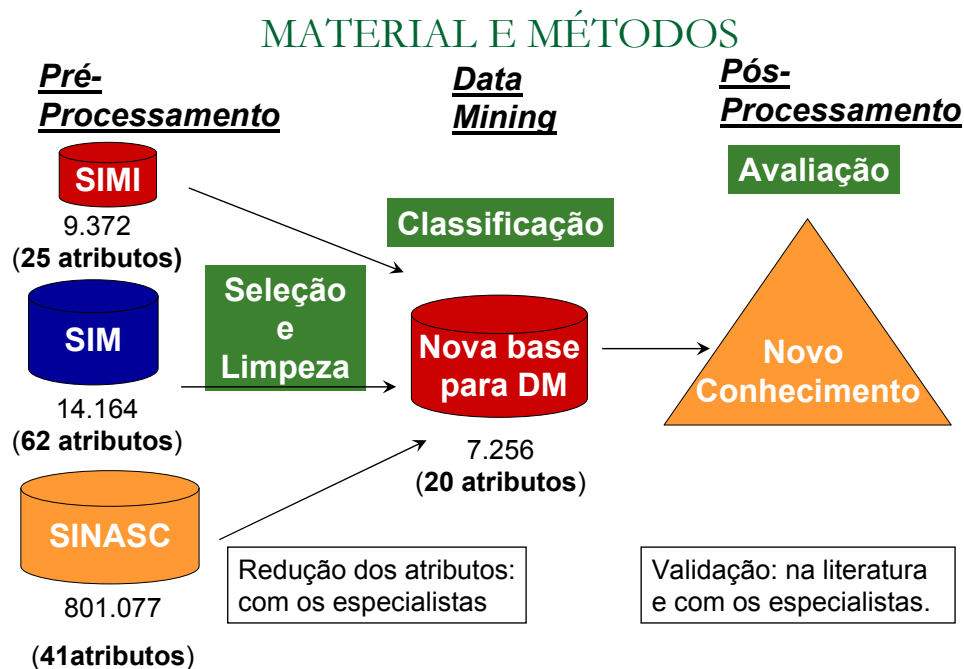


Figura 5 – Material e Métodos.

A fase seguinte do processo foi o *DM*, com a tarefa de classificação e a representação de regras, com o algoritmo C4.5 que pudessem ser analisadas pelos especialistas em mortalidade infantil. Como o processo de *DM* é iterativo, foram realizados vários experimentos com os atributos metas da evitabilidade dos óbitos e das alterações entre a causa do óbito na DO e após as análises dos Comitês. Durante este processo, o estudo foi avaliado pelos especialistas do grupo 1, para as escolhas das melhores opções das variáveis submetidas à mineração. Desta forma, a evitabilidade dos óbitos e a alteração entre as CID na DO e após as análises dos Comitês foram os atributos meta selecionados porque os objetivos principais dos Comitês são a caracterização da evitabilidade dos óbitos e a especificação da causa óbito, de maneira que se possa atuar em sua redução. Nesta fase, cumpriu-se o segundo objetivo específico do estudo: conhecer melhor as características relacionadas à mortalidade infantil.

Foi utilizada a tarefa de classificação que verifica os fatores e suas relações com o atributo meta, e, como eles influem no atributo meta para observar a ocorrência de novas características relacionadas à Mortalidade Infantil. A classificação utiliza dados do passado (treinamento) para induzir regras para o futuro, isto é, regras para prever o valor do atributo meta, objetivamente, preveem o futuro de um problema. Na classificação, ocorre a predição de uma classe de valor, que determina o atributo meta a partir dos valores dos atributos preditores. Pode-se dizer que a base de dados é percorrida em busca de regularidades (regras) ou combinações de valores para certos atributos. No aprendizado supervisionado a classe e sua descrição formam a regra onde:

Se “descrição” da regra,

então “pertence a esta classe”. Por exemplo:

Se peso $\leq$ 1500 g,
então baixo peso ao nascer.

Após cada uma destas tarefas de mineração de dados seguiu-se a avaliação dos especialistas do grupo 1 buscando relacionar os padrões encontrados no *DM* com a literatura existente e com a experiência dos mesmos em analisar óbitos infantis. Após a geração dos padrões foi elaborado um questionário com as melhores regras selecionadas pelos especialistas do grupo 1 que foram apresentadas então, aos especialistas do grupo 1 e do grupo 2 para suas análises e interpretação da importância ou não das regras geradas pelo *DM*, conforme observado na Figura 5.

Passou-se então, a última fase do processo: o pós-processamento, com a interpretação das regras, confirmando-se os conhecimentos gerados pelo *DM* com os existentes na literatura sobre a Mortalidade Infantil. Conclui-se, então, o terceiro objetivo específico que era identificar as relações existentes entre os padrões estabelecidos na literatura e os encontrados na pesquisa. A seguir, estão descritas com maior detalhamento as fases do processo do KDD, que podem ser observadas na Figura 5.

3.1. FASE DE PRÉ-PROCESSAMENTO

No pré-processamento foram selecionadas as três bases, com complementação dos dados existentes em cada uma delas para melhorar as informações coletadas e que não existiam nas análises dos Comitês de Prevenção da Mortalidade Infantil (CPMI). Estes Comitês só utilizam as informações do SIMI, faltando outras como, por exemplo, o apgar, o peso ao nascer, a raça/cor, a escolaridade da mãe, que só existem no Sistema de

Informação da Mortalidade (SIM) ou no Sistema de Informação de Nascidos Vivos (SINASC). Algumas destas características podem ser encontradas de maneira textual no SIMI, embora deveriam ser preenchido segundo um padrão pré-estabelecido durante o treinamento para as análises dos óbitos, o que nem sempre ocorre.

Após estes procedimentos iniciais procedeu-se a limpeza, enriquecimento e normalização destes dados, finalizando com a construção dos atributos baseados na integração das três bases de dados, conforme observado na Figura 5.

3.1.1. SELEÇÃO DOS DADOS

Para conhecer as informações obtidas do nascimento até o óbito das crianças, foi elaborada uma base de dados contendo as análises dos óbitos infantis contidas no SIMI, com as informações de nascimento contidas no SINASC e dos óbitos existentes no SIM. As três bases de dados escolhidas foram selecionadas recebendo para tanto o consentimento declarado da chefia imediata do Centro de Informação e Diagnóstico em Saúde (CIDS) ou Vigilância Epidemiológica, que consta no Anexo 8. Este trabalho foi aprovado no Comitê de Ética em Pesquisa (CEP) da PUCPR e pertence ao Grupo III, sob o registro no CEP número 957, conforme Anexo 9.

Para realizar a seleção dos dados foi preciso entender como funcionam as análises dos óbitos infantis nos CPMI. Historicamente, foi estabelecida como meta entre a SESA-PR, Regionais de Saúde e as Prefeituras Municipais de Saúde a análise de pelo menos 60% dos óbitos infantis ocorridos. Existe, portanto, uma diferença entre os óbitos infantis existentes no SIM (registro de todos os óbitos infantis selecionados na base geral da SESA-PR) e os óbitos infantis constantes

no SIMI, que totalizaram 66% dos óbitos infantis analisados no período deste estudo (sendo, portanto, superada a meta de análise proposta anteriormente). A partir dos óbitos analisados via SIMI, em um total de 9372, fez-se a inter-relação destas crianças com as bases de dados do SIM, que nos 5 anos cobertos no estudo, totalizam 14.164 e com o SINASC, do mesmo período, com 801.077 nascidos vivos, conforme observado na Figura 5.

Para a ligação das três bases de dados, utilizou-se o campo comum que era o número da DN e posteriormente o nome da mãe, chegando-se inicialmente a uma base de 3.933 casos. Realizaram-se várias integrações para aumentar o número de casos até que se chegou a uma base de 7256 casos, ou seja, 77% da base.

No processo de integração das bases, não foram considerados alguns casos do ano 2000 devido ao fato que as análises de óbitos deste ano contêm números de DN de 1999 em 2% da base (228), ficando impossível a integração das bases sem a característica comum escolhida que era o número da DN. Também não foram considerados 4% dos registros de 2004, pois são casos que foram a óbito, mas ainda não foram analisados quando da realização deste estudo. Durante o processo da seleção dos dados, concluiu-se que 2% (266) das DN não foram consideradas na junção das bases devido a erros que impossibilitaram esta integração; 2% (214) das DN estavam em branco, mesmo depois de reiteradas críticas (correções) no sistema; 2% (215 a 265) das DN e DO estavam com números diferentes do habitual.

Inicialmente, mantiveram-se dados como os códigos dos municípios e regionais de saúde, considerados importantes para a mineração. Desta forma, houve uma redução de valores nominais que, após uma avaliação pelos

especialistas do grupo 1, foram retirados devido à incongruência de alguns dados e ao parecer do Comitê de Ética.

3.1.2.LIMPEZA DOS DADOS

Foram realizadas várias seleções e identificações dos atributos, manuseio das bases de dados com testes e relacionamentos para averiguar qual a melhor seleção para posterior mineração. A primeira seleção dos atributos para conhecer melhor quais variáveis constam em cada base pode ser observada na Tabela 12, no Apêndice 1, onde “s” é utilizado para sim (existe na base) e “n” para não (não existe na base).

Durante esta limpeza foram corrigidos, por exemplo, valores de datas expressos de forma diferente do padrão dd/mm/aaaa.

3.1.3. ENRIQUECIMENTO E CONSTRUÇÃO DOS ATRIBUTOS

A seleção dos atributos foi realizada pelos especialistas do grupo 1, na qual se utilizam dados relacionados à melhor identificação dos riscos possivelmente relacionados a esse evento.

Foram retirados os atributos identificadores como nome da criança, nome da mãe, rua de residência, rua do óbito, código local de atendimento, conforme critérios estabelecidos pelo código de endereçamento postal. O atributo de qualidade da ficha que aparece na base do SIMI foi retirado porque a resposta costuma vir, em 90% dos casos, como “boa qualidade”, devido ao fato de a análise ser realizada pelos próprios preenchedores da ficha, tornando a resposta tendenciosa e de pouco valor para o trabalho em questão.

Descobriram-se formas diferentes no conteúdo da escrita dos atributos, mesmo nas mesmas bases de dados e em anos diferentes. Selecionou-se a escrita que mais aparecia e escreveram-se todos os atributos da mesma forma, excluindo-se os identificadores de local, códigos identificadores, atributos novos que entraram nas novas bases de DO de 2004.

Para enriquecimento e nova estratificação da base, obteve-se, por exemplo, a categorização das ocupações fornecidas pelo Ministério da Saúde, baseadas no Instituto Brasileiro de Geografia e Estatística (IBGE), juntamente com o programa do SIM/SINASC e especificadas na Tabela 13, no Apêndice 2, sendo categorizadas para o estudo em duas categorias de profissões. Posteriormente, este atributo foi retirado porque, em 72% da base (5227), os casos ficaram com a ocupação dona de casa ou doméstica, não identificando a profissão como característica importante, segundo análise dos especialistas do grupo 1 consultados. Também ocorre um viés de seleção nessa categorização, porque a dona de casa “em tese” tem condições de proporcionar melhores cuidados ao seu filho do que a doméstica que sai para trabalhar e deixa seu filho aos cuidados de terceiros que nem sempre irão fazê-lo de maneira adequada. E, na categorização do IBGE, as duas categorias aparecem agrupadas, sem quaisquer diferenciações.

3.1.4. NORMALIZAÇÃO DOS DADOS

Nesta nova base fez-se nova seleção e redução de dados dos atributos importantes (REZENDE, 2005), (WEISS, 1998) e (GLYMOUR et al., 1997). Categorizando-se a idade, por exemplo, conforme Tabela 14, no Apêndice 2, que são as categorias adotadas pela Organização Mundial de Saúde

(ORGANIZAÇÃO MUNDIAL DE SAÚDE, 2004), Ministério da Saúde (MINISTÉRIO DA SAÚDE, 2004) e Sociedades Médicas.

Com estas novas normalizações dos dados descobriu-se, por exemplo, que, apesar das críticas dos Sistemas de Informações (correções realizadas durante a construção da base), ainda existem dados errôneos como, por exemplo, natimortos (48 casos de crianças que não respiraram) com DN. A dúvida é: ocorreu erro de digitação ou erro de identificação ao nascer? Na seleção do atributo por idade, foi preenchida a idade erroneamente em cinco casos. Como o número de casos foi pequeno, optou-se pelo preenchimento estimado da data, seguindo a imputação de dados já relatados em trabalhos anteriores de Edwards e Khuri (KLUCK, 2004).

Para a normalização de todos os atributos foi feita a distribuição com a CID na DO e após a investigação. Não foram consideradas 15 CID (com numeração diferente do padrão, nos quais foi feita a aproximação para CID válido), ficando a base final com 7256 casos de CID válidos, distribuídos conforme está explícito na Tabela 15, no Apêndice 2. Nas identificações após a investigação dos Comitês, não foram consideradas mais 25 casos com CID não preenchido, permanecendo, então, 7214 CID válidos. Realizou-se, posteriormente, a aproximação da mesma forma, com mudanças na CID em 49% dos casos antes e após as análises, conforme segue na Tabela 16, no Apêndice 2.

Nesta distribuição de casos na DO e após as análises, 30% dos casos, por exemplo, foram melhor especificados, como os do grupo R (sintomas, sinais e achados anormais não classificados em outra parte). Após as análises, registraram-se outros grupos com diagnósticos de óbitos, por exemplo, de afecções maternas, mais específicas e de melhor atuação para sua redução, o

que coincide com os estudos realizados anteriormente pelos Comitês de Prevenção de Mortalidade Infantil do Estado do Paraná (VIANNA, 2004), (VIANNA 2004a), (VIANNA, 2005).

O atributo tempo de análise foi retirado porque se concluiu que este teria mais interesse para os serviços, fugindo do escopo deste trabalho. A categorização da idade da mãe, por exemplo, baseou-se nos critérios de risco biológico (MINISTÉRIO DA SAÚDE, 2004) e (FREITAS et al., 2001), dividida em $< = 16$ anos, entre 16 e 35 anos e $> = 35$ anos, conforme a Tabela 18, no Apêndice 3. O peso ao nascer foi categorizado em $< = 2500$ g, entre 2500 e 3500g e $> = 3500$ g, segundo estabelecido pelas sociedades de classe médica e pediátrica, especificado na Tabela 30, no Apêndice 3. Os registros que estavam sem preenchimento do peso foram categorizados como zero (RUMEL et al., 1992) e (MISRA et al., 2004).

Optou-se por nova categorização pela CID, com base nos especialistas da SESA-PR e nos resultados da CID mais encontrados na DO. A categorização das CID mais encontradas após a investigação dos Comitês, está apresentada na Tabela 16, no Apêndice 2.

Como o atributo apgar aparece na base com 11 categorias, optou-se por reduzi-lo para apenas 2 categorias, o que corresponde também, com a opinião dos especialistas do grupo 1 consultados, que concordaram que de 0 a 5 são os recém-nascidos que nascem mal e que, de 6 a 10, são os que nascem melhor. Visando um melhor desempenho do atributo minerador, foi possível reduzir o apgar para apenas estas 2 categorias, conforme a Tabela 27 no Apêndice 3 para o primeiro minuto e a Tabela 28, no Apêndice 3 para o quinto minuto. Concluiu-se também ser necessária a utilização da categorização dos critérios de evitabilidade

em grandes grupos A até I, sem suas subcategorias, conforme a Tabela 10, no Anexo 3.

A idade da mãe também foi categorizada conforme a Tabela 18, do Apêndice 3. O estado civil e a escolaridade da mãe já estão categorizados segundo o SIM/SINASC e aparecem nas tabelas 19 e 20, respectivamente, no Apêndice 3. Outra categorização baseada nas Sociedades médicas de classe foi a idade no óbito, que aparece especificada na Tabela 31, no Apêndice 3.

Após a normalização dos dados, foi feita uma nova seleção dos atributos com a validação dos especialistas do grupo 1, isto é: eles selecionaram o que consideravam importante segundo sua experiência após a explicação da necessidade de uma escolha efetiva para melhorar o desempenho da ferramenta computacional definida. Com isto definiram-se 20 atributos, conforme Tabela 4.

Tabela 4 - Seleção dos atributos da base para DM:

	* NOME DOS ATRIBUTOS	* DESCRIÇÃO DOS ATRIBUTOS
1	ClasseCodCID	Código da CID na DO (base do SIMI).
2	ClasseCodCIDInv	Código da CID após investigação (base do SIMI).
3	CodTipoMorteEvitavel2	Código do tipo de morte evitável, por grupos A até I (base do SIMI).
4	ClasseCodEvitabilidade	Código de evitabilidade (atributo meta, base do SIMI).
5	analiseobitoSEXO	Sexo em: masculino, feminino, ignorado (base do SIMI e confirmado na base do SIM)
6	ClasseIDADEMAE	Categorização da Idade da mãe em: <16 , 16 a 35 e >35 anos (base do SIM e do SINASC).
7	ESTCIVMAE	1-Solteiro, 2-Casado, 3- Viúvo, 4- Separado Judicialmente, 5- União consensual, 9- Ignorado (base do SINASC).
8	ESCMAE	1- Nenhuma, 2- De 1 a 3, 3- De 4 a 7, 4- De 8 a 11, 5- 12 e mais, 9- Ignorado (base do SIM e do SINASC).

