

Eliane de Freitas Drumond

SISTEMAS DE INFORMAÇÃO EM SAÚDE: POTENCIALIDADES  
METODOLÓGICAS NO MONITORAMENTO DA MORTALIDADE  
NEONATAL PRECOCE

Universidade Federal de Minas Gerais  
Programa de Pós-Graduação em Saúde Pública  
Belo Horizonte – MG  
2007

Eliane de Freitas Drumond

SISTEMAS DE INFORMAÇÃO EM SAÚDE: POTENCIALIDADES METODOLÓGICAS  
NO MONITORAMENTO DA MORTALIDADE NEONATAL PRECOCE

Tese apresentada ao Programa de Pós-Graduação em  
Saúde Pública da Universidade Federal de Minas Gerais,  
como requisito parcial para obtenção do título de Doutor  
em Saúde Pública (área de concentração em  
Epidemiologia).

Orientadora: Prof<sup>a</sup> Elisabeth Barboza França  
Co-orientadora: Prof<sup>a</sup> Carla Jorge Machado

Belo Horizonte  
2007

Drumond, Eliane de Freitas

D 795s    Sistemas de informação em saúde: potencialidades metodológicas no monitoramento da mortalidade neonatal precoce /  
Eliane de Freitas Drumond. Belo Horizonte, 2007.

xiii, 118f., il.

Tese.(doutorado) – Universidade Federal de Minas Gerais. Faculdade de Medicina,

Área de concentração: Epidemiologia

Orientadora: Elisabeth Barboza França

Co-orientadora: Carla Jorge Machado

1.Avaliação da pesquisa em saúde 2.Sistemas de informação/utilização  
3.Bases de dados/utilização 4.Indicadores de produção científica 5.Registro médico coordenado 6.Lógica Fuzzy 7.Aplicações da epidemiologia I.Título

NLM: WA 20.5

CDU: 614.2

## **UNIVERSIDADE FEDERAL DE MINAS GERAIS**

### **Reitor**

Prof. Ronaldo Tadêu Pena

### **Vice-Reitor**

Prof<sup>a</sup>. Heloisa Maria Murgel Starling

### **Pró-Reitor de Pós-Graduação**

Prof. Jaime Arturo Ramirez

### **Pró-Reitor de Pesquisa**

Prof. Carlos Alberto Pereira Tavares

## **FACULDADE DE MEDICINA**

### **Diretor**

Prof. Francisco José Penna

### **Chefe do Departamento de Medicina Preventiva e Social**

Prof<sup>a</sup>. Maria da Conceição Juste Werneck Côrtes

## **PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE PÚBLICA**

### **Coordenador**

Prof. Mark Drew Crosland Guimarães

### **Sub-Coordenador**

Prof<sup>a</sup>. Sandhi Maria Barreto

### **Colegiado**

Prof<sup>a</sup>. Ada Ávila Assunção

Prof<sup>a</sup>. Eli Iola Gurgel Andrade

Prof<sup>a</sup>. Elisabeth Barboza França

Prof. Fernando Augusto Proietti

Prof<sup>a</sup>. Maria Fernanda Furtado de Lima-Costa

Prof<sup>a</sup>. Mariângela Leal Cherchiglia

Prof<sup>a</sup>. Waleska Teixeira Caiaffa

Lorenza Nogueira Campos

Raquel Regina de Freitas Magalhães Gomes

Sarah de Araújo Carvalho

Roberto Marini Ladeira



UFMG

FACULDADE DE MEDICINA  
CENTRO DE PÓS-GRADUAÇÃO

Av. Prof. Alfredo Balena 190 / sala 7009  
Belo Horizonte - MG - CEP 30.130-100  
Fone: (031) 3248.9641 FAX: (31) 3248.9640



ATA DA VIGÉSIMA SEGUNDA DEFESA DE TESE DE DOUTORADO de **ELIANE DE FREITAS DRUMOND**. Às quatorze horas do dia quatro do mês de maio de dois mil e sete, reuniu-se na Faculdade de Medicina da UFMG a Comissão Examinadora de tese indicada pelo Colegiado do Programa para julgar, em exame final, o trabalho intitulado: **“SISTEMAS DE INFORMAÇÃO EM SAÚDE: POTENCIALIDADES E UTILIZAÇÃO NO MONITORAMENTO DA MORTALIDADE NEONATAL PRECOCE”** requisito final para a obtenção do Grau de Doutor em Saúde Pública, pelo Programa de Pós-Graduação em Saúde Pública – Área de Concentração em Epidemiologia. Abrindo a sessão, a Presidente da Comissão, Profa. Elisabeth Barboza França, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho final passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público para julgamento e expedição do resultado definitivo. Foram atribuídas as seguintes indicações:

Profa. Elisabeth Barboza França/orientadora  
Profa. Carla Jorge Machado/co-orientadora  
Profa. Sílvia Wanick Sarinho  
Profa. Ana Maria Pereira Cardoso  
Profa. Maria da Conceição Juste Werneck Côrtes  
Profa. Mariângela Leal Cherchiglia

Instituição: UFMG  
Instituição: UFMG  
Instituição: UFPE  
Instituição: PUC/MG  
Instituição: UFMG  
Instituição: UFMG

Indicação: Aprovada  
Indicação: Aprovada  
Indicação: Aprovada  
Indicação: Aprovada  
Indicação: Aprovada

Pelas indicações, a candidata foi considerada APROVADA.  
O resultado final foi comunicado publicamente à candidata pela presidente da comissão. Nada mais havendo a tratar a presidente encerrou a reunião e lavrou a presente ATA que será assinada por todos os membros participantes da comissão examinadora. Belo Horizonte, 04 de maio de 2007.

Profa. Elisabeth Barboza França/orientadora Elisabeth Barboza França

Profa. Carla Jorge Machado/co-orientadora Carla Jorge Machado

Profa. Sílvia Wanick Sarinho Silvia Wanick Sarinho

Profa. Ana Maria Pereira Cardoso Ana Maria Pereira Cardoso

Profa. Maria da Conceição Juste Werneck Côrtes Maria da Conceição Juste Werneck Côrtes

Profa. Mariângela Leal Cherchiglia Mariângela Leal Cherchiglia

Prof. Mark Drew Crosland Guimarães / Coordenador Mark Drew Crosland Guimarães

Obs.: Este documento não terá validade sem a assinatura e o carimbo do Coordenador  
PROF. MARK DREW CROSLAND GUIMARÃES  
Coordenador do Programa de Pós-Graduação em Saúde Pública  
Faculdade de Medicina/UFMG

## AGRADECIMENTOS

Gostaria de expressar meus agradecimentos e meu desagrado às pessoas que tornaram minha vida muito melhor.

Minha orientadora, professora e amiga Beth: agradeço a você pela paciência (em doses generosas!), atenção, carinho e dedicação.

Minha co-orientadora, professora amiga e “super ninja” Carla: agradeço a você por todo o carinho e apoio que você sempre me deu naquele monte de coisas relativas ao trabalho, mas acima de tudo, agradeço termos compartilhado muito mais do que o peso do trabalho.

Nosso Coordenador professor Mark: agradeço a você por ter nos estimulado a refletir sobre e a desejar construir uma PGSP ainda melhor.

Nosso Colegiado da PGSP: agradeço a vocês pela dedicação aos problemas comuns, quando já temos tão pouco tempo e tantos outros problemas a resolver.

Meus Professores e meus colegas da PGSP: agradeço por terem vivido comigo a alegria e os apertos (e põe aperto nisto!) de voltar a ser estudante. Foi muito bom!

Nossa Gloriosa Turma de 81: onde “muita gente boa pôs o pé na profissão”, no nosso Jubileu de Prata, agradeço a vocês pela emoção do antes e do agora. Nosso reencontro foi indescritível!

Minha queridíssima turma do SIM: agradeço pela solidariedade, apoio, carinho e por compartilharmos o peso e a alegria que é a construção diária do Sistema de Informação sobre Mortalidade nesta cidade. Para Lenice, amiga e exemplo de profissional, quero dizer meu especial muito obrigada.

Meus queridos e de longa data amigos da MOV: agradeço a vocês por manterem em mim a alegria e o desejo de continuar lá no nosso *front*. Ari, amigo, compadre e irmão querido; Mariano, parceirão de “lutar, lutar, lutar, com muita raça e orgulho prá vencer” e com quem a bola rola redonda desde sempre e Ruth Lira amiga fraterna e musa de todos nós.

Meus companheiros de Confraria: quero manifestar aqui o meu mais profundo e sincero desagrado a vocês: Peninha, Julinho, Regina, Ramon e Ana. Desagrado todas as vezes em que fui obrigada, sim obrigada, a não estudar para juntos tomarmos aqueles vinhos maravilhosos e compartilharmos de excelente mesa e melhor companhia. Encontros inesquecíveis!

Meu querido companheiro Paulo: agradeço a você por todo carinho, solidariedade e amor que temos compartilhado ao longo destes muitos anos de vida.

Meus queridos irmãos e, em especial, minha irmã Vânia: muito obrigada pelos muitos e memoráveis “momentos família” e pelos nossos sobrinhos-quase-filhos.

Minhas duas queridíssimas Marias: minha mãe e minha filha queridas, eu adoro estar no sanduíche de vocês duas, tão parecidas e tão geniais. Eu amo vocês.

Meu querido pai: obrigada.

À gloriosa Faculdade de Medicina da  
Universidade Federal de Minas Gerais,  
pública e de qualidade, que me deu  
régua e compasso.

“...Falo assim por saber  
Se muito vale o já feito  
Mais vale o que será  
E o que foi feito é preciso  
Conhecer para melhor prosseguir...”

(“O que foi feito deverá”  
Milton Nascimento)

## RESUMO

O objetivo desta tese de doutorado, apresentada no formato de quatro artigos, foi avaliar alternativas metodológicas para a utilização dos sistemas de informação em saúde (SIS) no monitoramento da mortalidade neonatal precoce e as potencialidades dos SIS como fontes de pesquisa científica no Brasil. Este objetivo se justifica face à importância atual do componente neonatal precoce na mortalidade infantil e ao grande volume e à relevância das informações disponíveis nos diferentes SIS. Optou-se pela abordagem quantitativa por meio de dois eixos: utilização concreta das bases de dados pelo autor e avaliação dos artigos científicos publicados em periódicos indexados no MEDLINE/PubMed. Foram eleitos dois SIS de base populacional (Sistemas de Informação sobre Mortalidade - SIM - e sobre Nascidos Vivos - SINASC) e um sistema de registro de atendimentos hospitalares da rede do SUS (Sistema de Informações Hospitalares - SIH).

No primeiro artigo foi descrita a metodologia de relacionamento probabilístico dos dados do SINASC e do SIH utilizada para avaliação de subenumeração de nascidos vivos no SINASC. No segundo artigo foram comparados os resultados obtidos por meio do linkage SINASC - SIH com os dados disponibilizados na Internet pelo DATASUS. O terceiro artigo referiu-se à análise de causas múltiplas de morte neonatal precoce por meio do método Grade of Membership (GoM). No quarto artigo, realizou-se estudo bibliométrico no MEDLINE/PubMed para avaliação do crescimento e de algumas características da publicação científica baseada em dados do SIM, SINASC e SIH no período 1990-2006.

Os principais resultados obtidos foram: 1) encontraram-se expressivas dificuldades para compatibilização e relacionamento das bases de dados; 2) a subenumeração de nascidos vivos ocorreu mesmo em municípios de médio e grande porte de Minas Gerais; 3) a análise de causas múltiplas de morte pelo método de GoM evidenciou diferenças no perfil de óbitos neonatais precoces segundo a natureza jurídica do hospital; 4) houve acelerado crescimento da produção científica brasileira que utilizou estes sistemas de informação, especialmente a partir do ano 2000. Nos comentários finais, foram discutidos os principais resultados e apontadas perspectivas para potencializar a utilização dos sistemas de informação nacionais como fontes de informação em pesquisa científica.

## ABSTRACT

This dissertation consists of four articles and aims at evaluating the use of healthcare information systems as sources for the monitoring of early neonatal death and scientific publication in Brazil. This aim is justified by the importance of the early neonatal component in infant mortality and by the large amount and relevance of information available in the systems. The quantitative analysis is divided in two axes: the authors own use of the databases and the assessment of scientific articles from MEDLINE/PubMed. Two population based systems were chosen: the Mortality Information System (SIM) and the Live Birth Information System (SINASC); the third system is the record hospital assistance form the SUS network, the Hospital Information System (SIH).

The first article describes how the probabilistic record linkage can be used to relate data from SINASC and SIH in order to asses underreporting of live births on SINASC. The second article compares the results of the linkage between SINASC and SIH with the data made available on the Internet by DATASUS. The third article analyses the multiple causes of early neonatal death by means of the Grade of Membership (GoM) method. The fourth article is a bibliometric study articles from MEDLINE/PubMed, and assesses the growth and some characteristics of the scientific publication based on data from SIM, SINASC and SIH in the years of 1990-2006.

The results indicate that: 1) the process of making compatible and linkage of database is still considerably difficult; 2) underreporting of live births occurs even in medium and large size cities in Minas Gerais; 3) the analyses of multiple causes of death by GoM can reveal differences in the profiles of early neonatal deaths according to the hospital's legal statuses; 4) there has been a considerable growth in Brazilian scientific production based on this information systems from the year 2000 onwards. The final comments discuss the main results and highlights possibilities to potentialize the use of the national information systems as sources for scientific research.

## APRESENTAÇÃO

Este trabalho é requisito parcial para a obtenção de título de doutor no Programa de Pós-Graduação em Saúde Pública, área de concentração em Epidemiologia do Departamento de Medicina Preventiva e Social da Faculdade de Medicina da Universidade Federal de Minas Gerais.

Segundo o regulamento do Programa, no seu Título IV – Do regime Didático, Capítulo V, Art. 55, “a tese será apresentada sob a forma de três artigos científicos, podendo ser um deles sob a forma de revisão da literatura ou metodologia de trabalho, e os outros referentes à apresentação dos resultados obtidos no desenvolvimento da pesquisa..., escritos de acordo com as normas de uma determinada revista científica de reconhecida qualidade na área, a partir de listagem definida pelo Colegiado do Programa, e estarem em condições de ser submetidos à mesma para publicação. Pelo menos um artigo já deverá estar aceito para publicação”. Optou-se pela apresentação de dois artigos de resultados, um artigo metodológico e um artigo de revisão da literatura.

Para os quatro artigos apresentados foram adotados dois eixos de abordagem e foram eleitos três sistemas de informação em saúde (SIS): o Sistema de Informação sobre Mortalidade (SIM), o Sistema de Informação sobre Nascidos Vivos (SINASC) e o Sistema de Informações Hospitalares (SIH). O primeiro eixo visou avaliar as dificuldades e facilidades observadas pelo pesquisador na utilização de metodologias ainda pouco adotadas com os sistemas de informação para o monitoramento da mortalidade neonatal precoce. Para atender a este objetivo, optou-se por duas abordagens distintas. Na primeira abordagem, realizou-se relacionamento probabilístico de dados (linkage) dos dados do SINASC, obtidos na Secretaria de Estado da Saúde de Minas Gerais e os do SIH, fornecidos pelo Departamento de Informática do SUS/DATASUS, ambos referentes ao ano de 2001. Na segunda abordagem, utilizou-se o SIM de 2000 a 2003 da Secretaria Municipal de Saúde de Belo Horizonte para o estudo de causas múltiplas de morte neonatal precoce. No segundo eixo, que visou avaliar a utilização dos sistemas de informação em pesquisas científicas, foi realizado estudo bibliométrico da produção científica brasileira publicada em periódicos indexados na base bibliográfica MEDLINE/PubMed a partir de dados secundários do Sistema de Informação sobre Mortalidade (SIM), do Sistema de Informação sobre Nascidos Vivos (SINASC) e do Sistema de Informações Hospitalares (SIH).

A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da UFMG (Nº 095/04) e contou com o apoio financeiro da Fundação de Apoio à Pesquisa Estado de Minas Gerais/FAPEMIG (EDT-1770/03) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico/ CNPq (403707/04-8).

Nas considerações iniciais são brevemente apresentadas algumas características da mortalidade neonatal precoce e da utilização dos sistemas de informação no monitoramento destas mortes, suas inter-relações e perspectivas.

O primeiro artigo descreve a metodologia de relacionamento probabilístico de dados do SINASC e do SIHSUS desenvolvida para avaliar a cobertura para nascimentos ocorridos na rede hospitalar do SUS em Minas Gerais (SUS/MG) - (publicado nos Cad Saúde Colet 2006; 14(2): 251-264).

No segundo artigo são comparados os resultados obtidos por duas técnicas de relacionamento de dados do SINASC e do SIH: a técnica ecológica e a probabilística (submetido à Revista de Saúde Pública).

No terceiro artigo são feitas descrição e análise - a partir do SIM - dos perfis dos óbitos neonatais precoces ocorridos em Belo Horizonte, no período de 2000 a 2003, utilizando o enfoque de causas múltiplas de morte a partir da teoria dos conjuntos nebulosos, por faixa de peso de nascimento dos recém-nascidos e pela natureza jurídica dos hospitais de ocorrência do óbito (publicado nos Cad Saúde Pública 2007; 23(1): 157-66).

E, finalmente, no quarto artigo se avalia o crescimento e algumas características da produção científica nacional com utilização de dados secundários provenientes do SIM, SINASC e/ou SIH no período de 1990-2006 (a ser submetido à Revista Brasileira de Epidemiologia).

Nas considerações finais e recomendações encontram-se reflexões sobre os resultados encontrados nos quatro artigos, as dificuldades e os problemas observados e são apontadas algumas contribuições para pesquisas epidemiológicas a partir de dados dos sistemas de informação em saúde.

Constam também deste volume o projeto de pesquisa apresentado e aprovado por banca examinadora como parte do processo de seleção para o doutorado (Anexo A); as aprovações dos projetos pelos Comitês de Ética em Pesquisa da UFMG e da Secretaria Municipal de Saúde de Belo Horizonte (Anexo B); o certificado de qualificação no doutorado (Anexo C) e as páginas iniciais dos dois artigos no formato original com o qual foram publicados (Anexo D).

## SUMÁRIO

1	CONSIDERAÇÕES INICIAIS.....	14
2	OBJETIVOS.....	18
3	ARTIGO 1: SIH-SUS E SINASC: UTILIZAÇÃO DO MÉTODO PROBABILÍSTICO PARA RELACIONAMENTO DE DADOS .....	19
3.1	INTRODUÇÃO.....	22
3.2	METODOLOGIA.....	24
3.3	RESULTADOS.....	27
3.4	DISCUSSÃO .....	29
	REFERÊNCIAS BIBLIOGRÁFICAS.....	32
4	ARTIGO 2: SUBNOTIFICAÇÃO DE NASCIDOS VIVOS NO SINASC: DOIS PROCEDIMENTOS DE MENSURAÇÃO A PARTIR DO SIH ....	35
4.1	INTRODUÇÃO.....	38
4.2	METODOLOGIA.....	39
4.3	RESULTADOS.....	42
4.4	DISCUSSÃO .....	45
	REFERÊNCIAS BIBLIOGRÁFICAS.....	49
5	ARTIGO 3: ÓBITOS NEONATAIS PRECOCES: ANÁLISE DE CAUSAS MÚLTIPLAS PELO MÉTODO DE <i>GRADE OF MEMBERSHIP</i> (GoM)..	55
5.1	INTRODUÇÃO.....	58
5.2	METODOLOGIA.....	59
5.3	RESULTADOS.....	63
5.4	DISCUSSÃO .....	66
	REFERÊNCIAS BIBLIOGRÁFICAS.....	70
6	ARTIGO 4: UTILIZAÇÃO DE DADOS SECUNDÁRIOS DO SIM, SINASC E SIH NA PRODUÇÃO CIENTÍFICA BRASILEIRA DE 1990 A 2006 .....	75
6.1	INTRODUÇÃO.....	78
6.2	METODOLOGIA.....	80
6.3	RESULTADOS.....	83
6.4	DISCUSSÃO .....	86
	REFERÊNCIAS BIBLIOGRÁFICAS.....	91
7	CONSIDERAÇÕES FINAIS .....	96
	ANEXOS.....	102
	ANEXO A: PROJETO DE PESQUISA.....	102
	ANEXO B: CERTIFICADO DE QUALIFICAÇÃO.....	113
	ANEXO C: APROVAÇÃO PELO CEP E COEP.....	114
	ANEXO D: ARTIGOS PUBLICADOS.....	116

## 1 CONSIDERAÇÕES INICIAIS

A Epidemiologia é uma ciência que aborda objetos complexos e que pode ser definida, segundo Last <sup>1</sup>, como o estudo da distribuição e dos determinantes dos estados e eventos relacionados à saúde em populações definidas e a aplicação deste conhecimento para a resolução dos problemas de saúde. Para Gordis <sup>2</sup> o principal objetivo da epidemiologia é ser a base para o desenvolvimento de políticas que afetam a saúde das pessoas, incluindo a prevenção e o controle das doenças. No Brasil, a Epidemiologia avançou com intenso vigor, que pode ser avaliado, especialmente nas duas últimas décadas, pelo crescimento da produção acadêmica epidemiológica <sup>3,4</sup> e também pelos avanços observados na utilização da epidemiologia nos serviços de saúde <sup>5</sup>.

Por sua relevância e complexidade, uma parcela significativa dos esforços e dos investimentos dos serviços de saúde e da pesquisa epidemiológica tem se dirigido à redução das mortes infantis, um dos Objetivos de Desenvolvimento do Milênio <sup>6</sup> para 2015. Essas mortes - em grande parte inaceitáveis - são conseqüências de uma combinação de fatores biológicos, sociais, culturais e de falhas do sistema de saúde.

Dentre as mortes infantis ocorridas no Brasil atualmente, o componente neonatal é responsável por cerca de dois terços do total. As mortes neonatais precoces concentram-se no momento do parto e nas primeiras horas de vida. Apresentam estreita relação com a assistência dispensada à gestante e ao recém-nascido no pré-parto e no parto e ligam-se, também, ao atendimento dado à criança na sala de parto e na unidade neonatal. Como a diminuição dessas mortes precoces depende em parte do reconhecimento dos seus riscos, os aspectos ligados à assistência hospitalar à gestante, ao parto e ao recém-nascido devem merecer atenção de gestores e pesquisadores. Entretanto, estudos publicados em revistas indexadas objetivando a avaliação da qualidade da assistência obstétrica e neonatal no âmbito

---

<sup>1</sup> Last JM. A Dictionary of Epidemiology. 4th ed. New York, NY: Oxford University Press; 2000.

<sup>2</sup> Gordis L. Epidemiology. 3rd ed. Philadelphia: Elsevier Inc; 2004.

<sup>3</sup> Barreto ML. Papel da epidemiologia no desenvolvimento do Sistema Único de Saúde no Brasil: histórico, fundamentos e perspectivas. Rev Bras Epidemiol 2002; 5 (sup 1): 4-17.

<sup>4</sup> Guimarães R, Lourenço de Oliveira R, Cosac S. Epidemiological research in Brazil. Rev Saúde Pública 2001; 35(4): 321-40.

<sup>5</sup> Drumond Jr M. Epidemiologia nos municípios muito além das normas. São Paulo. Editora Hucitec; 2003.

<sup>6</sup> United Nations. The Millenium Project.[acesso 2007 Jan 12]. Disponível em: <http://www.unmillenniumproject.org>.

hospitalar ainda são escassos <sup>7</sup> e seus resultados apontam para a existência de desigualdades na assistência hospitalar no país. Observou-se grande diferencial na mortalidade neonatal entre estabelecimentos das redes pública <sup>7,8</sup> e destes com a rede privada <sup>9</sup>. Portanto, o aprofundamento das análises sobre assistência hospitalar ao parto e ao recém-nascido pode contribuir para a melhor compreensão deste aspecto assistencial e para a mudança do quadro atual de índices persistentemente elevados de mortalidade neonatal.

Apesar de ainda se manter em níveis que podem ser considerados como inaceitáveis devido principalmente à ocorrência freqüente de morte em crianças maiores de 2.500 gramas e sem malformações congênitas, a taxa de mortalidade neonatal precoce no Brasil apresentou mudanças em pouco tempo e atualmente observam-se importantes diferenças nos níveis nacional e regional <sup>10</sup>. O uso de estimativas indiretas para o monitoramento da mortalidade não permitiria a detecção de mudanças ocorridas em curtos espaços de tempo e do reflexo das desigualdades - como as raciais e étnicas - nestas taxas. Dessa forma, as fontes de dados secundários sobre nascimentos, mortes e registros hospitalares, embora ofereçam limitações para a investigação de fatores de risco neonatais, são ferramentas fundamentais para as ações de vigilância e de monitoramento destes óbitos. Especialmente para períodos de tempo menores e em populações ou regiões específicas, as estatísticas vitais de boa qualidade sobre a ocorrência dos eventos e sobre a assistência hospitalar ao parto têm papel de destaque na análise de situação e na avaliação dos resultados de medidas de intervenção para a redução da mortalidade neonatal precoce.

Vários sistemas de informação são produzidos na rotina dos serviços de saúde. Entre eles estão os Sistemas de Informação sobre Nascidos Vivos (SINASC) e sobre Mortalidade (SIM), os quais contêm informações epidemiológicas sobre os eventos vitais (nascimentos e óbitos), estão implantados em todo o território nacional e são de cobertura universal. O Sistema de Internações Hospitalares (SIH), por outro lado, contém registros referentes às internações

---

<sup>7</sup> Lansky S, França E, César CC, Monteiro Neto LC, Leal MC. Mortes perinatais e avaliação da assistência ao parto em maternidades do Sistema Único de Saúde em Belo Horizonte, Minas Gerais, Brasil, 1999. *Cad. Saúde Pública* 2006; 22(1): 117-30.

<sup>8</sup> Gomes MASM, Lopes JMA, Moreira MEL, Gianini NOM. Assistência e mortalidade neonatal no setor público do Município do Rio de Janeiro, Brasil: uma análise do período 1994/2000. *Cad Saúde Pública*; 2005; 21(4): 269-77.

<sup>9</sup> Leal MC, Gama SGN, Campos MR, Cavalini LT, Garbayo LS, Brasil CLP, Szwarcwald CL. Fatores associados à morbi-mortalidade perinatal em uma amostra de maternidades públicas e privadas do Município do Rio de Janeiro, 1999-2001. *Cad. Saúde Pública* 2004; 20 (Suppl 1): S20-S33.

<sup>10</sup> Fonseca SC, Coutinho ESF. Pesquisa sobre mortalidade perinatal no Brasil: revisão da metodologia e resultados. *Cad Saúde Pública* 2004; (Suppl 1): S7-S19.

hospitalares ocorridas na rede própria e na rede contratada/conveniada do Sistema Único de Saúde (SUS) e cobre aproximadamente 70% das internações do país <sup>11</sup>. O uso das informações desses sistemas em pesquisas científicas sobre mortalidade neonatal precoce possibilita, ao mesmo tempo, a realização de estudos epidemiológicos e a avaliação da qualidade das informações dessas bases de dados.

Existem muitas alternativas para se avaliar a qualidade dos sistemas de informação e sua utilização em pesquisas científicas, dependendo do ponto de partida do pesquisador. Avaliar (do latim *valere* - digno ou merecedor de confiança) permite dignificar o saber, quando métodos confiáveis e sistemáticos são utilizados para se demonstrar como este saber vem se desenvolvendo e de que forma ele tem contribuído para resolver os problemas que se apresentam à sua área. As avaliações podem ser feitas por meio de metodologias qualitativa e/ou quantitativa. O pesquisador pode partir, por exemplo, de categorias relativas aos sistemas de informação (finalidade, cobertura, qualidade) e mesmo de doenças ou agravos específicos. Estudos recentes sobre uso dos sistemas de informação nas pesquisas epidemiológicas sobre mortalidade perinatal <sup>8,12</sup> e mortalidade neonatal precoce <sup>13</sup> enfocaram a frequência da utilização, as diferentes metodologias empregadas e as limitações apresentadas por estas bases de dados. Os autores são unânimes ao observar que a irregular qualidade dos dados dificulta o monitoramento dos óbitos. Por outro lado, Almeida e colaboradores <sup>12</sup> ressaltam que os dados poderiam ser facilmente obtidos, tendo em vista que 95% dos eventos ocorreram em hospitais.

Diante do exposto e levando-se em consideração a necessidade de incorporar a avaliação da assistência hospitalar nas atividades de monitoramento dos óbitos neonatais precoces desenvolveu-se a proposta central deste trabalho. Assim, este estudo visou utilizar dados dos sistemas de informação nacionais no monitoramento da mortalidade neonatal precoce e avaliar as potencialidades destas bases de dados secundários como fontes de pesquisas epidemiológicas. Partiu-se da premissa de que as pesquisas científicas sobre mortalidade neonatal precoce, se realizadas a partir dos dados dos sistemas de informação produzidos na

---

<sup>11</sup> Schramm JMA, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e a natimortalidade. Rev. Saúde Pública 2000; 34 (3): 272-79.

