

Eliane de Freitas Drumond<sup>I</sup>

Carla Jorge Machado<sup>II</sup>

Elisabeth França<sup>III</sup>

# Subnotificação de nascidos vivos: procedimentos de mensuração a partir do Sistema de Informação Hospitalar

## Underreporting of live births: measurement procedures using the Hospital Information System

---

### RESUMO

**OBJETIVO:** Avaliar a subnotificação de registros de nascidos vivos em sistemas de informação em saúde.

**MÉTODOS:** Foram utilizados dados secundários do Sistema de Informação Hospitalar (SIH) e Sistema de Informação sobre Nascidos Vivos (Sinasc) em municípios de Minas Gerais, 2001. A análise foi feita 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. Em ambos os casos, consideraram-se como indicadores de subnotificação as proporções de nascidos vivos informadas ao SIH e não obtidas no Sinasc. Os municípios foram posteriormente agregados em quatro faixas de tamanho populacional.

**RESULTADOS:** O relacionamento probabilístico identificou maior proporção de nascidos vivos subnotificados no Sinasc, relativamente à comparação do número de nascidos vivos nos municípios. As variações das diferenças entre os percentuais de subnotificação por procedimentos foram: 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.

**CONCLUSÕES:** A magnitude de acréscimos revelou-se sensível ao procedimento adotado. O relacionamento probabilístico incrementou certeza no pareamento e possibilitou identificação de maior proporção de casos não registrados no Sinasc, inclusive em grandes municípios. O SIH apresentou-se como importante indicador de subnotificação de nascidos vivos.

**DESCRITORES:** Nascido vivo. Sistema de registros. Sub-registro. Integração de sistemas. Estatísticas vitais.

<sup>I</sup> Gerência de Epidemiologia e Informação. Secretaria Municipal de Saúde. Belo Horizonte, MG, Brasil

<sup>II</sup> Departamento de Demografia. Faculdade de Ciências Econômicas. Universidade Federal de Minas Gerais (UFMG). Belo Horizonte, MG, Brasil

<sup>III</sup> Departamento de Medicina Preventiva e Social. Faculdade de Medicina. UFMG. Belo Horizonte, MG, Brasil

**Correspondência | Correspondence:**  
Carla Jorge Machado  
Av. Augusto de Lima 1376/ sala 908  
30190-003 Belo Horizonte, MG, Brasil  
E-mail: carla@cedeplar.ufmg.br

---

## ABSTRACT

**OBJECTIVE:** To assess underreporting of live birth records by health information systems.

**METHODS:** Secondary data of the Sistema de Informação Hospitalar (Hospital Information System – SIH) and of the Sistema de Informação de Nascidos Vivos (Information System on Live Birth – SINASC) were used in the state of Minas Gerais, Southeastern Brazil, in 2001. Two procedures were used in the analysis: the comparison between the number of live births per city and the probabilistic record linkage of individual data. For both procedures, indicators of underreporting considered were the proportion of live births presented at SIH system that were not obtained at SINASC. The municipalities were later added into four strips of population size.

**RESULTS:** The probabilistic linkage was able to identify a greater proportion of live births underreported at SINASC, relative to the comparison of live births in the municipalities. The variations of the differences among underreporting percentages per procedures were 9.4% in cities with population lower than 5,000 inhabitants; 9.1% in cities with population ranging from 5,000 and 9,999; and 8.0% in municipalities between 10,000 and 49,999 and over 50,000 inhabitants.

**CONCLUSIONS:** The amount of underreporting was sensitive to the procedures adopted. Probabilistic linkage reinforced the certainty of pairings, and also enabled to identify a greater proportion of cases not recorded at SINASC, also in greater cities. SIH was a strong indicator of underreporting of live births.