9	QTDFILVIVO	Quantidade de filhos vivos (base do SIM e do SINASC).
10	QTDFILMORT	Quantidade de filhos mortos (base do SIM e do SINASC).
11	GESTACAO	Categorizado em: 1- Menos de 22 semanas, 2- De 22 a 27 semanas, 3- De 28 a 31 semanas, 4- De 32 a 36 semanas, 5- De 37a 41 semanas, 6-42 e mais, 9- Ignorado (base do SIM).
12	GRAVIDEZ	1-Única, 2-Dupla, 3-Tripla e mais, 9- Ignorada (base do SIM e do SINASC).
13	PARTO	1-Vaginal, 2-Cesário, 9- Ignorado (base do SIM e do SINASC).
14	CONSULTAS	Categorizado em: 1, 2-4, 5-6, + 6, Ignorado e Sem preenchimento (base do SINASC).
15	ClasseAPGAR1	Nota de viabilidade dada no primeiro minuto de vida. Categorizado em: de 0 a 5 e de 6 a 10 (base do SINASC).
16	ClasseAPGAR5	Nota de viabilidade dada no quinto minuto de vida. Categorizado em: de 0 a 5 e de 6 a 10 (base do SINASC).
17	RACACOR	1-Branca, 2-Preta, 3-Amarela, 4-Parda, 5- Indígena (base do SIM e do SINASC).
18	ClassePESO	Categorizado em gramas, ao nascer: <2500g, de 2500 a 3500g, >3500g e sem preenchimento (base do SINASC).
19	IdadeClasse	Categorização da idade no óbito: <1 dia, de 1 a 6 dias, de 7 a 28 dias e de 29 a 365 dias (base do SIM).
20	Distancia CID	Capítulo: se alterou capítulo da CID10, Grupo: se alterou o grupo da CID10, Categoria: se alterou a categoria da CID10, Subcategoria: se alterou a subcategoria da CID10. (gerado para o trabalho e baseado no SIMI).

* O NOME e a DESCRIÇÃO DOS ATRIBUTOS foram grafados conforme aparecem na linguagem computacional das bases.

Após esta fase do pré-processamento iniciou-se o *DM* propriamente dito.

3.2. MINERAÇÃO DE DADOS

O *WEKA* (*Waikato Environment for Knowledge Analysis*) é uma ferramenta de mineração de dados, escolhida para esse trabalho devido à sua facilidade de instalação e do seu ambiente de exploração, pelas suas aplicações que podem ser automatizadas e também tem opções abertas para a pesquisa e utilização. Está disponível na internet (<http://www.cs.waikato.ac.nz/~ml/weka/>), e sua implementação feita em Java torna o sistema portátil. O algoritmo utilizado foi o C4. 5 *rules*, chamado J48 no ambiente WEKA, no qual o modelo é baseado num conjunto de dados de treinamento, posteriormente utilizado para classificar as instâncias do conjunto de testes.

Também interferiu nessa escolha a existência de um grupo de trabalho realizando outros estudos com a tarefa de classificação, com o algoritmo minerador C4. 5 (CARVALHO, 2002) e (STEINER et al., 2006). E também, na SESA já foi realizado trabalho com as hepatites virais utilizando o mesmo algoritmo minerador, já demonstrando sua utilidade na geração de padrões em grandes bases de dados, como a do SINAN (Sistema de Informação de Agravos de Notificação), nos serviços (TRINDADE, 2005).

Na revisão da literatura foi encontrada ainda uma justificativa para essa escolha, observadas as comparações de eficácia (que é a qualidade ou propriedade em obter os melhores resultados em condições ideais), rapidez (tempo gasto na realização dos testes com a ferramenta) e efetividade (que é a capacidade da ferramenta de obter os melhores resultados em condições reais) dessa tarefa (classificação) e desse algoritmo C4.5 (CUROTTO, 2003). Estas definições são coincidentes com as definições da ferramenta e da epidemiologia. Interferiu também nesta escolha da ferramenta do *WEKA* e da tarefa da

classificação a sua acurácia de predição que é a habilidade do modelo em prever corretamente a classe de amostras desconhecidas, o seu desempenho que está relacionada aos custos computacionais envolvidos na geração e na utilização do modelo, a sua robustez que é a habilidade do modelo em fazer previsões corretas em amostras com atributos faltantes ou ruídos, a sua escalabilidade que é a habilidade de construir um modelo eficiente a partir de grande quantidade de dados e na sua interpretabilidade que é a habilidade de tornar compreensível o conhecimento gerado pelo modelo.

O atributo meta inicialmente escolhido foi a evitabilidade dos óbitos, baseado nas análises dos Comitês em evitável e não evitável (VIANNA 2004, 2004a e 2005a). Baseou-se também no trabalho em que é realizada uma revisão da literatura sobre mortalidade perinatal com enfoque para a evitabilidade destes óbitos (LANSKI, FRANCA e LEAL, 2002). Com este atributo meta, foram geradas primeiramente 441 regras e posteriormente selecionadas apenas 225 regras que confirmaram os critérios de evitabilidade analisados pelos Comitês de Prevenção da Mortalidade Infantil (CPMI), comprovando, desta forma, a efetividade do Sistema de Informação estabelecido para estas análises: 78% dos casos foram classificados corretamente, com precisão de 83% para identificação dos óbitos evitáveis e 62% para os não evitáveis. A taxa de erro para as evitáveis foi de 12% e para as não evitáveis, de 44%. A taxa de erro é definida como a medida de flutuação dos resultados das medidas iteradas de uma grandeza realizada por um processo isenta de vícios, ou seja, a quantidade de erro que aparece, e, portanto quanto menor a taxa de erro, maior será a validade do estudo.

Durante o *DM* foram realizadas as tarefas de maneira interativa com os especialistas do grupo 1. Cada classificação era precedida por uma nova análise

da base, transferência e experimento no algoritmo C4.5, até que fossem selecionados padrões efetivos para a aquisição de conhecimento.

Foi estabelecido um novo atributo para esse estudo, que é a alteração entre a CID na DO e após as análises, utilizando o critério de distâncias (diferença da CID na DO e após as análises) (MORO, 2003), e, para o presente estudo comparou-se a alteração entre a CID na DO e após as análises dos Comitês. Foi constituída então uma nova base, sendo selecionadas apenas as CID alteradas, totalizando em 4230 instâncias. As regras foram definidas conforme o aparecimento e a taxa de erro maior que a menor frequência relativa de aparecimento nas regras geradas. Com 169 regras para este atributo meta, 88% de acertos, 0,095 de erro absoluto, 21% de erro relativo, 24 instâncias ignoradas, foram consideradas as regras como relevantes para a avaliação dos especialistas apenas as com taxa de erro $< 29\%$, que foi a menor taxa de erro encontrada pela ferramenta.

Estes atributos meta foram definidos com base nas principais dificuldades dos Comitês, que são: estabelecer o diagnóstico mais específico dos óbitos após as análises e os critérios de evitabilidade dos mesmos, para com essas dificuldades sanadas melhorar as análises dos óbitos infantis e propor as melhorias nos atendimentos, baseadas em evidências científicas. Concluiu-se que o SIMI está auxiliando as classificações das análises dos óbitos infantis no Estado do Paraná e as alterações são semelhantes às da literatura.

Após este procedimento, as 10 melhores regras foram selecionadas pelos especialistas do grupo 1, para cada atributo meta selecionado, e, foram mostradas para os especialistas do grupo 1 e 2 validarem se as mesmas eram relevantes ou surpreendentes, irrelevantes, se concordavam ou eram

redundantes. As 20 regras escolhidas pelos especialistas do grupo 1 para a análise dos especialistas do grupos 1 e 2 foram as que apresentaram maior relevância e frequência e menor taxa de erro, segundo as análises objetivas da ferramenta. As mesmas são relatadas nas Tabelas 5 e 6, no item 3.3.

3.3. PÓS-PROCESSAMENTO e AVALIAÇÃO

Elaborou-se, ainda, um questionário com as 20 regras (10 para cada atributo meta) incluídas nessa seleção para que os especialistas respondessem se eles concordavam com as mesmas, sob os critérios: relevante ou surpreendente, irrelevante, se eles não concordavam com as regras estabelecidas pelo DM ou se estas regras eram redundantes. Ou seja, foram selecionadas as 10 melhores regras, isto é, com maior relevância e frequência e menor taxa de erro, das 225 geradas com o atributo meta da evitabilidade. E também as 10 melhores regras das 169 geradas com o atributo meta da alteração da CID na DO e após as análises dos Comitês.

A avaliação objetiva foi realizada pelo *WEKA*, onde a ferramenta realiza a validação cruzada para estabelecer o aparecimento da regra e já faz o conjunto de treinamento e teste na própria ferramenta. Algumas medidas de suporte são geradas como a porcentagem de casos onde a regra apareceu e medidas de confiança como a porcentagem de acerto destas regras, bem como a taxa de erro e a consistência das regras. Todas estas definições baseadas nas análises objetivas da ferramenta como taxa de erro, já definida anteriormente. Precisão é a ausência de erros aleatórios (relacionada à inferência estatística do universo amostral), isto é, informa quanto a regra é correta. A sensibilidade está relacionada à validade do estudo, sendo sua capacidade em relacionar o que é

positivo como realmente positivo, ou seja, o que tem a doença como realmente portador da doença e especificidade como a capacidade do teste em identificar os realmente negativos em negativos, ou seja, quem não tem a doença realmente aparece como não portador da doença. Validez do estudo é definida como o critério de qualidade relacionado com a adequação das pontuações do teste para o objetivo que suscitou sua aplicação. Estas definições correspondem às definições estatísticas, matemáticas e epidemiológicas.

O questionário das regras foi estabelecido conforme Tabela 5, com o atributo meta “evitabilidade”. O questionário das regras com o atributo meta “alteração entre as CID na DO e após as análises dos Comitês” está na Tabela 6.

Tabela 5 - Questionário com as regras de mineração com o atributo meta de evitabilidade:

	REGRA	ANÁLISE OBJETIVA
1	Se a classe código de evitabilidade era do Grupo D (reduzível por adequada atenção na gestação, parto e ao RN e educação para a saúde) e a classe código da CID na DO era do grupo P2 (afecções originadas no período perinatal respiratórias, cardiovascular, hematológicas, digestivas ou da regulação térmica) e a escolaridade da mãe de 1-3 anos e a quantidade de filhos vivos = 1, então o óbito era evitável.	Taxa de erro 0%.
2	Se a classe código de evitabilidade era do Grupo D e o estado civil da mãe era o de união consensual e a classe código da CID após a investigação era do Grupo P1 (afecções originadas no período perinatal por fatores maternos, relacionados a gestação, trabalho de parto, parto e crescimento fetal) e nenhum filho morto e idade no óbito de 1-6 dias e nenhum filho vivo e a gestação de 28-31 semanas, então óbito evitável.	Taxa de erro 0%.

3	Se a classe código de evitabilidade era do grupo D e a classe código da CID na DO era do grupo P2 e a quantidade de filho morto era = 0 e a idade em classe era < 1 dia e a classe idade da mãe era 16 a 35 anos e a raça/cor era branca e a quantidade de filho vivo era = 0 e o estado civil da mãe era casada e a gravidez era única e a classe peso era < 2.500 g, então óbito evitável.	Taxa de erro 0%.
4	Se a classe código de evitabilidade era do grupo D e a classe da CID era P2 e a gravidez era única e a quantidade de filho morto era = 1 e a gestação era de 28 a 31 semanas e a classe do Apgar no primeiro minuto era < 5, então era um óbito evitável.	Taxa de erro 0%.
5	Se a classe código de evitabilidade é do grupo G (óbitos dificilmente reduzíveis) e a classe da CID após a investigação era do grupo Q (malformações congênicas) e o número de consultas foi de 5 a 6 consultas e a classe da CID na DO era do grupo Q e a escolaridade da mãe era de 1 a 3 anos, então o óbito era não evitável.	Taxa de erro 0%.
6	Se a classe da CID após a investigação era do grupo Q e o número de consultas foi > 6 e a quantidade de filho morto foi = 0 e a gravidez foi única e a gestação foi de 37 a 41 semanas e houve alteração de capítulo, então o óbito era não evitável.	Taxa de erro 0%.
7	Se a classe código de evitabilidade era do grupo G e a classe da CID após a investigação era do grupo Q e o número de consultas foi > 6 e a quantidade de filho morto foi = 0 e a quantidade de filho vivo foi = 1 e a classe peso foi entre 2.500 a 3.500 g	Taxa de erro 0%.

	e o estado civil da mãe era o de união consensual, então o óbito era não evitável.	
8	Se a classe código de evitabilidade era do grupo G e a classe da CID após a investigação era do grupo Q e o número de consultas foi de 5 a 6 consultas e a classe da CID na DO era do grupo Q e a gestação era de 37 a 41 semanas e a quantidade de filho morto era = 0 e a escolaridade da mãe era de 4 a 7 anos e o sexo era feminino, então o óbito era não evitável.	Taxa de erro 0%.
9	Se a classe do código de evitabilidade era do grupo G e a classe da CID após a investigação era do grupo Q e o número de consultas era > 6 e a quantidade de filhos mortos era = 1 e o sexo era feminino, então o óbito era não evitável.	Taxa de erro 0%.
10	Se a classe do código de evitabilidade era do grupo G e a classe da CID após a investigação era do grupo Q e o número de consultas foi > 6 e a quantidade de filhos mortos era = 0 e a quantidade de filhos vivos era = 1 e a classe peso era entre 2.500 e 3.500 g, então o óbito era não evitável.	Taxa de erro 0%.

Tabela 6 - Questionário com as regras de mineração com o atributo meta alteração da CID:

	REGRA	ANÁLISE OBJETIVA
1	Se a classe da CID na DO era do grupo P1 e a classe código da CID após a investigação foi para o Grupo P2, então alterou o grupo, especificando melhor a causa do óbito.	Taxa de erro 0%.
2	Se a idade estava entre 7 a 28 dias, a classe da CID após a investigação é do grupo R (sintomas, sinais e achados anormais) e a classe da CID na DO era do grupo P1, então alterou a categoria (especificando melhor da categoria geral para a mais específica).	Taxa de erro 3%.
3	Se a idade estava entre 7 a 28 dias e a classe da CID na DO era do grupo R,	Taxa de erro 0,5%.

	então alterou o capítulo (especificando melhor a causa do óbito).	
4	Se a classe da CID após a investigação era do grupo Q e a classe da CID era do grupo P2 na DO, então alterou o capítulo.	Taxa de erro 0%.
5	Se a classe da CID após a investigação era do grupo W (das causas externas), então alterou o capítulo, significando que é necessário especificar melhor as causas externas porque após a investigação consegue-se esclarecer essas causas externas e determinar outra causa de óbito em outro capítulo.	Taxa de erro 0%.
6	Se a classe da CID era do grupo P1, então alterou o capítulo.	Taxa de erro 0%.
7	Se a idade em classe era de 7 a 28 dias e a classe de evitabilidade era do grupo D e a classe da CID na DO era do grupo P2 e a classe da CID após a investigação era do grupo P1, então alterou o grupo, significando que é necessário esclarecer melhor as causas maternas relacionadas aos óbitos infantis.	Taxa de erro 1%.
8	Se a classe da CID após a investigação era do grupo R, então alterou o capítulo, significando que se conseguiu definir melhor os óbitos de causas indeterminadas.	Taxa de erro 0%.
9	Se a classe da CID na DO era do grupo E (doenças endócrinas, nutricionais e metabólicas), então alterou o capítulo da CID, indicando que é necessário para os médicos que atendem e atestam os óbitos conhecer melhor as causas endócrinas, nutricionais e metabólicas.	Taxa de erro 0%.
10	Se a classe idade era de 7 a 28 dias e a classe da evitabilidade era do grupo I (doenças do aparelho circulatório), então alterou o capítulo, especificando as mal definidas para outra causa melhor definida.	Taxa de erro 0%.

4. RESULTADOS E ANÁLISES

Para facilitar a compreensão e a análise dos resultados, foi necessária uma divisão segundo os objetivos específicos propostos. A primeira parte dos resultados é a da análise da integração das bases. A segunda parte indica os resultados do *DM* com o atributo meta da evitabilidade e da alteração entre as CID na DO e após as análises dos CPMI. A terceira parte corresponde aos resultados das análises dos especialistas com os mesmos atributos meta.

4.1. RESULTADOS E ANÁLISES DA INTEGRAÇÃO DAS BASES

Para responder ao primeiro objetivo específico, que era integrar três diferentes SIS, procedeu-se sua integração e análise, com os resultados apresentados a seguir, organizados sequencialmente por atributos analisados.