<sup>12</sup> Almeida MF, Alencar GP, Novaes HMD, Ortiz LP. Sistemas de informação e mortalidade perinatal: conceitos e condições de uso em estudos epidemiológicos. Rev Bras Epidemiol 2006; 9(1): 56-68.

<sup>13</sup> Pedrosa LDCO, Sarinho SW, Ordonha MAR. Óbitos neonatais: por que e como informar? Rev Bras Saúde Materno Infant 2005; 5(4): 411-18.

rotina dos serviços de saúde, podem contribuir tanto para a potencialização do uso das informações destas bases de dados quanto para qualificação destes sistemas de informação.

Para atingir estes objetivos foram eleitos dois sistemas de base populacional, os sistemas de estatísticas vitais - SIM e SINASC - e um de registro de atendimentos hospitalares da rede SUS (SIH). Optou-se pela adoção de metodologia quantitativa por meio de dois diferentes eixos. O primeiro eixo deste estudo visou a utilização de bases de dados nacionais para produção de pesquisa epidemiológica, como uma forma de explorar as possibilidades de utilização dos sistemas de informação no monitoramento ao óbito neonatal precoce. Os objetivos que guiaram esta opção foram analisar as facilidades e as dificuldades encontradas na pesquisa realizada com as bases de dados e explorar as potencialidades dos SIS por meio de metodologias ainda pouco utilizadas em nosso meio. Neste eixo - utilização das bases de dados – situam-se os três primeiros artigos aqui apresentados. Realizou-se relacionamento probabilístico dos dados do SINASC e do SIH para avaliação da subenumeração de nascidos vivos nos municípios de Minas Gerais. A metodologia do relacionamento probabilístico e a comparação de seus resultados são apresentadas nos dois primeiros artigos. Por sua vez, os resultados de análise de causas múltiplas dos óbitos neonatais precoces de residentes em Belo Horizonte pelo método de *Grade of Membership* (GoM) são objeto do terceiro artigo. Para atender ao segundo eixo – avaliação das potencialidades dos sistemas de informação como fontes de pesquisas epidemiológicas - optou-se pela realização de estudo bibliométrico da produção científica nacional a partir das três bases de dados utilizadas de forma isolada ou conjunta. Os resultados do estudo bibliométrico são apresentados no quarto artigo.

## 2 OBJETIVOS

### OBJETIVO GERAL

Avaliar as potencialidades da utilização dos sistemas de informação em saúde em pesquisas científicas e no monitoramento da mortalidade neonatal precoce.

### OBJETIVOS ESPECÍFICOS

**Artigo 1:** Descrever metodologia de relacionamento probabilístico de dados do SINASC e do SIHSUS utilizada para avaliar a cobertura para nascimentos ocorridos na rede hospitalar do SUS em Minas Gerais (SUS/MG).

**Artigo 2:** Comparar resultados obtidos por dois procedimentos de mensuração de subnotificação de nascidos vivos no SINASC a partir do SIH: o procedimento ecológico e o probabilístico.

**Artigo 3:** Descrever e analisar os perfis dos óbitos neonatais precoces ocorridos em Belo Horizonte, no período de 2000 a 2003, utilizando o enfoque de causas múltiplas de morte a partir da teoria dos conjuntos nebulosos, por faixa de peso e natureza jurídica do hospital.

**Artigo 4:** Avaliar o crescimento e algumas características da produção científica nacional com utilização de dados secundários provenientes do SIM, SINASC e/ou SIH no período de 1990 a 2006.

3 ARTIGO 1

SIH-SUS E SINASC: UTILIZAÇÃO DO MÉTODO PROBABILÍSTICO  
PARA RELACIONAMENTO DE DADOS

Publicado nos Cadernos de Saúde Coletiva

## RESUMO

**Introdução:** O SIH, única fonte de informações de morbidade hospitalar de abrangência nacional, é alimentado pelas autorizações de internação hospitalar (AIH), que fornecem dados de internações da rede do Sistema Único de Saúde (SUS), onde ocorrem aproximadamente 70% das internações no Brasil. O Sistema de Informação sobre Nascidos Vivos (SINASC), de base populacional, é alimentado pelas DN, contendo informações sobre nascimentos.

Propõe-se metodologia de relacionamento probabilístico entre SINASC e SIH para avaliar a cobertura para nascidos vivos na rede hospitalar SUS em Minas Gerais.

**Material e Métodos:** relacionaram-se probabilisticamente 293.213 DN da coorte de nascidos vivos em 2001 com 223.443 AIH de partos de nascidos vivos ocorridos em hospitais da rede pública mineira. Atribuíram-se valores de 0 (discordância total) a 100 (concordância total) de acordo com o percentual de concordância entre variáveis de relacionamento (nome da mãe e data da hospitalização).

**Resultados:** Cerca de 68% das DN foram relacionadas às AIH, percentual próximo aos 70% estimados como usuários exclusivos do SUS (a população chamada de “SUS-dependente”).

**Conclusões:** O resultado satisfatório obtido nesse relacionamento entre SINASC e SIH aponta para possibilidades de realização de análises a partir da combinação das variáveis de cada um dos sistemas. Novos estudos são necessários para investigar as razões pelas quais 12% das AIH não se relacionaram a qualquer DN.

**Palavras-chave:** SIH, SINASC, relacionamento de bancos de dados, partos hospitalares.

## **ABSTRACT**

**Introduction:** The hospital admittance authorization system (SIH) is a system fed with information from hospitalizations covered by SUS, the Brazilian Unified Healthcare System, which accounts for 70% of the hospitalizations countrywide. The live birth information system (Sistema de Informação sobre Nascidos Vivos – SINASC), a population-based system, records information from deliveries occurred in the State of Minas Gerais, in public and private hospitals.

The present study aims at evaluating the coverage of deliveries in public hospitals in State of Minas Gerais, based on the record linkage methodology.

**Material and Methods:** 293,213 birth records from the live cohort of 2001 covered by SINASC were probabilistically matched to 223,443 hospital authorizations from SIH. Scores given to the pairs ranged from 0 (total disagreement) to 100 (total agreement) based on the degree of agreement between the variables mother's name and date of hospitalization.

**Results:** About 68% of the birth records matched the hospital authorizations, a percentage close to what would be expected. However, around 12% of the hospitalizations could not be matched to any birth record.

**Conclusions:** The satisfactory results obtained from the matching of SIH and SINASC point out to the possibility of conducting studies combining variables from each system. Other studies are needed in order to investigate why some hospitals authorizations could not be related to any birth record.

**Key words:** hospital authorizations, birth records, record linkage, hospital deliveries

### 3.1 INTRODUÇÃO

Os Sistemas de Informação em Saúde (SIS) implantados no Brasil nas três últimas décadas têm ampliado sua cobertura e se consolidado, principalmente em resposta aos investimentos do Ministério da Saúde. Sua crescente utilização tanto pelos serviços de saúde quanto para fins acadêmicos, também incentivou os perceptíveis avanços observados nessas bases de dados. Entre os SIS de base populacional está o Sistema de Informação sobre Nascidos Vivos (SINASC), implantado oficialmente a partir de 1990 e que tem por objetivo propiciar “um aporte significativo de dados sobre nascidos vivos, com suas características mais importantes, como sexo, local onde ocorreu o nascimento, tipo de parto e peso ao nascer, entre outras”<sup>1</sup>. Outro importante produto do SINASC é o conhecimento do número de nascidos vivos a cada ano em uma dada região, que é ferramenta indispensável para o cálculo de dois indicadores, a taxa de mortalidade infantil e a razão de mortalidade materna, tradicionalmente utilizados no monitoramento das condições de vida de uma população. Tanto em relação à cobertura quanto sob o ponto de vista da qualidade das informações, o SINASC tem apresentado avanços importantes desde sua implantação <sup>2</sup>.

Apesar da sua importância para a construção de vários indicadores de saúde em algumas regiões do país o SINASC ainda apresenta muitas debilidades. Entre elas podem ser apontadas as falhas ligadas às tarefas de coleta, digitação e processamento de dados e as relacionadas a treinamento e sensibilização ainda deficientes das fontes notificantes. Em relação à caracterização da rede assistencial dos nascimentos, a principal desvantagem da utilização do SINASC como fonte de dados está na codificação do campo referente aos códigos dos estabelecimentos. Nesses campos são adotados códigos próprios desse sistema. Mas, o mais sério entre os reflexos das debilidades apresentadas pelo SINASC e que pode ser observado na maioria dos estados brasileiros <sup>3,4</sup> são as coberturas relativamente baixas apresentadas por esse SIS <sup>3,4</sup>. Em Minas Gerais, segundo avaliação do Ministério da Saúde, apesar da cobertura do SINASC estar crescendo desde sua implantação, ainda foi inferior a 85% no ano 2003.

O Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH) também foi criado no início dos anos 90, estruturado na lógica da avaliação e controle da produção, para pagamento das internações hospitalares. As unidades hospitalares participantes do SUS (públicas ou contratadas/conveniadas) enviam as informações das internações efetuadas

através das AIH - Autorização de Internação Hospitalar - para os gestores municipais ou estaduais. Estas informações são processadas no DATASUS, gerando créditos referentes aos serviços prestados e constituindo uma valiosa base de dados de morbidade hospitalar, que contém informações de grande parte das internações hospitalares realizadas no Brasil. Os dados provenientes do SIH estão, tradicionalmente, entre os mais fáceis e agilmente disponíveis. Inicialmente, as informações eram acessadas via CD-Rom e, atualmente, estão na Internet 5.

De forma alternativa, o SIH tem sido analisado, também, para além de seus fins contábeis 6, 7, 8, 9. E tem surpreendido positivamente em abordagens para avaliações de processos e resultados de hospitais da rede SUS 10, 11, 12 e como ferramenta auxiliar a outros SIS, como o SINAN (Sistema de Informações de Agravos Notificáveis) 13 e o SIM 14, 15.

Schramm & Szwarcwald 16 utilizaram o SIH de forma pioneira para estimar o número de nascidos vivos, natimortos e óbitos neonatais precoces e tardios em alguns estados brasileiros em 1995. Os dados obtidos no SIH foram comparados aos referentes ao SINASC e ao SIM de cada Estado, no mesmo ano. As autoras observaram, em alguns estados das Regiões Norte e Nordeste (como Maranhão, Ceará e Paraíba), um número maior de registros de natimortos e de óbitos neonatais no SIH que no SIM. Em relação aos nascidos vivos, tanto o SIH quanto o SINASC apresentaram baixa cobertura em algumas unidades federadas, em comparação com o número de estimado pelo método de Brass, muito utilizado por demógrafos para obtenção de estimativas indiretas de eventos. Constataram também que, ao lado da grande variação na cobertura apresentada pelo SIH em alguns estados da região Nordeste, o número de nascidos vivos registrados no SINASC foi menor que o número obtido pelo SIH. Para Minas Gerais, as autoras obtiveram a maior cobertura estimada de nascidos vivos do SIH: 95%.

Segundo dados do SINASC, no Brasil, mais de 95% dos nascimentos ocorrem em hospitais. Os elevados níveis das taxas de mortalidade neonatal precoce em nosso país tornam ainda mais relevantes os estudos dirigidos à avaliação da assistência prestada à parturiente e ao recém-nascido. Isto se deve especialmente ao fato de que as análises dos óbitos ocorridos no período neonatal precoce têm apontado para o grande percentual de mortes por causas evitáveis 17, 18, associados a falhas na assistência ao trabalho de parto e/ou ao parto. Assim, associar as informações das variáveis do SINASC, em especial o peso de nascimento, que é um importante fator para determinação do risco para mortes no primeiro ano de vida às



Primeira letra K e segunda A, O ou U - Primeira letra passa a C,

Primeira letra Y - Primeira letra passa a I,

Primeira letra C e segunda E ou I - Primeira letra passa a S,

Primeira letra G e segunda E ou I - Primeira letra passa a J;

Foi criado um campo CLSS para receber os campos NOME e DTSTRING concatenados, ou seja, o conteúdo de DTSTRING foi adicionado ao final do campo nome e este resultado foi armazenado no campo CLSS. E, finalmente, outro campo foi criado para a codificação dos estabelecimentos de saúde de ocorrência dos nascimentos com base em tabela fornecida pela SES/MG.

#### RELACIONAMENTO DE DADOS

Após a preparação das variáveis acima, iniciaram-se as cinco etapas do processo de relacionamento, denominadas “PASSO”, criando blocos lógicos de registros com os campos homólogos (campos que guardam informações de mesma natureza) e associação dos blocos combinando um ou mais campos, iniciando-se com uma chave mais restrita e posterior diminuição da restrição. Buscou-se, assim, minimizar a perda de pares. Nos cinco primeiros passos manteve-se relação de uma AIH para uma DN. Apenas no sexto passo foi admitida relação de uma AIH para uma ou mais DN.

#### DEFINIÇÃO DAS FUNÇÕES

**Soundex:** Algoritmo constituído por quatro dígitos, sendo o primeiro representado pela primeira letra da palavra a ser codificada e os demais por dígitos numéricos codificados, segundo regras que buscam minimizar erros. Essa função teve como objetivo a redução da perda de pares verdadeiros devido a problemas decorrentes de erros e/ou diferenças na grafia dos nomes das mães nos dois sistemas de informação.

**Difdata:** Função criada para retornar em dias a diferença entre duas datas. Foram utilizadas, neste caso, as datas da internação da AIH (DT\_INTER) e as datas de nascimento da DN (DTNASC). Tendo em vista que no SIH o cadastro é por internação e não por indivíduo e que se pode ter várias internações do mesmo indivíduo em um determinado ano, esta função visa minimizar a possibilidade de formação de pares não verdadeiros.

**Municnasc:** Função criada para comparar códigos de municípios de nascimento, CODMUNASC (DN) e MUNIC\_MOV (AIH), retornando igual para os casos coincidentes, ou diferente para os não coincidentes. Teve como objetivo principal a redução da formação de pares não verdadeiros que seria possibilitada pela presença de hospitais de mesmo nome em cidades diferentes.

**Municres:** Função criada para comparar códigos de municípios de residência da mãe, CODMUNRES (DN) e MUNIC\_RES (AIH), retornando, da mesma forma que na função anterior, igual ou diferente. Sua criação visava resgatar pares verdadeiros não formados em passos anteriores devido à ausência do código da variável município de ocorrência no SINASC.

**Estab:** Função criada para comparar os estabelecimentos de ocorrência do nascimento, também retornando igual ou diferente. Após a compatibilização entre os códigos dos locais de ocorrência (CODESTAB) utilizados pelo SINASC com o CGC do hospital (utilizado pelo SIH); esta função foi definida para aumentar a probabilidade de acerto da opção pelo par verdadeiro, entre os diversos pares formados para cada DN.

## PROCESSO DE PAREAMENTO

O processo de comparação entre as bases foi feito em blocos de um ou mais campos, iniciando-se com uma combinação mais restrita e posterior diminuição dessa restrição. Os relacionamentos foram feitos utilizando-se uma sentença em linguagem SQL denominada *datalink*, desenvolvida por Elan Martins<sup>20</sup> que faz a comparação direta entre o soundex do primeiro nome nos dois bancos. Os pares de registros que apresentaram coincidência neste campo foram então submetidos às seguintes funções, conforme definido na chave de blocagem (Quadro 1):

Difdata – variando no intervalo zero a quatro dias;

Municnasc – valor igual;

Municres – valor igual;

Estab – valor igual.

Quadro 1 – Blocos utilizados em cada passo de relacionamento de AIH e DN

Passo	Chave
1	<i>Soundex</i> (PrimeiroNOME)+Difdata+ Municnasc +Estab
2	<i>Soundex</i> (PrimeiroNOME)+Difdata+ Municnasc
3	<i>Soundex</i> (PrimeiroNOME)+Difdata+ Municres
4	<i>Soundex</i> (PrimeiroNOME)+Difdata
5	<i>Soundex</i> (PrimeiroNOME)
6 (somente para os gemelares)	<i>Soundex</i> (PrimeiroNOME)

O campo CLSS foi submetido à função Distância de Levenshtein (LD) e o resultado desta foi submetido a outra função denominada *Percentile Similarity* (SP) que determinava o percentual de semelhança entre *strings* comparadas, variando de 0% a 100%<sup>21</sup>. Esses percentuais de semelhança foram armazenados em uma variável nomeada SCORE, que foi indexada de maneira inversa para auxiliar na definição manual dos pares considerados verdadeiros. Entre cada um dos passos automatizados foi realizada a revisão manual dos pares formados, para sua admissão como pares verdadeiros e para verificação da adequação da mediana do campo SCORE como ponto de corte adotado. O uso da mediana, e não da média, se justifica pelo fato de não levar em conta os valores extremos de SCORE, especialmente os muito baixos, que poderiam enviesar a média para um valor artificialmente pequeno, visto que a maioria dos pares formados, em geral, são falsos. A exclusão dos pares verdadeiros foi feita entre cada um dos passos, para que se mantivesse a relação unívoca.

## ASPECTOS ÉTICOS

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFMG (Nº 095/04).

### 3.3 RESULTADOS

Os resultados obtidos em cada uma das etapas de relacionamento encontram-se na Figura 1.

No primeiro passo de relacionamento formaram-se 309.110 pares totais (média de 1,05 AIH para cada DN). A seguir, selecionaram-se os pares com o maior escore para cada DN, totalizando 146.359 pares distintos com relação unívoca. O cálculo da mediana dos escores obtidos nesta primeira etapa de relacionamento foi de 87,1 - valor este a partir do qual todos

os 123.369 pares formados foram considerados como pares verdadeiros. Destes, todos possuíam relação unívoca.

No segundo passo, do qual foram excluídos os pareamentos já formados, obtiveram-se 276.688 pares (média de 1,63 AIH por DN), dos quais foram selecionados os 77.668 pares distintos que apresentavam os maiores escores. A mediana, utilizada como ponto de corte a partir do qual os pares formados não foram considerados como verdadeiros, foi igual a 77,70. Nesta etapa, em 133 pares (escores de 87 a 100), não constava o estabelecimento de ocorrência do nascimento na DN. Foram considerados 56.740 pares verdadeiros, todos com relação unívoca.

No terceiro passo formaram-se 101.247 pares (média de 0,90 AIH por DN), sendo selecionados os 16.286 com os maiores escores. Nestes, a mediana foi de 58,55. Devido à baixa qualidade dos pares formados optou-se por um escore de 70, superior ao valor da mediana, como ponto de corte a partir do qual os 4.551 pares foram considerados verdadeiros.

No quarto passo foram formados 305.939 pares (2,81 AIH por DN). A mediana dos maiores escores dos 26.036 pares distintos foi de 57,48. Neste caso, o escore 90 foi considerado como limite para admissão de cada um dos 1.622 pares como verdadeiros.

O quinto passo foi subdividido em duas fases. Na primeira, foram formados 281.583 pares (5,39 AIH por DN). A mediana dos maiores escores dos 52.237 pares distintos foi de 47,47. Neste caso, o escore 90 foi considerado como limite para admissão de cada um dos 4.870 pares verdadeiros.

Na segunda fase do quinto passo, novamente apenas a variável nome da mãe foi utilizada para relacionar apenas as DN ainda não relacionadas a todas as AIH. Formaram-se 281.530 pares (2,63 AIH por DN), que (após seleção dos 24.392 pares com os maiores escores), tiveram o valor mediano dos escores de 57,75. Nesta etapa, os 2.107 considerados como verdadeiros foram selecionados um a um. O menor escore apresentado por um par verdadeiro foi de 35,89, segundo revisão manual. Observou-se, neste passo, que uma AIH previamente relacionada a uma DN em passos anteriores relacionou-se a uma nova DN nesta etapa. Admitiu-se, a partir desta constatação, a possibilidade de se tratarem de partos gemelares, onde a uma única AIH poderia estar relacionada duas ou mais DN. Nesse último passo, que

admitiu 6.977 pares nas suas duas fases, foi, então, considerado como último passo para admissão do total de 193.259 pares nos quais a relação unívoca se manteve.

A sexta e última etapa, dedicada ao relacionamento dos gemelares, também foi subdividida em duas fases. Na primeira, as AIH referentes a partos gemelares, identificadas pelo campo NASC\_VIVOS maior do que 1 foram relacionadas às DN ainda não pareadas. Foram utilizadas, para o relacionamento, as variáveis nome da mãe e a data da internação menor ou igual a 10 dias à data do nascimento. Na segunda fase, todas as AIH não relacionadas, independentemente do campo identificador de parto gemelar no SIH, foram relacionadas a todas as DN restantes. Formaram-se, no total, 96.390 pares (0,92 AIH por DN). Os 33.202 pares com os maiores escores obtiveram mediana igual a 56,02. A seleção um a um admitiu 3.145 pares (3.145 AIH e 6.316 DN), com menor escore igual a 24,39.

### **3.4 DISCUSSÃO**

A proposta de relacionamento probabilístico de dados, através de blocagem em cinco passos realizada no presente estudo, permitiu a identificação no SIH de cerca de 70% dos nascidos vivos do SINASC. É importante ressaltar que, diferentemente de outros estudos anteriormente realizados, nesse trabalho o pareamento foi feito através de variáveis individuais. A utilização de nomes das pacientes, chave essencial na estratégia de blocagem adotada e que são protegidos em arquivos específicos devido a questões éticas, só se tornou possível após liberação desses dados pelo DATASUS/Ministério da Saúde. Esses dados referiam-se, exclusivamente, a internações de mulheres para partos, o que impossibilitou a utilização da variável sexo na formação das chaves de pareamento.

Algumas dificuldades para definição das variáveis que comporiam as chaves para esse relacionamento, envolvendo duas grandes bases de dados, merecem ser destacadas. Além dos problemas referentes à completude de algumas variáveis do SINASC (como município de ocorrência e de residência), a utilização das variáveis CGC do estabelecimento (SIH) e códigos de estabelecimento definidos pela SES/MG (SINASC) apresentou dificuldade adicional. Para a identificação do estabelecimento de saúde onde ocorreu o parto cada um dos sistemas utiliza a sua própria forma, tanto em relação à identificação numérica (CGC ou códigos) quanto ao nome (razão social ou personalidade jurídica) utilizado pelos hospitais. Ressalte-se que, em Minas Gerais, observou-se um grande número de hospitais, de diferentes

municípios, com a mesma razão social (em sua maioria referente a santos ou santas da Igreja Católica). O não preenchimento do código dos estabelecimentos nas DN acrescentou dificuldade a esta compatibilização. Tendo em vista os problemas mencionados, a variável referente aos estabelecimentos de saúde de ocorrência dos partos foi excluída no segundo passo do relacionamento, apesar de sua importância. A codificação única dos estabelecimentos de saúde (Cadastro Nacional de Estabelecimentos de Saúde/CNES), em fase de implantação, apresenta grande potencial facilitador para o relacionamento de dados a partir do estabelecimento de ocorrência e, provavelmente, será adequada para superação dessa grande dificuldade.

As variáveis código do município de ocorrência do nascimento (SINASC) e código de localização do estabelecimento de saúde (SIH) também apresentaram limitações. Além da diferença entre esses códigos, que foram compatibilizados pela exclusão dos últimos dígitos dos códigos SIH, foram observados problemas relacionados ao preenchimento não completo dessa variável no SINASC. Optou-se, no entanto, pela sua utilização por ser reconhecido que, em muitos casos, há a chamada peregrinação das parturientes<sup>17</sup> em busca de melhores condições de atendimento. Assim, por medo de não serem atendidas em outro município no qual não residem de fato, as pacientes declaram erroneamente os seus municípios de residência. Desta maneira, poderia haver maior tendência de erro no relacionamento se a variável município de residência fosse a única a ser utilizada. Justifica-se, assim, porque os códigos dos municípios de ocorrência estiveram em uma chave e os códigos dos municípios de residência foram utilizados em outra chave.

Os partos gemelares foram outra fonte de dificuldade e provável geradora de falhas no pareamento. Se a cada nascimento vivo corresponde uma DN, uma única AIH pode estar relacionada a mais de um nascimento. Nos dois sistemas há um campo específico para essa informação que, no entanto, apresentava baixo percentual de preenchimento no SINASC.

Finalmente, devem ser destacadas as possibilidades de erros de preenchimento e de deficiências no controle da qualidade das datas das internações e dos nascimentos. Quando se procedeu ao processo de revisão manual dos pares absolutamente concordantes, observou-se que os intervalos entre as internações das mães e os nascimentos eram, em sua maioria, inferiores a quatro dias. Mas, buscando incluir o maior número de pares verdadeiros e admitindo que erros de digitação em datas sejam comuns, admitiram-se períodos de até dez

dias entre a internação da mãe e o parto para que os pares formados fossem admitidos como verdadeiros, nas fases iniciais. Posteriormente foram admitidos intervalos maiores, tendo em vista a baixa sensibilidade desta variável.

Além disso, os erros na grafia do nome da mãe, de transcrição das informações de data de nascimento e de internação e ausência de informações sobre os hospitais e os municípios podem ter contribuído para subestimar a cobertura da rede SUS para os nascimentos em Minas Gerais obtida no presente estudo. A estratégia de relacionamento adotada, com revisão manual de todos os pares formados em cada um dos passos, para verificação de sua qualidade e da adequação dos pontos de corte adotados, teve como um dos objetivos a redução dessas prováveis perdas. A revisão manual dos pares, no entanto, é um procedimento que requer critérios claramente definidos, além de consumir grande quantidade de tempo para sua execução.

Embora possa ser considerado satisfatório e próximo ao esperado, o percentual de pares obtidos através da metodologia utilizada nesse estudo (70%) foi bastante inferior aos 95% obtidos por Schramm & Szwarcwald<sup>16</sup> que, entretanto não utilizaram arquivos com nomes. A metodologia probabilística de relacionamento, com pareamento a partir da formação de blocos lógicos, apesar de sujeita a falhas, tem maior probabilidade de representar com maior nitidez a real cobertura do SUS/MG. Além dessa vantagem, e a despeito das dificuldades relatadas, principalmente atribuíveis às falhas na qualidade do preenchimento das variáveis, os procedimentos de relacionamento de bancos de dados de nascidos vivos e internações hospitalares adotados demonstraram ser uma estratégia viável e ainda pouco explorada em nosso meio. A agregação das variáveis do SINASC, especialmente o *peso de nascimento*, aos dados do SIH abre possibilidades para novos e diversos estudos analíticos longitudinais dos nascimentos na rede SUS, mesmo reconhecendo-se as limitações do SIH. Entre essas limitações estariam, principalmente, as ligadas à lógica contábil do sistema, voltado para o pagamento das internações. Portanto, as análises epidemiológicas daí decorrentes devem ser feitas com cautela. Dentro desse contexto, finalmente, cabe ressaltar que cerca de 12% de AIH não relacionadas a qualquer DN. A identificação das prováveis causas desse não pareamento requer que sejam realizados outros estudos.