**KEY WORDS:** Live birth. Registries. Underregistration. Systems integration. Vital statistics.

---

## INTRODUÇÃO

Os sistemas de informação em saúde nacionais contêm grande volume de dados para planejamento e avaliação de políticas públicas. Sua utilização permite que os possíveis problemas referentes à qualidade da informação sejam conhecidos e solucionados.<sup>12</sup> Reconhecem-se a importância e as dificuldades para obtenção das informações sobre o número de nascidos vivos, cuja fonte oficial é o Instituto Brasileiro de Geografia e Estatística (IBGE). Porém seu uso é limitado principalmente pelo sub-registro. Para enfrentar essa questão, em 1990 o Ministério da Saúde (MS) implantou o Sistema de Informação sobre Nascidos Vivos (Sinasc)<sup>a</sup> que contém importantes variáveis para análises epidemiológicas, como peso de nascimento. Mesmo apresentando significativas melhorias,<sup>3,12</sup> ainda existem deficiências na qualidade das informações do Sinasc.<sup>b</sup> Além disso, a subnotificação de nascidos vivos reflete a incapacidade de captação desses eventos pelo sistema de saúde. Essa realidade, encarada como um

desafio na saúde pública, motiva serviços de saúde e pesquisadores<sup>12</sup> a explorarem formas de mensurar e reduzir essa subnotificação.

Há diferenças locais, regionais e estaduais na cobertura e qualidade do Sinasc,<sup>12</sup> 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 em virtude da progressiva descentralização da gestão desse sistema. Portanto, estimar e localizar a subnotificação de nascidos vivos no nível municipal permitem a adoção de ações específicas e necessárias para sua redução em nível local.

Uma das técnicas para identificar as subnotificações de nascidos vivos é o relacionamento das informações do Sinasc às informações de outros sistemas de informação em saúde.<sup>9</sup> Com esse propósito já foram

---

<sup>a</sup> Ministério da Saúde. Informações de Saúde. Nascidos Vivos/Notas Técnicas. [acesso em 27/11/2007]. Disponível em <http://tabnet.datasus.gov.br/cgi/sinasc/nvdescr.htm#atdados>

<sup>b</sup> Indicadores Básicos para Saúde no Brasil (IDB 2006). Rede Interagencial de Informações para Saúde [acesso em 16/6/2007]. Disponível em: <http://tabnet.datasus.gov.br/cgi/idb2006/c01.htm>.

utilizados, por exemplo, o Sistema de Informação da Atenção Básica (SIAB)<sup>11</sup> e o Sistema de Informações Hospitalares (SIH).<sup>15</sup>

O SIH é o único sistema de informação em saúde sobre morbidade hospitalar e contém variáveis epidemiológicas,<sup>4</sup> além de informações com finalidade contábil. Refere-se apenas a registros de internações ocorridas nos hospitais próprios e contratados/conveniados da rede do Sistema Único de Saúde (SUS). Portanto, não são cobertos pelo SIH partos domiciliares, pagos por particulares ou pela assistência complementar. Dessa forma, ao comparar os registros do Sinasc – para o qual se pretende cobertura universal – aos 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>15</sup> de nascidos vivos no Sinasc.

Com base nessa premissa, Almeida & Alencar<sup>1</sup> propuseram a utilização do número de partos do SIH como parâmetro mínimo para captação de nascidos vivos pelo Sinasc. Schramm & Szwarcwald<sup>15</sup> utilizaram os dados do SIH para estimar também a natimortalidade e a mortalidade neonatal em estados brasileiros. Em ambos os estudos, embora não seja possível 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 permitiu concluir 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 sistema de informação em saúde refere-se ao registro do mesmo indivíduo em outro sistema de informação em saúde, é necessário que eles sejam relacionados, no mínimo, por meio de uma variável identificadora. O relacionamento probabilístico permite afirmar que os registros de diferentes bases de dados pertencem ao mesmo indivíduo, com base em probabilidades de concordância e discordância das variáveis-chave. Várias estratégias de relacionamento probabilístico de dados são utilizadas<sup>5,8,9</sup> para superar a ausência de uma variável identificadora única. Frequentemente, nos relacionamentos probabilísticos, os nomes dos pacientes são utilizados como a principal variável identificadora, associados ou não a outras variáveis.<sup>5,9</sup>

Visando explorar a idéia de que diferentes procedimentos para estimar subnotificação conduzem a diferentes resultados, o presente estudo teve por objetivo avaliar a subnotificação de registros de nascidos vivos no Sinasc obtida por meio de dois procedimentos.