1. O primeiro e o segundo atributos selecionados foram o código da CID na declaração de óbito e após as análises dos Comitês, categorizados em grupos devido ao grande número de possibilidades de diagnóstico. Observou-se que, no caso do óbito infantil, 67% dos óbitos na DO e 60% depois da análise dos Comitês foram classificados no grupo P, capítulo XVI da CID em sua décima revisão, das afecções originadas no período perinatal, que vai do grupo P00 ao P96. Eles estão relacionados à classe do código da CID10 na DO (que aparece como ClasseCodCID, retirado do SIMI e confirmado no SIM). Após a investigação dos Comitês, aparecem como ClasseCodCIDInv (retirado do SIMI), constando na base integrada com:

- 1625 casos (22% da base) na DO e 2270 casos (31% da base) após as investigações, aumentando a categorização da CID10 no grupo P1 que corresponde ao P00 ao P15, p.733 do volume 1 da Décima revisão da CID10.

- 2578 casos (35% da base) na DO e 2032 (28%) após as investigações com a divisão da CID10 no grupo P2, que corresponde ao P20 ao P83, p. 733 do volume 1 da Décima revisão da CID10. E assim sucessivamente conforme apresentado resumidamente na tabela 17, no Apêndice 2.

Na análise dos atributos selecionados da base integrada, gerada a partir dos três diferentes SIS (SINASC, SIM e SIMI), comprovou-se que, após as análises dos Comitês, podem-se especificar melhor os diagnósticos dos óbitos. Isto pode ser observado na Figura 6.

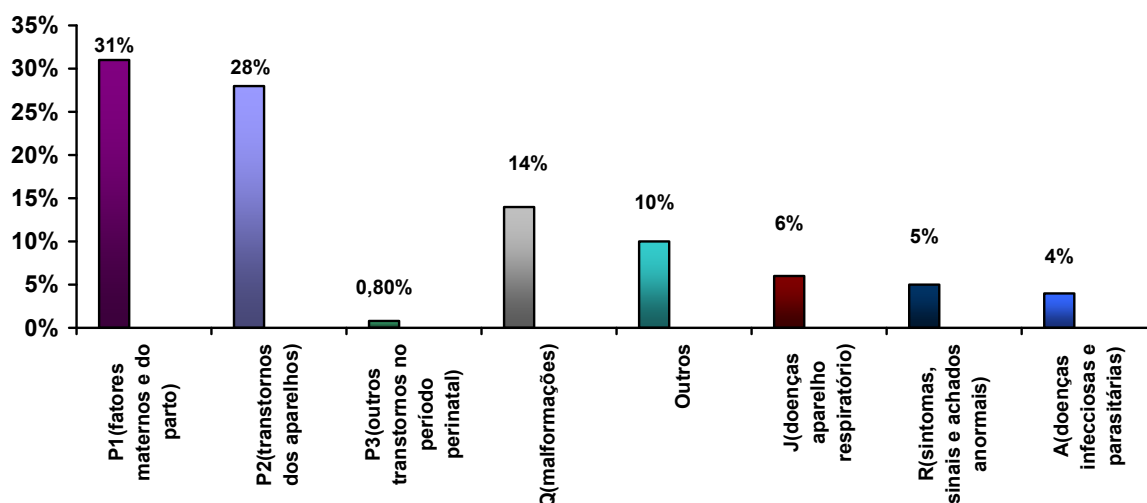


Figura 6 - Porcentagem da categorização da base segundo a CID após as análises.

Esse conhecimento gerado sugere a necessidade de melhor treinamento e atenção no preenchimento correto das DO. Demonstra também a eficácia do trabalho de análise dos óbitos para especificar melhor a causa do óbito, facilitando a tomada de decisões para a melhoria de ações na redução do problema. Reafirma igualmente a eficácia da ferramenta utilizada para facilitar esta compreensão da base, conforme exemplo da porcentagem das categorizações segundo a CID 10 após as análises dos Comitês na Figura 6.

2. O terceiro atributo analisado e um dos escolhidos como atributo meta foi o Código Tipo Morte Evitável (estabelecido como CodTipoMorteEvitavel), evitável ou não, atributo retirado do SIMI. Esse atributo meta foi escolhido porque uma das atribuições da investigação dos óbitos é descobrir se era um óbito evitável ou não. Pensou-se utilizar esta ferramenta não convencional em rotinas de Vigilância Epidemiológica, para estabelecer padrões na classificação de características relacionadas aos óbitos e para facilitar a descoberta de crianças suscetíveis ou não a essas ocorrências. Após as análises dos Comitês foram encontradas 5226 mortes (72%) evitáveis e 2030 mortes (28%) não evitáveis.

Sabe-se da fragilidade de caracterização da evitabilidade dos óbitos decorrentes das dificuldades de categorização. Muitas vezes, os analisadores deparam-se com situações como exemplificada a seguir: “como definir um óbito como evitável segundo a Associação Americana, na qual se baseou o padrão de definição de evitabilidade do SIMI, se no local onde a criança foi atendida não existia muitas vezes nem médico, nem transporte, nem aparelhagem suficiente para salvar esta vida” (MINISTÉRIO DA SAÚDE, 2004a).

Como um dos resultados práticos desse trabalho iniciou-se uma revisão em nível estadual e nacional, estabelecendo-se o consenso da mudança de óbitos evitáveis ou não para óbitos reduzíveis ou não.

3. O quarto atributo selecionado, retirado do SIMI (estabelecido como ClasseCodEvitabilidade), foi dos critérios da evitabilidade destes óbitos, que utiliza os critérios do Anexo 3, Tabela 10, já padronizado para análise dos óbitos nos Comitês, utilizando apenas as categorias de A a I, sem subdivisões, para facilitar a mineração.

- Foram 4033 (56%) no grupo D, o que corresponde aos óbitos evitáveis por adequada atenção na gestação, no parto e ao recém-nascido e na educação para a saúde.
- 1053 (14%) no grupo G, que são os óbitos dificilmente evitáveis. E assim sucessivamente conforme apresentado na Tabela 33, no Apêndice 3.

O grupo destes critérios de evitabilidade mais freqüente (4033 ou 56%) foi o grupo D que relaciona os óbitos reduzíveis por adequada atenção na gestação, no parto e ao recém-nascido e pela educação para a saúde. Está sendo desmembrado em:

- 1) um grupo para adequada atenção à gestação;
- 2) um grupo relacionado à atenção ao parto;
- 3) um grupo para a atenção ao recém-nascido e
- 4) um grupo para a educação para a saúde.

Esta readequação mostrou-se importante após a utilização dessa ferramenta, visto que com uma melhor especificação dos critérios pode-se saber mais fielmente onde ocorreu a falha e, assim, propor medidas de atenção adequadas para a redução destes óbitos.

Destaca-se, assim, a utilidade prática do uso da ferramenta, observando que a mesma facilita o conhecimento e relacionamento dos atributos contidos na base selecionada. A utilização do conhecimento adquirido ainda agiliza processos e pode ser transformada em benefício às populações atendidas pelos Sistemas de Saúde.

Observa-se na figura 7 que a utilização da ferramenta confirmou aspectos positivos da atenção prestada às crianças, pois com a disponibilidade e facilidade de acesso à imunização, conseguiu-se reduzir a freqüência dos óbitos infantis

relacionados a esta prática para apenas 0,1% dos casos, grupo A da evitabilidade. Confirma-se, desta forma, a importância da utilização de ferramentas computacionais aplicadas às bases de dados. A medicina baseada em evidências pode fazer uso desta tecnologia aplicada à área de saúde.

Toda esta análise baseada nos critérios de evitabilidade pode ser observada no exemplo da Figura 6, com a categorização da evitabilidade dos óbitos.

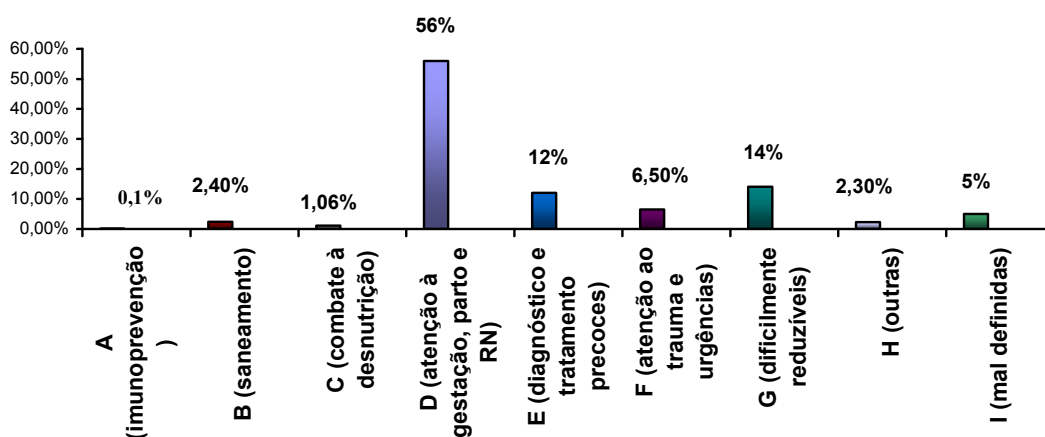


Figura 7 - Análise da evitabilidade dos óbitos.

4. O quinto atributo selecionado foi o sexo, estabelecido como `analiseobitoSEXO`, atributo retirado da base do SIMI. Foram 4199 (58%) casos masculinos, o que confirma o maior índice de nascimentos de sexo masculino e o maior número de óbitos neste sexo e nesta faixa etária, 3055 (42%) feminino e 2 não determinados.

5. O sexto atributo escolhido foi a classe idade da mãe, `ClasseldadeMae`, atributo do SINASC. Foram 6285 (87%) das mães com a idade entre 16 e 35 anos, 629 (9%) com mais de 35 anos e 342 (4%) das mães com menos de 16

anos. Foram 13% das mães em idade de risco para a gestação, com idade ≤ 16 e ≥ 35 anos de idade. Conforme se observa na Tabela 18, no Apêndice 3, com a representação da idade da mãe em categorias visualiza-se que as mesmas estão nos extremos da vida em apenas 13% dos casos, correspondendo aos casos de risco que devem ser avaliados com maior frequência durante a gestação, parto e puericultura.

6. O sétimo atributo selecionado foi o estado civil da mãe, ESTCIVMAE, já em categorias na DN (não ocorre união consensual nesta base até o ano de 2002) e na DO (com união consensual, onde foi escolhido para formação da nova base), retirado do SIM. Foram 2799 (39%) casadas, 2317 (32%) solteiras, 2000 (28%) casos em união consensual e assim sucessivamente conforme se pode observar na Tabela 19, no Apêndice 3, a representação do estado civil da mãe: 67% das mães caracterizadas com união estável (39% ou 2799 casadas e 28% ou 2000 em união consensual), foram caracterizadas apenas 32,4% (32% ou 2317 solteiras e 0,4% ou 30 viúvas) como em risco devido a falta de parceiro/pai para dividir e auxiliar nos cuidados com a criança.

7. O oitavo atributo escolhido foi a escolaridade da mãe, ESCMAE, retirado da base do SIM, com confirmação no SINASC. Foram 2901 (40%) com 4 a 7 anos de estudo, 2177 (30%) com 8 a 11 anos de estudo e assim sucessivamente conforme se observa na Tabela 20, no Apêndice 3, com a escolaridade da mãe, foram 70% (40% ou 2901 de 4-7 anos e 30% ou 2177 de 8 a 11 anos) das mães com razoável escolaridade, o que facilita a compreensão das medidas propostas nas avaliações e consultas durante a gestação, parto e puericultura e apenas 19% (16% ou 1132 de 1-3 anos e 3% ou 235 sem escolaridade) incluídas em critérios de risco da baixa escolaridade.

8. O nono atributo selecionado foi a quantidade de filhos vivos anteriores, QTDFILVIVO, retirado do SIM e confirmado no SINASC. Observado como atributo descritivo na DN e na DO. Foram 3011 (41%) sem nenhum filho, 1861 (26%) com 1 filho vivo anterior e assim sucessivamente conforme a Tabela 21, no Apêndice 3, com a quantidade de filhos vivos anteriores, observou-se 41% das mães sem nenhum filho vivo anterior, o que configura este critério de risco de outros filhos anteriores para cuidar como inexistente em quase metade dos casos desta base histórica. Foram, ao final das análises retirado, porque foram geradas poucas regras (5%), consideradas pelos especialistas como pouco relevantes.

9. O décimo atributo escolhido foi a quantidade de filhos mortos anteriores, QTDFILMORT, retirado do SIM e confirmado no SINASC. Foram 6098 (84%) sem nenhum filho morto anterior, 710 (10%) com 1 filho morto anterior e assim sucessivamente conforme observa-se na tabela 22, no Apêndice 3, 84% dos casos sem nenhum filho morto anterior, o que caracteriza apenas 13,5% dos casos com o critério de risco filho morto anterior.

10. O décimo primeiro atributo escolhido foi a gestação, retirado da base do SIM e confirmado na do SINASC, preenchido descritivamente na DN e na DO e categorizado em semanas de gestação. Foram 3389 (47%) com 37 a 41 semanas de gestação (consideradas a termo), 1401 (19%) com 32 a 36 semanas de gestação (consideradas pré-termo) e assim sucessivamente conforme se observa na Tabela 23, no Apêndice 3. Com as semanas de gestação, observou-se portanto, 47% da base com a gestação a termo e 20% nos critérios de risco onde 19% estavam com menos de 27 semanas de gestação e 0,9% com gestação acima de 42 semanas (pós - datismo).

11. O décimo segundo atributo escolhido foi a gravidez retirado da base do SIM e confirmado na base do SINASC. Aparece como GRAVIDEZ, classificada na DN e na DO. Foram 6623 (91%) com gravidez única e assim sucessivamente conforme se observa na Tabela 24, no Apêndice 3.

12. O décimo terceiro atributo escolhido foi o tipo de parto, retirado da base do SIM e confirmado na base do SINASC. Aparece como PARTO já categorizado na DN e na DO. Foram 4567 (63%) com parto vaginal, 2686 (37%) com parto cesáreo conforme observado na Tabela 25 com o tipo de parto, no Apêndice 3.

13. O décimo quarto atributo escolhido foi o número de consultas no pré-natal (durante a gestação), estabelecido como CONSULTAS, retirado do SINASC. Foram 2851 (39%) com mais de 6 consultas de pré-natal, 2599 (36%) com 5 a 6 consultas e assim sucessivamente conforme observa-se na Tabela 26, no Apêndice 3. Analisando-se o número de consultas na gestação 75% (36% com 5 a 6 e 39% com mais de 6) das gestantes com pelo menos 5, número considerado razoável. E 39% com o número de 6 consultas que é o preconizado pela OMS e pelas sociedades de classe como o necessário para um bom atendimento à gestante e ao seu filho, desde que realizado com qualidade segundo padronizações da OMS, MS e Sociedade de Ginecologia e Obstetria.

14. O décimo quinto atributo selecionado foi a média do Apgar no primeiro minuto, estabelecida para o estudo como ClasseApgar1, retirada do SINASC. Foram 3905 (53%) com Apgar > 5 no primeiro minuto, 3188 (44%) < = 5 no primeiro minuto e 163 (2%) sem preenchimento. Observa-se na Tabela 27, no Apêndice 3, a representação da média da classe Apgar no primeiro minuto. Foram mais da metade dos casos da base (53%) com Apgar bom, acima de 5 no primeiro minuto, o que demonstra boas condições ao nascer.

15. O décimo sexto atributo escolhido foi a média do Apgar no quinto minuto, categorizado para este estudo, retirada do SINASC, Classe APGAR5. Foram 5120 (71%) com Apgar > 5 no quinto minuto de vida, 1978 (27%) com Apgar ≤ 5 no quinto minuto de vida e 157 (2%) sem preenchimento. Observa-se na Tabela 28, no Apêndice 3, 71% com Apgar bom (>5) no quinto minuto de vida.

16. O décimo sétimo atributo escolhido foi o critério raça/cor, estabelecido como RACACOR, retirado do SIM e confirmado no SINASC, já categorizados na DN e na DO. Foram 6534 (91%) com raça branca, 520 (7,2%) com raça parda e assim sucessivamente conforme observado na Tabela 29, no Apêndice 3. Este atributo foi retirado porque 91% que aparecem como raça branca, é uma categorização duvidosa. Fica a pergunta: “por ser o sul do país, com a grande colonização européia, são realmente 91% brancos ou como é uma característica auto declarável, é questionável sua validade”? Geraram-se apenas duas regras com este atributo, sendo considerado pelos especialistas do grupo 1 como regras não relevantes.

17. O décimo oitavo atributo escolhido foi a categorização peso ao nascer, categorizada para este trabalho, ClassePeso, retirado do SIM e confirmado no SINASC. Foram 4084 (56%) com peso ao nascer ≤ 2.500 g, 2573 (35%) com peso ao nascer entre 2.500 a 3.500g e assim sucessivamente conforme observa-se na Tabela 30, no Apêndice 3. O peso baixo ao nascer (≤ 2.500 g) ocorreu em 56% dos óbitos. Desta forma, esta característica deve ser observada com maior frequência durante a puericultura para intervenção precoce em qualquer alteração, evitando a possibilidade da ocorrência de outros óbitos semelhantes.