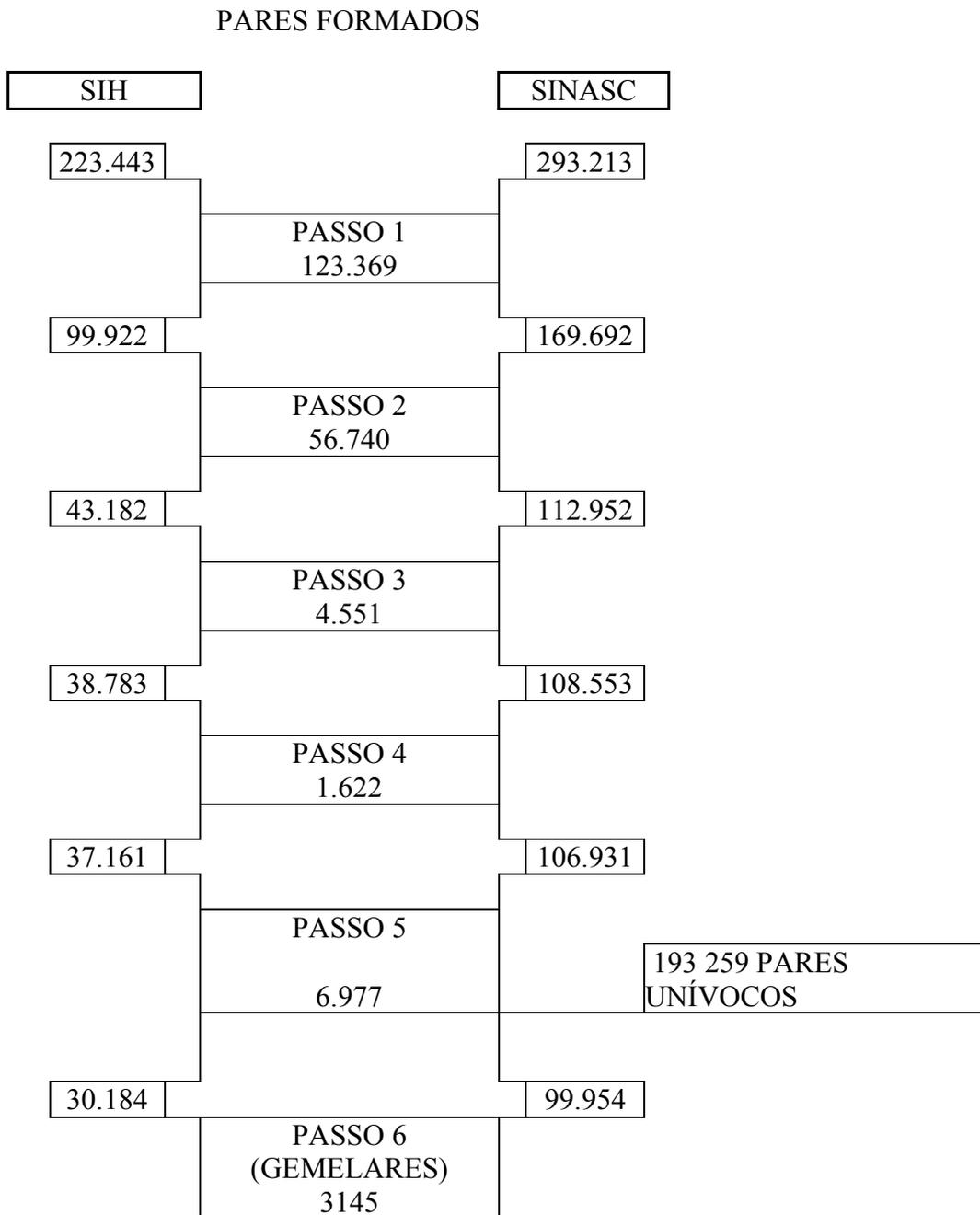
## REFERÊNCIAS BIBLIOGRÁFICAS

1. Departamento de Dados do Sistema Único de Saúde/ DATASUS. Sistema de Informação Sobre Nascidos Vivos. 2005. Disponível em: <http://datasus.gov.br/catalogo/sinasc.htm>. (acessado em 23/Dez/2005).
2. Theme Filha MM, Gama SGN, Cunha CB., Leal MC. Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no Município do Rio de Janeiro, 1999-2001. *Cad Saúde Pública* 2004; 20 (1 Suppl): 83-91.
3. Barros FC, Victora CG, Teixeira AMB, Filho MP. Mortalidade perinatal e infantil em Pelotas, Rio Grande do Sul: nossas estatísticas são confiáveis? *Cad Saúde Pública* 1985; 1 (3): 348-358.
4. Szwarcwald CL, Leal MC, Andrade CLT, Souza Jr PRB. Estimação da mortalidade infantil no Brasil: o que dizem as informações sobre óbitos e nascimentos do Ministério da Saúde? *Cad Saúde Pública* 2001; 18 (6): 1725-36.
5. Departamento de Dados do Sistema Único de Saúde/ DATASUS. Transferência de Arquivos. 2005. Disponível em: <http://datasus.gov.br/arquivos/ftpdata.htm>. (acessado em 22/Dez/2005).
6. Levcovitz E, Pereira TRC. SIH/SUS (Sistema AIH): 1983-1991. Série: Estudos em Saúde Coletiva, n 057. Universidade do Estado do Rio de Janeiro - UERJ, 1993.
7. Veras CMT, Martins MA. Confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH), Rio de Janeiro, Brasil. *Cad Saúde Pública* 1994; 10 (3): 339-55.
8. Ferreira VMB, Portela MC. Avaliação da subnotificação de casos de AIDS no município do Rio de Janeiro com base em dados do sistema de informações hospitalares do Sistema Único de Saúde. *Cad Saúde Pública* 1999; 15 (2): 317-24.
9. Escotesguy CC, Portela MC, Medronho RA, Vasconcelos MTL. Acute myocardial infarction. Clinical and epidemiological profile and factors associated with in-hospital death in the municipality of Rio de Janeiro. *Arq Bras Cardiol* 2003; 80 (6): 600-06.
10. Gouvêa CSD, Travassos C, Fernandes C. Produção de serviços e qualidade da assistência hospitalar no Estado do Rio de Janeiro, Brasil-1992 a 1995. *Rev Saúde Pública* 1997; 31 (6): 601-17.
11. Escotesguy CC, Portela MC, Medronho RA, Vasconcelos MTL. O Sistema de Informações Hospitalares e a assistência ao infarto agudo do miocárdio. *Rev Saúde Pública* 2002; 36 (4): 491-99.
12. Noronha JC, Travassos C, Martins M, Campos MR, Maia P, Panezzuti R. Avaliação da relação entre volume de procedimentos e a qualidade do cuidado: o caso de cirurgia coronariana no Brasil. *Cad Saúde Pública* 2003; 19 (6): 1781-89.

13. Lessa JJD, Mendes ACG, Farias SF, Sá DA, Duarte PO, Melo filho DA. Novas metodologias para vigilância epidemiológica: uso do Sistema de Informações Hospitalares-SIH/SUS. *Inf Epidemiol SUS* 2000; 9 (1 Suppl): .3-27.
14. Guerra HL, Barreto SM, Uchoa E, Firmo JOA, Lima e Costa MFF. A morte de idosos na Clínica Santa Genoveva, Rio de Janeiro: um excesso de mortalidade que o sistema público de saúde poderia ter evitado. *Cad Saúde Pública* 2000; 16 (2): 545-51.
15. Correa PRL, Assunção AA. A sub-notificação de mortes por acidentes de trabalho: estudo de três bancos de dados. *Epidemiol Serv Saúde* 2003; 12 (4): 203-12.
16. Schramm JMA, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e natimortalidade. *Rev Saúde Pública* 2000; 4 (3): 272-79.
17. Lansky S, França EB, Leal MC. Mortes perinatais evitáveis em Belo Horizonte, Minas Gerais, Brasil, 1999. *Cad Saúde Pública* 2002;18 (5): 1389-00.
18. Caldeira AP, França EB, Perpétuo IHO, Goulart EMA. Evolução da mortalidade infantil por causas evitáveis, Belo Horizonte, 1984 – 1998. *Rev Saúde Pública* 2005; 39 (1): 67-74.
19. Camargo KRJr, Coeli CM. *Reclink*: aplicativo para o relacionamento de bases de dados, implementando o método *probabilistic record linkage*. *Cad Saúde Pública* 2000; 16 (2): 439-47.
20. Martins ER. Manual de Documentação do *datalink* (em preparação), 2004.
21. Gilleland M. Levenshtein Distance, in three flavors. Merrian Park Software s/d. Disponível em; <http://www.merrianpark.com/ld.htm> (acessado em 09/Mar/2005).

Figura 1

Fluxograma do relacionamento dos dados entre SINASC e SIH.



#### 4 ARTIGO 2

### SUBNOTIFICAÇÃO DE NASCIDOS VIVOS NO SINASC: DOIS PROCEDIMENTOS DE MENSURAÇÃO A PARTIR DO SIH

Aceito para publicação na Revista de Saúde Pública

## RESUMO

Objetivou-se estimar a subnotificação de nascidos vivos no SINASC em Minas Gerais por meio de dois procedimentos – a comparação do número de nascidos vivos por município de residência e a técnica de relacionamento probabilístico de dados individuais. Foram utilizados o Sistema de Informação Hospitalar (SIH) e o Sistema sobre Nascidos Vivos (SINASC). Realizou-se a comparação do número de nascidos vivos por município de residência nas duas bases e também o relacionamento probabilístico de dados. Em ambos os casos, consideraram-se como indicadores de subnotificação as proporções de nascidos vivos informadas ao SIH e não obtidas no SINASC, segundo municípios de residência de Minas Gerais. Os municípios foram posteriormente agregados em quatro faixas de tamanho populacional. O relacionamento probabilístico identificou maior proporção de nascidos vivos subnotificados no SINASC, relativamente à comparação do número de nascidos vivos nos municípios. A variação das diferenças entre os percentuais de subnotificação por procedimentos foi de 9,4% nos municípios com população menor que 5.000 habitantes; 9,1% nos municípios com população entre 5.000 e 9.999 e 8,0% nos municípios com população entre 10.000 e 49.999 e com mais de 50.000 habitantes. É inegável a contribuição do SIH como indicador de subnotificação de nascidos vivos. A magnitude da contribuição revelou-se sensível ao procedimento adotado. O relacionamento probabilístico é tecnicamente mais complexo, mas incrementou certeza no pareamento e possibilitou identificação de maior proporção de casos não registrados no SINASC, inclusive em grandes municípios.

Palavras-chave: SINASC, SIH, nascido vivo, sistemas de informação, avaliação, relacionamento de dados, fontes de dados, notificação.

## ABSTRACT

This study aimed at an estimation of underreporting of live births in the SINASC system in the state of Minas Gerais, Brazil, by means of two procedures: a comparison between the number of live births in each town and the probabilistic linkage from individual data. Systems included the Hospital Information System (SIH) and the Live Birth System (SINASC). For both procedures, indices of underreporting were obtained from the proportion between live births present in the SIH system and those unavailable at the SINASC, according to each town of residence in Minas Gerais. The towns were later grouped into four ranges of population. The probabilistic linkage was able to identify a greater proportion of live births absent from SINASC, relative to the comparison of live births in each town. The percentages of underreporting between the procedures was of 9.4% in towns smaller than 5,000 inhabitants; 9.1% in those whose population ranged from 5,000 and 9,999; and 8.0% in those between 10,000 and 49,999 and over 50,000. SIH is a strong index of underreporting of live births. The amount of underreporting revealed to be sensitive to the procedures adopted. Probabilistic linkage is technically more complex; however, it reinforced the certainty in the pairings, and also made possible the identifications of a greater proportion of cases not recorded in the SINASC, mainly in bigger towns.

Keywords: SINASC, SIH, live birth, information systems, assessment, record linkage, underreporting.

## 4.1 INTRODUÇÃO

Os sistemas de informação em saúde (SIS) nacionais contêm grande volume de dados para planejamento e avaliação de políticas públicas e precisam ser utilizados para que os possíveis problemas referentes à qualidade da informação sejam conhecidos e solucionados <sup>1</sup>. Reconhece-se a importância e as dificuldades para obtenção das informações sobre o número de nascidos vivos, cuja fonte oficial é o IBGE mas que tem seu uso limitado pelo sub-registro, principalmente. Para o enfrentamento da questão, em 1990 o Ministério da Saúde (MS) implantou o Sistema de Informação sobre Nascidos Vivos (SINASC) <sup>1</sup> que contém importantes variáveis para análises epidemiológicas – como peso de nascimento. Mesmo apresentando significativas melhorias <sup>1</sup>, ainda existem deficiências na qualidade das informações <sup>2</sup> do SINASC e sabe-se que a subnotificação de nascidos vivos reflete a incapacidade de captação destes eventos pelo sistema de saúde. Esta realidade, encarada como um desafio na saúde pública, motiva os serviços de saúde e os pesquisadores <sup>1</sup> a explorar possíveis formas para mensurar e reduzir esta subnotificação.

Existem diferenças locais, regionais e estaduais na cobertura e qualidade do SINASC <sup>1,2</sup>, que é gerido pelas três esferas de governo (federal, estadual e municipal). Entretanto, o papel dos gestores municipais na qualificação do SINASC torna-se cada vez mais importante, com a progressiva descentralização da gestão deste sistema. Neste contexto, estimar e localizar a subnotificação de nascidos vivos no nível municipal permite a adoção de ações específicas e necessárias para a sua redução pelo nível local. Uma das técnicas para identificar as subnotificações de nascidos vivos é o relacionamento das informações do SINASC às informações de outros SIS <sup>3</sup>. Com este objetivo já foram utilizados, por exemplo, o Sistema de Informação da Atenção Básica (SIAB) <sup>4</sup> e o Sistema de Informações Hospitalares (SIH) <sup>5</sup>. O SIH, o único SIS sobre morbidade hospitalar, contém variáveis epidemiológicas <sup>6</sup>, além das informações com finalidade contábil. Refere-se apenas a registros de internações ocorridas nos hospitais próprios e contratados/conveniados da rede SUS. Portanto, não são cobertos pelo SIH partos domiciliares, pagos por particulares ou pela assistência complementar. Desta forma, ao serem comparados os registros do SINASC – para o qual se pretende cobertura universal – aos registros do SIH, espera-se que o SINASC contenha sempre igual ou maior número de notificações de nascidos vivos. Caso o SINASC não contenha volume igual ou superior de notificações que o SIH, a diferença entre os dois sistemas poderia ser considerada como indicativo de subnotificação <sup>6</sup> de nascidos vivos no SINASC. Com base nesta premissa,

Almeida & Alencar <sup>7</sup> propuseram a utilização do número de partos do SIH como parâmetro mínimo para captação de nascidos vivos pelo SINASC. Schramm & Szwarcwald <sup>5</sup> utilizaram os dados do SIH para estimar também a natimortalidade e a mortalidade neonatal em estados brasileiros. Em ambos os estudos, apesar de não se poder afirmar que se tratava do mesmo nascido vivo notificado – ou não – nos dois sistemas de informação, o maior número de registros de partos no SIH possibilitou aos autores concluírem tratar-se de uma fonte alternativa viável para captação de nascidos vivos em áreas de baixa cobertura do SINASC.

Para aumentar a probabilidade de acerto ao se afirmar que um registro em um SIS refere-se a registro de um mesmo indivíduo em outro SIS é necessário que eles sejam relacionados através de pelo menos uma variável identificadora. Por meio do relacionamento probabilístico pode-se afirmar, com base em probabilidades de concordância e discordância das variáveis-chave, que os registros de diferentes bases de dados pertencem ao mesmo indivíduo. Várias estratégias de relacionamento probabilístico de dados são utilizadas <sup>4</sup> para superar a ausência de uma variável identificadora única. Muito freqüentemente, nos relacionamentos probabilísticos os nomes dos pacientes são utilizados como a principal variável identificadora, associados ou não a outras variáveis <sup>8,9</sup>.

Visando explorar a idéia de que diferentes procedimentos para estimar subnotificação conduzem a diferentes resultados, este estudo objetivou avaliar os resultados obtidos pela comparação do número de nascidos vivos, por município de residência da mãe e disponibilizados em CD-ROM e no sítio da Internet pelo DATASUS <sup>11</sup> e sem variáveis individuais identificadoras, com aqueles obtidos por técnica de relacionamento probabilístico de registros, realizada a partir de dados disponibilizados pela Secretaria Estadual de Saúde de Minas Gerais (SES/MG) e com variáveis que permitiriam, de forma conjunta, identificar o indivíduo. Em cada um dos procedimentos buscou-se quantificar a contribuição do SIH como fonte de alternativa de notificação de nascidos vivos para o SINASC no ano 2001 em Minas Gerais. Considerou-se como indicador de subnotificação de nascidos vivos a diferença entre as ocorrências de partos de nascidos vivos informadas ao SIH e as de nascidos vivos obtidas por meio do SINASC.

## **4.2 METODOLOGIA**

### **FONTE DE DADOS**

Os dados do SINASC e do SIH referiam-se aos nascidos vivos e aos partos de nascidos vivos de residentes e ocorridos em Minas Gerais do ano de 2001. Para a comparação do número de eventos, utilizaram-se dados provenientes de CD-ROM disponibilizados pelo DATASUS. Para o relacionamento probabilístico, os dados obtidos junto à SES/MG continham variáveis individuais identificadoras, que possibilitaram a correção prévia de duplicidades e de erros na especificação dos municípios de residência antes da sua utilização.

#### COMPARAÇÃO DO NÚMERO DE EVENTOS

Os dados do SIH obtidos em CD-ROM referiam-se a procedimentos autorizados nos anos de 2001 e 2002. No subdiretório MA (Movimento da AIH), selecionaram-se os registros nos quais estivessem preenchidas a variável 54 (NASC\_VIVOS), a qual é preenchida apenas em casos de parto e que possibilitou a identificação das internações para partos de nascidos vivos. Realizou-se, a seguir, a comparação do número de nascidos vivos notificados ao SINASC e de partos de nascidos vivos notificados ao SIH por meio da variável município de residência do SINASC (CODMUNRES) e do SIH (MUNIC\_RES). Considerou-se haver subnotificação de nascidos vivos no SINASC quando o número de nascidos vivos obtido no SIH revelou-se superior ao obtido no SINASC por município de residência.

#### RELACIONAMENTO PROBABILÍSTICO

O relacionamento probabilístico foi adotado devido ao fato de não haver variável identificadora única nos dois bancos de dados. Incluíram-se todos os registros de nascidos vivos residentes em Minas Gerais no ano de 2001 do SINASC. No SIH foram utilizados três critérios para inclusão dos registros: 1) internações de mulheres; 2) internação no ano de 2001 e 3) campo 54 diferente de zero.

Foram necessários alguns ajustes nas bases de dados, para a adequada execução do relacionamento probabilístico <sup>12</sup>. Para isto, os campos foram mantidos de forma idêntica nos dois arquivos. Procedeu-se, por exemplo, à padronização das datas para que tivessem o mesmo formato (*string*). Também foi criado um campo CLSS originário dos campos NOME e DTSTRING concatenados, onde o conteúdo de DTSTRING foi adicionado ao final do campo nome e este resultado foi armazenado no campo CLSS.

Após a preparação das variáveis, iniciaram-se as seis etapas do processo de relacionamento <sup>12</sup>, criando blocos de registros com os campos homólogos (que guardam informações de mesma natureza) e associação dos blocos combinando um ou mais campos, iniciando-se com uma

chave mais restrita e posterior diminuição da restrição (Quadro 1). Buscou-se, desta forma, minimizar a perda de pares, ou seja, a ocorrência de falso-negativo. Nos cinco primeiros passos manteve-se relação de uma AIH para uma DN. Apenas no sexto passo foi admitida relação de uma AIH para uma ou mais DN.

As funções seguintes foram definidas a fim de serem utilizadas nas chaves de blocagem: a) *Soundex* do nome da mãe: criada para reduzir a perda de pares verdadeiros devido a problemas decorrentes de erros e/ou diferenças de grafia; b) *Difdata*: criada para retornar em dias a diferença entre as datas da internação e do nascimento, já que no SIH o cadastro é por internação e pode-se ter várias internações do mesmo indivíduo no mesmo ano; esta função visava minimizar a possibilidade de formação de pares falso-positivo, não-verdadeiros; c) *Municnasc*: criada para comparar códigos de municípios de nascimento, retornando igual para os casos coincidentes, ou diferente para os não-coincidentes e visava principalmente reduzir a formação de pares não-verdadeiros, possibilitada pela presença de hospitais de mesmo nome em cidades diferentes; d) *Municres*: criada para comparar códigos de municípios de residência da mãe, retornando, igual ou diferente; visava resgatar pares verdadeiros, não-formados em passos anteriores; e) *Estab*: função criada para comparar os estabelecimentos de ocorrência do nascimento, retornando igual ou diferente.

Os relacionamentos foram feitos utilizando-se uma sentença em linguagem SQL denominada *datalink*<sup>12</sup>, que faz a comparação direta entre o *soundex* do primeiro nome nos dois bancos. O único campo empregado para relacionamento, em todas as chaves de blocagem, foi o campo CLSS, que foi submetido à função Distância de Levenshtein (LD) e o resultado desta, foi submetido a outra função, denominada Percentual de Similaridade (SP), que determinava o percentual de semelhança entre *strings* comparadas, variando de 0% a 100%<sup>12</sup>. Esses percentuais de semelhança foram armazenados em uma variável nomeada SCORE.

Entre cada um dos seis passos automatizados foi realizada a revisão manual dos pares formados, para sua admissão como pares verdadeiros e para verificação da adequação da mediana do campo SCORE como ponto de corte adotado. Quando se procedeu à revisão manual dos pares absolutamente concordantes, observou-se que os intervalos entre as internações das mães e os nascimentos eram, em sua maioria, inferiores a quatro dias. Mas, buscando-se incluir o maior número de pares verdadeiros foram admitidos intervalos maiores.

Assim, após as seis etapas de relacionamento, considerou-se haver subnotificação no SINASC quando uma internação para parto de nascido vivo registrada no SIH não havia sido relacionada a pelo menos um registro de nascido vivo do SINASC <sup>5</sup>.

### ANÁLISE ESTATÍSTICA

Com as bases de dados provenientes do DATASUS (comparação do número de eventos) e da SES/MG (relacionamento probabilístico) foram calculados os números totais de nascidos vivos e a razão entre os nascidos vivos do SIH e do SINASC segundo faixas populacionais dos municípios de residência. Embora não tenham sido encontrados critérios para se proceder à classificação dos municípios por faixas de população, os 853 municípios mineiros foram classificados segundo quatro faixas populacionais <sup>13</sup>: municípios com população menor que 5.000 habitantes (pequenos municípios); população entre 5.000 e 9.999 habitantes (porte intermediário 1); população entre 10.000 e 49.999 habitantes (porte intermediário 2) e população maior ou igual a 50.000 habitantes (grande porte). Optou-se por definir uma faixa de pequenos municípios (com população menor que 5000 habitantes) porque um grande número de municípios se incluía nesta categoria (n=246). Outros 547 municípios foram subdivididos em duas faixas semelhantes: uma com 272 municípios e população de 5.000 a 9.999 habitantes e outra com 275 municípios com 10.000 a 49.999 habitantes. E, para os 60 municípios com mais de 50.000 habitantes adotou-se a classificação (grande porte) proposta por Andrade & Szwarcwald <sup>14</sup>.

Para processamento e análise dos dados utilizou-se o programa SPSS 12.0. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (ETIC nº. 095/04).

### 4.3 RESULTADOS

Nas bases de dados provenientes do DATASUS e da SES/MG foram identificados no SINASC 298.515 e 293.213 nascidos vivos e no SIH 237.441 e 223.443 internações para partos de nascidos vivos, respectivamente.

#### COMPARAÇÃO DO NÚMERO DE EVENTOS

Nos municípios de pequeno porte foram encontrados 11.715 nascidos vivos notificados no SINASC e 12.258 partos de nascidos vivos registrados no SIH. Assim, foram obtidos 543

registros do SIH que não constavam no SINASC. Considerou-se que, após correção feita pelo acréscimo dos nascidos vivos captados apenas pelo SIH, haveria um total de 12.258 nascidos vivos neste grupo de municípios. Conseqüentemente, o percentual de nascidos vivos subnotificados no SINASC e conhecidos por meio de registros do SIH foi de 4,43%, ou seja, 543 em relação a um total corrigido no SINASC de 12.258 (Tabela 1).

Nos municípios de porte intermediário 1 foram encontrados 28.213 nascidos vivos notificados no SINASC e 28.912 partos de nascidos vivos no SIH. Assim, para pelo menos 699 nascidos vivos não havia notificação no SINASC, mas apenas no SIH. Da mesma forma, após correção, considerou-se haver 28.912 nascidos vivos nestes municípios e que o percentual de casos conhecidos através do SIH foi igual a 2,4%, ou seja, 699 em relação a um total corrigido no SINASC de 28.912 de nascidos vivos (Tabela 1).

Já nos municípios de porte intermediário 2 foram identificados 2.343 casos não notificados no SINASC, com registro no SIH. O total de nascidos vivos, calculado após correção, foi igual a 94.831 (92.488 do SINASC além dos 2.343 a mais, identificados no SIH). Neste grupo, o percentual de nascidos vivos conhecidos pelo SIH e não conhecidos pelo SINASC foi de 2,5%.

Finalmente, nos municípios de grande porte não foi identificada qualquer subnotificação no SINASC. Ao contrário, foram identificados menos notificações no SIH (101.440) do que no SINASC (166.099).

Assim, obteve-se, pela comparação do número de eventos, 3.585 partos de nascidos vivos do SIH não notificados entre os 302.100 nascidos vivos do SINASC. Pela definição de subnotificação aqui adotada foi igual a 1,2% o percentual de nascidos vivos subnotificados no SINASC e conhecidos apenas pelo SIH. Estas subnotificações ocorreram em 189 municípios, localizados principalmente nas regiões Norte de Minas e Jequitinhonha.

#### RELACIONAMENTO PROBABILÍSTICO

Pelo relacionamento probabilístico foram obtidos 193.259 pares unívocos de nascidos vivos - com relação de uma DN para uma AIH. Foram também obtidos 6.316 pares de gemelares, nos quais para 3.145 AIH foram relacionadas 6.316 DN, resultando, em média, em 2,01 DN para cada AIH.

A Tabela 2 indica a distribuição dos pares e não pares para as notificações do SIH e do SINASC, por faixa de população. Em relação ao SINASC, de um total de 293.213 DN, obteve-se ao menos um par para 199.565 notificações, o que resultou em um total de 68,0% de DN pareadas a uma AIH. No caso do SIH, das 223.443 AIH, foi possível obter par para 87,9% destas (n=196.404). Assim, para 27.039 AIH não foi possível formar par entre as DN disponíveis. Estes registros do SIH não pareados a registros do SINASC foram, pelo critério de subnotificação adotado neste estudo, considerados como subnotificações de nascidos vivos no SINASC.

Observou-se que nos municípios de pequeno porte havia 9.995 partos de nascidos vivos no SIH, que foram relacionados a 8.111 notificações entre os 11.277 nascidos vivos do SINASC. Ou seja, nesses municípios, 1.884 notificações de partos do SIH não formaram par com qualquer DN e foram considerados subnotificações do SINASC. Nos municípios com porte intermediário 1, dos 23.000 partos de nascidos vivos no SIH, 19.404 formaram par com 19.711 registros do SINASC. Assim, neste grupo de municípios, para 3.596 registros do SIH não se encontrou par dentre os registros do SINASC e, também neste caso, foram consideradas subnotificações no SINASC. Para os municípios de porte intermediário 2, não se encontrou par e foram consideradas subnotificações no SINASC em 10.689 registros de partos de nascidos vivos do SIH. Finalmente, nos municípios de grande porte, foram subnotificados no SINASC 10.870 registros do SIH.

Os resultados do relacionamento probabilístico por faixa de população residente apresentaram semelhanças nas distribuições de pares do SINASC e do SIH, mas diferenças entre os percentuais de não pares. Em relação ao SINASC, observou-se maior concentração de DN não pareadas nos municípios de grande porte. E, diferentemente, em relação ao SIH, observou-se maior concentração de AIH não pareadas nos municípios de portes intermediário e pequeno, especialmente (Tabela 2).

Com base nos registros do SIH não pareados foi calculado o número esperado de notificações de nascidos vivos no SINASC, por faixa de população. Obteve-se, por exemplo, um total corrigido de 13.161 nascidos vivos no SINASC nos municípios de pequeno porte, por relacionamento probabilístico (Tabela 2).

## DIFERENÇA ENTRE OS DOIS PROCEDIMENTOS

Observaram-se diferenças entre as proporções de nascidos vivos conhecidos por meio do SIH conforme o procedimento adotado. O relacionamento probabilístico identificou 8,4% e a comparação de eventos por município identificou 1,2% de subnotificações no SINASC. As subnotificações de nascidos vivos foram menores nos municípios de grande porte, com base nos dois procedimentos, sendo que, nesses municípios, as subnotificações foram identificadas apenas pelo relacionamento probabilístico (Tabela 3).

### 4.4 DISCUSSÃO

A principal motivação para a realização deste estudo foi a necessidade de explorar técnicas de mensuração direta das subnotificações de nascidos vivos por unidades geográficas cada vez mais desagregadas, como os municípios. A obtenção de informações de boa qualidade sobre nascidos vivos que permitam, por exemplo, o cálculo da taxa de mortalidade infantil de forma direta, é um passo fundamental para a adequação das políticas públicas de saúde com vistas à redução destas mortes, um dos Objetivos do Milênio <sup>15</sup>.

Dois resultados obtidos neste estudo merecem ser destacados. O primeiro refere-se à potencialidade, ainda pouco explorada, do SIH como fonte de captação de nascidos vivos para o SINASC, evidenciada em ambos os procedimentos aqui realizados. Este resultado é compatível com os de estudos que utilizaram estratégias de relacionamento de dados do SIH a outros sistemas de informação em saúde para avaliações de subnotificações de eventos e agravos <sup>6</sup>. O segundo refere-se às diferenças nos resultados obtidos, conforme o procedimento adotado. Assim, independentemente da faixa de população considerada, o relacionamento probabilístico identificou maior número de nascidos vivos subnotificados no SINASC, ao obter maior proporção de nascidos vivos conhecidos apenas pelo SIH. Uma possível explicação para a diferença observada entre os dois procedimentos diz respeito ao tipo de erro a que cada um dos procedimentos está mais sujeito. A comparação do número de eventos estaria mais sujeita a erros do tipo falso-positivo. Ao utilizar apenas a variável município de residência poder-se-ia, equivocadamente, considerar dois registros como pertencentes ao mesmo indivíduo – quando de fato eles não pertencem. Por outro lado, o relacionamento probabilístico, ao utilizar seis variáveis de relacionamento, seria relativamente mais sujeito a erros do tipo falso-negativo, ou seja, poder-se-ia, também de forma equivocada, considerar dois registros como não pertencentes ao mesmo indivíduo – quando de fato eles pertencem.

Assim, o maior número de registros encontrados no SIH pode não significar, necessariamente, maior proporção de subnotificação de nascidos vivos no SINASC.