## MÉTODOS

Foram utilizados dados do Sinasc e do SIH que se referiam aos nascidos vivos e aos partos de nascidos vivos de residentes e ocorridos em Minas Gerais no ano 2001. Para a comparação do número de eventos, utilizaram-se dados em CD-ROM disponibilizados pelo Departamento de Informática do SUS (Datusus).<sup>4</sup> Para o relacionamento probabilístico, os dados foram obtidos na Secretaria de Estado da Saúde (SES) de Minas Gerais. Eles continham variáveis individuais identificadoras que possibilitaram a correção prévia de duplicidades e erros na especificação dos municípios de residência.

Os dados do SIH referiam-se a procedimentos autorizados em 2001 e 2002. No subdiretório MA (Movimento da AIH), selecionaram-se os registros nos quais estivesse preenchida a variável 54 (NASC\_VIVOS), que é preenchida apenas em casos de parto e que possibilitou a identificação das internações para partos de nascidos vivos. Em seguida, comparou-se o 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).

O relacionamento probabilístico foi adotado devido ao fato de não haver variável identificadora única nos dois bancos de dados. Do Sinasc foram incluídos todos os registros de nascidos vivos, residentes em Minas Gerais. Para inclusão dos registros do SIH foram utilizados três critérios: 1) internações de mulheres; 2) internação em 2001 e 3) campo 54 diferente de zero.

Foram necessários alguns ajustes nas bases de dados para a adequada execução do relacionamento probabilístico.<sup>7</sup> Para isso, os campos foram mantidos de forma idêntica nos dois arquivos. Procedeu-se à padronização das datas para que tivessem o mesmo formato (*string*) e foi criado um campo CLSS originário dos campos NOME e DTSTRING concatenados, onde o conteúdo de DTSTRING foi adicionado ao final do campo nome. Esse resultado foi armazenado no campo CLSS.

Após a preparação das variáveis, iniciaram-se as seis etapas do processo de relacionamento.<sup>7</sup> Foram criados 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 com posterior diminuição da restrição (Tabela 1). Assim, buscou-se minimizar a perda de pares, ou seja, a ocorrência de falso-negativo. Nos cinco primeiros passos manteve-se relação de uma autorização de internação hospitalar (AIH) para uma declaração de nascimento (DN). Ape-

<sup>4</sup> Ministério da Saúde. Secretaria de Vigilância em Saúde. Banco de dados dos Sistemas de Informações sobre Mortalidade (SIM) e Nascidos Vivos (Sinasc), 1997 a 2003. [CD-ROM]. Brasília; 2005.

**Tabela 1.** Passos para relacionamento dos bancos de dados do Sistema de Informação Hospitalar (SIH) e Sistema de Informação sobre Nascidos Vivos (Sinasc). Estado de Minas Gerais, 2001.

Passo	Chave de relacionamento
1	Soundex (primeiro nome) + Difdata + Municnasc + Estab
2	Soundex (primeiro nome) + Difdata + Municnasc
3	Soundex (primeiro nome) + Difdata + Municres
4	Soundex (primeiro nome) + Difdata
5	Soundex (primeiro nome)
6 (gemelares)	Soundex (primeiro nome)

nas no sexto passo foi admitida relação de uma AIH para uma ou mais DN.

As funções seguintes foram definidas para utilização nas chaves de blocagem:

1. Soundex do nome da mãe: criada para reduzir a perda de pares verdadeiros devido a problemas decorrentes de erros e/ou diferenças de grafia;
2. Difdata: criada para retornar em dias a diferença entre as datas da internação e do nascimento, uma vez que no SIH o cadastro é por internação e pode haver várias internações do mesmo indivíduo no mesmo ano; esta função visava minimizar a possibilidade de formação de pares falso-positivo, não-verdadeiros;
3. 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;
4. Municres: criada para comparar códigos de municípios de residência da mãe, retornando, igual ou diferente; visava resgatar pares verdadeiros, não-formados em passos anteriores;
5. 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>7</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 CLSS, submetido à função distância de Levenshtein.

Esse resultado foi submetido a outra função, denominada percentual de similaridade, que determinava o percentual de semelhança entre *strings* comparadas, variando de zero a 100%.<sup>7</sup> Esses percentuais de semelhança foram armazenados em uma variável nomeada SCORE.

Entre cada passo automatizado, foi realizada 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. No entanto, buscando-se incluir o maior número de pares verdadeiros, foram admitidos intervalos maiores.