18. O décimo nono atributo escolhido foi a categorização para idade na época do óbito, calculada com a diferença da idade ao nascer e no óbito. Estabelecida

como IdadeClasse, retirada do SIM e confirmada no SINASC. Foi categorizada para o trabalho. Foram 2272 (31%) com idade no óbito entre 1 e 6 dias e assim sucessivamente conforme observa-se na Tabela 31, no Apêndice 3. Em 54% dos casos o óbito ocorreu na primeira semana de vida (23% $\leq$ 1 dia e 31% entre 1 e 6 dias). Este conhecimento gerado sugere que os serviços devem realizar esforços na melhoria das condições de atendimento nessa época da vida, intimamente relacionada às condições da gestação e do parto.

19. O vigésimo atributo escolhido foi a alteração entre as Classificações segundo a CID10 na DO e após as análises dos Comitês, quando ocorreu mudança no capítulo, no grupo, na categoria, ou na subcategoria (MORO, 2003). Descrita na base como Distancia. Foram 3026 (42%) com a CID igual antes e após as análises, 1538 (21%) mudando de categoria, 1494 (20%) mudou de grupo e 1174 (16%) mudou de capítulo, coincidindo com as análises dos Comitês (VIANNA, 2004), (VIANNA, 2004a), (VIANNA, 2005a) e (VIANNA e MORO, 2006a). Observa-se na Tabela 32, no Apêndice 3, com a representação das alterações entre as CID10 antes e após as análises dos Comitês, 42% das CID sem alteração antes e após as análises dos Comitês, o que caracteriza um bom preenchimento das DO e 16% com mudança de capítulo caracterizando uma dificuldade na hora do óbito da caracterização da real causa do óbito. Este fato pode ser melhorado após as análises dos atendimentos prestados, do histórico do paciente e, em alguns casos, das análises pelos Serviços de Verificação de Óbito (SVO) ou pelos Institutos de Medicina Legal (IML), que realizam as necropsias de alguns destes óbitos que ficaram duvidosos, a partir dos quais aparece na DO como aguardando exames, somente especificados após os resultados dos exames dos IML. Este fato confirma a necessidade das análises pelos SVO, o

que está sendo viabilizado a partir de portaria firmada em 2006, após a confirmação de mais este estudo.

4.2. RESULTADOS E ANÁLISES DO DM

Essa parte dos resultados foi dividida em duas, segundo os dois atributos meta selecionados.

4.2.1. ATRIBUTO META EVITABILIDADE

Quando o atributo meta escolhido foi a evitabilidade dos óbitos, foram geradas 441 regras, com precisão de 80%. Obteve-se uma sensibilidade de 83% para identificação dos evitáveis e 62% dos não evitáveis e uma especificidade de 62%. A taxa de erro para as evitáveis foi de 12% e para as não evitáveis, de 44% conforme Tabela 7.

Tabela 7 - Regras Meta Evitabilidade:

Regras	Precisão	Sensibilidade	Especificidade	Taxa de erro
441	80%	83%	62%	12%
		evitáveis		evitáveis
78%		62%		44%
classificadas		não evitáveis		não evitáveis
corretamente				

Escolheram-se as 4 primeiras regras relacionadas aos óbitos evitáveis, que estão no questionário da tabela 5, do item 3.3, para os especialistas do grupo 1 e 2 avaliarem, com nenhuma taxa de erro e que apareceram em pelo menos 29% dos casos, isto é, as mais frequentes. E também as 6 últimas regras relacionadas aos óbitos não evitáveis que apareceram em no máximo 12% dos casos, isto é,

as mais raras, que estão no questionário da tabela 5, no item 3.3. Estas regras foram escolhidas obedecendo a critérios de relevância segundo a experiência dos especialistas do grupo 1 consultados (análise qualitativa). Pensou-se em validar os extremos, porque, dessa forma pôde-se estabelecer a validade da utilização desta técnica não convencional de identificação de características relacionadas à mortalidade infantil no máximo de acertos, validando também o SIMI e os erros para poder estabelecer como resultado desse trabalho a melhoria das análises baseadas na utilização da ferramenta proposta. As regras foram escolhidas conforme Tabela 8.

Tabela 8 - Escolha das regras, meta evitabilidade:

<i>Regras</i>	<i>Taxa de erro</i>	<i>Frequência</i>	<i>Validação</i>	<i>Avaliação</i>
4	nenhuma	29%	literatura	especialistas
óbitos evitáveis		mais frequentes		
6	nenhuma	12%	literatura	especialistas
óbitos não evitáveis		mais raras		

4.2.2. ATRIBUTO META ALTERAÇÃO ENTRE AS CID

Com o atributo meta alteração entre as CID, que comparou a CID existente na DO e após as investigações, as instâncias ou número de casos utilizados no DM foram apenas os que sofreram alteração da CID após a investigação dos Comitês que foram 4230 (58%) casos dos 7256, sendo portanto, gerada nova base para esta mineração. Desta forma, foram geradas 169 regras com 88% de acertos, precisão de 93%, sensibilidade de 88% e uma especificidade de 49% e foram consideradas as taxas de erro de 0% em 7 regras, conforme questionário na tabela 6, no item 3.3 . Nas outras 3 regras, conforme questionário na tabela 6,

no item 3.3, foram consideradas taxas de erro de até 3% nos casos de maior frequência de aparecimento e com um valor menor que a taxa de erro gerada na matriz de confusão e que não deve ultrapassar 27%, valor este estabelecido de maneira objetiva pela ferramenta. Estas regras foram escolhidas também segundo critérios de relevância, após a avaliação dos especialistas do grupo 1 consultados, conforme Tabela 9.

Tabela 9 - Escolha das regras, meta alteração entre as CID:

Regras	Acertos	Precisão	Sensibilidade	Especificidade	Taxa de erro
169	88%	93%	88%	49%	0%
					(em 7 regras)
					3%
					(em 3 regras)

Agora serão descritos os resultados após as análises dos especialistas.

4.3. RESULTADOS E ANÁLISES DOS ESPECIALISTAS

As regras analisadas pelos especialistas foram divididas segundo os atributos meta evitabilidade e alteração entre as CID na DO e após as análises dos Comitês, segundo os questionários nas tabelas 5 e 6, no item 3.3.

4.3.1. ATRIBUTO META EVITABILIDADE

Para as regras relacionadas ao atributo meta da evitabilidade, após as análises dos 25 especialistas do grupo 1 e 2 consultados, os resultados foram os seguintes:

- Para a regra 1, na tabela 5, no item 3.3, 96% (24) dos especialistas consultados, relataram ser uma regra relevante e não surpreendente, visto que historicamente 54% dos casos dos óbitos infantis no Paraná são evitáveis por uma melhor atenção na gestação, parto e ao RN, relatado também em outros trabalhos (VICTORA et al., 1996 e 2005) e (LANSKI, FRANCA e LEAL, 2002), estando também relacionados a esta regra a baixa escolaridade da mãe e a existência de outro filho para cuidar como dificuldades no cuidado a esta criança que foi a óbito.

- Para a regra 2, na tabela 5, no item 3.3, 92% (23) dos especialistas observaram como regra relevante e não surpreendente pois a prematuridade está relacionada a afecções originadas no período perinatal por fatores maternos, relacionados a gestação, parto e crescimento fetal e muitas vezes podem ser evitadas por uma atenção mais efetiva durante o planejamento familiar, a gestação e o parto (VIANNA, 2004), (VIANNA, 2004a), (VIANNA, 2005a).

- Para a regra 3, na tabela 5, no item 3.3, 100% (25) dos especialistas concordaram na relevância e em seu caráter não surpreendente, visto que, o baixo peso está relacionado a óbitos infantis evitáveis na idade perinatal; desde que, melhorem-se as condições da gestação e de um planejamento familiar efetivo, observado também por (VICTORA et al, 1996), (VICTORA et al, 2005) e (LANSKI, FRANCA e LEAL, 2002).

- Para a regra 4, na tabela 5, no item 3.3, 96% (24) dos especialistas relataram a relevância e seu caráter não surpreendente porque a prematuridade e o Apgar baixo no primeiro minuto também são importantes características relacionadas ao óbito infantil evitável por uma adequada atenção na gestação,

parto e RN. Fato este também observado nos trabalhos de (VICTORA et al, 1996), (VICTORA et al., 2005) e (LANSKI, FRANCA e LEAL, 2002).

- Para a regra 5, na tabela 5, no item 3.3, 96% (24) dos especialistas concordaram na relevância e seu caráter não surpreendente pois os óbitos de malformações são dificilmente evitáveis, pois geralmente são incompatíveis com a vida, principalmente quando a escolaridade da mãe é baixa de 1 a 3 anos, mesmo com número adequado de consultas no pré-natal (5-6).

- Para a regra 6, na tabela 5, no item 3.3, 100% (25) dos especialistas concordaram que as malformações são não evitáveis, mesmo com número maior que 6 consultas no pré-natal e gestação a termo (37-41 semanas).

- Para a regra 7, na tabela 5, no item 3.3, 100% (25) dos especialistas concordaram na relevância e em seu caráter não surpreendente, pois as malformações, mesmo com número de consultas no pré-natal e peso adequados, geralmente são dificilmente evitáveis.

- Para a regra 8, na tabela 5, no item 3.3, 100% (25) dos especialistas relataram a relevância da regra, apesar de sua redundância e seu caráter não surpreendente, pois estava também relacionada as malformações, que por si só são dificilmente evitáveis. Mas, a característica do sexo feminino não tinha sido observada pelos especialistas como relevante anteriormente ou de maior relevância visto que historicamente em 58% dos casos os óbitos infantis ocorrem mais no sexo masculino, surgindo como característica surpreendente. Sugerindo desta forma, maiores estudos para estabelecer a relação entre o sexo feminino e as malformações que mais levaram à óbito, se elas realmente são relevantes ou não.

- Para a regra 9, na tabela 5, no item 3.3, 96% (24) dos especialistas relataram a redundância da regra, pois estava relacionada às malformações, apenas com número maior de consultas no pré-natal, aparecendo novamente o sexo feminino como maior ocorrência, sugerindo também seu caráter surpreendente e que merece maiores estudos para revelar sua relevância ou não.

- Para a regra 10, na tabela 5, no item 3.3, 96% (24) relataram sua redundância, pois mais uma vez aparecem as malformações como não evitáveis, apesar do peso e número de consultas de pré-natal ser adequados.

4.3.2. ATRIBUTO META ALTERAÇÃO ENTRE AS CID

Para os resultados do atributo meta alteração entre as CID, após as análises dos 25 especialistas do grupo 1 e 2, os resultados foram os seguintes:

- Para a regra 1, na tabela 6, no item 3.3, 100% (25) acharam a regra irrelevante e não surpreendente apesar da melhor especificação da causa do óbito, alterando do grupo das afecções maternas para as afecções relacionadas aos aparelhos.

- Para a regra 2, na tabela 6, no item 3.3, 96% (24) acharam a regra relevante e não surpreendente devido ao fato de especificar melhor a causa óbito de sintomas, sinais e achados anormais para afecções originadas no período perinatal afetados por fatores maternos.

- Para a regra 3, na tabela 6, no item 3.3, 96% (24) dos especialistas consultados concordaram na relevância e em seu caráter não surpreendente da regra pois, nos óbitos neonatais conseguiu-se especificar melhor a causa óbito que estava na DO no grupo de sintomas, sinais e achados anormais inespecíficos para outro capítulo mais específico.

- Para a regra 4, na tabela 6, no item 3.3, 100% (25) relataram a relevância e seu caráter não surpreendente. Referem também a necessidade de um melhor preenchimento das DO, pois, os prestadores dos atendimentos acharam na hora do óbito uma alteração de aparelhos que justificou aquele óbito e após as análises dos especialistas observaram-se malformações que justificavam este óbito.

- Para a regra 5, na tabela 6, no item 3.3, 100% (25) acharam a regra relevante e surpreendente, pois esta regra sugere a necessidade de especificar melhor as causas de óbito relacionadas às causas externas, pois após as análises conseguiu-se especificar melhor estas causas de óbitos.

- Para a regra 6, na tabela 6, no item 3.3, 100% (25) dos especialistas acharam a regra redundante e surpreendente, pois como o grupo P é abrangente, pode ser melhor especificado e alterar o capítulo após as análises, mas deveria ser bem conhecida na hora da definição do óbito.

- Já para a regra 7, na tabela 6, no item 3.3, 100% (25) dos especialistas acharam a regra relevante e não surpreendente, pois nos óbitos neonatais, podem estar na DO como relacionadas aos aparelhos e após as investigações observou-se sua relação com as afecções maternas, do parto e do RN, sugerindo a necessidade de melhor atenção neste período da vida.

- Para a regra 8, na tabela 6, no item 3.3, 96% (24) dos especialistas consultados relataram sua relevância e seu caráter surpreendente pois, conseguiu-se após as investigações especificar melhor as causas dos óbitos o que já poderia ser melhor especificado no óbito.

- Para a regra 9, na tabela 6, no item 3.3, 96% (24) dos especialistas acharam esta regra irrelevante e não surpreendente pois, sendo as causas

endócrinas responsáveis por aproximadamente 1% dos óbitos é uma regra de pouca ocorrência e como alterou o capítulo após as análises faz-se necessário conhecer melhor as causas endócrinas, nutricionais e metabólicas na hora do óbito.

- Para a regra 10, na tabela 6, no item 3.3, 100% (25) dos especialistas relataram ser regra surpreendente estar entre 7 e 28 dias como causa do óbito o grupo I da CID 10 das doenças do aparelho circulatório e após as análises dos óbitos alterar para outro capítulo, sugerindo a necessidade de melhorar o conhecimento de causas de óbitos relacionados a este aparelho.

Os vinte e cinco especialistas do grupo 1 e 2 observaram que as regras geradas objetivamente pela ferramenta confirmam e validam o SIMI. Foram unânimes na importância da integração das bases para conhecer algumas características faltantes no referido sistema.

5. DISCUSSÃO E CONCLUSÃO

Após a integração dos três SIS, descobriram-se características que confirmam o conhecimento anteriormente observado e relatado pelos Comitês sobre a ocorrência dos óbitos infantis no Estado, (VIANNA, 2004), (VIANNA, 2004a), (VIANNA, 2005a), (VIANNA, 2005b).

A integração e relacionamento (*linkage*) de bases de dados diferentes auxiliam a complementar, rever e recuperar informações (CAVALCANTE, RAMOS e PONTES, 2005). Apresentando-se como uma estratégia adequada para aprimorar a qualidade das informações, permite, dessa forma, uma aproximação da situação epidemiológica real existente na população estudada (ALMEIDA, 1996).

Esta integração e relacionamento podem ser utilizados também para o uso de outras ferramentas. Como, por exemplo, de georreferenciamento, tal como ocorreu na prevenção da dengue em Porto Alegre. Nessa experiência foi feito a *linkage* da base de dados dos casos investigados de dengue registrados no SINAN com as informações do Sistema de Informações de Febre Amarela e Dengue. Criou-se, então, uma nova base no *software Microsoft Access*, que foi exportada para o MAPINFO para ser georreferenciadas. Desta forma, abrem-se caminhos para trabalhos futuros com a utilização das informações geradas pelo DM em georreferenciamento (PUSTAI et al., 2005).

A integração das informações contidas em várias bases de dados pode ser um importante instrumento de regulação para internações hospitalares obstétricas no SUS. Por exemplo, em um estudo no qual foram utilizados para análise as

informações do SINASC, do SIH-SUS e do SIM, com a observação das potencialidades do SINASC (CUNHA e TEIXEIRA, 2005).

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 dos mesmos. Esta importância também é relatada no trabalho que discorre sobre a boa seleção e preparação dos dados para obtenção de uma boa mineração de dados (PYLE, 1999).

A consolidação dos dados para o KDD também é tarefa importante para chegar a resultados significativos (SETH, GUATAM e BALINT, 2002).

Considerando-se a evolução histórica na qualidade das informações de mortalidade na população idosa residente no Município de Maringá, em um período de vinte e um anos (MATHIAS et al., 2005).

A coleta correta dos dados de óbito, por exemplo, foi importante para a mudança e especificação das causas de óbitos, até com a mudança do perfil epidemiológico da população. Antes figurava a morte sem assistência como a principal causa de óbito no município, e, após ações educativas e de criação do Comitê de Prevenção de Mortalidade Materna e Infantil a partir de 2003, passou a figurar como quinta causa de óbitos. Estas análises dos Comitês possibilitaram, além da mudança do perfil epidemiológico de mortalidade no município, o cruzamento de dados com a DN, o que melhorou seu preenchimento (SOUZA e TOSETTO, 2005).

Com a análise após a *linkage*, pode-se melhorar o entendimento do perfil predominante da criança, relacionada à mortalidade infantil no Estado do Paraná, no período de 2000 a 2004, destacando-se as probabilidades condicionadas aos

óbitos; não quer dizer, portanto, que aumentam o risco do óbito, pois não foi feita comparação com as crianças da população das que não foram à óbito:

- 55% destes óbitos poderiam ser evitáveis, por uma adequada atenção à gestação, parto e ao RN.

- O sexo era masculino (58%).
- A idade da mãe de 16-35 anos (87%).
- A mãe casada (39%).
- Com escolaridade de 4-7 anos de estudo (40%).
- Sem nenhum filho vivo (41%).
- E nenhum filho morto (84%).
- Com gestação a termo (47%).
- Gravidez única (91%).
- Parto vaginal (63%).
- Com mais de 6 consultas de pré-natal (39%).
- Apgar 1>5 (53%).
- E apgar 5>5 (71%).
- Raça/cor branca (91%).
- Peso ao nascer ≤ 2.500 g (56%).
- Idade do óbito < 1 mês (54%).