A qualidade da informação registrada nos dois sistemas também pode afetar a formação de pares verdadeiros nos relacionamentos probabilísticos, pois a deficiência de qualidade aumenta a probabilidade de erros homônimos <sup>16</sup>, nos quais alguns pares podem ser classificados como verdadeiros quando na realidade as notificações referem-se a indivíduos diferentes. Nos casos, como o deste estudo, onde a base de dados na qual o par é procurado (SINASC) possui um elevado número de notificações, são sempre adotadas regras para minimizar o número de pares falso-positivo, e que podem, por outro lado, induzir ao aumento do número de resultados falso-negativo. Assim, parece razoável supor que a real magnitude da proporção de subnotificações seja maior que a observada com a comparação de eventos, mas inferior à observada com o relacionamento probabilístico. Desta forma, os dois métodos, se utilizados em conjunto, podem fornecer um limite superior e um inferior para a real magnitude da subnotificação.

Pela comparação do número de eventos, foi possível a identificação de pequenas proporções de subnotificações de nascidos vivos no SINASC, inferiores a 5%, apenas em 189 municípios de portes intermediários e pequenos. A ausência de subnotificação no SINASC em municípios de grande porte parece improvável <sup>17</sup>. Mas o procedimento apresenta vantagens como a facilidade para obtenção dos dados, a rapidez de sua execução e sua facilidade metodológica. Portanto, é factível mesmo quando se dispõe de poucos recursos técnicos. Entre as suas principais desvantagens estão a impossibilidade de se afirmar se se trata do mesmo indivíduo registrado em cada um dos SIS e a impossibilidade de identificação de subnotificações de nascidos vivos nos municípios com menores coberturas do SIH. Assim, com base nos resultados obtidos, infere-se que a comparação do número de eventos pode ter desempenho ótimo nos municípios onde a cobertura de partos da rede SUS é próxima ao total de nascimentos do município. As atividades de acompanhamento e supervisão dos sistemas de estatísticas vitais, necessárias especialmente neste grupo de municípios, podem se beneficiar do cálculo e utilização deste parâmetro.

Por meio do relacionamento probabilístico foi possível a identificação de nascidos vivos pelo SIH, não notificados ao SINASC, em 852 dos 853 municípios mineiros, incluída a capital. Obteve-se número sete vezes maior de nascidos vivos identificados apenas no SIH do que

pela comparação do número de eventos. O percentual de 12,1% de nascidos vivos subnotificados ao SINASC obtido pelo relacionamento probabilístico é bastante próximo daquele estimado pelo Ministério da Saúde <sup>18</sup> para Minas Gerais e corrobora a adequação dos resultados obtidos pelo relacionamento probabilístico. Também corroboram esta adequação a maior proporção de registros do SINASC não pareados a registros do SIH em municípios de grande porte, nos quais se concentra a população com melhores condições socioeconômicas e não usuária do SUS <sup>19</sup> e onde um maior percentual de partos é pago pela assistência complementar. Em Minas Gerais, especificamente, para o ano de 2006 a cobertura populacional da assistência complementar estimada é de 18,9%, sendo que na capital é de 45,5% e na região metropolitana de 31,6% <sup>19</sup>. Embora estejam excluídos do SIH os partos ocorridos em domicílio ou em trânsito, os quais são mais freqüentes entre os grupos socialmente menos favorecidos, a busca das subnotificações de nascidos vivos no SINASC por meio do SIH traz para o foco da análise a população usuária da rede SUS, entre a qual também se encontram os mais vulneráveis socialmente.

Entre as principais vantagens do relacionamento probabilístico está a considerável redução da incerteza do pareamento, gerando a possibilidade de detecção de subnotificações de nascidos vivos no SINASC, mesmo em municípios com baixa cobertura de partos pela rede SUS. Confirma-se a importância dos hospitais SUS como fontes notificadoras de nascidos vivos em todos os municípios, independentemente do tamanho de sua população. Mas, o relacionamento probabilístico também apresenta dificuldades. Entre elas, há os problemas referentes à completitude de variáveis e à utilização de códigos diferentes em cada um dos sistemas. Para a identificação do estabelecimento de saúde onde ocorreu o parto, por exemplo, cada um dos sistemas utiliza a sua própria forma. O Cadastro Nacional de Estabelecimentos de Saúde (CNES), em fase de implantação, apresenta grande potencial facilitador para o relacionamento de dados a partir do estabelecimento de ocorrência.

A variável município de ocorrência apresentou limitações ligadas à confiabilidade e completitude <sup>12</sup>. Porém, optou-se pela sua utilização, pois, em decorrência da chamada peregrinação das parturientes <sup>20</sup>, estas em muitos casos, por receio de não serem atendidas em município diferente daquele onde não residem, não declaram corretamente os seus municípios de residência. Poderia haver, assim, maior tendência de erro se a variável município de residência fosse a única a ser utilizada e justifica-se porque os códigos dos municípios de ocorrência estiveram em diferentes chaves de blocagem. A estratégia de relacionamento de

revisão manual de todos os pares formados teve como um dos objetivos a redução de prováveis falso-negativo. No entanto, este é um procedimento que requer critérios claramente definidos e que consome muito tempo para sua execução.

A análise de variáveis possivelmente relacionadas à subnotificação de nascidos vivos no SINASC identificada pelo relacionamento probabilístico foge ao escopo deste trabalho. Mas, entre os aspectos relacionados à subnotificação poderiam ser citados os problemas de entendimento de conceitos como aborto, óbito fetal e nascido vivo. Mello-Jorge e colaboradores<sup>20</sup> observaram que em casos de nascidos vivos considerados como pouco viáveis (de baixo peso, por exemplo) e nos casos de neomortos que vêm a óbito nos primeiros minutos pós-parto (asfixiados, por exemplo) houve aumento da subnotificação de nascidos vivos. Outro aspecto liga-se às evasões de partos, que podem resultar em falhas na retroalimentação de dados entre os municípios<sup>1</sup>.

Finalmente, um aspecto importante a ser considerado em estudos baseados em dados secundários é a possibilidade de viés de informação. Neste estudo os registros foram obtidos do SINASC e do SIH sem que tenha sido realizada validação independente. Em relação ao SINASC, estudos de avaliação das suas cobertura e completitude, disponíveis desde o início dos anos 90<sup>20</sup>, demonstram sua progressiva qualificação<sup>1,10</sup>. Veras & Martins<sup>21</sup> verificaram ser a confiabilidade dos dados do SIH superior àquela que lhe era imputada – inclusive para a variável “nascido vivo”. Novas análises de validade e confiabilidade dos dados do SIH contribuirão para ele possa ser cada vez mais utilizado como fonte de dados confiável e oportuna para pesquisa, planejamento e avaliação em saúde.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Mello Jorge MHP, Laurenti R, Gottlieb SLD. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC. *Ciênc & Saúde Coletiva* 2007; 12 (3): 643-654.
2. Indicadores Básicos para Saúde no Brasil (IDB 2006). Rede Interagencial de Informações para Saúde. Disponível em URL: <http://tabnet.datasus.gov.br/cgi/idb2006/c01.htm>. [2007 jun 16].
3. Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad Saúde Pública*. 2004; 20(2): 362-71.
4. Mello-Jorge MHP, Gottlieb SLD. O Sistema de Informação de Atenção Básica como fonte de dados para os Sistemas de Informações sobre Mortalidade e sobre nascidos Vivos. *I E SUS*. 2001; 10(10): 7-18.
5. Scharmm JMA, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e natimortalidade. *Rev Saúde Pública*. 2000; 34(3): 272-79.
6. Bittencourt SA, Camacho LAB, Leal MC. O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva. *Cad Saúde Pública*. 2006; 22(1)19-30.
7. Almeida MF, Alencar GP. Informações em saúde: necessidade de introdução de mecanismos de gerenciamento dos sistemas. *IESUS*. 2000; 9(4): 241-49.
8. Camargo Jr KR, Coeli CM. ReLink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. *Cad Saúde Pública*. 2000; 16(2): 439-47.
9. Machado CJ. Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. *Cad Saúde Pública*. 2004; 20(4): 915-25.
10. Almeida MF, Alencar GP, Novaes HMD, Ortiz LP. Sistemas de informação e mortalidade perinatal: conceitos e condições de uso em estudos epidemiológicos. *Rev Bras Epidemiol*. 2006; 9(1): 56-68.
11. Ministério da Saúde. SVS. Banco de dados dos Sistemas de Informações sobre Mortalidade (SIM) e Nascidos Vivos (SINASC), 1997 a 2003, 'CD-ROM'. Brasília: MS/SVS: 2005.
12. Drumond EF, Machado CJ, França E. SIHSUS e SINASC: utilização do método probabilístico para relacionamento de dados. *Cad Saúde Coletiva*. 2006; 14(2): 251-264.
13. Ministério da Saúde. DATASUS. Informações de Saúde - Demográficas e Socioeconômicas. Disponível em URL: (<http://tabnet.datasus.gov.br/cgi/defthtm.exe?ibge/cnv/poptmg.htm>.) [2007 jul 03]
14. Andrade CL, Szwarcwald CL. Socio-spatial inequalities in the adequacy of Ministry of Health data on births and deaths at the municipal level in Brazil, 2000-2002. *Cad. Saúde*

Pública 2007; 23(5): 1207-1216.

15. Organização Mundial de Saúde. Objectifs du Millénaire pour le développement.

Disponível em URL: <http://www.un.org/french/millenniumgoals/index.html> [2007 jun 26].

16. Coutinho ESF, Coeli, C M. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevida. *Cad. Saúde Pública* 2006; 22 (10): 2249-2252.

17. Perpétuo IHO, França-Mendonça E. Avaliação das estatísticas de nascimentos em Belo Horizonte 1974-1994. Encontro Nacional de Estudos Populacionais; 07 a 11 de outubro de 1996. Anais. Belo Horizonte: ABEP,1996.

18. Ministério da Saúde. Agência Nacional de Saúde Suplementar. Caderno de Informação de Saúde Suplementar: beneficiários, operadoras e planos. Disponível em URL: [www.ans.gov.br/portal/upload/informacoess/caderno\\_informaca\\_12\\_2006.pdf](http://www.ans.gov.br/portal/upload/informacoess/caderno_informaca_12_2006.pdf). [2007 jun 16].

19. Menezes DCS, Leite IC, Schramm JMA, Leal MC. Avaliação da peregrinação anteparto numa amostra de puérperas no Município do Rio de Janeiro, Brasil, 1999/2001. *Cad Saúde Pública*. 2006; 22(3): 553-59.

20. Mello-Jorge MHP, Gotlieb SLD, Soboll MLMS, Almeida MF, Latorre MRDO. Avaliação do Sistema de Informação sobre Nascidos Vivos e o uso de seus dados em epidemiologia e estatísticas de saúde. *Rev Saúde Pública*. 1993; 27 (Supl 6): 19-30.

21. Veras CMT, Martins M. A confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH), Rio de Janeiro, Brasil. *Cad Saúde Pública*. 1994; 10(3): 339-55.

Quadro 1: Blocos utilizados em cada passo de relacionamento dos bancos SIH e SINASC.

<b>Passo</b>	<b>Chave</b>
<b>1</b>	<i>Soundex</i> (primeiro nome) + Difdata + Municnasc+Estab
<b>2</b>	<i>Soundex</i> (primeiro nome) + Difdata + Municnasc
<b>3</b>	<i>Soundex</i> (primeiro nome) + Difdata + Municres
<b>4</b>	<i>Soundex</i> (primeiro nome) + Difdata
<b>5</b>	<i>Soundex</i> (primeiro nome)
<b>6 (gemelares)</b>	<i>Soundex</i> (primeiro nome)

Tabela 1: Distribuição absoluta e percentual dos registros do SINASC e SIH, subnotificações no SINASC e nascidos vivos corrigidos no SINASC segundo o procedimento de comparação do número de eventos por faixa de população dos municípios de MG. 2001

		<b>Faixas de população</b>				
		<b>&lt; 5.000</b>	<b>5.000 - 9.999</b>	<b>10.000 - 49.999</b>	<b>&gt;50.000</b>	<b>Total</b>
<b>SINASC</b>						
<b>n (1)</b>		11.715	28.213	92.488	166.099	298.515
<b>%</b>		4,1	9,6	31,4	55	100
<b>SIH</b>						
<b>n (2)</b>		12.258	28.912	94.831	101.440	237.441
<b>%</b>		5,2	12,2	39,9	42,7	100
<b>Subnotificação SINASC</b>						
<b>n (3)</b>		543	699	2.343	0	3.585
<b>%</b>		15,2	19,5	65,4	0	100
<b>SINASC corrigido</b>						
<b>n (1+3)</b>		12.258	28.912	94.831	166.099	302.100
<b>%</b>		3,9	9,5	31	100	100

Fonte: DATASUS

Tabela 2: Distribuição absoluta e percentual dos pares e dos não pares de registros do SINASC e SIH e dos nascidos vivos corrigidos no SINASC segundo o relacionamento probabilístico por faixa de população dos municípios de MG. 2001

	<b>Faixas de população</b>				<b>Total</b>
	<b>&lt; 5.000</b>	<b>5.000 - 9.999</b>	<b>10.000 - 49.999</b>	<b>&gt;50.000</b>	
<b>Pares SINASC</b>					
n (1)	8.253	19.711	65.101	106.510	199.404
%	4,1	9,9	33	53	100
<b>Pares SIH</b>					
n (2)	8.111	19.404	64.119	104.770	196.404
%	4,1	9,9	32,6	53,3	100
<b>Não pares SINASC</b>					
n (3)	3.024	7.499	25.092	58.023	93.638
%	3,2	8	26,8	62	100
<b>Não pares SIH</b>					
n (4)	1.884	3.596	10.689	10.870	27.039
%	7	13,3	39,5	40,2	100
<b>SINASC corrigido</b>					
n (1+3+4)	13.161	30.806	100.882	175.403	320.252
%	4,1	9,6	31,5	54,8	100

Fonte: SES/MG

Tabela 3: Distribuição dos números absolutos e dos percentuais de nascidos vivos subnotificados no SINASC e de nascidos vivos estimados no SINASC após correção segundo faixa populacional dos municípios de MG. 2001

	<b>Faixa populacional</b>				<b>Total</b>
	<b>&lt;5.000</b>	<b>5.000-. 9.999</b>	<b>10.000-49.999</b>	<b>&gt;50.000</b>	
<b>NV não notificados</b>					
Comparação de eventos	543	699	2.343	0	3.585
Probabilístico	1.884	3.695	10.689	10.870	27.039
<b>NV notificados</b>					
Comparação de eventos	11.715	28.213	92.488	166.099	298.515
Probabilístico	11.277	27.210	90.193	164.533	293.213
<b>Total após correção</b>					
Comparação de eventos	12.258	28.912	94.831	166.099	302.100
Probabilístico	13.161	30.305	100.882	175.403	320.252
<b>Percentual de NV em relação ao total corrigido</b>					
Comparação de eventos	4,4	2,4	2,5	0	1,2
Probabilístico	14,3	11,7	10,6	6,2	8,4

Fonte: SES/MG e DATASUS

5 ARTIGO 3: ÓBITOS NEONATAIS PRECOSES: ANÁLISE DE CAUSAS  
MÚLTIPLAS PELO MÉTODO DE *GRADE OF MEMBERSHIP* (GoM)

Publicado nos Cadernos de Saúde Pública

## RESUMO

Estudo de base populacional com objetivo de determinar perfis de óbitos neonatais precoces ocorridos em Belo Horizonte, Minas Gerais, Brasil, de 2000 a 2003. A definição dos perfis de causas amplia a possibilidade de análise dos resultados sob enfoque de evitabilidade, justificada pela persistência de altas taxas de mortalidade neonatal precoce no município.

Três perfis foram gerados, a partir do enfoque de causas múltiplas, sob a perspectiva dos conjuntos nebulosos (*fuzzy sets*), por meio da técnica de GoM (*Grade of Membership*). Esses perfis foram então relacionados ao peso de nascimento e a natureza jurídica do hospital de ocorrência do óbito.

Nos hospitais da rede privada ocorreram os “óbitos dificilmente preveníveis, com menção de malformação congênita” (perfil 2). Já aos hospitais da rede do Sistema Único de Saúde (SUS) se associaram dois perfis distintos. Nos hospitais contratados/conveniados ocorreram os chamados “óbitos passíveis de prevenção” (perfil 1) e na rede própria os “óbitos de prematuros” (perfil 3).

A tipologia observada aponta para a necessidade de adoção de políticas diferenciadas na rede SUS: prioritariamente, na rede contratada/conveniada, ações voltadas para credenciamento e avaliação da qualidade da assistência; e, em toda rede, adoção rotineira de protocolos assistenciais e medidas profiláticas, redutores da morbimortalidade neonatal.

Palavras-chave: mortalidade infantil, evitabilidade, causas múltiplas, Grade of Membership (GoM).

## ABSTRACT

Population-based study aimed at determining profiles of early neonatal deaths occurred in Belo Horizonte, Minas Gerais, Brazil, from 2000 to 2003. The generated profiles were analyzed under the perspective of avoidability, justified by the persistence of high levels of early neonatal mortality in the city.

Three profiles were generated from multiple death causes, under the perspective of fuzzy sets, by the Grade of Membership (GoM) method. Birth weight and hospitals' legal statuses were also related to the three profiles. In private hospitals occurred the so-called "hardly preventable deaths, with congenital malformation" (profile 2). The *Brazilian National Unified Health System (SUS)* generated two distinct profiles. In private maternity centers contracted out by the SUS occurred the "preventable deaths" (profile 1); "premature deaths" (profile 3) occurred in the stratum consisting of public Federal and State maternity hospitals.

This typology elucidates the necessity of adopting differentiated policies in the *Brazilian National Unified Health System (SUS)*, which should focus, in the case of the maternity centers contracted out by the SUS, on evaluation and credential; and throughout the system, on the routine adoption of protocols of assistance and prophylactic measures, known to reduce both neonatal morbidity and mortality.

Keywords: infant mortality, avoidability, multiple causes of death, Grade of Membership (GoM).

## 5.1 INTRODUÇÃO

O Brasil é signatário dos Objetivos de Desenvolvimento do Milênio (ODM), compromisso proposto às nações pela Organização Mundial de Saúde no ano 2000 <sup>1</sup>. A redução da taxa de mortalidade infantil (TMI) até o ano 2015, uma das metas propostas pelo ODM, depende essencialmente da redução do componente neonatal precoce, atualmente responsável por mais da metade dos óbitos de crianças brasileiras no primeiro ano de vida <sup>2</sup> e estreitamente ligado a problemas na atenção à saúde da gestante e do recém-nascido <sup>3</sup>. No entanto, na complexa rede causal de determinação destes óbitos estão também envolvidos outros fatores, diretos e indiretos. Dentre esses, o peso de nascimento é a variável direta mais consistentemente associada a maior risco de morte neonatal e deve ser considerada nos estudos sobre mortalidade infantil <sup>4</sup>. A desigualdade socioeconômica, principal fator determinante distal, tem sido relacionada a dificuldades de acesso e pior qualidade de atenção em saúde perinatal <sup>5</sup>. A adequação dos processos assistenciais pode ser avaliada por sua capacidade de interferir e reduzir os riscos de morte neonatal precoce, diretos ou indiretos. Esses processos assistenciais vão desde o atendimento individual até a organização, de forma hierarquizada e regionalizada, da rede de assistência à gestante, ao parto e ao recém-nascido. O elevado percentual de mortes nas primeiras horas de vida extra-uterina, em locais onde virtualmente todos os óbitos são hospitalares é, portanto, um indicador da importância de assistência hospitalar oportuna e adequada <sup>6</sup> ao binômio mãe-filho.

A abordagem segundo causas de morte é um dos mais importantes aspectos das análises de mortalidade. As causas básicas de morte utilizadas na divulgação das estatísticas de mortalidade são de fácil compreensão, mas podem representar significativa perda de informações, se comparadas à abordagem de causas múltiplas <sup>7</sup>. No caso específico dos óbitos neonatais precoces, esta abordagem pode ampliar a compreensão dos complexos patológicos que envolveram os binômios mãe-filho na gravidez e no parto. E também, ao possibilitar a avaliação da presença de exposição simultânea do recém-nascido a mais de um fator de risco, auxiliar as abordagens sob enfoque de evitabilidade <sup>8</sup>, especialmente aplicáveis no Brasil <sup>9</sup>.

A Conferência Internacional para a Décima Revisão da Classificação Internacional de Doenças <sup>10</sup>, ao expressar seu estímulo à utilização da abordagem segundo causas múltiplas, não recomendou qualquer regra particular ou método de análise a serem seguidos. Os

métodos devem se basear no pressuposto de que cada óbito é consequência de um conjunto de causas de morte, e não apenas de uma causa básica. A partir da teoria dos conjuntos nebulosos (“fuzzy sets”) <sup>11</sup> foi desenvolvido o método GoM ou “*Grade of Membership*”, que permite a análise de todas as causas de óbito mencionadas. Por esta lógica, pode-se assumir que uma população caracteriza-se por perfis extremos, representados por um grupo de causas de morte. Contudo, nem todos os indivíduos possuem exatamente as mesmas causas de morte do perfil extremo e isto é o mesmo que dizer que nem todos os indivíduos possuem inclusão – ou pertencimento – total a um dos perfis extremos característicos daquela população. Se um indivíduo tem pertencimento total a um perfil extremo, ele passa a ser considerado um tipo puro deste perfil. Para atender à diversidade de possibilidades de não pertencimento total a um perfil extremo assume-se, pelo método de *GoM*, que um indivíduo pode apresentar características de mais de um perfil e que é possível quantificar o seu grau de pertencimento a cada um dos perfis traçados para a população em estudo.

No município de Belo Horizonte, a redução da taxa de mortalidade neonatal precoce tem sido irregular e ela se mantém em níveis considerados ainda elevados (taxa de 7,4 por mil nascidos vivos em 2003) <sup>12</sup>. A importância de se abordar este componente da TMI, onde há boa cobertura pré-natal e virtualmente todos os partos são hospitalares, foi evidenciada pelo elevado percentual de óbitos por causas consideradas evitáveis <sup>13</sup>, atribuídos especialmente a hipóxia <sup>14</sup> no município. Neste sentido, o objetivo deste trabalho é descrever e analisar os perfis dos óbitos neonatais precoces ocorridos em Belo Horizonte, no período de 2000 a 2003, utilizando o enfoque de causas múltiplas de morte a partir da teoria dos conjuntos nebulosos, por faixa de peso e natureza jurídica do hospital.

## 5.2 METODOLOGIA

Analisaram-se inicialmente todos óbitos neonatais precoces de residentes em Belo Horizonte, ocorridos no período de 2000 a 2003, obtidos no Sistema de Informações sobre Mortalidade (SIM) <sup>15</sup> da secretaria de saúde do município. O Sistema de Informação sobre Nascidos Vivos (SINASC) foi utilizado como fonte para resgate de pesos de nascimento em branco e para validação dos pesos registrados do SIM. Todos os óbitos excluídos tiveram peso de nascimento inferior a 500 gramas. A codificação das causas do óbito baseou-se na Décima Revisão da Classificação Internacional de Doenças (CID10) <sup>10</sup>. Foram consideradas causas múltiplas todas as causas mencionadas em cada uma das declarações de óbito, sem distinção

da sua posição.

As variáveis utilizadas para determinação dos perfis de causas de morte foram as causas mencionadas nas declarações de óbito. Posteriormente, sobre os perfis determinados, foram relacionados as frequências de pesos de nascimento e as naturezas jurídicas dos estabelecimentos de ocorrência do óbito. Para análise da frequência dos diagnósticos e da relação de suas menções como causa básica ou associada, todas as afecções mencionadas na declaração de óbito foram consideradas e arquivadas em campos distintos, de forma que cada linha reproduziu o conteúdo de uma declaração de óbito. Em seguida, foram excluídas as causas que, na mesma declaração de óbito, foram mencionadas mais de uma vez (duplicações) e as que eram semelhantes (multiplicações). Buscava-se, também, através desta avaliação inicial, que o maior conhecimento do padrão de causas de óbito da população possibilitasse a elaboração de uma lista reduzida dos grupos de causas, com as seguintes características: 1) ser definida a partir dos códigos da CID10; 2) refletir a frequência e relevância dos diagnósticos mencionados nos atestados de óbito; 3) ter uma denominação a mais informativa possível sobre as diversas patologias e os códigos nela incluídos; 4) permitir a comparabilidade com outros dados e, 5) não conter número excessivo de categorias. A partir destas premissas e de adequação de proposta internacional para definição de grupos de causas<sup>16</sup>, foram obtidos seis grupos: *imaturidade* (códigos P00 a P15), *hipóxia* (códigos P20 a P21), *outras respiratórias* (códigos P22 a P28), *infecciosas* (códigos A50 a A64 e P35 a P39 e P77), *malformações* (códigos Q00 a Q99) e *demais causas* (códigos P29 e P50 a P96).

Para simplificar, os grupos de causas assim definidos e que compuseram a lista reduzida de causas serão chamados de causas de morte. Para evitar a multiplicação de diagnósticos causada pela agregação de causas, as duplicações foram eliminadas. Cada causa de morte recebeu, então, o código “0” ou “1” em cada uma das declarações de óbito. O código “1” indicou a menção de uma das causas de óbito daquele grupo de causas naquela declaração de óbito. Nos casos de declarações onde houve ausência de menção do grupo, o código foi igual a “0”.

Para o peso de nascimento foram definidas três categorias: menor que 1500 gramas; de 1500 gramas a 2499 gramas e igual ou superior a 2500 gramas.

Os hospitais de ocorrência dos óbitos foram categorizados, segundo sua natureza jurídica,

como privados, públicos/universitários e contratados/conveniados ao Sistema Único de Saúde (SUS). Foram classificados na categoria privados, nove estabelecimentos de saúde que tinham como fontes de financiamento os pagamentos particulares por serviço e aqueles advindos dos chamados convênios de pré-pagamento. Entre os dezesseis hospitais categorizados como contratados/conveniados, estão os privados ou filantrópicos e que prestam serviço, sob forma de pós-pagamento, ao SUS. Na categoria públicos/universitários, foram consideradas seis unidades (municipais, estaduais e federais) que compõem a rede própria do SUS de Belo Horizonte (SUSBH).

O método de *GoM*, um procedimento multivariado, foi utilizado para examinar perfis latentes de associação entre as causas de morte, para caracterizar em que medida os indivíduos se assemelham a estes perfis. Ele possibilita que as causas de morte de cada indivíduo sejam descritas de acordo com o grau de similaridade ou pertencimento aos aglomerados de causas comuns (ou típicos) encontrados<sup>17</sup>. Os graus de similaridade entre aglomerados de causas e indivíduos somam a unidade para cada indivíduo e variam entre 0 e 1, necessariamente<sup>18</sup>.

O ponto de partida do método é assumir a existência de  $K$  conjuntos nebulosos (*fuzzy sets*) ou perfis extremos a serem definidos, tal que  $k=1, 2, 3, 4, \dots, K$ <sup>17,18</sup>. Cabe notar que estes conjuntos são chamados difusos por permitirem graus de pertencimento que podem ser parciais<sup>11</sup>. A população estudada compreende  $I$  indivíduos ( $I=1, 2, 3, \dots, I$ ), com  $J$  variáveis categóricas, onde a  $j$ -ésima variável possui  $l_j$  níveis de resposta<sup>17,18</sup>. Assume-se ainda que  $\lambda_{kjl}$  é a probabilidade de que a  $l$ -ésima resposta à variável  $j$ -ésima esteja associada com o  $k$ -ésimo conjunto nebuloso e que  $g_{ik}$  é o parâmetro que mensura o grau de similaridade das causas múltiplas de morte específicas de um dado indivíduo em relação às características de cada um dos  $K$  perfis extremos, com as seguintes restrições:

$$0 \leq \lambda_{kjl} \leq 1,$$

$$\sum_k \lambda_{kjl} = 1,$$

$$0 \leq g_{ik} \leq 1,$$

$$\sum_k g_{ik} = 1$$

Assim, observa-se que o primeiro tipo de coeficientes,  $\lambda_{kjl}$  relaciona os  $K$  tipos puros às  $J$  variáveis observadas, enquanto que o segundo tipo de coeficientes,  $g_{ik}$ , descreve a distância ou similaridade entre as características observadas de um indivíduo em termos de causas múltiplas e as características de cada perfil extremo  $K$ <sup>17,18</sup>.

O produto  $g_{ik}\lambda_{kjl}$  revela a probabilidade de que um indivíduo que pertence inteiramente ao perfil  $k$  (ou seja, apenas aqueles que se associam unicamente a um perfil extremo ou tipo puro) possua a resposta  $l$  para a questão  $j$ . A probabilidade que um indivíduo tenha uma resposta  $l$  à questão  $j$ ,  $p(x_{ijl}) = 1$ , é a soma deste produto para todos os perfis de *GoM* gerados ( $K$  perfis)<sup>17,18</sup>. Esta probabilidade é dada por

$$P(x_{ijl} = 1,0) = \sum_k g_{ik} \lambda_{kjl}$$

E, finalmente, assumindo independência das observações individuais, condicionalmente a  $\lambda_{kjl}$  e  $g_{ik}$ , a função de verossimilhança fica assim definida:

$$L = \prod_i \prod_j \prod_l \left( \sum_k g_{ik} \lambda_{kjl} \right)^{x_{ijl}}$$

Essa função fica sujeita às seguintes restrições:

$$0 \leq \lambda_{kjl} \leq 1,$$

$$\sum_k \lambda_{kjl} = 1,$$

$$0 \leq g_{ik} \leq 1,$$

$$\sum_k g_{ik} = 1$$

Os parâmetros  $g_{ik}$  e  $\lambda_{kjl}$  foram estimados, de forma iterativa, utilizando o software GOM3.EXE, executável em ambiente DOS. O programa encontra-se disponível na internet (<http://www.stat.unipg.it/stat/statlib/DOS/general/>). Estimou-se o modelo em dois níveis<sup>19</sup>. No primeiro nível, estimaram-se os parâmetros  $\lambda_{kjl}$  e  $g_{ik}$  a partir das causas de morte (variáveis internas), ou seja, excluindo as variáveis natureza do estabelecimento e peso (variáveis externas). Em seguida, a matriz de  $g_{ik}$  foi fixada e um novo modelo ajustado, com novos valores de  $\lambda_{kjl}$  obtidos para as variáveis natureza do estabelecimento e peso ao nascer (variáveis externas ao modelo). Assim, as causas de morte fazem parte da conformação dos perfis e dos graus de pertencimento, enquanto que as estimativas de  $\lambda_{kjl}$  obtidas para as variáveis remanescentes foram empregadas como uma aproximação de variáveis de estratificação de cada um dos perfis extremos<sup>17,18</sup>. A idéia é que se deseja conhecer as probabilidades de que cada um dos indivíduos tipos puros de um perfil extremo  $K$  tenham um nível  $l$  de uma variável externa  $J$ . Cabe comentar que as categorias de peso ao nascer e de natureza do estabelecimento foram definidas como variáveis externas porque o objetivo do estudo é entender como as causas múltiplas aparecem de forma conjunta. Assim, desejava-se encontrar "fuzzy sets" para os quais os graus de pertencimento pudessem ser especificados

apenas em termos das variáveis internas (as condições mórbidas no momento da morte). As variáveis externas são aquelas que, ainda que relacionadas aos conjuntos formados, são de outra natureza e não entram na conformação dos graus de pertencimento. A utilização de uma ou mais variáveis externas se justifica como forma de validação dos perfis encontrados<sup>20</sup>. Quanto à variável sexo, deve ser ressaltado ainda que, nesse estudo, a não utilização desta variável foi uma decisão tomada, também, com base nos seguintes aspectos: não ter havido diferença na taxa de mortalidade entre os sexos, por faixa de peso, conforme tabulações preliminares dos dados realizadas pelos autores; pelo fato de os critérios de evitabilidade comumente adotados não considerarem essa variável<sup>21</sup> e, tendo em vista que neste estudo se propôs, principalmente, a discussão de variáveis assistenciais envolvidas na determinação dos óbitos. Assim, a abordagem dos mecanismos biológicos ultrapassaria seu escopo.

O número ótimo de perfis pode ser determinado tanto por quão interpretável é um determinado conjunto de K perfis, do ponto de vista teórico e conceitual, quanto por um critério estatístico – o teste da razão de verossimilhança, que mede a variação no poder explicativo entre os modelos com K e K+1 perfis extremos. Esta razão segue uma distribuição qui-quadrado, com número de graus de liberdade igual à diferença no número de parâmetros a serem estimados entre os modelos<sup>18</sup>

$$df = (I - 1) + \sum_j L_j$$

A hipótese nula, ao proceder ao teste da razão de verossimilhança, é que o um modelo com K+1 perfis não adiciona poder explicativo a um modelo com K perfis.

### 5.3 RESULTADOS

Do total de 998 óbitos neonatais precoces ocorridos no período, foram selecionados 875 para inclusão no estudo. Nos hospitais da rede privada ocorreram 11% (n=92), na rede contratada/conveniada 43% (n=379) e nos hospitais da categoria público/universitário 46% (n=404) dos óbitos. Em relação ao sexo e ao peso, 53,3% (n=467) dos óbitos foram de recém-nascidos do sexo masculino e 59,2% (n=518) apresentaram peso ao nascer inferior a 1500 g. Encontraram-se 2481 menções de causas de morte, com variação entre uma e seis causas de morte por declaração e média de 2,8.

De acordo com o método de *GoM* e os critérios de perfis interpretáveis e estatístico, concluiu-

se que três perfis representavam adequadamente o panorama de co-morbidade, a partir das causas de morte mencionadas. A Tabela 1 mostra as relações entre os perfis extremos e as variáveis indicadoras de presença das causas de morte através das probabilidades  $\lambda_{kjl}$ . Nas primeiras colunas estão o número e a proporção de recém nascidos que faleceram com e sem menção de cada uma das causas. As demais colunas são as associadas com cada perfil extremo especificamente e revelam  $\lambda_{kjl}$ , a probabilidade de que a presença ou ausência da causa de morte esteja associada com um determinado perfil extremo. Os perfis podem ser descritos através da comparação da diferença em valores entre as frequências relativas da população de recém nascidos em estudo na primeira coluna com os  $\lambda_{kjl}$  nas demais colunas. As últimas duas variáveis apresentadas são as variáveis externas ao modelo, com suas respectivas probabilidades de que a presença ou ausência da causa múltipla esteja associada com um determinado perfil extremo, ou seja,  $\lambda_{kjl}$ .

Em relação aos perfis, observa-se que os recém nascidos tipos puros do perfil extremo 1 ou, em outras palavras, com grau de pertencimento total ao Perfil 1 não possuíam menção de prematuridade e nem de malformação congênita. Tenderam a ter mais menções de hipóxia, de outras respiratórias, de doenças infecciosas e de demais afecções, tanto em relação a qualquer recém-nascido tomado ao acaso desta população quantos em relação aos tipos puros de qualquer dos outros dois perfis. Denominou-se este perfil de “*óbitos passíveis de prevenção*”.

Os recém-nascidos com grau de pertencimento total ao Perfil 2 não possuíam menção de hipóxia. Todos tiveram menção de malformação congênita. Tenderam a ter um menor número de menções de outras afecções respiratórias e infecciosas, relativamente à população como um todo e aos tipos puros dos demais perfis extremos. Devido à presença de malformação congênita em todos os tipos puros deste perfil, este foi denominado de “*óbitos dificilmente preveníveis, com menção de malformação*”.

Finalmente, os recém nascidos com grau de pertencimento total ao Perfil 3 eram invariavelmente prematuros. Não possuíam menção de malformação congênita. Tenderam a ser hipóxicos e a ter outras afecções respiratórias. Este perfil foi denominado de “*óbitos de prematuros*”.

Observa-se, a seguir, que todos os “*óbitos passíveis de prevenção*”, dos tipos puros do Perfil 1, tinham peso de nascimento acima de 2500 gramas e ocorreram em hospitais contratados/conveniados. Já os tipos puros do Perfil 2 tenderam a ser óbitos ocorridos em

hospitais privados ou em hospitais públicos, em detrimento de hospitais contratados/conveniados. Assim, interpreta-se que na população em estudo 11% dos óbitos ocorreram em hospitais privados, 43% em hospitais contratados/conveniados e 46% em hospitais públicos, mas se um recém-nascido é um tipo puro do Perfil 2, o óbito tinha 32% de probabilidade de ter ocorrido em hospital privado e 68% de probabilidade de ter ocorrido em hospital público, sendo nula a probabilidade de ter ocorrido em hospital contratado/conveniado. Todos os tipos puros deste perfil extremo nasceram com peso entre 1500 gramas e 2499 gramas. Finalmente, os tipos puros do Perfil 3 tiveram maior probabilidade de ocorrer em hospitais públicos, em detrimento tanto dos privados, quanto dos contratados/conveniados. Todos os óbitos foram de recém-nascidos cujo peso ao nascimento foi inferior a 1500 gramas.

A Tabela 2 mostra a distribuição dos escores  $g_{ik}$  na população estudada, para cada perfil. Observa-se que 48,4% (n=424) dos óbitos infantis teve pertencimento total ao Perfil 3, o que implica nas seguintes probabilidades, associadas aos tipos puros do Perfil 3 (Tabela 1): probabilidade de 100% de menção de prematuridade e probabilidade nula de menção de malformação congênita; probabilidade de 50% de possuir menção de hipóxia, 43% de doenças infecciosas e de demais afecções e 54% de outras afecções respiratórias. Em termos das variáveis externas ao modelo, estes recém-nascidos apresentam 7% de probabilidade de terem nascido em hospital privado, 37% de probabilidade de terem nascido em hospital contratado/conveniado e 56% de terem nascido em hospital público. Além disso, há uma probabilidade de 100% destes recém nascidos terem peso inferior a 1500 gramas. Ou seja, as características do Perfil 3 prevalecem totalmente em 424 recém-nascidos falecidos. Mas há prevalências parciais do Perfil 3 sobre os demais recém nascidos: 15,3% (n=134) dos recém nascidos possuem um pertencimento alto, mas não total, ao perfil (escores GoM entre 0,75 e 0,99) e 32,8% (n=287) apresentam pertencimento nulo ao Perfil 3, possuindo características apenas de um ou outro dos outros dois perfis (ou de ambos). Já o Perfil 1, por sua vez, concentra o maior número de recém nascidos com pertencimento nulo ao perfil, pois 81,3% (n= 711) dos óbitos têm pertencimento nulo ao Perfil 1. No caso do Perfil 2, observa-se que 61,7% dos recém nascidos têm pertencimento nulo ao perfil, o que equivale a 540 óbitos. Calculou-se também, com base nos pertencimentos totais, parciais e nulos dos recém-nascidos a cada um dos perfis, uma média ponderada dos graus de pertencimento dos indivíduos a cada um dos perfis, onde os pesos são dados pelo grau de pertencimento dos óbitos, o que equivale a um escore médio para cada um dos perfis. Pode-se dizer que esses escores médios

corresponderam, então, à prevalência ponderada de cada um dos perfis na população e que, desta forma, o Perfil 3 é o de maior prevalência, seguido pelo Perfil 2 e pelo Perfil 1 (Tabela 2).

#### 5.4 DISCUSSÃO

Análises de causas básicas de óbitos neonatais precoces, comumente classificadas segundo grandes agrupamentos como as afecções perinatais e as malformações congênitas, podem ser pouco elucidativas por não se relacionarem à possibilidade de sua prevenção<sup>5,10</sup>. A utilização de todas as causas mencionadas nas declarações de óbito minimiza dificuldades, muitas vezes relatadas pelos médicos, na identificação da melhor seqüência que descreveria a sucessão de eventos que culminou na morte. Essas dificuldades para a ordenação das causas têm reflexo sobre as regras de seleção de causas básicas, internacionalmente adotadas, que levam em conta a posição nas quais as afecções foram citadas. Por isto, é possível que não sejam selecionadas como causas básicas, após aplicação das regras de seleção, aquelas causas que seriam eleitas pelo médico declarante.

Para Santo & Pinheiro<sup>22</sup> a análise do número médio de diagnósticos mencionados na declaração de óbito é de fundamental importância já que, além de outros aspectos, poderá revelar tanto a qualidade da assistência médica recebida como a qualidade do preenchimento da declaração. Neste estudo, após a exclusão das duplicidades e multiplicações de causas, verificou-se uma média de 2,8 diagnósticos por declaração de óbito, que é comparável às médias obtidas para as doenças crônico-degenerativas<sup>23</sup>, para as quais a abordagem de causas múltiplas tem sido mais freqüentemente adotada. Por outro lado, em estudo de confiabilidade da causa básica de óbito realizado em 13 cidades da região metropolitana de Belo Horizonte, os autores observaram que havia subestimação do agrupamento relativo às afecções maternas<sup>24</sup>. Nesse agrupamento se incluem causas de morte que são passíveis de prevenção por adequada atenção pré-natal. Assim, sob o enfoque da evitabilidade, os resultados atualmente obtidos poderiam estar subestimando este agrupamento de causas. Portanto, a qualificação dos dados de mortalidade e em especial das causas de morte, uma das ações prioritárias dos gestores dos sistemas de informação em saúde, é ferramenta básica para análises da situação de saúde de uma população.

A utilização da teoria dos conjuntos nebulosos (*fuzzy sets*), através da aplicação da metodologia do *GoM* sobre os grupos de causas de óbitos, com a posterior agregação do peso de nascimento e da natureza jurídica dos hospitais de ocorrência a esses perfis, possibilitou a maximização da utilização das informações contidas nas declarações de óbito. Assim, delinear-se três distintos perfis de causas, com grandes diferenças nos padrões de frequência de patologias que levaram a morte e aos quais os recém-nascidos podiam ter diferentes graus de pertencimento. A incorporação das informações relativas as variáveis externas “peso de nascimento” e “natureza jurídica do estabelecimento de saúde”, tornou ainda mais nítidas as diferenças entre os três perfis e também foi útil no sentido de validar os perfis de causas encontrados. A título de exemplo, observa-se que um dos perfis encontrados foi definido como sendo de prematuros (todos os óbitos com pertencimento total a este perfil eram de prematuros). Dado que os prematuros tendem a ter baixo peso, o fato de tipos puros deste perfil terem sido todos de menos de 1500 gramas ao nascimento é uma evidência de que o perfil encontrado é condizente com o esperado.

Para análises de causas múltiplas há principalmente duas propostas <sup>22,25</sup>. A primeira refere-se às análises da frequência de cada diagnóstico como causa básica ou associada, mais facilmente acessíveis em termos de saúde pública. Na outra proposta é ainda de uso restrito. São os estudos de associações de causas nos quais se parte da premissa de interação entre as causas. Ainda não há conceituação e procedimentos para codificação e tabulação internacionalmente padronizadas e os métodos de análise baseados na independência das variáveis não se aplicam às análises de associações de causas. A opção aqui adotada, o método de *GoM*, permitiu a sistematização das causas múltiplas de morte em conjuntos de enfermidades que se associavam entre si, desdobrando-se em manifestações de doenças que conduziram à morte. Observou-se com mais clareza e simplicidade os conjuntos de enfermidades que levaram ao óbito. Para abordagens em Saúde Pública é mais elucidativo entender como a confluência de doenças compõem mecanismos de causação de mortes e, uma vez conhecidas estas etapas, tornam-se mais evidentes as medidas a serem tomadas para evitar o óbito. Esta abordagem vai além de análises conhecidas. Podem-se citar análises baseadas no cálculo do “Q de Yule” para verificação da magnitude da associação de causas (básicas e associadas) <sup>26, 27</sup>, que parte da pressuposição de ocorrência independente das doenças, o que claramente não é verdadeiro no caso das causas múltiplas para um indivíduo.

Através do método de *GoM* também foi possível definir para cada óbito seu grau de pertencimento a cada um dos conjuntos formados. Isso indicou a menor ou maior prevalência dos perfis na população. É importante frisar que a obtenção de parâmetro individual como este implica no reconhecimento das diferenças entre indivíduos e dos diferentes caminhos e associações de mesma causa de morte. Ou seja, admite-se que há uma heterogeneidade mensurável na população em estudo.

É importante ressaltar também que são encaminhadas gestantes, parturientes e recém-nascidos de maior risco para os ambulatórios e hospitais públicos da rede SUS de Belo Horizonte. Esses hospitais são também referências para tratamentos de infertilidade e patologias fetais. Já os partos de risco habitual são mais freqüentemente realizados na rede contratada/conveniada. Chamam atenção, especialmente, as características dos recém-nascidos pertencentes ao Perfil 1 (n=164), de “*óbitos passíveis de prevenção*”: pesos de nascimento maiores que 2500 gramas, elevada probabilidade de ocorrência de hipóxia, todos ocorridos em hospitais contratados/conveniados. As freqüentes associações de menções de infecção e demais causas não foram suficientes para elucidar a etiologia dos quadros de hipóxia, e não contribuíram para melhor localização das possibilidades de falhas, se na integração da rede de assistência à gestante e ao recém-nascido ou na identificação e minimização de riscos de morte. Vale ressaltar que óbitos em maiores de 2500 gramas são considerados eventos-sentinelas em saúde, relacionados a prováveis falhas no manejo obstétrico ou na assistência neonatal <sup>21</sup>. Problemas no atendimento à parturiente podem ser devidos a dificuldades de acesso tempestivo à assistência obstétrica, em consequência da falta de percepção da gravidade do caso pela parturiente e sua família ou pela já conhecida peregrinação em busca de assistência obstétrica qualificada a que as parturientes são submetidas <sup>28</sup>. Podem, também, ser decorrentes da subestimação de riscos já detectados no pré-natal ou à inexistência/inadequação de protocolos hospitalares para avaliações durante o trabalho de parto, capazes de alertar quando há aumento do risco no pré-parto.

Por outro lado, em cerca da metade das mortes incluídas no Perfil 3, “*óbitos de prematuros*”, houve associação de causas típica de países onde a mortalidade é elevada: hipóxia, outras doenças respiratórias e infecção, principalmente. Esses óbitos ocorreram na rede SUS, com maior proporção em hospitais públicos. Apesar da observada tendência de aumento de nascimentos pré-termo no mundo e no Brasil <sup>29,30</sup>, as dificuldades inerentes à correta determinação da idade gestacional podem ter levado à inclusão neste grupo de crianças a

termo, com retardo de crescimento intra-uterino. Um acompanhamento pré-natal adequado pode diminuir a incidência de baixo peso ao nascer, reduzindo tanto os casos de retardo do crescimento intra-uterino quanto os de prematuridade evitável. A associação de menções de doenças respiratórias e infecciosas com o perfil “óbitos de prematuros” faz supor que, mesmo com a comprovação dos efeitos da corticoterapia antenatal <sup>31</sup> e da antibioticoterapia profilática intraparto <sup>32</sup> na redução da morbimortalidade neonatal, a cobertura nessa população seja baixa. Finalmente, a impossibilidade de disponibilização de suporte avançado de vida para estas crianças poderia estar ligada às características da estruturação das unidades de terapia intensiva neonatal relacionadas por Barbosa <sup>33</sup>: falta de equidade na distribuição dos leitos, limitação do acesso para parcela mais carente da população e qualidade irregular dos serviços prestados.

Apesar da classificação dos hospitais aqui utilizada poder representar uma limitação para essa análise, pois poderiam estar incluídos serviços com características diferentes em uma mesma categoria, a observação desta tipologia na rede SUS de Belo Horizonte aponta para a necessidade de abordagens específicas para prevenção dos óbitos neonatais precoces, segundo sua natureza jurídica. Nos hospitais públicos, referências para partos de risco, deve ser incentivado o uso rotineiro de medidas profiláticas de reconhecido impacto sobre a mortalidade de prematuros. A associação entre óbitos de crianças de baixo risco aos hospitais contratados/conveniados ao SUS aponta para a necessidade de reavaliação destas instituições, visando especialmente à adequação da utilização de tecnologias, incluídos aqui tanto as estruturas hospitalares quanto os protocolos de atendimento ao parto e ao recém-nascido.

Para que o efeito protetor da ampla cobertura da assistência pré-natal em Belo Horizonte se estenda a todo ciclo gravídico-puerperal, as intervenções oportunas e qualificadas no trabalho de parto e na assistência imediata ao recém-nascido também devem estar disponíveis para todos. A ocorrência da quase totalidade dos partos em ambiente hospitalar no período estudado é evidência da suficiência do número de leitos e do desejo da população de buscar os hospitais como locais com potencial para reduzir os riscos de morbi-mortalidade do binômio mãe-filho. É eticamente inaceitável, portanto, a ocorrência de óbitos potencialmente evitáveis na proporção detectada neste estudo.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Organização Mundial de Saúde. Objectifs du Millénaire pour le développement. <http://www.un.org/french/millenniumgoals/index.html> (acessado em 02/Abr/2005).
2. Ministério da Saúde. Saúde Brasil 2004 uma análise da situação de saúde. Brasília DF: Ministério da Saúde; 2004.
3. Tinkar A, Hoop-Bender P, Azfar S, Bustreo S, Bell R. A continuum of care to save newborn lifes. <http://www.activemag.co.uk/BookReader/getBook.asp?Path=LAN/2005/03/01&BookCollection=GL> (acessado em 06/Mar/2005).
4. Wilcox AJ. On the importance -and the unimportance- of the birthweight. *International Journal of Epidemiology* 2001; 30:1233-1241.
5. Lawn JE, Cousens S, Zupan J for the Lancet Neonatal Survival Steering Team. <http://www.activemag.co.uk/BookReader/getBook.asp?Path=LAN/2005/03/01&BookCollection=GL> (acessado em 06/Mar/2005).
- 6- Leal MC, Gama SGN, Campos MR, Cavalini LT, Garbayo LS, Brasil CLP, Szwarcwald CL. Fatores associados aa morbi-mortalidade perinatal em amostra de maternidades públicas e privadas do Município do Rio de Janeiro, 1999-2001. *Cad. Saúde Pública* 2004; (20 suppl.1):20-33.
7. Niobey FML, Cascão AM, Duchiate ML, Sabroza PC. Qualidade do preenchimento dos atestados de óbitos de menores de um ano na região metropolitana do Rio de Janeiro. *Rev Saúde Pública* 1990; 24(4): 311-318.
8. Machado CJ. Mortalidade infantil no Estado de São Paulo, 1999: uma análise das causas múltiplas de morte a partir dos componentes principais. *Rev Bras Epidemiol* 2004; 7(1): 11-21.
9. Lansky S, França E, Leal MC. Mortes perinatais evitáveis em Belo Horizonte, Minas Gerais, Brasil, 1999. *Cad Saúde Pública* 2002; 18: 1389-00.
10. Organização Mundial de Saúde. Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde - 10º revisão, v1. São Paulo: Edusp; 2000.
11. Zadeh LA. Probability measures and fuzzy events. *J Math Anal Appl* 1968; 23: 421-7.
12. SMSA/PBH (Secretaria Municipal de Saúde de Belo Horizonte), 2004. Boletim Epidemiológico: Mortalidade Infantil em Belo Horizonte. Belo Horizonte: SMSA/PBH.
13. Caldeira AP, França E, Perpétuo IH, Goulart EMA. Evolução da mortalidade infantil por causas evitáveis, Belo Horizonte, 1984-1998. *Rev Saúde Pública* 2005; 39(1): 67-74.
14. Lansky S, França E, Leal MC. Mortes perinatais evitáveis em Belo Horizonte, Minas Gerais, Brasil, 1999. *Cad. Saúde Pública* 2002; 18 (5): 1389-1400.

15. Ministério da Saúde, 1996. Sistema de Informação sobre Mortalidade, 1979-1992. Brasília: Departamento de Informática do Sistema Único de Saúde.
16. Center for Disease Control. National Center for Health Statistics. Proceedings of the International Collaborative Effort on Perinatal and Infant Mortality. Maryland, 1992; 3.
17. Manton, KG, Woodbury, MA, and Tolley, HD 1994. Statistical Applications Using Fuzzy Sets, John Wiley & Sons Inc., New York, USA.
18. McNamee P. A comparison of the grade of membership measure with alternative health indicators in explaining costs for older people. *Health Economics* 2004; 13 (4): 379 – 395.
19. Sawyer DO, Leite IC, Alexandrino R. Perfis de utilização de serviços de saúde no Brasil. *Ciência & Saúde Coletiva* 2002; 7(4): 757-776.
20. Cassidy F, Pieper CF, Carroll BJ Subtypes of mania determined by grade of membership analysis. *Neuropsychopharmacology*, 2001; 25: 373-383.
21. Wigglesworth JS. Monitoring perinatal mortality: a pathophysiological approach. *Lancet* 1980; 2: 684-6.
22. Santo AH, Pinheiro CE. Multiple causes-of-death tabulator. *Rev Bras Epidemiol* 1999; 2 (1-2):90-97.
23. Rezende EM, Sampaio IBM, Ishitani LH. Causas múltiplas de morte por doenças crônico-degenerativas: uma análise multidimensional. *Cad Saúde Pública* 2004; 20(5): 1223-1231.
24. França E, Goulart EMA, Machado JAD. Confiabilidade da declaração de causa básica de mortes infantis em região metropolitana do sudeste do Brasil. *Rev Saúde Pública* 1994; 28(5): 385-391.
25. Laurenti R, Buchalla CM. Elaboração de estatísticas de mortalidade segundo causas múltiplas. *Rev Bras Epidemiol* 2000; 3(1-3):21-28.
26. Saad PM. Mortalidade infantil por causas no Estado de São Paulo (Brasil) em 1983: análise sob a perspectiva das causas múltiplas de morte. *Rev Saúde Pública*, 1986; 20(6): 481-488.
27. Niobey CS. A Análise da Mortalidade Infantil Pós-Neonatal Sob a Perspectiva das Causas Múltiplas [Dissertação de Mestrado]. Rio de Janeiro: Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz; 1994.
- 28- Tanaka ACA. Maternidade dilema entre nascimento e morte. São Paulo - Rio de Janeiro: Editora Hucitec; 1995.
29. Barbieri MA, Silva AAM, Bettiol H, Gomes UA. Fatores de risco para a tendência ascendente do baixo peso ao nascer em nascidos vivos de parto vaginal no Sudeste do Brasil. *Rev Saúde Pública* 2000; 34(6): 596-602.

30. Barros F, Diaz-Rossello JL. Redes multicêntricas e qualidade da atenção neonatal. *Jornal de Pediatria* 2004; 80(4): 254-5.
31. Halliday HL. Use of steroids in the perinatal period. *Paediatric Respiratory Reviews* 2004; 5(SupplA):S321-S327.
32. Lukacs SL, Schoendorf KC, Schuchat A. Trends in sepsis-related neonatal mortality in the United States, 1985-1998. *The Pediatric Infectious Disease Journal* 2004; 23(7): 599-603.
33. Barbosa AP. Terapia intensiva neonatal e pediátrica no Brasil: o ideal, o real e o possível. *Jornal de Pediatria* 2004; 80(6): 437-8.

Tabela 1  
 Coeficientes ( $\lambda_{kjl}$ ) das variáveis internas e externas para cada perfil de óbito neonatal precoce.  
 Belo Horizonte, 2000-2003.

Variável	Frequência absoluta	Frequência relativa	Perfil 1	Perfil 2	Perfil 3
			$l_{1j}$	$l_{2j}$	$l_{3j}$
<i>Grupos de Causas</i>					
<i>Prematuridade</i>					
Sim	620	0,71	0	0,48	1
Não	255	0,29	1	0,52	0
<i>Hipóxia</i>					
Sim	132	0,15	0,54	0	0,50
Não	743	0,85	0,46	1	0,50
<i>Outras Respiratórias</i>					
Sim	539	0,60	0,52	0,37	0,54
Não	336	0,40	0,48	0,63	0,46
<i>Infecciosas</i>					
Sim	210	0,24	0,49	0,32	0,43
Não	665	0,76	0,51	0,68	0,57
<i>Malformações</i>					
Sim	171	0,19	0	1	0
Não	704	0,81	1	0	1
<i>Demais</i>					
Sim	310	0,35	0,51	0,49	0,43
Não	565	0,65	0,49	0,51	0,57
<i>Variáveis Externas</i>					
<i>Natureza Jurídica</i>					
Privado	96	0,11	0	0,32	0,07
Contrat/Conv	376	0,43	1	0,00	0,37
Público	403	0,46	0	0,68	0,56
<i>Peso (em gramas)</i>					
≤ 1500	516	0,59	0	0	1
1500 a 2499	175	0,20	0	1	0
≥ 2500	184	0,21	1	0	0

Fonte: SIM/SMSAPBH

Tabela 2

Distribuição dos óbitos segundo intervalos de escores (gik) para cada perfil de óbitos neonatal precoce. Belo Horizonte, 2000-2003

<b>Intervalos de gik*</b>	<b>Perfil 1</b>	<b>Perfil 2</b>	<b>Perfil 3</b>
0	81,3	61,7	32,8
0,01 a 0,24	0,1	18,9	2,6
0,25 a 0,49	0	2,6	0
0,50 a 0,74	1,7	0	0,9
0,75 a 0,99	3,5	2,7	15,3
1	13,4	14,1	48,4
Total	100	100	100
Prevalência (%)**	17,4	18,9	63,7

Fonte: SIM/SMSA/PBH

\* Intervalos de graus de pertencimento dos óbitos a cada um dos perfis

\*\* Prevalência ponderada pelo grau de pertencimento dos indivíduos ao perfil (%)

6 ARTIGO 4: UTILIZAÇÃO DE DADOS SECUNDÁRIOS DO SIM,  
SINASC E SIH NA PRODUÇÃO CIENTÍFICA BRASILEIRA DE 1990 A  
2006

Submetido aos Cadernos de Saúde Pública

## RESUMO

Trata-se de estudo exploratório da distribuição e de algumas características dos artigos científicos publicados de 1990 a 2006 com utilização de dados secundários dos Sistemas de Informação sobre Mortalidade (SIM), Nascidos Vivos (SINASC) e/ou Informações Hospitalares (SIH).

Após pesquisa bibliográfica na base de dados MEDLINE/PubMed utilizando do MeSH (Medical Subject Headings) selecionaram-se 294 artigos. Foram analisadas variáveis referentes aos artigos, aos autores e aos periódicos.