Estas mesmas características relacionadas após a integração das bases, foram as que mais apareceram nas regras geradas pelo *DM*. Completou-se, desta forma, o segundo objetivo específico de conhecer as principais características relacionadas a mortalidade infantil encontradas pelo *DM*. Estas características coincidem com as observadas pelas coortes históricas de 1982, 1993 e 2004,

observadas na mortalidade neonatal (VICTORA et al., 2005). Esse autor também relata, em outro estudo, a importância de algumas características, como, por exemplo, o baixo peso ao nascer relacionado ao óbito (VICTORA et al., 1996) e (ORGANIZAÇÃO PANAMERICANA DA SAÚDE, 1999). Coincidem também com o uso de *linkage* de sistemas de informação relacionados à estudos de coorte sobre mortalidade neonatal (ALMEIDA e JORGE, 1996). Também coincidem com o estudo de características relacionadas à mortalidade perinatal (LANSKY, FRANCA e LEAL, 2002). No estudo da investigação dos óbitos perinatais e infantis para o planejamento de políticas públicas de saúde aparecem características semelhantes (AERTS, 1997). Coincidem também com os estudos da Organização Pan-americana de la Salud em seu Manual sobre el enfoque de riesgo en la atención materno infantil (1999), principalmente o baixo peso ao nascer que aparece com o maior risco relativo no exemplo à página 184. Coincidem também com os estudos sobre as variáveis de condições de vida no Estado do Paraná relacionadas à mortalidade e desigualdades sociais onde é feita uma análise ecológica com base nas variáveis do Censo Demográfico de 2000 (ANDRADE et al., 2004). No trabalho de desigualdades na mortalidade, espaço e estratos sociais, já relacionavam estas variáveis (SILVA et al., 1999). No trabalho de fatores de risco associados à mortalidade infantil em duas áreas da região metropolitana de São Paulo também já havia relacionado as mesmas variáveis (CÉSAR, 1990).

As características geradas pelo DM são importantes também para a construção de mapas de risco. 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-setoriais e intersetoriais no território (LYRA et al., 2005).

O trabalho da evolução da mortalidade infantil por causas evitáveis, já observou estas mesmas causas (CALDEIRA et al., 2005). Os fatores de risco associados à mortalidade infantil em duas áreas da Região Metropolitana de São Paulo (Brasil), 1984-1985, com sua proposta de instrumentos preditivos, já relatava fatores de risco semelhantes aos encontrados pelo *DM* (CHESTER, 1990).

No estabelecimento do estudo sobre a mortalidade no Brasil, em períodos recentes de crise econômica, já se observou fatores semelhantes relacionados à mortalidade (COSTA et al., 2003). No estudo sobre a evolução da mortalidade infantil na Região Metropolitana de São Paulo, já se relatou fatores semelhantes aos aqui encontrados (HOLCMAN, LATORRE e SANTOS, 2004). No trabalho sobre a evolução da mortalidade fetal e infantil já se relataram fatores semelhantes (KOONTZ, KATHLEEN e RUDERMAN, 2004).

Nas recomendações sobre assistência materno-infantil, baseadas no estudo sobre a evolução da mortalidade fetal e infantil já se observaram casos semelhantes aos aqui encontrados (MISRA et al., 2004). Na dissertação de mestrado com aplicação do *DM* na busca de um modelo de prevenção da mortalidade infantil, os resultados foram semelhantes aos encontrados neste estudo (OLIVEIRA, 2001).

No estudo sobre a acurácia dos indicadores de risco do programa de defesa da vida dos lactentes em regiões do estado de São Paulo, Brasil já se observou fatores de risco semelhantes (RUMEL et al., 1992). No estabelecimento das

relações entre baixo peso e mortalidade infantil, observaram-se fatores de risco semelhantes (SILVA et al., 2003).

No artigo sobre capital social realizou-se uma meta-análise com critérios de risco relacionados às características sociais envolvidas com a saúde e a mortalidade, aparecendo também critérios como escolaridade, gestação na adolescência, desemprego e renda, risco nutricional e sexo (PATTUSSI et al., 2006). Alcança-se, desta forma, o terceiro objetivo específico: identificar as relações existentes entre os padrões encontrados pelo *DM* e a literatura.

A conclusão do *DM*, na validação cruzada, permite inferir que foram geradas as melhores regras de maneira objetiva e, juntamente com a análise qualitativa dos 25 especialistas três chamam a atenção pelo que representam:

- A relacionada à mãe adolescente (≤ 16 anos), sugerindo que mesmo com bom peso ao nascer (2500 a 3500g) a mãe adolescente surge como alto critério de risco para o óbito perinatal.
- A regra seguinte, 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 RN.
- A próxima regra, que chama a atenção, revela que o baixo peso ao nascer (≤ 2500 g), com pós - datismo (41 semanas de gestação ao nascer) e feto afetado por afecções maternas, está relacionado ao óbito em 30 dias de vida.

Dessa forma, conclui-se que, para melhorar a atenção aos menores de 1 ano de idade no Paraná, de acordo com as informações fornecidas pelas bases de dados dos 5 anos selecionados e analisados, e após o estabelecimento dos padrões gerados pelo *DM*, devem ocorrer ações voltadas:

- Às adolescentes, principalmente as que já tem outro filho.

- E às mães com problemas na gestação.
- Às mães com filhos de baixo peso ao nascer.
- E às mães com pós - datismo.

Os 25 especialistas consultados foram unânimes na necessidade de maior integração dos especialistas em tecnologia e os da área da saúde para validarem melhor esta e outras ferramentas computacionais que possam vir a auxiliar no estabelecimento ou confirmação de características de grandes bases de dados que dificilmente são analisadas por outras ferramentas computacionais devido ao grande volume dos dados e da falta de integração destes setores.

Após a avaliação e análise desses resultados, confirmou-se que os instrumentos (SIMI) de análise utilizados pelos Comitês de Prevenção da Mortalidade Infantil no Estado do Paraná, são eficazes e podem ser confirmados com outros estudos, e, com mais esta ferramenta computacional utilizada neste estudo. Confirma-se 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 prestadores de serviços de saúde à população infantil (MINISTÉRIO DA SAÚDE, 2003).

Como contribuições deste trabalho observaram-se:

- A integração de bases reais.
- A importância dessa integração de maneira efetiva, como também, a construção dos Sistemas de Informação que deve ser realizada de maneira a poder obter características importantes para estudos futuros com a população abrangida no SI, já observado também em outro trabalho (GREIN, 2005).

- 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 MI, melhorando, dessa forma, as características de evitabilidade, já implementadas por esse estudo.
- Que é necessário melhorar as informações prestadas nas DO, inclusive com o incentivo da implantação dos SVO.
- A conclusão de que para melhorar a aceitação do DM no meio médico (COLLAZOS, BARRETO e ROISEMBERG, 2002), é necessário o estabelecimento de procedimentos para reduzir os problemas de definição e resultados, apesar de sua boa acurácia preditiva (de 78 a 90%), constatado também por (SHEARER, 2000) e (BOSTWICK e BURKE, 2001).
- Durante as entrevistas, provavelmente devido a falta de familiaridade com a ferramenta, percebeu-se alguma resistência dos especialistas em entender o funcionamento da mesma, observado com os relatos sobre o entendimento da linguagem das regras. Conclui-se, portanto, que as ferramentas computacionais facilitam a análise dos dados de grandes bases, desde que haja maior integração entre os profissionais da informática e da área da saúde para elaborar ferramentas mais específicas e fáceis de entender a um grupo maior de usuários.
- É também considerada como uma pesquisa estratégica (MINAYO, 2006), (PELLEGRINI, 2000) para lançar luz sobre o aspecto da realidade relacionada aos óbitos infantis, juntando a investigação de conceitos básicos para a formulação de políticas públicas ou de processos de desenvolvimento.

Como trabalhos futuros, sugerem-se:

a) a utilização desta ferramenta para auxiliar as identificações e especificações de características relacionadas à saúde das populações em outras áreas, e não somente, relativas à mortalidade infantil.

b) Relacionado ainda à mortalidade infantil, pode-se especificar melhor as características identificadas neste estudo e estabelecer, se elas realmente são significativas ou se são devidas ao acaso.

c) A possibilidade de utilizar o mapeamento das ocorrências com a aplicação de outras ferramentas computacionais específicas.

d) A construção de um *Data warehouse*³, que já está em andamento na SESA-PR, juntamente com a sala de situação para facilitar o acesso de todos os usuários a todas as informações de todas as bases de dados existentes no Estado, que também foram sugeridas por este trabalho.

³ *Data warehouse*: armazém de dados inter - relacional, em várias dimensões e não apenas, de maneira linear.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AERTS, D. R. G. de C. **Investigação dos óbitos perinatais e infantis: seu uso no planejamento de políticas públicas de saúde.** Jornal de. Pediatria. (Rio de Janeiro.) 1997; 73(6): 364-366.

AGRAWAL, R. S.; IMIELÍNSKI, T.; SWAMI, A. **Mining Association Rules between sets of itens in databases.** U.S.A. ACM Press In Proceedings of the 1993 ACM SIGMOD International Conference on Management of data, 1993. ISBN: 0-89791-592-5.

ALMEIDA, M. F.; JORGE, M. H. P. de M. **O uso da técnica de “Linkage” de sistemas de informação em estudos de coorte sobre mortalidade neonatal.** Revista de Saúde Pública v. 30 n. 2 SP, abr. 1996.

ANDRADE, S. M.; MATSUO, T.; SOARES, D.A.; SOUZA, R.K.T.; MATHIAS, T.A. de F.; IWAKURA, M. L. H.; ZEQUIM, M.A. **Condições de vida no Estado do Paraná: Análise Ecológica com base em variáveis do Censo Demográfico de 2000.** Semina: Ciências Biológicas e da Saúde, Londrina, v.25, p.73-80, Jan./Dez 2004.

ÁVILA, B. C. **Data Mining.** Evento da VI Escola de Informática da Sociedade Brasileira de Computação – Regional Sul, 1998.

BARRETO, J. M.; PELEGRINI, G. F.; COLLAZOS, K. L. **Extração do Conhecimento a partir dos Sistemas de Informação.** Anais VII Congresso Brasileiro de Informática em Saúde e II Simpósio Internacional de Sistemas de Informação Hospitalar (CBIS'2000 SISIH'2000), SP-BR, 14-18 out. 2000.

BARRETO, J. M.; FALQUATO, J.; BORGES, P. S. S. **Amplification of Perspectives in the use of Evolutionary Computation.** IEEE International Symposium on Bioinformatics and Biomedical Engineering BIBE 2000, Arlington, VA., 8-10, p. 150-157, nov. 2000a.

BERRY, M. J. A; LINOFF, G. **Data Mining techniques for marketing, sales, and costumer support**. John Willey & Sons. New York, 1997.

BERRY, M.; LINOFF, G. **Mastering Data Mining- Art and Science of Customer Relationship Management**. Ed. Wiley. 2000.

BERSON, A.; SMITH, S. J. **Data Warehousing, Data Mining and OLAP**. N.Y., U.S.A. McGraw-Hill, ISBN0-07-006272-2. 1998.

BOSTWICK, D.G.; BURKE, H. B. **Prediction of individual patient outcome in cancer**. Cancer supplement, 2001; 91:1643-6.

BRAUNER, D. F. **O Processo de Descoberta de Conhecimento em Banco de Dados: um estudo de caso sobre os dados da UFPEL**. Pelotas. Universidade Federal de Pelotas, 2003.

CABENA, P.; STADLER, R.; ZANASI, A. **Discovering data mining: From concept to implementation**. Upper Saddle River, NJ: Prentice Hall, 1998.

CALDEIRA, A. P.; FRANÇA, E.; PERPÉTUO, I. H. O.; GOULART, E. M. A. **Evolução da mortalidade infantil por causas evitáveis, Belo Horizonte, 1984-1998**. Revista de Saúde Pública. V. 39, n. 1. S. P. Feb. 2005.

CARROLL, P.W; YASNOFF, W.A; WARD, M.E.; RIPP, L.H.; MARTIN, E.L. **Public Health Informatics and Information Systems**. USA. Springer-Verlag. New York, Inc. 2003.

CARVALHO, D. R. **Um método híbrido árvore de decisão/ algoritmo genético para data mining**. Tese de Doutorado. Pontifícia Universidade Católica do Paraná. Curitiba, 2002.

CAVALCANTE, M. S.; RAMOS JR., A. N.; PONTES, L. R. S. K. **Linkage of Health Information Systems: a Strategy to optimize the Surveillance of**

Pregnant women infected by HIV. Epidemiologia e Serviços de Saúde. Revista do Sistema Único de Saúde do Brasil. Volume 14 - Nº 2 – abril / junho de 2005. ISSN 1679-4974.

CESAR, C. L. G. **Fatores de risco associados à mortalidade infantil em duas áreas da região metropolitana de São Paulo (Brasil), 1984-1985. Proposta de instrumentos preditivos.** Revista de Saúde Pública, SP, 24: 300-10, 1990.

CHESTER, L.G. C. **Fatores de risco associados à mortalidade infantil em duas áreas da região metropolitana de São Paulo (Brasil), 1984-1985. Proposta de instrumentos preditivos.** Revista de Saúde Pública. V. 24, n 4 São Paulo ago. 1990. ISSN 0034-8910.

CIOS, K.; PEDRYCZ, W; SWINIARSKI, R. **Data Mining: Methods for Knowledge Discovery.** Kluwer Academic Publishers. Boston, Dordrecht, London, 1998.

COLLAZOS, K.; BARRETO, J. M. **KDD para o estudo epidemiológico das malformações.** In I Congresso Peruano de Ingenieria Biomédica, TUMI'99, P.U. C, Peru, 8-9/09 1999.

COLLAZOS, K.; BARRETO, J. M. **KDD em base de dados da Área Médica.** In XVII Congresso Brasileiro de Engenharia Biomédica. CBEB 2000, Florianópolis, p. 765-768, 11-13 set. 2000.

COLLAZOS, K.; BARRETO, J. M. e ROISEMBERG, M. **Dificuldades na aplicação de KDD na medicina.** In II Workshop de Informática aplicada à Saúde-CBComp 2002.

COSTA, M. C. N.; MOTA, E. L. A.; PAIM, J. S.; SILVA, L. M. V da; TEIXEIRA, M da G.; MENDES, C. M. C. **Mortalidade no Brasil em períodos recentes de crise econômica.** In Revista de Saúde Pública. 2003. 37 (6): 699-709.

CUNHA, M. G. F. da; TEIXEIRA, A. M. da S. **Potencialidades do Sistema de Informações sobre Nascidos Vivos (SINASC) como instrumento de regulação para internações hospitalares obstétricas no Sistema Único de Saúde (SUS)**. Anais da 4ª EXPOEPI. Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças. Brasília – DF, 2005.

CUROTTO, C. L. **Integração de recursos de Data Mining com gerenciadores de bancos de dados relacionais**. Tese de Engenharia Civil da Universidade Federal do Rio de Janeiro, R.J., BR, 2003.

DHAR, V.; STEIN, R. **Seven methods for transforming corporate data into business intelligence**. Upper Saddle River, NJ: prentice Hall, 1997.

FAUSETT, L. V. **Fundamentals of neural net works: architectures, algorithms, and applications**. ISBN-13-334186-0 Prentice- Hall Inc., 1994.

FAYYAD, U.; WEIR, N.; DJORGOVSKI, S. **Automated Analysis of a Large-Scale Sky Survey: The SKICAT System**. In **Knowledge Discovery in Databases Workshop 1993**, Working Notes, 1993, p. 1-13.

FAYYAD, U. M.; SHAPIRO, G. P.; SMYTH, P. **From Data Mining to Knowledge Discovery: an Overview**. In **Advances in Knowledge. Discovery and Data Mining**. U.M. Fayyad et al. California Editor. Cap. 1, p.1-34, 1996.

FAYYAD, U. M.; SHAPIRO, G. P.; SMYTH, P.; UTHURUSAMY, R. **Advances in Knowledge Discovery and Data Mining**. California America Association for Artificial Intelligent, California, 1996a.

FAYYAD, U.; SHAPIRO, G. P.; SMYTH, P. **The KDD Process for Extracting Useful Knowledge from Volumes of Data**. Association for Computing Machinery, p. 27. Nov. 1996b; 39,11.

FREITAS, F.; COSTA, S. H. M.; RAMOS, J. G. L.; MAGALHÃES, J. A. **Rotinas em Obstetrícia**. Artmed Ed. SP, 2001.

FUNDAÇÃO NACIONAL DE SAÚDE, **Manual de procedimentos do Sistema de Informações sobre Mortalidade**. FUNASA. Brasília, abril de 2003.

FUNDAÇÃO NACIONAL DE SAÚDE, **Manual de instruções para o preenchimento da Declaração de Óbito**. FUNASA. Brasília, agosto de 2001.

FUNDAÇÃO NACIONAL DE SAÚDE, **Manual de Procedimentos do Sistema de Informações sobre Nascidos Vivos**. FUNASA. Brasília, agosto de 2001a.