No período analisado observou-se aumento de cerca de quatro vezes no volume de publicações e nas publicações em inglês. Também houve aumento no número de periódicos utilizados; instituições de origem dos pesquisadores; utilização conjunta de duas bases de dados e nas pesquisas dirigidas às doenças e agravos não transmissíveis (especialmente doenças cardiovasculares e causas externas).

Os resultados apontam expressivo aumento da produção científica brasileira que utiliza dados do SIM, SINASC e/ou SIH. A melhoria de qualidade dessas bases de dados, necessárias para a gestão local, foi confirmada pelo crescimento acelerado da sua utilização em pesquisa epidemiológica observado nessa avaliação bibliométrica.

Palavras-chave: Bibliometria; Estatísticas Vitais; Pesquisa/ tendências; Indicadores de Produção Científica; Mortalidade.

## **ABSTRACT**

This study aims at evaluating the growth and some characteristics of the national scientific production which uses secondary data from the information systems SIM, SINASC, or SIH in 1990-2006.

Bibliographic research from MEDLINE/PubMed using MeSH retrieved 294 articles. Variables analyzed included articles, authors and periodicals.

Between 1990 and 2006 publications and publications in English increased four times. There was also an increase in the number of periodicals featured; the number of host institutions increased; more research was done using two coordinated databases; more was published on non-transmissible diseases and afflictions, especially cardiovascular disease and external causes.

The results point to a significant increase in the Brazilian scientific production based on SIM, SINASC, or SIH. The need to improve these databases, necessary for local management, was confirmed by the growing demand in epidemiological research as observed in this bibliometric analysis.

Keywords: Journal article. Publications. Research, trends. Information systems.

## 6.1 INTRODUÇÃO

Os Sistemas de Informação em Saúde (SIS) desenvolvidos e implantados pelo Ministério da Saúde em todo o Brasil têm se tornado, cada vez mais, ferramentas necessárias para gestão loco-regional do sistema de saúde e vêm apresentando consolidação e a qualificação progressivas <sup>1, 2, 3, 4</sup>. Cada um dos SIS atende a diferentes finalidades, além de ter sido desenvolvido, implantado e ser operacionalizado de forma distinta dos demais. No Sistema de Informação sobre Mortalidade (SIM) e no Sistema de Informação sobre Nascidos Vivos (SINASC), por exemplo, são armazenados e processados dados sobre as estatísticas vitais. O SINASC e o SIM disponibilizam, além dos números de nascidos vivos e óbitos que possibilitam o cálculo direto de importantes indicadores de saúde como a mortalidade infantil, outras informações fundamentais como o peso ao nascer e a causa básica da morte, respectivamente. Já no Sistema de Informação Hospitalar (SIH), único sistema de informação de abrangência nacional sobre morbidade hospitalar, estão os registros sobre internações hospitalares nos hospitais públicos municipais, estaduais ou federais e nos hospitais contratados/conveniados integrantes da rede SUS. A disponibilização no *site* do Departamento de Informática do Sistema Único de Saúde (DATASUS) dos dados do SIM, SINASC e SIH desagregados até o nível dos municípios impulsionou a sua utilização para análises epidemiológicas. O acesso gratuito e em meio eletrônico a essas bases de dados também tem facilitado as avaliações de cobertura, de completitude do preenchimento das variáveis e de confiabilidade desses SIS.

O SIM foi desenvolvido em 1976 e sua implantação iniciou-se em 1977. Encontra-se disponível na Internet a série histórica, que se inicia em 1979, com informações importantes e necessárias para as estatísticas de mortalidade, inclusive as causas básicas de morte. As causas de morte declaradas pelo médico foram codificadas segundo regras estabelecidas pela Organização Mundial da Saúde, por meio da Classificação Estatística Internacional de Doenças e Problemas Relacionados (CID). Embora a qualidade das informações do SIM ainda seja controversa <sup>4</sup>, observa-se crescente ampliação da cobertura <sup>5, 6</sup> e de confiabilidade <sup>7</sup> desse sistema.

No SINASC, implantado pelo Ministério da Saúde a partir de 1990, podem ser obtidas informações sobre condições de saúde materno-infantil, por meio de variáveis com as características da mãe - como parturição, duração da gravidez e tipo de parto, e de variáveis

com características do recém-nascido – como peso de nascimento e Apgar. O SINASC é reconhecido pela boa cobertura <sup>8</sup>, pela maior completude <sup>9</sup> e confiabilidade <sup>1, 3,8</sup> das suas informações em relação às do SIM. A maior qualidade do SINASC pode ligar-se às diferenças nas rotinas de registro dos seus dados pelos serviços de saúde e na forma de coleta e processamento pelas secretarias municipais de saúde <sup>8</sup>.

O SIH foi desenvolvido principalmente para o processamento das informações administrativas sobre as internações hospitalares na rede SUS. Nele também podem ser obtidas informações epidemiológicas e que possibilitam a caracterização da população atendida e da morbi-mortalidade nos hospitais da rede. No entanto, as possibilidades de exploração do SIH para ampliação das análises epidemiológicas focadas no atendimento hospitalar da rede SUS têm sido pouco utilizadas devido, principalmente, às controvérsias sobre a qualidade dos dados sobre o diagnóstico da internação e ao pequeno número de estudos sobre a confiabilidade das suas informações <sup>11</sup>. Estima-se que, devido às suas características contábeis, o sub-registro de internações neste sistema seja pequeno <sup>11</sup> e que a média nacional de cobertura do SIH varie entre 70-80% <sup>11, 12</sup>, sendo inversamente proporcional ao desenvolvimento socioeconômico de cada localidade.

A descentralização das ações, serviços e da gestão impulsionada na década de 1990 no Brasil teve como um de seus resultados a evolução nos conceitos de informação em saúde e de epidemiologia nos serviços de saúde. A epidemiologia passou a ser reconhecida, cada vez mais, como recurso estratégico para a construção e consolidação do SUS e as informações contidas nas bases de dados nacionais valorizaram-se, ao serem cada vez mais necessárias para as formulações e avaliações de políticas públicas voltadas para a melhoria das condições de saúde da população. Conseqüentemente aumentaram as necessidades de qualificação das informações e de conhecimento e uso de ferramentas epidemiológicas pelos serviços de saúde <sup>13</sup>.

Embora nos estudos realizados a partir dos SIS se observem as desvantagens ligadas à utilização de dados secundários em pesquisas científicas (como o viés de informação, por exemplo), uma das maiores vantagens dessa utilização é permitir aos pesquisadores a realização, a baixo custo, de estudos de base populacional e com abrangência nacional. Obtidos por meio de metodologias que não são rotineiramente utilizadas pelos serviços de saúde, os resultados destes estudos científicos possibilitam e disponibilizam novas formas de

análises da situação de saúde e de avaliação dos serviços. Assim, a complexidade do sistema de saúde brasileiro tem motivado pesquisadores de universidades e de instituições de pesquisa a produzir metodologias e indicadores que norteiam os serviços de saúde na formulação e implementação de políticas públicas orientadas pelos princípios do SUS <sup>14</sup>.

Artigos científicos publicados em periódicos indexados representam apenas uma parcela desta produção científica. Reconhece-se, no entanto, a importância dessas publicações para a divulgação do conhecimento científico, caracterizadas pelo controle de qualidade feito por pares e pelo acesso ampliado devido à democratização do uso de micro-computadores e à maior facilidade de utilização da Internet.

Por meio da Internet, pode-se ter acesso imediato, parcial ou integral, a artigos científicos publicados em periódicos indexados e disponibilizados por bases de dados bibliográficas. A MEDLINE/PubMed é a maior base mundial de dados na área da saúde. Coordenada pela *National Library of Medicine* (NLM), a MEDLINE/PubMed contém *abstracts* e citações bibliográficas de mais de 5.000 periódicos biomédicos de áreas da medicina, enfermagem, odontologia e veterinária, entre outros e é atualizada semanalmente. Além de possibilitar a recuperação de artigos científicos, a MEDLINE/PubMed oferece a possibilidade de estimar e avaliar a evolução da produção científica em determinada área, identificar o ano da publicação, o periódico e o assunto, entre outros. Por outro lado, estudos de citações e de afiliação dos autores não podem ser realizados por meio dessa base de dados. Aumentos de volume das publicações científicas brasileiras em geral e das áreas das ciências da saúde em especial em bases de dados bibliográficas como ISI (*Institute for Scientific Information*) e MEDLINE/PubMed têm sido observados <sup>15, 16, 17</sup>, a despeito das limitações orçamentárias e da ainda pequena comunidade científica nacionais, se comparadas aos países desenvolvidos.

O objetivo do presente estudo foi realizar análise exploratória sobre a distribuição e algumas características da produção científica nacional com utilização de dados secundários provenientes do SIM, SINASC e/ou SIH referenciadas na base de dados MEDLINE/PubMed no período de 1990-2006, situando essa produção no cenário nacional e internacional.

## **6.2 METODOLOGIA**

Realizou-se levantamento do universo de registros da produção científica internacional e nacional e dos registros nos quais as fontes de dados utilizadas foram o SIM, o SINASC e/ou o SIH recuperados a partir da base de dados MEDLINE/PubMed em janeiro de 2007, correspondente ao período de dezessete anos - entre 1990 e 2006.

#### FONTE DE INFORMAÇÃO

Os artigos foram recuperados na base MEDLINE, interface PubMed, subconjunto da base PubMed, que indexava 5.164 periódicos em janeiro de 2007. Devido às suas atualizações semanais, o total de registros da MEDLINE/PubMed para determinado período pode mudar a cada atualização<sup>18</sup>. Os descritores, que podem ser utilizados em pesquisas bibliográficas e estudos bibliométricos, representam os assuntos discutidos nos artigos e são extraídos do vocabulário MeSH – *Medical Subject Headings*.

#### FILTROS TEMÁTICOS

O ano de início do período do estudo – 1990 coincidiu com o ano da implantação do SINASC. Os dados foram extraídos para cada ano, de primeiro de janeiro a 31 de dezembro do período de 1990 a 2006. Cada ano foi definido por meio da opção *Limits* da interface PubMed.

A pesquisa foi realizada em duas etapas. Na primeira etapa, destinada à obtenção da proporção de participação anual de artigos com estudos brasileiros em relação a todos os artigos da base de dados bibliográfica, foi feita a pesquisa do universo de citações e das citações de artigos referentes ao país Brasil foi realizada no campo de palavras *Brazil*.

A segunda etapa teve dois objetivos. O primeiro objetivo foi caracterizar se um artigo utilizou dados secundários provenientes do SIM, do SINASC e/ou do SIH. O segundo objetivo foi obter a proporção de participação dos artigos selecionados no universo da produção científica brasileira, resultado obtido na etapa anterior.

Para consecução do primeiro objetivo, qual seja, caracterizar se um artigo utilizou dados secundários provenientes do SIM, do SINASC e/ou do SIH foram utilizados, concomitantemente, os descritores do MESH *Mortality OR Pregnancy OR Pregnancy Rate OR Pregnancy Outcome OR Live Birth OR Infant,newborn OR Hospital Information Systems OR Information Systems OR Databases,factual e no campo de palavras SIM OR SINASC OR*

*SIH* combinados (AND) com *Brazil* no campo de palavras. Foram incluídos apenas artigos originais.

## ANÁLISE DOS DADOS

O software SPSS versão 10.0 para Windows (SPSS Inc., Chicago, Estados Unidos) foi utilizado para armazenamento e análise dos dados. Para fins de análise, os dezessete anos do estudo foram divididos em dois períodos, segundo a década de publicação do artigo: de 1990 a 1999 (dez anos) e de 2000 a 2006 (sete anos).

As variáveis selecionadas em cada um dos artigos selecionados foram o ano de publicação, idioma e as bases de dados secundários utilizadas, faixa etária da população estudada e classificação do conteúdo. Dos periódicos foram selecionados os nomes e o país de origem. Em relação ao primeiro autor analisaram-se as instituições de origem e as regiões do país onde se localizavam. As faixas etárias da população do estudo foram classificadas como perinatal (de 22 semanas a sete dias de vida), neonatal (de zero a 27 dias de vida), infantil (de zero a um ano de vida), criança (de zero a 10 anos), adolescente (de 11 a 18 anos), adultos (de 19 a 60 anos) e idosos (mais de 60 anos). Quando na metodologia do estudo não havia especificação sobre a idade da população estudada, classificou-se a faixa etária como não especificada. Para a classificação do seu conteúdo os artigos foram agrupados em cinco categorias temáticas, obtidas a partir dos descritores de cada artigo e consideradas como mutuamente excludentes, as quais foram desagregadas em subcategorias. As categorias e subcategorias utilizadas foram: (1) causas de morbi-mortalidade (doenças transmissíveis, doenças crônicas, causas externas); (2) saúde materno-infantil (mortalidade perinatal, neonatal e infantil e mortalidade materna); (3) sistemas de informação (SIM, SINASC, SIH); (4) assistência à saúde (cuidado pré-natal/tocologia, outras). A quinta categoria, outros (5) (composta das subcategorias desigualdade em saúde, evitabilidade, poluição, doenças profissionais, etnia e saúde e demais), foi criada para enquadrar artigos que não pertenciam às categorias anteriores.

Os resultados foram tabulados pela evolução ano a ano do total de artigos recuperados por sistema de informação utilizado e a sua representação no total dos artigos indexados; faixa etária estudada e temas em destaque, a partir dos demais descritores utilizados; periódicos onde foram publicados os artigos; instituição de origem do primeiro autor e sua localização

nas grandes regiões geográficas do país; idioma das publicações. Finalmente, realizou-se a comparação dos resultados obtidos entre os dois períodos selecionados.

### 6.3 RESULTADOS

Em janeiro de 2007, na primeira etapa da busca foram recuperados 8.522.732 artigos publicados na base MEDLINE/PubMed no período de 1990 a 2006, dos quais 0,29% (n=24.351) foram referentes ao Brasil. A partir de janeiro de 2007, novos artigos publicados no mesmo período do estudo ingressaram na base de dados, mas não foram incluídos neste estudo.

Na Tabela 1 se apresenta a distribuição anual do total de artigos e dos artigos brasileiros recuperados nesta base de dados bibliográficos. A proporção anual de artigos brasileiros em relação ao total de artigos aumentou de 1,3% no ano de 1990 para 4,0% no ano de 2006. O número absoluto de artigos brasileiros teve aumento de 453,5% - passando de 516 artigos em 1990 para 2856 em 2006. O crescimento do total de artigos publicados na mesma base foi de 78,9% no período 1990-2006.

Por meio da combinação *booleana* de descritores do *MESH* foram recuperados 2.373 artigos. Os resumos foram disponibilizados em 2.156 (90,8%) dos artigos, entre os quais em 1.712 artigos (75%) o texto completo também estava disponível. Quando a leitura do resumo não permitia a inclusão/exclusão do artigo, os artigos foram lidos na íntegra. Por meio da leitura dos resumos e/ou dos textos completos foram selecionados 294 artigos. Na Tabela 1 observa-se a distribuição anual dos artigos selecionados. A média anual de artigos publicados no período foi igual a 16,3. Observa-se, no ano de 1990, um desproporcionalmente elevado número de artigos - se comparado aos demais anos. Mas, apesar disto, obtém-se um aumento do número anual de publicações igual a 287,5%, que pode ser observado pela diferença entre o número de artigos publicados em 1990 (n=8) e os de 2006 (n=31).

No Gráfico 1 observa-se a distribuição anual de artigos científicos recuperados na base MEDLINE/PubMed no período de 1990 a 2006 segundo o SIS utilizado. A base de dados mais utilizada foi o SIM (n=232), seguida pelo SINASC (n=63) e pelo SIH (n=59). Um único SIS foi utilizado isoladamente em 80,3% (n=236) dos artigos e em 58 (19,7%) houve

utilização simultânea do SIM/SINASC (n=37), SIM/SIH (n=19), SINASC/SIH (n=1) e SIM/SINASC/SIH (n=1).

Aproximadamente 70% dos textos publicados foram escritos exclusivamente em português. Cinquenta e quatro periódicos publicaram um ou mais artigos sobre o tema. Os periódicos nacionais veicularam 81% (n=238) do total de textos selecionados. Entre os periódicos nacionais que mais publicaram artigos na área estão, com números bastante similares, a Revista de Saúde Pública (n=98) e os Cadernos de Saúde Pública (n=94). Os outros periódicos com maior volume de publicações foram os Arquivos Brasileiros de Cardiologia (n=16), Arquivos de Neuropsiquiatria (n=5), *São Paulo Medical Journal* e Revista da Associação Médica Brasileira (4 artigos em cada um), Jornal de Pediatria (3 artigos) e as Revistas da Sociedade Brasileira de Medicina Tropical e Brasileira de Enfermagem (2 artigos em cada um). A *Revista Panamericana de Salud Pública* (n=8), o *Journal of Epidemiology and Community Health* (n=4) e o *International Journal of Epidemiology* (n=3) foram as revistas internacionais que publicaram o maior número de artigos.

Sessenta e quatro instituições de origem dos primeiros autores foram citadas pelo menos uma vez. Entre elas encontravam-se instituições de ensino, de pesquisa ou serviços de saúde públicos ou privados. Em 72,1% (n=212) dos casos eram instituições localizadas na Região Sudeste, seguidas da Região Nordeste (n=42) e da Região Sul (n=31). Na Região Sudeste foram expressivos os números de trabalhos publicados pela Universidade de São Paulo (n=59), Escola Nacional de Saúde Pública/ Centro de Informação Científica e Tecnológica/ Fundação Oswaldo Cruz (n=50), Universidade Federal do Rio de Janeiro (n=13) e Universidade Federal de Minas Gerais (n=12). Também puderam ser observadas publicações que tiveram como primeiros autores pesquisadores originários dos serviços de saúde como as secretarias estaduais de saúde do Rio de Janeiro (n=13) e de São Paulo (n=12), e as secretarias municipais de saúde do Rio de Janeiro (n=7) e de Belo Horizonte (n=5). Nas regiões Nordeste e Sul destacaram-se a Universidade Federal da Bahia (n=17) e a Universidade Federal do Rio Grande do Sul (n=7), respectivamente. O maior número de artigos da região Norte foi de pesquisadores da Universidade Federal do Maranhão (n=3).

A Tabela 2 apresenta os resultados encontrados na classificação dos temas dos artigos. Considerando-se o enfoque dado pelos artigos a partir da análise dos seus conteúdos observou-se um maior número de estudos sobre causas de morbi-mortalidade, na qual se

enquadraram 133 textos. As doenças cardiovasculares (n=31), as causas externas de morte (n=23), os cânceres (n=14), os homicídios (n=13) e a AIDS (n=10) foram as causas de morte mais encontradas nesta categoria. A categoria outros respondeu pela segunda maior frequência de artigos (n=61), principalmente devido ao significativo número de artigos que abordaram o tema das desigualdades em saúde (n=26). A terceira maior frequência de artigos foi a da categoria referente à saúde materno-infantil. Dos 40 artigos sobre saúde materno-infantil, 70% (n=28) estudaram os períodos perinatal, neonatal e/ou infantil. A categoria de artigos que realizaram avaliações dos SIS ocupou o quarto lugar em número de publicações (n=36). Entre estes, 24 (66,6%) realizaram avaliações da cobertura, completitude e/ou de confiabilidade dos dados. E, finalmente, na categoria assistência à saúde foram selecionados 26 textos que, em sete casos, abordavam questões relacionadas à assistência pré-natal e/ou ao parto. As faixas etárias do estudo não foram especificadas em 129 artigos. Nos casos com faixas etárias especificadas os textos eram dedicados ao período perinatal (n=7), ao período neonatal (n=20), aos menores de um ano (n=30), às crianças (n=14), aos adolescentes (n=13), aos adultos (n=37) e aos idosos (n=14).

A divisão da série histórica estudada em dois períodos, de 1990 a 1999 e de 2000 a 2006, possibilitou a observação de algumas diferenças entre eles. O número médio do total de artigos brasileiros publicados em cada período evoluiu de 830 artigos, no período 1990-1999, para 2.292 no período seguinte, o que representou um aumento de 2,8 vezes. Em relação aos textos selecionados para este estudo este aumento foi igual a 4,6 vezes, tendo em vista que a média do primeiro período foi de 6,9 artigos e a do segundo período foi igual a 32,1 artigos. No período 2000-2006 observou-se concentração dos artigos (91,3%) que utilizaram simultaneamente mais de um dos SIS.

Houve um aumento de mais de quatro vezes no número de artigos em inglês no período 2000-2006, se comparado ao período anterior. A partir do ano 2000 foram recuperadas 13 publicações bilíngües português/inglês, uma publicação em francês e duas em espanhol. Ressalte-se a diversificação dos periódicos utilizados a partir do ano 2000 quando foram observados os primeiros artigos científicos a partir de dados do SIM, SINASC e/ou SIH em 36 revistas. Algumas destas revistas são dirigidas a clínicos, como os Arquivos Brasileiros de Cardiologia (n=16) e o Jornal de Pediatria (n=3). A partir do ano 2000 também pela primeira vez foram citadas 39 novas instituições, podendo ser destacadas entre elas a Secretaria de Estado da Saúde de São Paulo (n=12) e a Secretaria Municipal de Saúde de Belo Horizonte

(n=5). A produção científica das regiões Centro-Oeste e Norte também ganharam impulso a partir do ano 2000, com a publicação de sete artigos.

Em relação às faixas etárias estudadas, destacam-se os significativos aumentos observados nas publicações dedicadas ao período perinatal (n=7), ao adolescente (n=13) e à população idosa (n=14) a partir do ano 2000. O interesse pelas populações infantil e neonatal foi observado por meio dos 12 artigos publicados no período 1990-1999 e dos 50 artigos publicados no período seguinte.

Alguns aspectos foram enfocados pela primeira vez em artigos publicados a partir do ano 2000. Entre eles podem ser citadas as avaliações da assistência ao pré-natal e/ou ao parto (n=7) e o suicídio (n=3). Houve aumento no número de artigos dedicados às doenças crônicas, especialmente as cardiovasculares (n=27) e às causas externas (n=31), embora as doenças transmissíveis (especialmente a AIDS) também tenham permanecido em foco. Os SIS também foram utilizados maior número de vezes em abordagens sobre desigualdade em saúde (n=22), evitabilidade (n=6) e poluição ambiental (n=7) no período 2000-2006 (Tabela 2).

#### **6.4 DISCUSSÃO**

Os resultados obtidos indicam que as pesquisas que se utilizaram das bases de dados secundárias do SIM, SINASC e/ou SIH apresentaram tendência crescente no período 1990-2006. Estes resultados corroboram evidências observadas em outros estudos sobre o crescimento da produção científica em epidemiologia <sup>16</sup>, da utilização do SIM e do SINASC em pesquisas brasileiras sobre mortalidade perinatal <sup>19</sup> e do SIH na Saúde Coletiva <sup>11</sup>.

Coura & Wilcox <sup>20</sup> observaram que a publicação de artigos de autores brasileiros em periódicos internacionais indexados pelo ISI quintuplicou no período 1981-2001 e que o número de pessoas qualificadas (doutores e mestres) triplicou na última década. A limitação da busca à base de dados MEDLINE/PubMed não possibilitou o acesso a periódicos da área que também são muito consultados e que não estão indexados nesta base bibliográfica, como a Revista Brasileira de Epidemiologia e o periódico Epidemiologia e Serviços de Saúde. No entanto, a MEDLINE/PubMed é a fonte bibliográfica mais consultada no campo da saúde <sup>16</sup> e nela estão indexadas os Cadernos de Saúde Pública e a Revista de Saúde Pública, nas quais se

concentrou grande parte das publicações na área. Assim, os resultados aqui apresentados cumpriram o objetivo de evidenciar o crescimento das publicações na área, embora a ampliação da busca em outras bases bibliográficas pudesse resultar na obtenção de maior magnitude de crescimento.

Mais de 70% da produção científica capturada na base de dados MEDLINE/PubMed por este estudo foram escritos em português, o que pode representar uma barreira para a leitura destes artigos por pesquisadores de outras nacionalidades. Artigos publicados em línguas diferentes do inglês são menos frequentemente lidos e citados, mesmo que seja possível o acesso ao artigo completo, via SciELO <sup>21</sup>. Após a publicação de um artigo, deseja-se que ele atinja o maior número de leitores e que seu conteúdo promova e/ou participe do diálogo da comunidade científica nacional e internacional. Em busca dessas metas os conselhos editoriais das revistas científicas trabalham intensamente para que seus periódicos tenham maior visibilidade internacional, por meio da indexação em bases bibliográficas de acesso internacional. Neste sentido, Blank e colaboradores <sup>17</sup> avaliaram que a indexação do Jornal de Pediatria ao MEDLINE/PubMed aumentou significativamente as submissões de artigos brasileiros e estrangeiros a este periódico. A tendência de aumento da pluralidade da língua (ou internacionalização) das publicações foi observada neste estudo por meio de duas importantes iniciativas. Uma delas foi observada nos Arquivos Brasileiros de Cardiologia e no Jornal de Pediatria que, ao disponibilizarem na Internet a opção da publicação bilíngüe (português-inglês), permitiram acesso aos artigos completos à audiência internacional e a todo público brasileiro. A outra iniciativa que também deve ser ressaltada refere-se ao aumento dos artigos em inglês publicados pela Revista de Saúde Pública e pelos Cadernos de Saúde Pública

A maior utilização do SIM como fonte de pesquisa é coerente com a reconhecida importância das estatísticas de mortalidade, embora não tenha sido observada diferença significativa na tendência de utilização do SIM em relação aos demais SIS a partir do ano 2000. Recentes avaliações do SIM observaram que este sistema apresentou ampliação da cobertura de óbitos <sup>6</sup> e redução do número de óbitos por causas mal definidas <sup>7</sup>, embora ainda persistam muitas deficiências na qualidade da informação sobre mortalidade em nosso país <sup>9</sup>. Deve ser ressaltado o grande potencial demonstrado pela utilização conjunta de duas ou mais bases de dados, que foi evidenciado neste estudo pelo aumento do número de artigos que cruzaram

informações de diferentes origens, principalmente por meio de *linkages* ou de relacionamentos probabilísticos de dados.

A concentração da produção científica em instituições dos estados da Região Sudeste, notadamente em São Paulo e no Rio de Janeiro observada neste estudo corroborou os resultados obtidos por pesquisas em outras áreas<sup>15, 17, 18, 22, 24</sup>. Essa concentração foi ainda mais acentuada nas publicações com dados do SIH, conforme já relatado por Bittencourt e colaboradores<sup>11</sup>. Em levantamento sobre a produção científica em saúde em seis países da América Latina no período 1973-1992, Pellegrini Filho e colaboradores<sup>24</sup> demonstraram predomínio de trabalhos produzidos por autores de uma única instituição (73,4%), mas também observaram tendência de aumento dos trabalhos em colaboração. Ao se limitar a considerar apenas a instituição de origem do primeiro autor, os resultados obtidos neste estudo podem ter subestimado o papel de colaboradores de instituições de pesquisa e/ou centros formadores de docentes e pesquisadores.

Verificou-se haver ampliação na variedade temática nos artigos analisados. Ao longo dos anos foram sendo abordados novos temas, como as doenças crônicas (as cardiovasculares, principalmente), a violência (em especial os homicídios), as desigualdades em saúde (inclusive as raciais), e as abordagens sob as óticas da evitabilidade, dos riscos ambientais e da saúde ocupacional. As faixas etárias extremas da vida (perinatal e idosos) e a adolescência também surgiram como novos temas. Embora fuja ao escopo desse estudo, deve ser citada a riqueza de alternativas metodológicas adotadas para ampliação da compreensão sobre os problemas de saúde da população. Em especial, devem ser referidos os crescimentos das análises de tendências temporais de causas de morte, das abordagens ecológicas com análises espaciais, das análises baseadas em diferentes técnicas de relacionamento de dado e, dos estudos caso-controle e de coorte.

Entre as limitações desse estudo, que se propôs explorar quantitativamente a distribuição da produção de artigos científicos a partir de três sistemas de informação selecionados – SIM, SINASC e SIH, deve ser citada a utilização apenas da MEDLINE/PubMed. Sem dúvida, a incorporação da base de dados LILACS (Literatura Latino-americana e do Caribe de Informação em Ciências da Saúde) poderia trazer resultados ainda mais expressivos, tendo em vista que assim o conjunto mais relevante da produção científica nacional estaria melhor representado. Além disso, não foram feitas avaliações mais abrangentes da qualidade dos

artigos publicados (por meio de análises qualitativas e/ou de avaliações de impacto, por exemplo).

Finalmente, deve-se ressaltar que o presente estudo, ao avaliar a tendência de utilização dos sistemas de informação em saúde em pesquisas científicas, visou prioritariamente valorizar e estimular o uso dessas bases de dados. Como resultado da progressiva descentralização da gestão, as análises de situação de saúde são realizadas pelos serviços de saúde em níveis cada vez mais desagregados, tendo como metas as intervenções específicas a serem adotadas pelos níveis locais. A demanda pela aproximação entre os diagnósticos de situação e os níveis mais descentralizados da atenção criou a necessidade de construção de indicadores de saúde adequados para intervenções específicas. Conseqüentemente, aumentou também a demanda pela qualificação das bases de dados, com ampliação da cobertura e maior qualidade no preenchimento das variáveis. Para atender a estas demandas, observam-se nos três níveis de governo que os setores responsáveis pela coleta, processamento e divulgação das informações das bases de dados nacionais têm realizado grandes investimentos em qualificações das equipes técnicas e das suas infra-estruturas. Assim, além dos ganhos mais facilmente mensuráveis que podem advir da utilização das bases de dados como fonte de pesquisa científica soma-se mais um, qual seja, o maior estímulo dado aos profissionais destas equipes técnicas resultante da ampliação do significado do produto de seu trabalho.

As avaliações críticas da qualidade dos dados e análises de feitas em ambientes acadêmicos, nos quais podem ser utilizadas metodologias sofisticadas e dificilmente acessíveis aos serviços de saúde também repercutem na qualificação destas bases de dados, ao criarem a possibilidade de ampliação do conhecimento pelos serviços e pela academia dos pontos fracos e fortes dos SIS. Assim, é desejável que a pesquisa científica a partir das bases de dados secundárias, como é o caso do SIM, SINASC e SIH se consolide ainda mais e atinja maior velocidade de crescimento e de internacionalização, como observado em outras práticas científicas brasileiras<sup>15, 16</sup>. Para que esta meta seja atingida alguns caminhos podem ser apontados. Entre eles se incluem a necessidade de melhora da qualidade das informações que pode ser obtida, por exemplo, com maior clareza e adequada padronização nas instruções orientadoras dadas pelos manuais nos diversos passos necessários para o preenchimento, coleta e processamento dos dados; com a ampliação dos investimentos feitos pelos três níveis de gestão dos SIS na qualificação técnica dos profissionais – incluindo as fontes primárias e

os processadores das informações e também na infra-estrutura dos setores de epidemiologia e informação dos serviços de saúde.

Outro passo fundamental para que as bases de dados nacionais sejam cada vez mais utilizadas como fontes para pesquisas científicas, importantes para o avanço e consolidação do SUS, os serviços de saúde, em todos os níveis, devem possibilitar que o acesso dos pesquisadores interessados em criar uma agenda comum entre gestores, pesquisadores e cidadãos a estas bases seja cada vez mais ágil e democrático.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Levy S. Pense grande faça pequeno. *Rev. Bras. Epidemiol.* 2007; 10(1): 127-128.
2. Paes NA. Qualidade das estatísticas de óbitos por causas desconhecidas dos Estados brasileiros. *Rev. Saúde Pública* 2007, 41 (3): 436-445.
3. Theme Filha MM, Gama SGN, Cunha CB, Leal MC. Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no Município do Rio de Janeiro, 1999-2001. *Cad. Saúde Pública* 2004, 20 (suppl.): S83-S91.
4. Romero DE, Cunha CB. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de menores de um ano registrados no Sistema de Informações sobre Mortalidade (1996/2001). *Cad Saúde Pública* 2006; 22(3): 673-84.
5. Cardoso AM, Santos RV, Coimbra JR, Carlos EA. Mortalidade infantil segundo raça/cor no Brasil: o que dizem os sistemas nacionais de informação? *Cad. Saúde Pública.* 2005; 21(5): 1602-1608.
6. Paes NA. Avaliação da cobertura dos registros de óbitos dos Estados brasileiros em 2000. *Rev Saúde Pública* 2005; 39(6): 882-90.
7. Laurenti R; Mello-Jorge MHP, Lebrão ML; Gotlieb SLD; Márcia Furquim de Almeida MF. Estatísticas Vitais: contando os nascimentos e as mortes. *Rev Bras Epidemiol* 2005, 8 (2): 108-110.
8. Haraki CAC, Gotlieb SLD, Laurenti R. Confiabilidade do Sistema de Informações sobre Mortalidade em município do sul do Estado de São Paulo. *Rev Bras Epidemiol* 2005; 8(1): 19-24.
9. Almeida MF, Alencar GP, Novaes HMD, Ortiz LP. Sistemas de informação e mortalidade perinatal: conceitos e condições de uso em estudos epidemiológicos. *Rev Bras Epidemiol* 2006; 9(1): 56-68.
10. Silva AAM, Ribeiro VS, Borba Júnior AF, Coimbra LC, Silva RA. Avaliação da qualidade dos dados do Sistema de Informações sobre Nascidos Vivos em 1997-1998. *Rev Saúde Pública* 2001; 35(6): 508-14.
11. Bittencourt SA, Camacho LAB, Leal MC. O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva. *Cad Saúde Pública* 2006; 22(1): 19-30.
12. Ministério da Saúde. Agência Nacional de Saúde Suplementar. Caderno de Informação da Saúde Suplementar: beneficiários, operadoras e planos. Disponível em: [http://www.ans.gov.br/portal/upload/informacoess/caderno\\_informaca\\_03\\_2007.pdf](http://www.ans.gov.br/portal/upload/informacoess/caderno_informaca_03_2007.pdf) (acessado em 02/jun/2007).
13. Drumond Jr M. Epidemiologia em serviços de saúde. In Campos GWS, Minayo MCS, Akerman M, Drumond Jr M, Carvalho YM (Org). *Tratado de Saúde Coletiva*. São Paulo: Hucitec; Rio de Janeiro: Fiocruz, 2006.

14. Paim JS. Atenção à Saúde no Brasil. In Paim JS. Desafios da Saúde Coletiva no Século XXI. Salvador, EDUFBA, 2006.
15. Rodrigues OS, Fonseca L, Chaimovich H. Mapping câncer, cardiovascular and malaria research in Brazil. *Braz J Med Biol Res* 2000; 33:853-867.
16. Barreto ML. Crescimento e tendência da produção científica em epidemiologia no Brasil. *Rev Saúde Pública* 2006; 40(N Esp): 79-85.
17. Blank D, Rosa LO, Gurgel RQ, Goldani MZ. Brazilian knowledge production in the field of child and adolescent health. *J Pediatr (Rio J)* 2006; 82(2): 97-102.
18. Packer AL, Tardelli AO, Castro RCF. A distribuição de conhecimento científico público em informação, comunicação e informática em saúde indexado nas bases de dados MEDLINE/PubMed e LILACS. *Ciência & Saúde Coletiva*, 12(3): 587-599 2007.
19. Fonseca SC, Coutinho ESF. Pesquisa sobre mortalidade perinatal no Brasil: revisão da metodologia e resultados. *Cad Saúde Pública* 2004; 20: Sup 1:S7-S19.
20. Coura JR, Wilcox LCB. Impact factor, scientific production and quality of brazilian medical journals. *Mem Inst Oswaldo Cruz* 2003; 98(3): 293-297.
21. Cunha-Melo JR, Santos GC, Andrade MV. Brazilian medical publications: citation patterns for Brazilian-edited and non-Brazilian literature. *Braz J Med Biol Res* 2006; 39: 997-1002.
22. Mansur JA, Abud AS, Albuquerque AP. Tendências das publicações na fase trimestral, bimensal e mensal de cinco décadas dos Arquivos Brasileiros de Cardiologia. *Arq Bras Cardiol* 2000; 75 (1): 1-7.
23. Guimarães JA. A pesquisa médica e biomédica no Brasil. Comparações com o desempenho científico brasileiro e mundial. *Ciência & Saúde Coletiva* 2004; 9 (2): 303-27.
24. Pellegrini Filho A, Goldbaum M, Silvi J. Production of scientific articles on health in six Latin American countries, 1973-1992. *Rev Panam Salud Publica* 1997; 2:121-9.

Tabela 1

Distribuição do anual total de artigos, dos artigos referentes ao Brasil e dos artigos selecionados na base MEDLINE/PUBMED, 1990 a 2006.

Ano da publicação	Total de artigos (a)	Artigos brasileiros (b)		Artigos selecionados (c)	
		n	Proporção (a X 100/b)	n	Proporção (b X 100/c)
1990	399.926	516	0,13	8	1,55
1991	401.431	642	0,16	4	0,62
1992	405.750	768	0,19	2	0,26
1993	413.360	715	0,17	4	0,56
1994	423.609	733	0,17	5	0,68
1995	434.111	760	0,18	3	0,39
1996	444.408	834	0,19	5	0,60
1997	442.652	898	0,20	10	1,11
1998	461.410	1186	0,26	10	0,84
1999	479.798	1253	0,26	18	2,08
2000	521.197	1536	0,29	32	1,98
2001	533.456	1666	0,31	33	1,46
2002	553.508	2050	0,37	30	1,32
2003	584.071	2341	0,40	31	1,16
2004	624.265	2664	0,43	31	1,26
2005	684.156	2933	0,43	37	1,09
2006	715.624	2856	0,40	31	1,40
1990-2006*	8.522.732	24.351	0,29	294	1,2
$\Delta$ %**	78,94	453,49		287,50	

\* Número total de artigos de 1990 a 2006.

\*\* Porcentagem diferencial entre os números de 2006 e 1990.

Gráfico 1

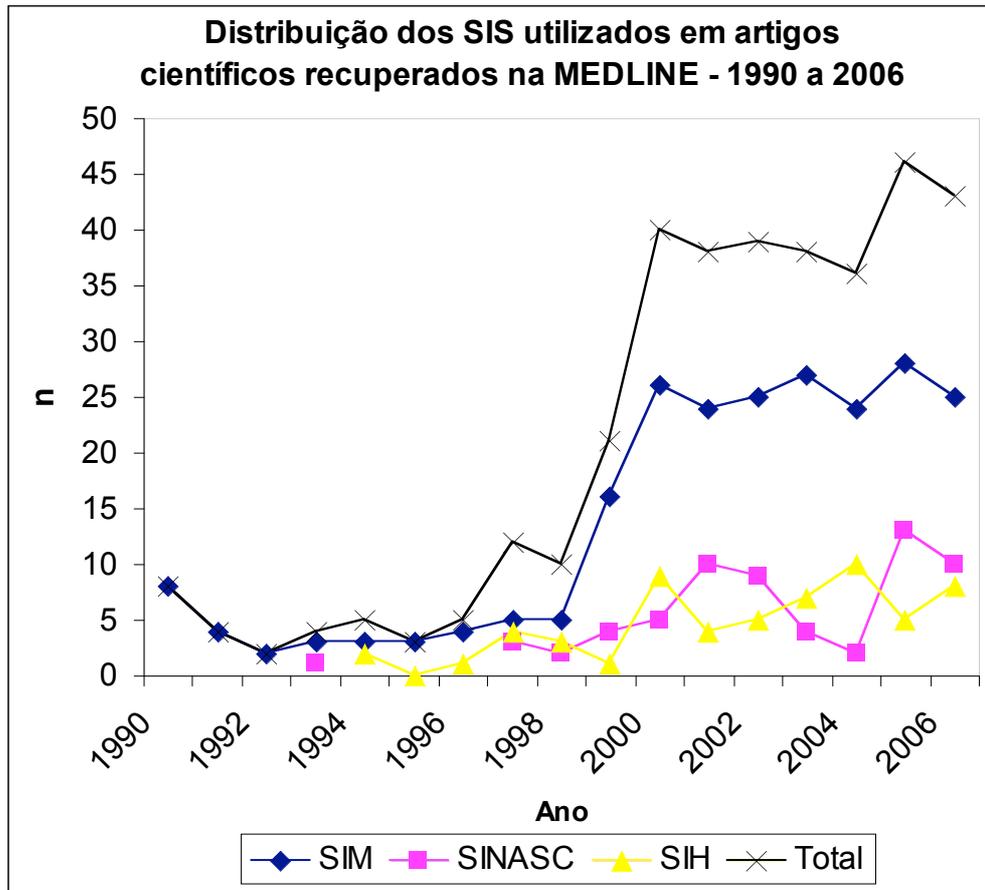


Tabela 2

Distribuição dos artigos segundo classificação do conteúdo e por período de publicação. MEDLINE/PUBMED, 1990-2006.

<b>Conteúdo</b>	1990-1999	2000-2006	Total
<b>Causas</b>			
<i>Doenças transmissíveis</i>			
AIDS	1	9	10
Tuberculose	0	6	6
Diarréias	0	5	5
Sífilis	0	2	2
Outras	2	5	7
Total	3	27	30
<i>Doenças crônicas</i>			
D Cardiovasculares	4	27	31
Câncer	3	11	14
D Respiratórias/Asma	1	5	6
Diabetes	2	3	5
Outras	0	5	5
Total	10	51	61
<i>Causas externas</i>			
Homicídios	1	12	13
Suicídios	0	3	3
Ac. Trânsito	1	2	3
Todas as causas ext.	10	14	23
Total	12	31	42
<b>Materno-infantil</b>			
Perinatal/neonatal/infantil	5	23	28
Maternas/aborto/ fertilidade	4	8	12
Total	9	31	40
<b>Sistemas de informação</b>			
SIM	12	12	24
SINASC	4	4	8
SIH	3	1	4
Total	19	17	36
<b>Assistência à saúde</b>			
Assist. pré-natal/ parto	0	7	7
Outras	2	17	19
Total	2	24	26
<b>Outros</b>			
Desigualdade em saúde	4	22	26
Evitabilidade	1	5	6
Poluição	2	7	9
Doenças profissionais	1	2	3
Etnia e saúde	2	1	3
Demais	4	7	11
Total	14	47	61
<b>Total</b>	<b>69</b>	<b>225</b>	<b>294</b>

## 7 CONSIDERAÇÕES FINAIS

Os resultados obtidos neste estudo mostraram algumas potencialidades e dificuldades decorrentes da utilização de sistemas de informação no monitoramento do óbito neonatal precoce. As maiores potencialidades do relacionamento probabilístico ligam-se à incorporação das variáveis do SINASC (socioeconômicas, maternas, da gravidez e do recém-nascido, por exemplo) às variáveis do SIH (caráter da internação, tempo de permanência, especialidade do médico assistente, natureza jurídica do hospital, por exemplo). Por outro lado, em relação ao SIM, o enfoque de causas múltiplas de morte possibilitou uma visão mais ampla sobre a crescente complexidade e o aumento das inter-relações patológicas envolvidas nos processos mórbidos do período neonatal. Observou-se neste estudo que as análises de causas múltiplas de morte no período neonatal precoce são possíveis, além de necessárias. O crescente desenvolvimento da Perinatologia, observado por meio da sofisticação das ações e das tecnologias de suporte básico, médio e avançado, disponibilizados para recém-nascidos cada vez mais prematuros, demanda a ampliação das maneiras de analisar as causas de morte destes recém-nascidos. Por sua vez, os resultados do estudo bibliométrico confirmaram o potencial de crescimento das pesquisas científicas a partir de dados secundários, para atender demandas das instituições de pesquisas e também no âmbito do SUS.

O projeto de compatibilizar bases de dados por meio de relacionamento é antigo e é compartilhado por todos aqueles que desejam ver estes dados cada vez mais utilizados em todo o seu potencial. As principais vantagens obtidas pelo relacionamento são o aprimoramento das informações e a ampliação das possibilidades de utilização das bases de dados devido à incorporação de novas variáveis, com relativo baixo custo. Diversas técnicas de relacionamento de dados vêm sendo crescentemente adotadas em pesquisas e na rotina dos serviços de saúde. O relacionamento SIM e SINASC é o mais realizado em pesquisa e no serviço, principalmente para as ações de vigilância ao óbito infantil e materno. Os estudos com o objetivo de avaliar a qualidade das bases de dados (como subenumeração de eventos, completitude, confiabilidade) ainda são pouco freqüentes. A utilização da base de dados do SIH com vistas à sua compatibilização com outras bases é praticamente restrito a instituições de pesquisa do Rio e de São Paulo, onde vem ocorrendo, principalmente, por meio do relacionamento com o SIM e o SINAN (Sistema de Informações de Agravos de Notificação). Este trabalho avançou ao compatibilizar o SIH com o SINASC no estado de Minas Gerais e mostrar a sua viabilidade. Avançou, também, ao realizar esta compatibilização visando

identificar, pela primeira vez, a subenumeração de nascidos vivos. A compatibilização entre estas duas bases de dados, já realizada em estudos desenvolvidos nos estados do Rio de Janeiro e São Paulo, objetivou a identificação exclusiva de mortes maternas. O relacionamento de dados feito neste estudo, ao incorporar o peso de nascimento – o mais importante fator de risco isolado para o óbito infantil – aos dados do SIH cria muitas e novas perspectivas de análise. Possibilita, também, a identificação clara dos hospitais da rede SUS que apresentam piores coberturas e qualidade no preenchimento do SINASC e que poderiam ser alvos da atuação dos serviços de controle e avaliação hospitalar, por meio de ações visando melhorias na cobertura e qualidade do preenchimento das variáveis.

O relacionamento probabilístico dos dados do SINASC e do SIH apresentou várias dificuldades para a avaliação da subenumeração de nascidos vivos nos municípios de Minas Gerais. Entre estas, pode ser citada a execução do relacionamento, tarefa árdua, principalmente devido às diferenças de lógicas dos dois sistemas e à irregular qualidade dos dados. A compatibilização da codificação dos estabelecimentos de saúde (que recebem números e/ou nomes diferentes em cada um dos sistemas) e os erros decorrentes do menor percentual de preenchimento de algumas variáveis (como município de residência) aumentaram as dificuldades. Existe também a possibilidade de erro advinda dos problemas no entendimento das diferenças existentes nos conceitos de aborto, nascido morto e nascido vivo.

A utilização de sentença em linguagem SQL (*Structured Query Language*), desenvolvida especificamente para este estudo, aumentou a agilidade e a eficiência na realização do relacionamento dos dados. Mas a boa qualidade dos dados é um aspecto determinante para que se possam formar pares corretos. No processo de relacionamento constroem-se escores, que sintetizam o grau de concordância global entre registros de um mesmo par. Cabe ao pesquisador a definição de escores limiares a partir dos quais os pares serão classificados como verdadeiros, falsos ou duvidosos. Esta é uma das grandes vantagens do relacionamento probabilístico, bastante adequada à situação deste estudo, qual seja, admitir que os possíveis erros dos registros sejam considerados para avaliar a qualidade dos pares formados. Mas como proceder diante da incerteza se o registro proveniente de cada base de dados pertence ou não à mesma pessoa? Não há uma forma estabelecida para se resolver esta dúvida. A revisão manual dos pares duvidosos, estratégia que se adotou neste estudo, revelou-se demorada e sujeita a erros, devido ao imenso volume de pares considerados como duvidosos e ao amplo espectro de possibilidades de erros e omissões que podem ocorrer no SINASC e no

SIH. Apesar de sua importância, ainda são poucos os estudos brasileiros que informam sobre o percentual de pares formados ao fim do processo de relacionamento. Por outro lado, algumas questões devem ser respondidas para que sejam adequadas as avaliações de subnumeração de nascimentos no SINASC. Entre elas estão as perguntas referentes aos 12% de registros de internações para parto do SIH que, diferentemente do esperado, não se relacionaram a registros do SINASC em se partindo do princípio de que a relação entre os registros dos dois bancos é unívoca. São necessárias análises cuidadosas deste conjunto de registros de internações para partos visando a identificação de características ligadas à subnumeração de nascidos vivos no SINASC e, também, a avaliação da possibilidade da presença de erro sistemático ligado ao não pareamento destes dados.

As dificuldades operacionais e as dúvidas quanto aos resultados obtidos pelo relacionamento probabilístico seriam as principais fontes de desestímulo para a realização de maior número de pesquisas utilizando esta técnica? A ausência específica de trabalhos de relacionamento de dados do SIH e do SINASC e que poderiam ter sido utilizados na avaliação da validade dos resultados obtidos neste estudo também levantou algumas questões. Esta ausência estaria ligada ao fato de que a subnumeração de nascidos vivos ocorre principalmente nas regiões brasileiras – notadamente Norte e Nordeste - onde é menor a produção científica e que, portanto, não despertaria interesse em pesquisadores de instituições do Sul e do Sudeste? Ou estaria ligada às freqüentes dificuldades no acesso a tecnologias de informação nas regiões onde é maior a subnumeração de eventos vitais? Ou, por outro lado, estaria ligada às dificuldades técnicas ligadas ao processo de relacionamento? Ou, ainda, seria a atribuída deficiência de qualidade e/ou a escassez de estudos de confiabilidade dos dados do SIH que estaria desestimulando sua utilização? Ou, finalmente, poderia a ainda pequena produção científica brasileira a partir de dados gerados pelo enorme contingente de internações hospitalares na rede SUS ser decorrente das dificuldades observadas para obtenção de dados junto ao DATASUS?

As reconhecidas vantagens obtidas com o relacionamento de registros poderiam ser utilizadas em maior escala em pesquisa e no serviço de saúde se maiores facilidades técnicas fossem incorporadas a este processo. A implantação do Código Nacional de Estabelecimentos de Saúde (CNES) e do Cadastro Único de Saúde (o esperado Cartão SUS) que atribuirão número único a cada estabelecimento e a cada cidadão, respectivamente, poderão resultar em maiores facilidades e redução do grau de incerteza dos relacionamentos.

Com os dados do SIM, que é o sistema de informação mais frequentemente utilizado como fonte de dados em pesquisa, realizou-se análise de causas múltiplas dos óbitos neonatais precoces ocorridos em Belo Horizonte, a partir da teoria dos conjuntos nebulosos (*fuzzy sets*). As informações do SIM no município apresentam qualidade e coerência suficientes para serem submetidas a esta análise. A complexidade dos processos mórbidos dos recém-nascidos reflete-se parcialmente nas causas de morte. Observa-se, no entanto, muito pequeno número de citações das intercorrências patológicas da gravidez (como hipertensão, diabetes), o que poderia representar uma perda das informações relacionadas à gravidez e/ou ao parto que repercutem em afecções do recém-nascido. A análise de causas múltiplas de morte, mais frequentemente utilizada na abordagem de doenças crônico-degenerativas, e a lógica *fuzzy* revelaram-se muito adequadas na abordagem de um evento complexo, como são os óbitos neonatais precoces na cidade de Belo Horizonte. A sua aplicação propiciou, por meio da maximização da utilização das informações contidas nas declarações de óbito, que se desvendassem diferenças nos perfis de causas das mortes evitáveis, segundo a natureza jurídica dos hospitais. A classificação dos hospitais por natureza jurídica representa uma limitação desta análise, diante das possíveis desigualdades entre os estabelecimentos de uma mesma natureza jurídica. Mas, a despeito desta limitação, a diferente ocorrência de óbitos potencialmente evitáveis entre as categorias deve merecer a atenção dos gestores, com vistas à adoção de abordagens específicas segundo a natureza jurídica do hospital para a redução destas mortes.

Finalmente, para avaliar a utilização das bases de dados em todos os artigos publicados em revistas indexadas por pesquisadores brasileiros – e não apenas os referentes ao período neonatal precoce – procedeu-se ao estudo bibliométrico. Os resultados obtidos por este estudo demonstraram vigoroso aumento – de 4,6 vezes – na média anual dos artigos selecionados no período 2000-2006 em relação ao período 1990-1999. Este resultado coloca uma nova perspectiva sobre um argumento frequentemente repetido, segundo o qual a má qualidade dos dados disponíveis nos sistemas de informação nacionais seria um empecilho para a sua utilização em pesquisa científica. A pujança demonstrada pelo volume de trabalhos selecionados demonstra o contrário.

Nos trabalhos selecionados, refletiu-se o empenho dos pesquisadores brasileiros em se debruçar sobre questões relacionadas ao mosaico epidemiológico decorrente da superposição

de etapas e do prolongamento da transição epidemiológica brasileira. A singularidade do nosso perfil epidemiológico resulta em demandas múltiplas sobre o sistema de saúde. O sistema de saúde torna-se, então, fonte de um grande volume de achados e de questões e, por meio de fluxo inverso, também pressiona os pesquisadores para realização de análises com metodologias apropriadas, que possam auxiliar na maior compreensão das inter-relações entre o amplo espectro de variáveis que devem ser contempladas na área da saúde. Prováveis resultados deste círculo talvez possam ser observados entre as publicações sobre doenças infecciosas e crônico-degenerativas, desigualdade social, riscos ambientais, envelhecimento e iniquidades no acesso aos serviços de saúde, por exemplo, que procuram trazer à tona mecanismos de associação e/ou causalidade enfocando variáveis de ordem socioeconômica, epidemiológica, assistencial e biológica, entre outras.

Além de permitir ampliar o conhecimento de nossa realidade de saúde, o amplo espectro de temas abordado possibilita que os pesquisadores brasileiros possam dialogar tanto com os cientistas do mundo desenvolvido quanto com os cientistas dos chamados países em desenvolvimento. E, embora fuja ao escopo deste trabalho, deve-se ressaltar que foram identificadas pesquisas realizadas com metodologias de uso corrente na literatura internacional, que potencializaram a utilização das informações. A análise das metodologias utilizadas ofereceria contribuição significativa para a maior compreensão da riqueza contida nestes trabalhos e poderia ser tema de estudos futuros. Por outro lado, não surpreende a concentração institucional e geográfica das publicações nas regiões Sul e Sudeste, que, além de corroborar achados anteriores, está de acordo com o padrão de distribuição das leis e princípios bibliométricos que segue a máxima poucos com muito e muitos com pouco.

Entre as limitações deste trabalho deve ser considerada a ausência da avaliação da confiabilidade dos dados utilizados para o relacionamento e para a análise de causas múltiplas, o que poderia afetar os resultados. Outra limitação destes estudos refere-se à restrição do número e do tipo de indicadores utilizados e à ausência de um padrão ouro de referência. Mas, a despeito de suas limitações, percebe-se que os sistemas de informação têm suas potencialidades crescentemente exploradas em pesquisas científicas e que há um campo aberto para a realização de outras análises, utilizando novas e diferentes formas de avaliação, que poderiam estimular o debate e a produção de respostas criativas para os problemas de produção, qualificação e utilização dos sistemas de informação em saúde para o monitoramento das mortes neonatais precoces.

A valorização das informações contidas nos sistemas de informação nacionais por meio da sua utilização em pesquisas científicas - especialmente no monitoramento do óbito neonatal precoce - é o principal intuito deste trabalho. As pesquisas científicas podem contribuir para a discussão do importante papel destes sistemas, para sua avaliação e divulgação de seus pontos fortes e fracos, aspectos fundamentais ao seu aprimoramento. Por meio da utilização dos sistemas de informação nacionais a comunidade científica brasileira estaria enfrentando duplo desafio: fortalecer a sua capacidade de produção a custos reduzidos e criar espaços democráticos que possibilitem que esta capacidade seja colocada a serviço da busca de maior equidade em saúde.

## ANEXOS

### **ANEXO A: PROJETO DE PESQUISA**

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**FACULDADE DE MEDICINA**  
**DEPARTAMENTO DE MEDICINA PREVENTIVA E SOCIAL**

Sistema de Informação Hospitalar: complemento para  
vigilância ao óbito neonatal precoce?

Doutoranda: **Eliane de Freitas Drumond**  
Orientadora: **Profa. Elisabeth Barboza França**

Belo Horizonte/MG

2004

## INTRODUÇÃO

A persistentemente elevada taxa de mortalidade neonatal precoce no Brasil, complexa em seus fatores causais, também é reflexo de nossa sociedade injustamente desigual<sup>3, 5, 9, 10, 14, 19, 32, 34, 44, 46, 48, 51</sup>.

Hospitais e maternidades são locais privilegiados para a adoção de medidas redutoras do número de óbitos neonatais precoces<sup>16, 18, 34, 42</sup>. Ali podem ser garantidos acessos à assistência hospitalar ao parto qualificada, hierarquizada e regionalizada, complemento fundamental para as políticas de priorização da atenção básica pré-natal<sup>45,46</sup>. Também é ali que se produzem as informações sobre os óbitos neonatais precoces. A melhora da qualidade destas informações, geradoras de indicadores de saúde, é um passo indispensável na busca da equidade em saúde<sup>29</sup>. Através de avaliação e monitoramento contínuos de hospitais e maternidades poderão ser adotadas medidas para redução da heterogeneidade das qualidades da assistência e das importantes informações por eles produzidas<sup>1, 2, 8, 31, 35, 38</sup>.

O déficit de qualidade na informação em saúde liga-se tanto ao profissional de saúde<sup>4, 8, 19, 38</sup> quanto às dificuldades e particularidades decorrentes do recente processo de municipalização da gestão dos Sistemas de Informação em Saúde, com as diferenças locais de infra-estrutura e de políticas de investimento em produção de informação<sup>47-49</sup>.

Segundo o Ministério da Saúde, o Estado de Minas Gerais - com mais de 800 municípios - tem o pior desempenho da Região Sudeste tanto na cobertura quanto na regularidade de envio dos dados de mortalidade através do Sistema de Informação de Mortalidade (SIM). Minas é o único estado da região a ter estatísticas de mortalidade calculadas por métodos indiretos e divulgadas com atraso. As declarações de óbito geram os dados para o sistema.

As técnicas de mensuração indireta da mortalidade utilizadas pela Fundação IBGE e outras propostas na literatura<sup>42,47</sup> são passíveis de críticas. As iniciativas visando obtenção destas taxas por métodos diretos têm sido estimuladas. O Ministério da Saúde calcula que o SIM de Minas Gerais cobre apenas cerca de 80% dos óbitos infantis.