FUNDAÇÃO NACIONAL DE SAÚDE, **Manual de Instruções para o preenchimento da Declaração de nascido Vivo**. FUNASA. Brasília, agosto de 2001b.

GLYMOUR, C.; MADIGAN, D.; PREGIBON, D; SMYTH, P. **Statistical themes and lessons for data mining**. Journal of Data Mining and Knowledge Discovery. V.1, n 1, p 11-26, 1997.

GOLDSCHIMIDT, R; PASSOS, E. **Data Mining um guia prático**. Editora Campus/ Elsevier, SP, 2005.

GREIN, D. **Uma contribuição para a integração de sistemas legados da Saúde Pública do Brasil usando agentes de software**. Dissertação de mestrado de Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná, Curitiba, 2005.

GROTH, R. **Data Mining: hands on approach for business professionals**. ISBN-13-756412-0 New Jersey, USA. Prentice Hall, 1998.

HAN, J; KAMBER, M. **Data Mining: Concepts and Techniques**. Califórnia, USA. Morgan Kaufmann Publishers, 2001.

HAND, D. J.; MANNILA, H.; SMYTH, P. **Principles of data mining**. Cambridge. MA: MIT Press, 2001.

HARRICSON, T. H. **Intranet Data Warehouse**. São Paulo. Ed. Berkeley, 1998.

HOLCMAN, M. M.; LATORRE, M. R. D. O; SANTOS, J. L. F. **Evolução da mortalidade infantil na região metropolitana de São Paulo, 1980-2000**. Revista de Saúde Pública v. 38 n. 2. SP., abril 2004. ISSN 0034-89, 10.

IAN, H. W. e FRANK E. **Data Mining Practical Machine Learning Tools and Techniques**. Segunda edição. Morgan Kaufmann Publishers, U. S. A, 2005.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Anuário Estatístico**. 1999.

IMIELINSKI, T.; MANNILA, H. **A Database Perspective on Knowledge Discovery**. Association for Computing Machinery. p. 58. Nov. 1996; 39-11.

KLUCK, M.M. **Metodologia para ajuste de desfechos hospitalares por risco prévio do paciente**. Tese Doutorado. Programa de Pós-Graduação em Epidemiologia, UFRGS, Porto Alegre, dezembro 2004.

KOONTZ, A.M.; KATHLEEN, A.B.; RUDERMAN, M. **The Evolution of Fetal and Infant Mortality** Review as a Public Health Strategy. USA Maternal and Child Health Journal, Vol.8, n.4. Dec. 2004, p. 195-203. ISSN 1092-7875 (paper).

KUSIAK, A; KERN, J. A.; KERNSTINE, K.H.; TSENG, B. T. L. **Autonomous Decision Making a Data Mining Approach**. IEEE. Transactions on Information Technology in Biomedicine, vol. 4, número 4. Dec. 2000.

LANSKY, S.; FRANCA, E.; LEAL, M. C. **Mortalidade perinatal e evitabilidade: revisão da literatura**. Revista de Saúde Pública; 36 (6): 759-772, dez 2002. tab.

no projeto da World Health Organization Pan-Americana da Saúde. AMR/99/078643-01.

LAURENTI, R.; JORGE, M. H. P de M.; LEBRÃO, M. L.; GOTLIEB, S. L. D. **Estatísticas de Saúde**. Ed. EPU. SP, 1987.

LAURENTI, R.; HARAKI, C. A C.; GOTILIEB, S. L. D. **Confiabilidade do Sistema de Informações sobre Mortalidade em município do sul do Estado de São Paulo**. Revista Brasileira de Epidemiologia. Vol.8 n 1. SP Mar. 2005.

LINARES, K.S.C. **Aspectos Teóricos do Data Mining: Descoberta do Conhecimento em Medicina**. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Elétrica, UFSC, Florianópolis, junho 2003.

LYRA, T. M.; GUIMARÃES, M. J. B.; MELO FILHO, D. A.; NASCIMENTO JUNIOR, J. A. do; LIMA, A. A. F.; MONTENEGRO, D.; CAMARÃO, F.; ACIOLI, R.V.; MELO, N. G. **Utilização de mapas de risco para vigilância em saúde: a experiência do Recife**. Anais da 4ª EXPOEPI. Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças. Brasília – DF, 2005.

LIU, H.; LU, H.; YAO, J. **Toward Multidatabase Mining. Identifying relevant database**. IEEE. Transactions on Knowledge and Data Engineering. 13 (4). 541-553. July, August, 2001.

MANSANO, N. H.; MAZZA, V. de A.; SOARES, V. M. N.; ARALDI, M. A. R.; CABRAL, V. L. M. **Comitês de prevenção da mortalidade infantil no Paraná, Brasil: implantação e operacionalização**. Cadernos de Saúde Pública. V. 20, n 1. Fevereiro 2004. ISSN 0102-311X.

MARTIN, G. B.; CORDONI JR. L.; BASTOS, Y. G. L. **Aspectos demográficos do processo de envelhecimento em cidade do sul do Brasil**. Epidemiologia e

Serviços de Saúde. Revista do Sistema Único de Saúde do Brasil. Volume 14 – N 3 – julho / setembro de 2005. ISSN 1679-4974.

MARTIN, J.; MCLURE, C. **Técnicas estruturadas e case**. McGraw Hill, 1991.

MATHIAS, T. A. de F.; JORGE, M. H. P. de M.; LAURENTI, R.; AIDAR, T. **Considerações sobre a qualidade de informações de mortalidade na população idosa residente no Município de Maringá, Estado do Paraná, Brasil, no período de 1979 a 1998**. Epidemiologia e Serviços de Saúde. Revista do Sistema Único de Saúde do Brasil. Volume 14 - Nº 3 – julho / setembro de 2005. ISSN 1679-4974.

McDONALD, J.M; BROSSETTE, S.; MOSER, S.A. **Pathology information systems; data mining leads to knowledge discovery. Data mining and knowledge discovery techniques have great potential**. Arch Pathol Lab Med 1998 May; 122(5):409-11. Informatics Review Vol. 2 n 12.

MEDRONHO, R. **A Epidemiologia**. S. P. Editora Atheneu, 2003.

MELLO JORGE, H. **Registro dos eventos vitais: sua importância em Saúde Pública**. Centro da Organização Mundial de Saúde (OMS) para classificação de Doenças em Portugal. Série Divulgação, número 2. São Paulo, 2001.

MELLO JORGE, M. H. P de; GOTLIEB, S. L. D; SOBOLL, M. L. de M. S.; ALMEIDA, M. F. de; LATORRE, M. do R. D. de O. **Avaliação sobre o sistema de informação sobre nascidos vivos e o uso de seus dados em epidemiologia e estatística de saúde**. Revista de Saúde Pública; 27 (supl): 1-46, dez. 1993. tab.

MENESES, C.J.; GRINSTEIN, C.G. **Categorization and Evaluation of Data Mining Techniques**. In Data Mining Proceedings of the Int'l Conference on Data Mining, R.J. WIT Press, p. 53-80, 1998.

MINAYO, M.C.de S. **O desafio do conhecimento**. Nona edição. São Paulo, 2006.

MINISTÉRIO DA SAÚDE. **Informe Epidemiológico SUS (Brasil)-Ano 4**. Brasília: CENEPI. Sistemas de Informação em Saúde, 1995.

MINISTÉRIO DA SAÚDE. **SINASC: Manual de Procedimento**, Terceira Edição, Brasília, 1999. **SIM: Manual de Procedimento**, Terceira Edição, Brasília, 1999.

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**. Primeira edição. Brasília, 2003.

MINISTÉRIO DA SAÚDE. **Saúde Brasil 2004. Uma análise da Situação de Saúde**. Brasília. Maio de 2004. p. 94 e p.120 a 132.

MINISTÉRIO DA SAÚDE. **Manual do Comitê de Prevenção do Óbito Infantil e fetal**. Brasília, 2004a.

MINISTÉRIO DA SAÚDE. **DataSUS**, MS. Brasília, 2005.

MISRA, D. P.; GRASON, H.; LIAO, M.; STROBINO, D. M.; McDONNELL, K.A; ALLSTON, A. A. **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 Journal; 8(4): 217-29, 2004 Dec.

MORO, C. M. C. **Medidas de similaridade entre conceitos da Saúde**. Tese de Doutorado. Escola Politécnica da Universidade de São Paulo - Poli/USP, 2003.

OLIVEIRA, I.C. **Aplicação de Data Mining na Busca de um Modelo de Prevenção da Mortalidade Infantil**. Dissertação de Mestrado do Programa de

Pós-graduação em Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina. Florianópolis, 2001.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Dados estatísticos**. ONU, Genebra, 2004.

ORGANIZAÇÃO MUNDIAL DE SAÚDE 1994 In **Manual dos Comitês de Prevenção do Óbito Infantil e Fetal**. Ministério da Saúde. Brasília, 2004a.

ORGANIZAÇÃO MUNDIAL DE SAÚDE. **Classificação Estatística Internacional de Doenças (CID) 10**. OMS. p. 136 a 142. Editora da Universidade de S. P., 2004b.

ORGANIZAÇÃO PANAMERICANA DE LA SALUD. **Manual sobre el enfoque de riesgo en la atención materno infantil**. OPS. p. 184. Editor Dr. Carlos Castillo-Salgado. Washington, D.C. 20037, 1999.

PARSAYE, K. **Intelligent Data Base Object-Oriented, deductive and hypermedia technologies**. New York. John Wiley, 1989.

PATTUSSI, M. P.; MOYSÉS, S. J.; JUNGES, J. R.; SHETHAM, A. **Capital social e a agenda de pesquisa em epidemiologia**. Cadernos de Saúde Pública, v. 2, n 8. Rio de Janeiro, Aug. 2006.

PELLEGRINI, G. F.; COLLAZOS, K. **Extração de Conhecimentos a partir dos Sistemas de Informações**. SC Artigos da Universidade Federal de Santa Catarina, 1999.

PELLEGRINI, A. **Ciência en Pro de la Salud**. Washington: OPAS, 2000.

PEREIRA, B. B. **Estatística em Psiquiatria**. Revista Brasileira de Psiquiatria, 23 (3): 178-70, 2001.

PIATETSKY-SHAPIRO, G.; W. J. Frawley, editors. **Knowledge discovery in databases**. Menlo Park, CA: AAAI Press/MIT Press, 1991.

PIATETSKY-SHAPIRO. **The Data-Mining Industry Coming of Age**. IEEE Intelligent Systems, p. 32-34, 1999.

PUSTAI, A. K.; BRITO, M. R. V.; WEBER, M.A.; BELLO, M.I.; FETZER, L.; BRITTO, M.J.; BARCELLOS, C. **Uso do georreferenciamento na prevenção do dengue em Porto Alegre – uma proposta de ferramenta para uso local**. Anais da 4ª EXPOEPI. Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças. Brasília – DF, 2005.

PYLE, D. **Data preparation for Data Mining**. U.S.A. Morgan Kaufman Publishers, 1999.

QUINLAN, J. R. **Introduction of Decision Trees**. Machine Learning. Boston, 1986. V.1. p.81-106.

QUONIAM, L.; TARAPANOFF, K.; ARAÚJO JÚNIOR, R. H.; ALVARES, L. **Inteligência obtida pela aplicação de Data Mining em base de teses francesas sobre o Brasil**. Ciência da Informação. Brasília, v. 30, número 2, p. 20-28. Maio, Agosto, 2001.

REZENDE, S. O. **Sistemas Inteligentes Fundamentos e Aplicações**. Editora Manole. SP, 2005.

RICHARS, G.; RAYWARD-SMITH, V. J.; SÖNKSEN, P. H.; CAREY, S.; WENG, C. **Data Mining for indicators of early mortality in a database of clinical records**. Artificial Intelligence in Medicine, 2001, Jun.; 22 (3): 215-31.

ROSEN, G. **Primeira História da Saúde Pública**. Segunda Edição Hucitec Ltda. SP, 1958.

RUMEL, D; COSTA FILHO, D. C.; AIELLO, S.; KOYAMA, S. M.; ROSENBAUM, W. **Acurácia dos Indicadores de Risco do Programa de Defesa da Vida dos Lactentes em regiões do estado de São Paulo, Brasil.** Revista de Saúde Pública, 6-11, 1992.

SETH, P.; GUATAM, N. T.; BALINT, R. **Preparing and Mining Data with Microsoft[®] SQLServer[™] 2000 and Analysis Services.** Online ed. U.S.A. Microsoft Online Books, 2002.

SHEARER, R. C. The CRISP - **DM model the new blueprint for data mining.** J. Data Warehousing 2000; 5:13-22.

SHORTLIFFE, E. H.; PERREAULT, L. E. **Medical Informatics Computer Applications in Health Care and Biomedicine.** Springer Ed. Canadá, 2000.

SILVA, A. A.; BETTIOL, H.; BARBIERI, M. A.; RIBEIRO, V. S.; ARAGÃO, V. M. de F.; BRITO, L. G. O.; PEREIRA, M. M. **Infant mortality and low birth weight in cities of Northeastern and Southeastern Brazil.** Revista de Saúde Pública v.37 n.6 São Paulo dez. 2003. ISSN 0034-8910.

SILVA, L. M. V. da; PAIM, J.S.; COSTA, M. da C. **Desigualdades na mortalidade, espaço e estratos sociais.** Revista de Saúde Pública. Journal of Public Health, 33 (2): 187-97, 1999.

SILVER, D. L.; **Knowledge Discovery and Data Mining**, MBA course notes of Dalhousie University Nova Scotia, Canada, 1998.

SINASC-MINISTÉRIO DA SAÚDE, SIM-MINISTÉRIO DA SAÚDE e SIMI-SECRETARIA DE ESTADO DA SAÚDE-PR **Sistemas de Informação em Saúde** in DataSUS e SESA-PR , 2005.

SOUZA, P. T. A. de; TOSETTO, V. G. **Mudança na estratégia de coleta de informações com alteração do perfil epidemiológico de mortalidade.** Anais

da 4ª EXPOEPI. Mostra Nacional de Experiências Bem-Sucedidas em Epidemiologia, Prevenção e Controle de Doenças. Brasília, DF, 2005.

SRIKANT, R.; AGRAWAL, R. **MINING GENERALIZED ASSOCIATION RULES**. In Proceedings of the 21st Very Large Databases Conference. Zurique, Suíça, 1995.

STEINER, M. T. A.; SOMA, N, Y.; SHIMIZU, T.; NIEVOLA, J. C. **Abordagem de um problema médico por meio de um processo de KDD com ênfase à análise exploratória de dados**. Gestão & Produção v.13 n.2 p. 325-337, maio-ago, 2006.

TRINDADE, C. M. **Descoberta de perfis epidemiológicos sobre as Hepatites Virais a partir da base de dados de saúde pública**. Dissertação de Mestrado. Programa de Pós-Graduação de Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná, Curitiba, 2005.

VAN BEMMEL, J. H.; MUSEN, M. A. **Handbook of Medical Informatics**. Springer Helder Ed. Germany, 1997.

VIANNA, R. C. X. F. **Situação da Mortalidade Infantil no triênio de 2000 a 2002**. Anais do VI Congresso Brasileiro de Epidemiologia - EPIRECIFE. 2004.

VIANNA, R. C. X. F **Situação da Mortalidade Infantil no Paraná em 2003**. Anais do IV Simpósio Internacional de Vigilância Sanitária em 2004a.

VIANNA, R. C. X. F **Sistema de Investigação da Mortalidade Infantil (SIMI)**. Anais de IV Congresso de Pediatria-Pediatria Social, Curitiba, 2005.

VIANNA, R. C. X. F **Análise da Situação da Mortalidade Infantil no Estado do Paraná, Brazil em 2004**. Anais do II Congresso Internacional de Especialidades Pediátricas, Criança 2005, em 2005a.

VIANNA, R. C. X. F.; MORO, C. M. C. **Aplicação da Descoberta de Conhecimento em Base de Dados para a identificação de características relacionadas à Mortalidade Infantil.** In Proceedings da Conferência IADIS (International Association for Development of the Information Society) Íbero-Americana, Lisboa, Portugal, 2005b.

VIANNA, R. C. X. F.; MORO, C. M. C. **Histórico da Mortalidade Infantil no Estado do Paraná, Brazil, antes e após o Sistema de Investigação da Mortalidade Infantil (SIMI).** Anais do XI Congresso Mundial de Saúde Coletiva, Rio de Janeiro, 2006.

VICTORA, C. G.; HALPERN, R.; BARROS, F. C.; HORTA, B. L. **Baixo peso ao nascer em duas coortes de base populacional no sul do Brasil.** Cadernos de Saúde Pública. S. P. v. 12 supl. 1 R.J., 1996.

VICTORA, C. G.; BARROS, F.; BARROS, A.; SANTOS, I.; ALBERNAZ, E.; MATIJASEVICH, F.A.; DOMINGUES, M.; SCLOWITZ, I.; HALLAL, P.; SILVEIRA, M. **The challenge of reducing neonatal mortality in middle income countries: findings from three Brazilian birth cohorts in 1982, 1993 and 2004.** The Lancet, 2005; 365: 847-854. DOI: 10.1016/S0140-6736(05)71042-4.

WEISS, S. M.; INDURKHA, N. **Predictive data mining: A practical guide.** Morgan Kaufmann Editors, San Francisco, 1998.

WIGGLESWORTH, J. S.; **Monitoring perinatal mortality: a pathophysiological approach.** Lancet 1980; 27:684-6.

WITTEN, I. H.; FRANK, E. **Data Mining practical Machine Learning Tools and Techniques with Java Implementations.** ISBN1-55860-552-5. Morgan Kaufmann Publishers, USA, 2000.

7. ANEXOS

Anexo 1: Declaração de Nascido Vivo

Modelo da Declaração de Nascido Vivo

República Federativa do Brasil
Ministério da Saúde
1º Via - Secretaria de Saúde

Declaração de Nascido Vivo Nº _____

I	Cópis	1) Cartório	Código	2) Registro	3) Data	
		4) Município				5) UF
II	Local da Ocorrência					
		6) Local da ocorrência 1 - Hospital 2 - Outros estabelecimento 3 - Domicílio 4 - Outros 5 - Ignorado	7) Estabelecimento	Código		
		8) Endereço da ocorrência, se fora do estado ou da residência da mãe (Rua, praça, avenida, etc)	Número	Complemento	9) CEP	
		10) Bairro/Distrito	Código	11) Município da ocorrência	Código	12) UF
III	Mãe					
		13) Nome da mãe	14) RG			
		15) Mãe 1 - Solteira 2 - Casada 3 - Viúva 4 - Sep. Judicialmente 5 - Ignorado	16) Estado civil	17) Escolaridade 1 - Nenhuma 2 - De 1 a 3 3 - De 4 a 7 4 - De 8 a 11 5 - 12 e mais 6 - Ignorado	18) Ocupação habitual e ramo de atividade	19) Num. de filhos vivos, em gestações anteriores (Caso utilize RG ou Ignorado) Nascidos vivos Nascidos mortos
		20) Nascimento da mãe				
		21) Legitimidade	Número	Complemento	22) CEP	
		23) Bairro/Distrito	Código	24) Município	Código	25) UF
IV	Gestação e Parto					
		26) Duração da gestação em semanas	27) Tipo de gravidez	28) Tipo de parto	29) Número de consultas de pré-natal	
		Semanas 30 - Ignorado	1 - Única 2 - Única 3 - Única e mais 4 - Ignorado	1 - Vaginal 2 - Cesárea 3 - Ignorado	31 - Ignorado	
V	Nascimento					
		32) Nascimento	Data	33) Sexo	34) Índice de Apgar	
		35) Hora		M - Masculino F - Feminino 1 - Ignorado	1º minuto 5º minuto	
		36) Raça/cor	37) Peso ao nascer	38) em gramas		
		1 - Branca 2 - Parda 3 - Amarela 4 - Preta 5 - Indígena				
		39) Detectada alguma malformação congênita, erro anatômico ou congênito?	Código			
		1 - Sim 2 - Não 3 - Ignorado				
VI	Identificação					
		40) Polgar direito da mãe	41) Pê direito da criança			
VII	Assinatura					
		42) Responsável pelo preenchimento	43) Função	44) Identidade	45) Órgão Emissor	46) Data
		Morta				
ATENÇÃO : ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO O Registro de Nascimento é obrigatório por lei. Para registrar esta criança, o pai ou responsável deverá levar este documento ao cartório de registro civil.						

Anexo 2: Declaração de Óbito

República Federativa do Brasil
Ministério da Saúde
1ª VIA - SECRETARIA DE SAÚDE

Declaração de Óbito

I Cartório

1) Cartório Código 2) Registro 3) Data

4) Município 5) UF 6) Cartório

II Identificação

7) Tipo de Óbito 8) Cód. Data Hora 9) R.C. 10) Naturalidade

11) Nome do falecido 12) Nome da mãe

13) Data de nascimento 14) Sexo 15) Estado civil 16) Escolaridade (Em anos de estudo concluídos) 17) Ocupação habitual e ramo de atividade (Se aposentado, indicar a ocupação habitual anterior) 18) Raciocínio

19) Logradouro (Rua, praça, avenida etc.) 20) Código 21) Número 22) Complemento 23) CEP

24) Bairro/Distrito 25) Município de residência 26) UF

III Residência

27) Local de ocorrência do óbito 28) Estabelecimento 29) Código

30) Endereço da ocorrência, se fora do estabelecimento ou da residência (Rua, praça, avenida etc.) 31) Número 32) Complemento 33) CEP

34) Bairro/Distrito 35) Código 36) Município de ocorrência 37) UF

IV Ocorrência

V Fetal ou menor que 1 ano

38) PREENHIMENTO EXCLUSIVO PARA ÓBITOS FETAIS E DE MENORES DE 1 ANO

39) Nome 40) Escolaridade (Em anos de estudo concluídos) 41) Ocupação habitual e ramo de atividade da mãe 42) Número de filhos vivos (Indicar se falecido)

43) Duração da gestação (Em semanas) 44) Tipo de Gravidez 45) Tipo de parto 46) Morte em relação ao parto 47) Peso ao nascer 48) Número da Declaração de Nascido Vivo

49) A morte ocorreu durante a gravidez, parto ou parto? 50) A morte ocorreu antes ou durante o parto? 51) Recurso assist. médica durante a doença que ocasionou a morte?

52) Exame complementar? 53) Cirurgia? 54) Necropsia?

VI Causas da morte

55) Causas da morte (Parte I) 56) Causas da morte (Parte II)

VII Médico

57) Nome do médico 58) CRM 59) O médico que assinou atendeu ao falecido?

60) Meio de contato (Telefone, fax, e-mail etc.) 61) Data do atestado 62) Assinatura

VIII Causas externas

63) Prováveis circunstâncias de morte não natural (informações de caráter estritamente epidemiológico)

64) Tipo 65) Acidente do trabalho 66) Fonte de informação

67) Descrição sumária do evento, incluindo o tipo de local de ocorrência

68) SE A OCORRÊNCIA FOR EM VIA PÚBLICA, ANEXAR O ENQUILÓ

69) Logradouro (Rua, praça, avenida etc.) 70) Código

IX Local de ocorrência

71) Declarante 72) Testemunhas

Versão 0000-01

Anexo 3: Tabela 10 - Critérios de Evitabilidade

Tabela 10 - Critérios de Evitabilidade:

Grupo A: óbitos reduzíveis por imunoprevenção		
Número Doença	Doença	CID10
01	Difteria	A36
02	Coqueluche	A37
03	Hepatite	B15-B19
04	Poliomielite	A80
05	Rubéola Congênita	P35.0
06	Sarampo	B05
07	Tétano	A35
08	Tétano Neonatal	A33
09	Tuberculose	A15-A19
10	Varicela	B01
Grupo B: óbitos reduzíveis por saneamento básico, terapia de reidratação oral, incentivo ao aleitamento materno e educação para a saúde.		
01	Enteroinfecções, diarreias e Desidratação	A00-A09 E86
Grupo C: óbitos reduzíveis por medidas de combate à desnutrição, incentivo ao aleitamento materno e educação para a saúde.		
01	desnutrição	E40-E46; E63;
Grupo D: óbitos reduzíveis por adequada atenção na gestação, no parto e ao recém-nascido e educação para a saúde.		
01	Afecções respiratórias do RN	P22-P28
02	Prematuridade e baixo peso ao nascer	P05, P07;
03	Asfixia ao nascer	P20, P21;
04	Óbito causado por afecções maternas	P00-P03
05	Sífilis congênita	A50
06	Outros problemas do RN	P04; P08; P15; P24-P27; P29; P35; P37-P94; P95; P96
Grupo E: óbitos reduzíveis por diagnóstico e tratamento precoces e educação para a saúde.		
01	Infecções respiratórias (pneumonia, gripe)	J10-J16; J18; J20-21; J06; J85; J86
02	Outras afecções do aparelho respiratório	J45; J46; J31; J91-99;
03	Meningites	A39; G00; G03; B45.1; A54.8; A87
04	Septicemias (Infecções generalizadas)	P36; A40; A41;
05	Outras doenças infecciosas (malária, hepatite)	B50-B54; B15; B17-19; A27; B20-24; A46; H66
Grupo F: óbitos reduzíveis por atenção adequada ao trauma e urgências e educação.		
01	Afogamento	W65-70; W73; W74; X92; Y21;
02	Queimadura	X00-X06; X08; X09; X12; X48; X97;
03	Quedas	W00-19; V80; Y01; Y30;
04	Acidentes de trânsito (inclui atropelamentos)	V01-99;
05	Outras mortes violentas	W85-W99; X43; X59; X91; X99; Y06; Y02-Y20; Y22- 29; Y31 98;

06	Broncoaspiração (conteúdo gástrico, corpo estranho)	P24; W75; W78-84; J69;
Grupo G: óbitos dificilmente reduzíveis.		
01	Anomalias congênitas	Q00-99
02 Neoplasias	(câncer, leucemia)	C00-99
Grupo H: óbitos devidos a outras causas de morte.		
01	Outras causas de morte	A86; D00-D89; E03; E24; E25; E84; G00-G99;
Grupo I: causas mal definidas.		
01 E87; E88; R00-R99;	Causas mal definidas (com assistência médica)	E87; E88; R00-R99;

Anexo 4: Tabela 11 – Medidas de prevenção e intervenção da MI

Tabela 11 - Medidas de prevenção e intervenção da mortalidade infantil:

CÓDIGO	TIPO
	1. Medidas para atenção ambulatorial
1.A	Garantir acesso da gestante ao pré-natal
1.B	Qualidade no pré-natal
1.C	Acesso a tratamento adequado em serviços especializados em gestação de alto risco
1.D	Realizar vigilância do risco gestacional
1.E	Busca ativa à criança de risco (vigilância e visita domiciliar)
1.F	Acesso ao tratamento adequado das complicações a nível ambulatorial
1.G	Outras
	2 Medidas para atenção hospitalar
2.A	Garantir acesso a internamento hospitalar
2.B	Melhor diagnóstico a nível hospitalar
2.C	Acesso a tratamento adequado à mãe e criança
2.D	Assistência adequada ao parto
2.E	Acesso a medicação
2.F	Acesso à referência secundária e ou terciária
2.G	Outras
	3 Medidas de Educação em Saúde aos familiares
3.A	Melhor acesso à informação em saúde
	4 Medidas Sociais
4 A	Melhores condições de renda familiar
4.B	Melhores condições de escolaridade
4.C	Melhores condições de infra-estrutura básica (saneamento, habitação, etc.)
4.D	Outras

Anexo 5:Tela do SIMI

Análise de Óbito Infantil

Alterar Incluir Excluir Cancelar Confirmar Consultar Imprimir Primeiro Anterior Próximo Último Sair **Total de Análises : 8686**

Análise do Óbito Infantil

Comitê: ☐ Municipal ☐ Regional ☐ Estadual

Município de Residência:

Análise Óbito

Mãe 22a, alcoista, péssimas condições sócio econômicas, 3a. gestação, 1 filho vivo, Pré natal tardio e deficitário. Gestação a termo, parto normal, RN em boas condições, baixo peso, amamentado no seio. Após amamentá-lo, colocou

DO

Numero: 0000000 Data / Hora do Nascimento: 26/05/2003 21:00 Data/Hora do Óbito: 05/06/2003 04:00 Encerramento Análise: 12/08/2003

Causa Básica na DO

CID: R98 Descrição da Causa Básica na DO: Morte sem assistência

Causa Básica após Investigação

CID: P243 Descrição da Causa Básica após Investigação: Aspiração neonatal de leite e alimento regurgitados

Morte Evitável

Morte Evitável: Sim

Código de Evitabilidade

Código: F06 Descrição: BRONCOASPIRAÇÃO (CONTEÚDO GÁSTRICO CO

Responsabilidade pelo Óbito

Família: ☐ Social: ☐

Analisadores

Membro Presente: Profissão:

Medidas de Prevenção

Código	Prevenção	Descrição
1e	Busca ativa à criança de risco (vigilância e visita domiciliar)	Medidas para atenção ambulatorial
3a	Melhor acesso à informação em saúde	Medidas de Educação em Saúde aos família

Qualidade da Ficha

☐ Satisfatória ☐ Pouco Satisfatória ☐ Insatisfatória

Última Atualização

13/08/2003

13:24 10 07/03/2005

Fonte: SESA PR BR

Figura 8 - Tela da Ficha de Análise do Óbito Infantil.

Anexo 6: Descrição da Ficha de Análise de Óbito Infantil

Comitê de Análise do Óbito: Distrital, Municipal, Regional ou Estadual-Comitê que efetuou a análise do óbito.

Regional de Saúde e Município de residência: município e regional de residência da criança.

Número da Declaração de Óbito (DO): dados da Declaração de Óbito como data e hora do óbito. Estabelecido pelo Código Sanitário do Estado do Paraná 2004. São dados que localizam a declaração de óbito da criança, pelo número da DO é possível o cruzamento com informações disponibilizadas no Sistema de Informação de Mortalidade (SIM). Por meio destes dados é possível também avaliar os dias ou horários mais precários no atendimento de saúde.

Data e hora do nascimento e do óbito: completados com dados da DN e da DO.

Data de Encerramento da Análise: quando foi encerrada a análise do óbito.

Causa básica do óbito na DO: Esta parte da investigação é muito importante, pois um dos objetivos da investigação é definir corretamente a causa básica para ter-se conhecimento da real causa da morte. A causa básica tem sido definida como: “a doença ou lesão que iniciou a cadeia de acontecimentos patológicos que conduziram diretamente à morte, ou as circunstâncias do acidente ou violência que produziram a lesão fetal.” Prematuridade é uma causa que tem aparecido sempre no mapa de investigação de óbitos infantis. Esclarece-se que só deverá ser considerada causa básica se tiver somente ela nas informações disponíveis, mas não deverão constar como causa básica, nestes casos os comitês analisam a declaração de óbito, dados do prontuário de atendimento, do pré-natal,

entrevista com a família para chegar a uma causa básica melhor codificada que é colocada na DO: Segundo manual da Fundação Nacional de Saúde (2001).

Causa básica do óbito após a investigação: após a investigação é confirmado ou não a Classificação Internacional de Doenças (CID-10) contida na DO, sendo apresentada a real causa do óbito conforme codificação e orientações contidas no CID vol. 2 onde informa como preencher corretamente a DO e no Manual de Morte Evitável: este campo permite avaliar o número de óbitos que poderiam ser evitados. Onde, os óbitos evitáveis são aqueles que, levando em consideração a ciência e tecnologias existentes atualmente, poderiam ser evitadas ou não deveriam ter ocorrido. Este é um critério dinâmico, mutável no tempo. E os inevitáveis são aqueles que, considerando o estado atual das ciências e da tecnologia não puderam ser impedidos. Considerar todas as ciências (sociais, biológicas, etc.), e todas as técnicas políticas, médicas, administrativas, etc. Considerado também por (CALDEIRA et al, 2005). preenchimento da DO da Fundação Nacional de Saúde (FUNASA, 2003).

Critérios de evitabilidade: quando evitáveis, são relacionados a CID-10 após a investigação alguns critérios de evitabilidade, de acordo com os grupos descritos na Tabela 10 no Anexo 3.

Medidas de prevenção: foram elencadas as seguintes medidas de prevenção e intervenção da mortalidade infantil, conforme Tabela 11 no Anexo 4. As medidas de prevenção voltadas à assistência médica estão relacionadas às mortes evitáveis por maiores estudos, ponderação e cuidados prestados pelo médico, o que evitaria repetição de circunstâncias semelhantes. Estes erros deveriam merecer maior atenção dos diretores de hospitais e dos órgãos de classe (Conselhos Regionais e Sociedades Científicas). São recomendados também:

estímulos à atualização do conhecimento e troca de opiniões e experiências e organização de grupos de estudo da mortalidade nos serviços para discutir todos os casos de óbito. Quanto à assistência hospitalar são: em relação às condições do hospital, seus equipamentos e funcionamento, estão interligados com a equipe técnica de profissionais que administra, coordena, supervisiona, providencia recursos e pessoal, supre deficiências e aponta soluções. Os funcionários da maternidade ou hospital, médicos em particular, não podem fugir à responsabilidade do funcionamento da instituição. As medidas educativas e informativas referem-se à responsabilidade dos pais para evitar que os mesmos se recusem em participar dos tratamentos indicados. As medidas de educação em saúde são importantes para que os pais percebam a importância do acompanhamento dos profissionais de saúde durante o período de tratamento da criança. Observados também por (SILVA et al, 2003). Já as medidas sociais são: melhoria de acesso aos serviços de saúde, saneamento básico e melhores condições de vida; podem reduzir certo número de mortes infantis que praticamente ocorrem nas camadas mais empobrecidas da população.

Análise do Óbito: neste campo é feito um descritivo resumido da história do paciente.

Número da Declaração de Nascido Vivo (DN): número da DN, data e hora do nascimento: estes dados permitem avaliar se é um óbito em menor de um ano e também propicia o cruzamento com dados que estejam disponibilizados nas declarações de Nascidos Vivos, armazenados no SINASC. Observando Manual da Fundação Nacional de Saúde (2001).