Obter taxa de mortalidade neonatal precoce por método direto, a partir de dados e informações reais que direcionem prioridades para investigações e intervenções<sup>15, 19, 34, 38, 48, 51</sup>, é uma meta a ser alcançada em nosso Estado, para melhor compreensão de nosso “mosaico

epidemiológico.” Mas, para utilização do método direto a cobertura dos sistemas de informação deve ser completa. A estratégia de agregação de diferentes bases de dados permite ampliar a cobertura, além da checagem das fontes de informação. O impacto desta agregação pode ser sentido na melhora da qualidade da assistência e da informação<sup>1, 2, 11, 12, 15, 21-30, 35-37, 40</sup>.

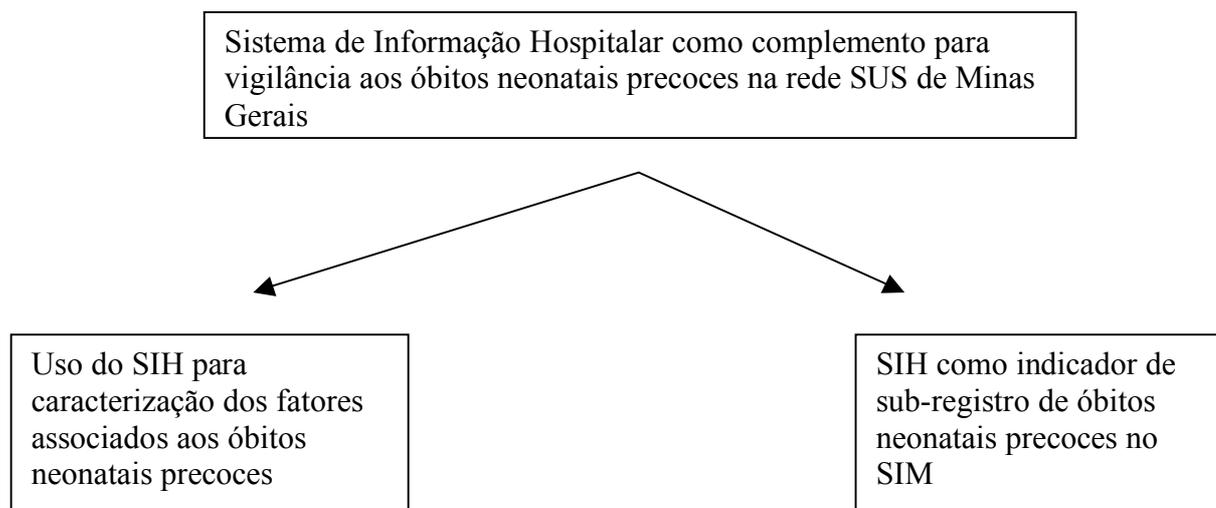
Minas tem 490 hospitais e maternidades do Sistema Único de Saúde (SUS/MG), de diferentes naturezas jurídicas e distribuídos heterogeneamente entre os municípios. Cerca de 100% de todos os partos de MG são hospitalares. Destes, cerca de 70% ocorrem na rede SUS/MG (MS, 2003).

Cada parto na rede SUS gera uma Autorização de Internação Hospitalar (AIH), documento com fins contábeis que alimenta o Sistema de Informação Hospitalar (SIH) e permite a identificação do paciente e do prestador da assistência. A utilização do SIH na avaliação da qualidade e no monitoramento da assistência hospitalar, e não apenas para fins contábeis, é preconizada por vários autores<sup>7, 11-13, 27-29, 36, 39, 40, 43</sup>.

O SIH mostrou-se adequado como complemento no monitoramento e avaliação da sub-enumeração de diversos agravos<sup>11-13, 24, 27-30</sup>, devido à sua agilidade na produção de informação e ao baixo custo de sua utilização. Quanto à confiabilidade de suas informações, demonstrou ter melhor qualidade do que lhe é imputada<sup>26</sup>.

A utilização inovadora do SIH como complemento à vigilância aos óbitos infantis mostra-se vantajosa em relação a agilidade, baixo custo e disponibilidade da mensal da informação via Internet. Permite, também, o estudo dos fatores hospitalares relacionados à assistência ao parto e aos óbitos neonatais precoces. Na literatura, não há relato destes trabalhos desenvolvidos em Minas Gerais.

Este estudo visa identificar as potencialidades do SIH como complemento ao SIM na vigilância e monitoramento dos óbitos neonatais precoces em MG, conforme o diagrama a seguir:



Considerando que as qualidades da informação e da assistência podem estar relacionadas entre si e a diferenças regionais, a características do serviço de saúde, aos riscos da gravidez e do neonato, no **primeiro componente** se propõe o estudo de possíveis fatores associados aos óbitos neonatais precoces na rede hospitalar SUS/MG.

O **segundo componente** será constituído pela estimação das diferenças entre os coeficientes de mortalidade neonatal precoce obtidos através do SIM e do SIH no Estado de Minas Gerais.

## OBJETIVOS

### OBJETIVO GERAL

Avaliar a potencialidade do SIH como complemento do SIM para vigilância ao óbito neonatal precoce em Minas Gerais.

### OBJETIVOS ESPECÍFICOS

- Caracterizar possíveis fatores associados aos óbitos neonatais precoces nos hospitais da rede SUS em Minas Gerais;
- Avaliar a sub-enumeração dos óbitos neonatais precoces em Minas Gerais, por município de residência da mãe, no ano de 2001.

### PROPOSTA METODOLÓGICA

1º OBJETIVO ESPECÍFICO: Caracterizar possíveis fatores associados aos óbitos neonatais precoces nos hospitais da rede SUS em Minas Gerais

DESENHO DO ESTUDO: de coorte retrospectivo.

VARIÁVEL DEPENDENTE: óbitos neonatais precoces (até 07/01/02)

HIPÓTESE: os óbitos neonatais precoces associam-se a determinadas características dos hospitais da rede SUS/MG.

ANÁLISE: para verificação de associação entre variáveis selecionadas e o óbito será realizada análise multivariada para controle de variáveis de confusão. Será utilizado o risco relativo (e IC 95%) como medida da magnitude da associação.

2º OBJETIVO ESPECÍFICO: Avaliar a sub-enumeração dos óbitos neonatais precoces em Minas Gerais, por município de residência da mãe, no ano de 2001.

DESENHO DO ESTUDO: de corte transversal

VARIÁVEL DEPENDENTE: óbitos neonatais precoces não identificados pelo SIM.

HIPÓTESE: há óbitos neonatais precoces informados ao SIH e não informados ao SIM.

ANÁLISE: os bancos de dados preparados serão cruzados através do *software Reclink*, para avaliação das diferenças obtidas em cada um dos sistemas. Para avaliar possíveis fatores associados às diferenças na qualidade da informação será utilizada análise multivariada para controle de possíveis fatores de confusão. Como medida da magnitude da associação entre variáveis selecionadas e o subregistro de óbitos será utilizado o *odds ratio* ( e IC95%).

## **VIABILIDADE**

A viabilidade do presente projeto liga-se, principalmente a:

- Utilização de dados secundários, disponíveis no DATASUS.
- Disponibilização da ferramenta de relacionamento dos bancos de dados pelo Ministério da Saúde.
- Banco de dados QUALIMAT já validado.
- Financiamento aprovado pela FAPEMIG, para desenvolvimento no ano de 2004.

**CRONOGRAMA**

Atividades	2004 (trimestre)				2005 (trimestre)				2006 (trimestre)			
	1	2	3	4	1	2	3	4	1	2	3	4
Revisão bibliográfica	X	X	X	X	X	X	X	X	X	X		
Seleção dos casos e preparação dos dados	X	X	X	X								
Cruzamento dos bancos de dados		X	X	X	X							
Análise dos dados					X	X	X	X	X			
Redação de artigos			X				X				X	
Submeter artigos para publicação				X				X				X

## REFERÊNCIAS BIBLIOGRÁFICAS

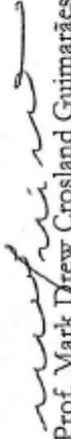
1. Almeida MF, Mello Jorge MHP. O uso da técnica de “linkage” de sistemas de informação em estudos de coorte sobre mortalidade neonatal. *Rev Saúde Pública* 1996; 30 (2): 141-147.
2. Almeida MF, Alencar GP. Informações em saúde: necessidade de introdução de mecanismos de gerenciamento dos sistemas. *Inf Epidemiol SUS* 2000; 9 (4) :241-50.
3. Barcellos CC, Sabroza PC, Peiter P, Rojas LI. Organização espacial, saúde e qualidade de vida: análise espacial e uso de indicadores na avaliação de situações de saúde. *IESUS* 2002; 11 (3);129-38.
4. Barros FC, Victora CG, Teixeira AMB, Puerto Filho M. Mortalidade perinatal e infantil em Pelotas, rio Grande do Sul: nossas estatísticas são confiáveis? *Cad Saúde Pública* 1985; 1: 348-58.
5. Borba, MCLM. Fatores socioeconômicos e mortalidade infantil em Minas Gerais: um estudo ecológico. Dissertação apresentada ao Programa de Pós-Graduação em Saúde Pública – Epidemiologia, Mestrado, Departamento de Medicina Preventiva e Social, Faculdade de Medicina da UFMG. 2000.
6. Camargo KR, Coeli CM. *Reclink*: aplicativo para relacionamento de bases de dados, implementando o método *probabilistic record linkage*. *Cad Saúde Publica* 2000; 16(2): 439-47.
7. Campos MR, Martins M, Noronha JC, Travassos. Proposta de integração de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH-SUS) para pesquisa. *IESUS* 2000; 9 (1):51-58.
8. Carvalho ML, Silver LD. Confiabilidade da declaração da causa básica de óbitos neonatais: implicações para o estudo da mortalidade prevenível. *Rev Saúde Pública* 1995; 29: 292-97.
9. Costa MCN, Azi PA, Paim JS, Silva LMV. Mortalidade infantil e condições de vida: a reprodução das desigualdades sociais em saúde na década de 90. *Cad Saúde Publica* 2001; 17(3): 555-67.
10. Costa MCN, Mota ELA, Paim JS, Silva LMV, Teixeira MG, Mendes CMC. Mortalidade infantil no Brasil em períodos recentes de crise econômica. *Rev Saúde Pública* 2003; 37(6): 101-109.
11. Escosteguy CC, Portela MC, Medronho RA, Vasconcellos MTL. O Sistema de Informações Hospitalares e a assistência ao infarto agudo do miocárdio. *Rev Saúde Pública* 2002; 36(4): 491-499.
12. Escosteguy CC, Medronho RA, Portela MC. Avaliação da letalidade hospitalar do infarto agudo do miocárdio no estado do Rio de Janeiro através do uso do Sistema de Informações Hospitalares/SUS. *Cad Saúde Coletiva* 1999; 7: 739-59.

13. Ferreira VMB, Portela MC. Avaliação da subnotificação de casos de AIDS no município do Rio de Janeiro com base em dados do sistema de informações hospitalares do Sistema Único de Saúde. *Cad Saúde Pública* 1999; 15(2): 413-25.
14. Ferreira VMB, Portela MC, Vasconcellos MTL. Fatores associados à subnotificação de pacientes com AIDS, no Rio de Janeiro, 1996. *Rev Saúde Pública* 2000; 34(2): 234-242.
15. Goldani MZ, Benatti R, Silva AAM, Bettiol H, Correa JCW, Tietzmann M, Barbieri MA. Narrowing inequalities in infant mortality in Southern Brazil. *Rev Saúde Publica* 2002; 36(4):478-483.
16. Gomes FBC. Abordagem epidemiológica dos sistemas de informação no Brasil: gerenciamento, realização de pesquisas e divulgação no IESUS - Editorial. *IESUS* 2001; 10 (3): 109-12.
17. Gouvêa CSD, Travassos C, Fernandes C. Produção de serviços e qualidade da assistência hospitalar no Estado do Rio de Janeiro, Brasil -1992 a 1995. *Rev. Saúde Publica* 1997; 31(6): 601-17.
18. Lansky S, França E, Leal MC. Mortalidade infantil e evitabilidade: revisão da literatura. *Rev Saúde Pública* 2002; 36(6): 759-772.
19. Lansky S. Avaliação de alguns indicadores de qualidade da assistência pré-natal em Belo Horizonte. Belo Horizonte: UFMG. Projeto de pesquisa (doutorado em Saúde Pública), 2003.
20. Laurenti R. Mortalidade infantil nos Estados Unidos, Suécia e Estado de São Paulo. *Rev Saúde Pública* 1987; 21:268-73.
21. Leal MC, Szwarcwald CL. Evolução da mortalidade neonatal no Estado do Rio de Janeiro, Brasil, de 1979 a 1993. *Rev Saúde Pública* 1996; 30(5): 403-12.
22. Lebrão M. Análise da fidedignidade dos dados estatísticos hospitalares disponíveis na Secretaria da Saúde de São Paulo em 1974. *Rev Saúde Pública* 1978; 12: 234-49.
23. Lemos KRV, Valente JG. A declaração de óbito como indicador de sub-registro de casos de AIDS. *Cad Saúde Pública* 2001; 17 (3): 309-17.
24. Lessa FJD, Mendes ACG, Farias SF, Sá DA, Duarte PO, Melo Filho DA. Novas metodologias para vigilância epidemiológica: uso do Sistema de Informação Hospitalar-SIH. *IESUS* 2000; 9(supl): 3-27.
25. Lyra TM, Mendes ACG, Silva Júnior JB, Duarte PO, Melo Filho DA, Albuquerque PC. Sistema de Informações Hospitalares: fonte complementar na vigilância e monitoramento de doenças imunopreveníveis. *IESUS* 2000; 9 (2):87-110.
26. Magalhães ML, Carvalho MS. Atenção hospitalar perinatal e mortalidade neonatal no município de Juiz de Fora, Minas Gerais. *Rev Bras Mater Inf* 2003; 3(3): 329-337.
27. Mathias TAF, Soboll MLMS. Confiabilidade de diagnósticos nos formulários da autorização de internação hospitalar. *Rev Saúde Pública* 1998; 32 (6): 493-501.

28. Melione LPR. Utilização das informações hospitalares do Sistema Único de Saúde para vigilância epidemiológica e avaliação de serviços ambulatoriais em São José dos Campos – São Paulo *Inf Epidemiol SUS* 2002; 11 (3/4): 215-25.
29. Mendes ACG, Silva Júnior JB, Medeiros KR, Lyra TM, Melo Filho DA, Sá DA. Avaliação do Sistema de Informações Hospitalares (SIH-SUS) como fonte complementar na vigilância e monitoramento de doenças de notificação compulsória. *Inf Epidemiol SUS* 2000; 9(2): 67-86.
30. Mendes ACG, Albuquerque PC, Lessa FD, Maciel Filho R, Farias SF, Montenegro TO. Sistema de Informações Hospitalares: fonte complementar na vigilância e monitoramento de doenças de transmissão vetorial. *Inf Epidemiol SUS* 2000; 9 (2):125-36.
31. Mendes ACG, Medeiros KR, Farias SF, Lessa FD, Carvalho CN, Duarte PO. Sistema de Informações Hospitalares fonte complementar na vigilância e monitoramento das doenças de veiculação hídrica. *Inf Epidemiol SUS* 2000; 9 (2):111-24.
32. Mendonça EF, Goulart EMA, Machado JAD. Confiabilidade da declaração de causa básica de mortes infantis em região metropolitana do sudeste do Brasil. *Rev Saúde Pública* 1994; 5: 385-91.
33. Menezes AMB, Barros FC, Victora CG, Tomasi E, Oliveira ALB. Fatores de risco para mortalidade perinatal em Pelotas, RS, 1993. *Rev. Saúde Publica.* 1998; 32: 209-16.
34. Moraes IHS, Santos SRFR. Informações para gestão do SUS: necessidades e perspectivas. *Inf Epidemiol SUS* 2001; 10 (1): 49-56.
35. Nascimento LFC, Gotlieb SLD. Fatores de risco para o baixo peso ao nascer, com base nas informações da declaração de nascido vivo em Guaratinguetá, SP, no ano de 1998. *Inf Epidemiol Sus.* 2001; 10(3): 113-20.
36. Nobre LC, Victora CG, Barros FC, Lombardi C, Teixeira AMB, Fuchs SC. Avaliação da qualidade da informação sobre a causa básica de óbitos infantis no Rio Grande do Sul (Brasil). *Rev Saúde Pública.* 1989; 23: 207-13.
37. Prade SS, Vasconcellos MM. Identificando fontes de dados nos sistemas de informações do SUS para vigilância epidemiológica das infecções hospitalares. *Inf Epidemiol SUS.* 2001; 10 (3): 121- 28.
38. Puccini RF, Pedroso GC, Silva EMK, Araújo NS, Sila NN. Equidade na atenção pré-natal e ao parto em área da região Metropolitana de São Paulo, 1996. *Cad Saúde Pública.* 2003; 19(1); 35-45.
39. Santa Helena ET, Rosa MB. Avaliação da qualidade das informações relativas aos óbitos em menores de um ano em Blumenau, 1998. *Rev Bras Saúde Mater Infant* 2003; 3(1): 75-83.

40. Scatena JHG, Tanaka OU. Utilização do Sistema de Informações Hospitalares (SIH-SUS) na análise de descentralização da saúde em Mato Grosso. *Inf Epidemiol SUS* 2001; 10(1): 19-30.
41. Schramm J, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e a natimortalidade. *Rev Saúde Pública* 2000; 34(3):272-91.
42. Schramm J, Szwarcwald CL. Diferenciais nas taxas de mortalidade neonatal e natimortalidade hospitalares no Brasil: um estudo com base no Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH). *Cad Saúde Pública* 2000;16(4):39-48.
43. Schramm J, Szwarcwald CL, Esteves MAP. Assistência obstétrica e risco de internação na rede de hospitais do Estado do Rio de Janeiro. *Rev Saúde Pública* 2002; 36 (5): 590-97.
44. Silva Júnior JB, Mendes ACG, Campos Neta TJ, Medeiros JR, DA. Sistema de Informações Hospitalares: fonte complementar na vigilância e monitoramento de doenças transmitidas entre pessoas. *Inf Epidemiol SUS* 2000; 9 (2):137-62.
45. Silva LMV, Paim JS, Costa MCN. Desigualdades na mortalidade, espaço e estratos sociais. *Rev Saúde Pública* 1999; 33 (2): 187-97.
46. Silveira DS, Santos IS, Costa JSD. Atenção pré-natal na rede básica: uma avaliação da estrutura e do processo. *Cad Saúde Publica* 2001; 17(1): 131-39.
47. Souza ACT, Cufino E, Peterson KE, Gardner J, Amaral MIV, Ascherio A. Variations in infant mortality rates among municipalities in the state of Ceara, Northeast Brazil: an ecological analysis. *Int J Epidemiol* 1999; 28:267-75.
48. Szwarcwald CL, Leal MC, Andrade CLT, Souza Júnior, PRB. Estimação da mortalidade infantil no Brasil: o que dizem as informações sobre óbitos e nascimentos do Ministério da Saúde? *Cad Saúde Pública* 2002; 18 (6): 1725-36.
49. Szwarcwald CL, Leal MC, Andrade CLT. Mortalidade infantil no Brasil: Belíndia ou Bulgária? *Cad Saúde Pública* 1997; 13: 503-16.
50. Travassos C, Viacava F, Fernandes C, Almeida CM. Desigualdades geográficas e sociais na utilização de serviços de saúde no Brasil. *Ciênc Saúde Colet* 2001; 5:133-149.
51. Viacava F. Informações em saúde: a importância de inquéritos populacionais. *Ciênc. Saúde Colet* 2002; 7: 607-22.
52. Victora CG, Barros FC. Infant mortality due perinatal causes in Brasil: trends, regional patterns and possible interventions. *São Paulo Med J* 2001; 19(1); 33-42.

## ANEXO B: CERTIFICADO DE QUALIFICAÇÃO

 <p>DEPARTAMENTO DE SAÚDE PÚBLICA DEPS - UFMG</p>	<p>Universidade Federal de Minas Gerais Faculdade de Medicina Programa de Pós-Graduação em Saúde Pública <i>Seminários em Saúde Coletiva</i></p>	
<p><b>Certificado</b></p>		
<p>Certifico que Eliane de Freitas Drumond participou dos <i>Seminários em Saúde Coletiva</i> promovido pelo Programa de Pós-Graduação em Saúde Pública, apresentando o projeto de tese intitulado <b>Óbitos Neonatais Precoces: análise de causas múltiplas de morte pelo método <i>Grade of Membership (GoM)</i></b>, perante banca examinadora.</p>		
<p>Belo Horizonte, 18 de novembro de 2005</p>		
<p> Prof. Mark Drew Crosland Guimarães Coordenador do Programa de Pós-Graduação em Saúde Pública</p>		

**ANEXO C: APROVAÇÃO PELO CEP E COEP**

UFMG

Universidade Federal de Minas Gerais  
Comitê de Ética em Pesquisa da UFMG - COEP**Parecer nº. ETIC 095/04****Interessada: Profa. Dra. Elizabeth Barbosa França**  
**Faculdade de Medicina - UFMG****DECISÃO**

O Comitê de Ética em Pesquisa da UFMG – COEP, aprovou no dia 19 de maio de 2004 o projeto de pesquisa intitulado « **Sistema de Informação Hospitalar: Complemento para Vigilância ao Óbito Neonatal Precoce?**, » e o Termo de Consentimento Livre e Esclarecido do referido projeto.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto.

  
**Profa. Dra. Maria Elena Lima Perez Garcia**  
**Presidente do COEP/UFMG**



**COMITÊ DE ÉTICA EM PESQUISA-SECRETARIA MUNICIPAL DE SAÚDE DE  
BELO HORIZONTE (CEP-SMSA/PBH)**

Avaliação de projeto de pesquisa – Protocolo 0022005

Título do projeto: "Perfil dos óbitos neonatais precoces sob a perspectiva das causas múltiplas, BH, 2000 a 2003"

Nome do pesquisador: Eliane de Freitas Drumond

Elizabeth Barboza França

O projeto acima referido cumpriu os requisitos da resolução 196/96 da CONEP, tendo sido aprovado na reunião do Comitê de Ética em Pesquisa da Secretaria Municipal de Saúde de Belo Horizonte realizada no dia 14 de abril de 2005.

O relatório final ou parcial deverá ser encaminhado ao CEP um ano após início do projeto ou ao final desde, se em prazo inferior a um ano.

Celeste de Souza Rodrigues

Coordenadora do CEP-SMSA/PBH

Belo Horizonte, 26 de abril de 2005.

## ANEXO D: ARTIGOS PUBLICADOS

ARTIGO ARTICLE 157

### Óbitos neonatais precoces: análise de causas múltiplas de morte pelo método *Grade of Membership*

Early neonatal mortality: an analysis of multiple causes of death by the *Grade of Membership* method

Eliane de Freitas Drumond <sup>1,2</sup>  
Carla Jorge Machado <sup>3</sup>  
Elizabeth Franço <sup>3</sup>

#### Abstract

*This population-based study aimed to determine the profile of early neonatal deaths in Belo Horizonte, Minas Gerais, Brazil, from 2000 to 2003. Profiles were analyzed from the perspective of avoidability, justified by persistently high early neonatal mortality rates in the city. Three profiles were generated for multiple causes of death from the perspective of fuzzy sets, using the Grade of Membership method. Birth weight and the hospital's corporate status were also related to the three profiles. Private hospitals were characterized by so-called "difficult-to-prevent deaths, with mention of congenital malformations" (profile 2). The Unified National Health System (SUS) generated two distinct profiles. Private maternity facilities contracted out by the SUS showed "preventable deaths" (profile 1), while "premature deaths" (profile 3) occurred in the public Federal and State maternity hospitals. This typology highlights the need to adopt differential policies in the SUS, focusing on evaluation and accreditation for maternity facilities contracted out by the SUS and – for the system as a whole – on the routine adoption of protocols for childbirth care and prophylactic measures that are known to reduce neonatal morbidity and mortality.*

*Infant Mortality; Early Neonatal Mortality; Multiple Causes of Death*

#### Introdução

O Brasil é signatário dos *Objetivos de Desenvolvimento do Milênio* (ODM), compromisso proposto às nações pela Organização Mundial da Saúde no ano 2000 <sup>1</sup>. A redução da taxa de mortalidade infantil (TMI) até o ano 2015, uma das metas propostas pelos ODM, depende essencialmente da redução do componente neonatal precoce, atualmente responsável por mais da metade dos óbitos de crianças brasileiras no primeiro ano de vida <sup>2</sup> e estreitamente ligado a problemas na atenção à saúde da gestante e do recém-nascido <sup>3</sup>. No entanto, na complexa rede causal de determinação destes óbitos estão também envolvidos outros fatores, diretos e indiretos. Dentre esses, o peso do nascimento é a variável direta mais consistentemente associada a maior risco de morte neonatal e deve ser considerada nos estudos sobre mortalidade infantil <sup>4</sup>. A desigualdade sócio-econômica, principal fator determinante distal, tem sido relacionada a dificuldades de acesso e pior qualidade de atenção em saúde perinatal <sup>5</sup>. A adequação dos processos assistenciais pode ser avaliada por sua capacidade de interferir e reduzir os riscos de morte neonatal precoce, diretos ou indiretos. Esses processos assistenciais vão desde o atendimento individual até a organização, de forma hierarquizada e regionalizada, da rede de assistência à gestante, ao parto e ao recém-nascido. O elevado percentual de mortes nas primeiras horas de vida extra-ute-

<sup>1</sup> Secretaria Municipal de Saúde de Belo Horizonte, Belo Horizonte, Brasil.  
<sup>2</sup> Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.  
<sup>3</sup> Faculdade de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

Correspondência:  
E. Franço  
Departamento de  
Medicina Preventiva e Social,  
Faculdade de Medicina,  
Universidade Federal  
de Minas Gerais,  
Av. Alfredo Balzan 150,  
30130-100,  
Belo Horizonte, 301  
30130-100, Brasil.  
efranco@medicina.ufmg.br

## SIH-SUS e SINASC: UTILIZAÇÃO DO MÉTODO PROBABILÍSTICO PARA RELACIONAMENTO DE DADOS

*SIH-SUS and SINASC information systems: an approach using the  
probabilistic method to link databases*

Eliane de Freitas Drumond<sup>1</sup>, Elizabeth Barboza França<sup>2</sup>, Carla Jorge Machado<sup>3</sup>

### RESUMO

**Introdução:** O Sistema de Internação Hospitalar (SIH-SUS), única fonte de informações de morbidade hospitalar de abrangência nacional, é alimentado pelas autorizações de internação hospitalar (AIH), que fornecem dados de internações da rede do Sistema Único de Saúde (SUS), onde ocorrem aproximadamente 70% das internações no Brasil. O Sistema de Informação sobre Nascidos Vivos (SINASC), de base populacional, é alimentado pelas Declarações de Nascido Vivo (DN), contendo informações sobre nascimentos. Propõe-se metodologia de relacionamento probabilístico entre SINASC e SIH-SUS para avaliar a cobertura de nascidos vivos na rede hospitalar SUS em Minas Gerais. **Material e Métodos:** relacionaram-se probabilisticamente 293.215 DN da coorte de nascidos vivos em 2001 com 223.443 AIH de partos de nascidos vivos ocorridos em hospitais da rede pública mineira. Atribuíram-se valores de 0 (discordância total) a 100 (concordância total) de acordo com o percentual de concordância entre variáveis de relacionamento (nome da mãe e data de hospitalização). **Resultados:** cerca de 68% das DN foram relacionadas às AIH, percentual próximo aos 70% estimados como usuários exclusivos do SUS (a população chamada de "SUS-dependente"). **Conclusões:** o resultado satisfatório obtido nesse relacionamento entre SINASC e SIH-SUS aponta para possibilidades de realização análises a partir da combinação das variáveis de cada um dos sistemas. Novos estudos são necessários para investigar as razões pelas quais 12% das AIH não se relacionaram a qualquer DN.

### PALAVRAS-CHAVE

Sistemas de informação, relacionamento probabilístico de registros, bases de dados, partos hospitalares

### ABSTRACT

**Introduction:** The hospital admittance authorization system (SIH-SUS) is a system fed with information from hospitalizations covered by SUS, the Brazilian Unified Healthcare System, which accounts for 70% of the hospitalizations countrywide. The live birth information system (Sistema de Informação sobre Nascidos Vivos - SINASC), a population-based system, records information from deliveries occurred in

<sup>1</sup> Mestre. Gerência de Epidemiologia e Informação, Secretaria Municipal de Saúde de Belo Horizonte.

<sup>2</sup> Doutorado. Professora Adjunta do Departamento de Medicina Preventiva / UFMG.

<sup>3</sup> Ph.D. Professora Adjunta do Departamento de Demografia/FACEUFMG. End.: Av. Augusto de Lima 1378/ s/ 908, Belo Horizonte - MG. CEP: 30190-003 - e-mail: cjeschado@terra.com.br