Sexo: é armazenado o sexo da criança neste campo para avaliação sobre a relação de óbitos infantis e sexo.

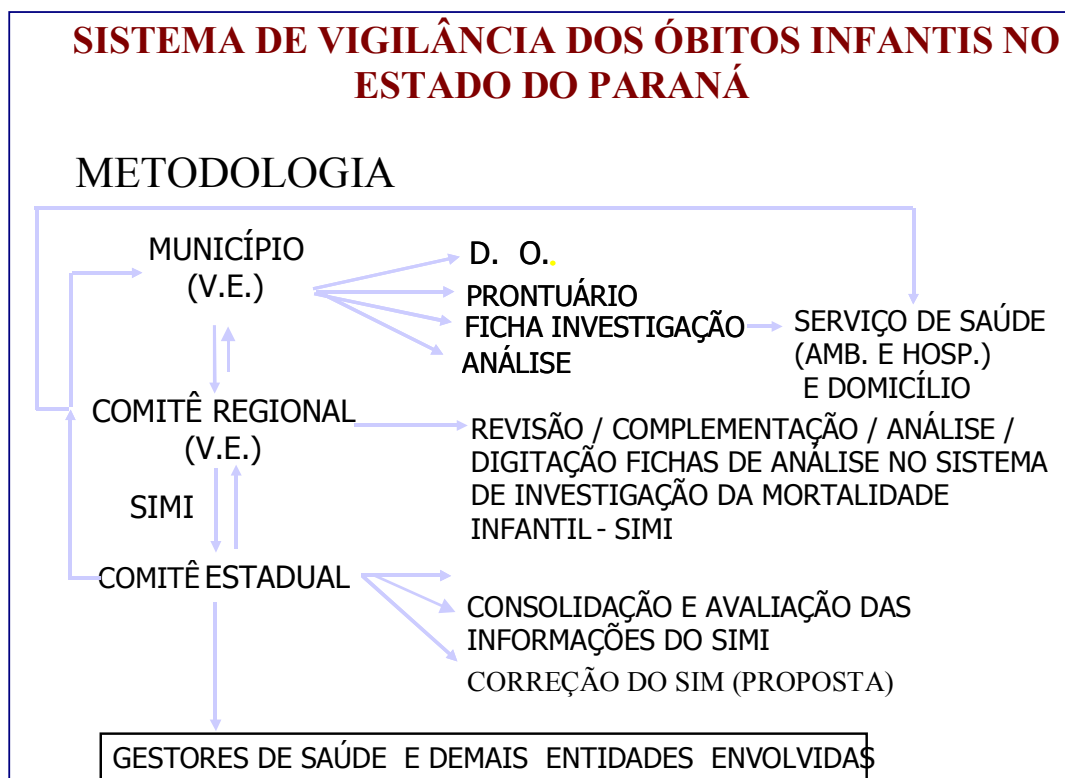
Responsabilidade pelo óbito: neste campo é armazenada a provável responsabilidade pelo óbito, isto é, se o óbito está vinculado à responsabilidade familiar, social, assistência médica, hospitalar, ambulatorial ou caso tenha sido inevitável quando não tem critério de responsabilidade. Sendo assim é de responsabilidade da Assistência Médica quando ocorre erro por não obediência às normas assistenciais, falta de conhecimento científico do médico que presta assistência, ou seja, imperícia ou imprudência do clínico, do obstetra ou do pediatra ou do profissional de qualquer natureza, aplicação incorreta de medicamentos ou tratamentos por pessoas não qualificadas. Estes são os chamados erros grosseiros. Como de responsabilidade da assistência hospitalar e ambulatorial sendo também chamados erros administrativos, quando as condições hospitalares e ambulatoriais são insuficientes ou inadequadas. São ditos de responsabilidade da família quando os mesmos recebem informações e não levam a criança à Unidade de Saúde ou não realizam os procedimentos necessários prescritos por negligência ou ignorância dos familiares. E são atribuídos como de responsabilidade social quando são de responsabilidade coletiva de toda a comunidade ou nação, envolvendo as condições sociais, econômicas, políticas. Incluem-se os casos decorrentes da miséria, do analfabetismo, da falta de informação, desnutrição, epidemias, etc. A classificação do óbito quanto a sua evitabilidade e responsabilidade, muitas vezes têm aspectos subjetivos. Para aumentar a sua validade os Comitês devem ser multiprofissionais. A multiplicidade de opiniões diminui a subjetividade e reforça as decisões.

Analísadores: Membros presentes e profissão: neste campo são armazenados os dados referentes aos membros presentes na investigação e a profissão destes.

Qualidade dos dados contidos na ficha de investigação: se foi: satisfatória, pouco satisfatória ou mesmo insatisfatória (CHESTER, 1990).

Última atualização: quando são necessárias atualizações conforme se adquire maior número de informações, melhorando desta forma a qualidade da investigação.

Anexo 7: Sistema de Vigilância dos Óbitos Infantis no Estado do Paraná



Fonte: SESA PR BR

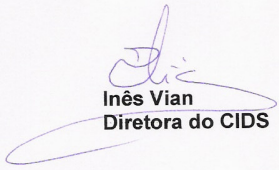
Figura 9 - Sistema de Vigilância dos Óbitos Infantis no Estado do Paraná-Brasil.

Anexo 8: Declaração de Consentimento Declarado para Liberação de Bancos de Dados



Declaração de Consentimento Declarado
Curitiba, 30 de novembro de 2005.

Eu, Inês Vian,
Diretora do CIDS (Centro de Informação e Diagnóstico em Saúde do Centro de Epidemiologia) da SESA-PR (Secretaria Estadual de Saúde do Estado do Paraná), declaro estar ciente e liberar a utilização dos bancos de dados do SIM (Sistema de Informação de Mortalidade), SINASC (Sistema de Informação de Nascidos Vivos) e do SIMI (Sistema de Investigação de Mortalidade Infantil) dos meses de janeiro de 2000 a dezembro de 2004 para a realização da pesquisa de dissertação de mestrado da funcionária Rossana Cristina Xavier Ferreira Vianna, desde que garantidos os princípios éticos de não identificação das pessoas inscritas nestas bases de dados.


Inês Vian
Diretora do CIDS

SECRETARIA DA SAÚDE/ISEP
Centro de Informações e Diagnóstico em Saúde / VEOMI
Rua Piquiri, 170 Curitiba Paraná CEP 80230 140
Fone (41) 330-4568 Fax (41) 330-4571 e-mail: leiko@pr.gov.br

Anexo 9: Liberação do CEP para Realização da Pesquisa



Pontifícia Universidade Católica do Paraná
Pró-Reitoria de Pesquisa e Pós-Graduação

Curitiba, 08 de dezembro de 2005.
2ª via - Of. 568/05/CEP-PUCPR

**Ref. "Aplicação de KDD relacionado a mortalidade infantil.
Pesquisa sem patrocinador"**

Prezado (a) Pesquisador

Venho por meio deste, informar a Vossa Senhoria que o Comitê de Ética em Pesquisa da PUCPR, no dia 07 de dezembro do corrente ano aprovou a continuidade da pesquisa intitulada **"Aplicação de KDD relacionado a mortalidade infantil. Pesquisa sem patrocinador"** pertencente ao Grupo III, sob o registro no CEP nº 957 e será encaminhado a CONEP para o devido cadastro. Lembro ao senhor (a) pesquisador (a) que é obrigatório encaminhar relatório anual parcial e relatório final a este CEP.

Atenciosamente,


Profª M. Sc Ana Cristina Miguez Ribeiro
Coordenadora do Comitê de Ética em Pesquisa - PUCPR

Ilma Sra
Rossana Cristina Xavier Ferreira Vianna

8. APÊNDICES

Apêndice 1: Tabela 12 - Seleção dos atributos das três bases de dados SINASC/SIM/SIMI

Tabela 12 - Seleção dos atributos das três bases de dados SINASC/ SIM/SIMI:

Atributos selecionados	DN	DO	SIMI
Número	s	s	s
Município residência	s	s	s
Município ocorrência	s	s	n
Nome mãe	s	s	n
Idade mãe	s	s	n
Estado civil mãe	s	s	n
Escolaridade mãe	s	s	n
Ocupação mãe	s	s	n
Data nascimento	s	s	s
Hora nascimento	s	n	s
Data óbito	n	s	s
Hora óbito	n	s	s
Sexo	s	s	s
Idade (meses, dias, horas, minutos)	n	s	s
raça/cor	s	s	n
número filhos gest. Ant.(vivos/mortos)	s	s	n
Dur. gest.(sem)	s	s	n
Tipo grav.(ún, dupla, tripla/+)	s	s	n
Tipo parto(vag., ces., ign.)	s	s	n
Número cons. Pré-natal	s	n	n
Apgar(1,5)	s	n	n
peso ao nascer	s	n	n
Malformações	s	n	n
CID DO	n	s	s
CID pós análise	n	n	s
Morte em relação ao parto	n	s	n
Evitável	n	n	s
Critérios evitabilidade	n	n	s
Responsabilidade	n	n	s
Prevenção	n	n	s

Apêndice 2: Categorizações normalizadas

Tabela 13 - Códigos das ocupações:

	CÓDIGO	OCUPAÇÃO
1	Código 00800 do IBGE (5227 casos).	Atividade dona-de-casa/doméstica (72% dos casos da base).
2	Outros códigos do IBGE (< ou = 198 casos para cada ocupação).	Outras profissões (28% dos casos da base).

Tabela 14 - Categoria para idade da criança:

	IDADE	CATEGORIA
1	menor de 1 dia	perinatal
2	1 a 6 dias	neonatal precoce
3	7 a 28 dias	neonatal tardio
4	29 a 360 dias	tardio

Tabela 15 - Distribuição da CID na DO:

	Número de DO no Grupo da CID	Grupo da CID
1	4599 no Grupo P	Afecções no Período Perinatal
2	1031 no Grupo Q	Malformações
3	541 no Grupo R	Afecções não classificadas em outra parte
4	497 no Grupo J	Doenças do aparelho respiratório
5	319 no Grupo A	Doenças infecciosas e parasitárias
6	222 no Grupo W	Trauma
7	47 ou menos casos	Demais casos

Tabela 16 - Distribuição da CID após a investigação:

	Número de DO no Grupo da CID	Grupo da CID pós investigação
1	4368 no Grupo P	Diminuindo portanto as causas perinatais em relação ao que foi relatado nas DO.
2	1031 no Grupo Q	Permanecendo igual os casos de malformações após as análises.
3	464 no Grupo J	Diminuindo ao que foi encontrado na DO.
4	378 no Grupo R	Diminuindo do que aparece na DO, confirmando que após análise consegue-se especificar melhor as causa de óbito.
5	319 no Grupo A	Igual ao que aparece na DO.
6	222 no Grupo W	Igual a DO.
7	474 em outros	Maior do que o da DO.

Tabela 17 - Categoria CID mais encontradas:

	CATEGORIAS	GRUPO DA CID
1	Três divisões para o CID mais encontrado do Grupo P	4368 casos (60% dos casos da base)
2	Uma categoria do Grupo Q	1031 casos (14% dos casos da base)
3	Uma categoria do Grupo J	463 casos (6,3% dos casos da base)
4	Uma categoria do Grupo R	377 casos (5% dos casos da base)
5	Uma categoria do Grupo A	319 casos (4,3% dos casos da base)
6	Uma categoria do Grupo W	224 casos (3% dos casos da base)
7	Uma categoria dos outros grupos que apareceram em menor número ou eram não especificados ou não classificados em outra parte	474 casos. (6,5% dos casos da base)

Apêndice 3 : Tabelas das análises da junção das bases

Tabela 18 - Categoria para idade da mãe:

IDADE DA MÃE	NÚMERO	PORCENTAGEM
< e = 16 anos	342	4%
16 a 35 anos	6285	87%
> = 35 anos	629	9%

Tabela 19 - Estado Civil da mãe:

ESTADO CIVIL MÃE	NÚMERO	PORCENTAGEM
Solteira	2317	32%
Casada	2799	39%
Viúva	30	0,4%
Separada	64	0,9%
União Consensual	2000	28%
Ignorado	37	0,5%
Sem preenchimento	9	0,1%

Tabela 20 - Escolaridade da mãe:

ESCOLARIDADE DA MÃE	NÚMERO	PORCENTAGEM
Sem estudo	235	3,2%
1-3 anos de estudo	1132	16%
4-7 anos de estudo	2901	40%
8-11 anos de estudo	2177	30%
12/+ anos de estudo	712	10%
Ignorado	76	1%
Sem preenchimento	23	0,3%

Tabela 21 - Quantidade de filhos vivos:

QUANTIDADE DE FILHOS VIVOS	NÚMERO	PORCENTAGEM
Sem	3011	41%
1	1861	26%
2	1095	15%
3	582	8%
4	282	4%
5	141	2%
6	92	1,2%
sem preenchimento	89	1,2%
7	44	0,6%
com mais de 7	59	0,8%

Tabela 22 - Quantidade de filhos mortos:

QUANTIDADE DE FILHOS MORTOS	NÚMERO	PORCENTAGEM
Sem	6098	84%
1	710	10%
sem preenchimento	186	2,5%
2	180	2,4%
3	48	0,7%
4	15	0,2%
Outros (> 4 e com preenchimento incorreto)	11	0,1%
Ignorado	8	0,1%

Tabela 23 - Número de semanas de gestação:

NÚMERO DE SEMANAS DE GESTAÇÃO	NÚMERO	PORCENTAGEM
< 22 semanas	106	1,5%
22 a 27 semanas	1311	18%
28 a 31 semanas	940	13%
32 a 36 semanas	1401	19%
37 a 41 semanas	3389	47%
42e +	66	0,9%
Ignorado	29	0,3%

Tabela 24 - Tipo de gravidez:

GRAVIDEZ	NÚMERO	PORCENTAGEM
única	6623	91%
dupla	580	8%
tripla e mais	52	0,7%
Ignorado	1	0,01%

Tabela 25 - Tipo de Parto:

TIPOS DE PARTO	NÚMERO	PORCENTAGEM
Vaginal	4567	63%
Cesariano	2686	37%
Ignorado	2	0,02%
Sem preenchimento	1	0,01%

Tabela 26 - Número de consultas de pré-natal:

CONSULTAS DE PRÉ-NATAL	NÚMERO	PORCENTAGEM
1	418	6%
2-4	1307	18%
5-6	2599	36%
+ 6	2851	39%
Ignorado	67	0,9%
Sem preenchimento	14	0,1%

Tabela 27 - Média da classe Apgar no primeiro minuto:

MÉDIA DA CLASSE APGAR NO PRIMEIRO MINUTO	NÚMERO	PORCENTAGEM
>5	3905	53%
< e = 5	3188	44%
Sem preenchimento	163	2%

Tabela 28 - Média da classe Apgar no quinto minuto:

MÉDIA DA CLASSE APGAR NO QUINTO MINUTO	NÚMERO	PORCENTAGEM
>5	5120	71%
< e = 5	1978	27%
Sem preenchimento	157	2%

Tabela 29 - Raça/Cor:

RAÇA/COR	NÚMERO	PORCENTAGEM
Branca	6534	91%
Preta	85	1%
Amarela	17	0,2%
Parda	520	7,2%
Indígena	23	0,3%
Sem preenchimento	77	0,9%

Tabela 30 - Categoria do peso ao nascer:

PESO	NÚMERO	PORCENTAGEM
= < 2.500 g	4084	56%
Entre 2.500 e 3.500g	2573	35%
> = 3.500 g	570	8%
Sem preenchimento	29	0,4%

Tabela 31 - Categoria da idade no óbito:

IDADE	NÚMERO	PORCENTAGEM
Entre 1 e 6 dias	2272	31%
Entre 29 e 365 dias	2245	31%
< e = 1 dia	1704	23%
Entre 7 e 28 dias	1035	14%

Tabela 32 - Alteração entre as CID10 antes e após as análises dos Comitês:

DIFERENÇA	NÚMERO	PORCENTAGEM
CID igual	3026	42%
Muda Categoria	1538	21%
Muda Grupo	1494	20%
Muda Capítulo	1174	16%

Tabela 33 - Critérios de evitabilidade:

GRUPO	NÚMERO	PORCENTAGEM
D (óbitos evitáveis por adequada atenção na gestação, no parto e ao recém-nascido e na educação para a saúde).	4033	56%
G (óbitos dificilmente reduzíveis)	1053	14%
E (óbitos reduzíveis por diagnóstico e tratamento precoces e pela educação para a saúde)	900	12%
F (óbitos reduzíveis por atenção ao trauma e urgências e pela educação)	473	6,5%
I (mal definidas, do grupo R da CID).	372	5%
B (óbitos reduzíveis por saneamento básico, terapia de reidratação oral, incentivo ao aleitamento materno e educação para a saúde).	171	2%
H (óbitos devidos a outras causas de mortes, do grupo A, D, E, G da CID).	166	2%
C (óbitos reduzíveis por medidas de combate à desnutrição, incentivo ao aleitamento materno e pela educação para a saúde,	76	1%

grupo E da CID).		
A (óbitos reduzíveis por imunoprevenção)	11	0,1%
Sem critérios	1	0,01%