Com as bases de dados provenientes do Datasus (comparação do número de eventos) e da SES (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.<sup>a</sup> Embora não tenham sido encontrados critérios para se proceder à classificação dos municípios por faixas de população, os 853 municípios foram classificados segundo quatro faixas populacionais: 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 em razão do grande número de municípios incluídos 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>3</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).

## RESULTADOS

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

<sup>a</sup> Ministério da Saúde. DATASUS. Informações de Saúde - demográficas e socioeconômicas [acesso em 3/7/2007]. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/poptmg.htm>

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 2).

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, pelo menos para 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 por meio do SIH foi 2,4%, ou seja, 699 em relação a um total corrigido no Sinasc de 28.912 de nascidos vivos (Tabela 2).

Nos municípios de porte intermediário 2 foram identificados 2.343 casos não notificados no Sinasc, com registro no SIH. O total de nascidos vivos, após correção, foi 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%.

Nos municípios de grande porte não foi identificada qualquer subnotificação no Sinasc. Ao contrário, foram identificadas menos notificações no SIH (101.440) do que no Sinasc (166.099).

**Tabela 2.** 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. Estado de Minas Gerais, 2001.

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

SIH: Sistema de Informação Hospitalar

Sinasc: Sistema de Informação sobre Nascidos Vivos

Fonte: Departamento de Informática do SUS – Datasus

**Tabela 3.** 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. Estado de Minas Gerais, 2001.

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

Fonte: Secretaria de Estado de Saúde de Minas Gerais-SES/MG

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 adotada, o percentual de nascidos vivos subnotificados no Sinasc e conhecidos apenas pelo SIH foi igual a 1,2%. Essas subnotificações ocorreram em 189 municípios, localizados principalmente nas regiões Norte de Minas e Jequitinhonha.

Pelo relacionamento probabilístico, foram obtidos 193.259 pares unívocos de nascidos vivos – 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 3 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, resultando em 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, 27.039 AIH registros do SIH não foram pareados a registros do Sinasc e, pelo critério de subnotificação adotado, foram 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 consideradas 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, nesse grupo de municípios, para 3.596 registros do SIH não se encontrou par dentre os registros do Sinasc, sendo também 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. 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 principalmente nos municípios de portes intermediário e pequeno (Tabela 3).

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 assim um total corrigido de 13.161 nascidos vivos no Sinasc nos municípios de pequeno porte, por relacionamento probabilístico (Tabela 3).

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. Nos dois procedimentos, as subnotificações de nascidos vivos foram menores nos municípios de grande porte, sendo identificadas apenas pelo relacionamento probabilístico (Tabela 4).

**Tabela 4.** 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. Estado de Minas Gerais, 2001

Variável	Faixa populacional				Total
	<5.000	5.000 - 9.999	10.000 - 49.999	>50.000	
NV não notificados					
Comparação de eventos	543	699	2.343	0	3.585
Probabilístico	1.884	3.695	10.689	10.870	27.039
NV notificados					
Comparação de eventos	11.715	28.213	92.488	166.099	298.515
Probabilístico	11.277	27.210	90.193	164.533	293.213
Total após correção					
Comparação de eventos	12.258	28.912	94.831	166.099	302.100
Probabilístico	13.161	30.305	100.882	175.403	320.252
Percentual de NV em relação ao total corrigido					
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  
NV: nascidos vivos

## DISCUSSÃO

O principal interesse para a realização do presente 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 permite, por exemplo, o cálculo da taxa de mortalidade infantil de forma direta. Esse é um passo fundamental para a adequação das políticas públicas de saúde com vistas à redução dessas mortes.<sup>a</sup>

Dentre os resultados obtidos no presente estudo, destacam-se: 1) a potencialidade ainda pouco explorada do SIH como fonte de captação de nascidos vivos para o Sinasc, evidenciada em ambos os procedimentos aqui realizados. Esse 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>4</sup> 2) as 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 refere-se 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. 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, considerar dois registros como não pertencentes ao mesmo indivíduo. 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>6</sup> Alguns pares podem ser classificados como verdadeiros quando na realidade as notificações referem-se a indivíduos diferentes. Assim, como no presente estudo, nos casos em que a base de dados na qual o par é procurado (Sinasc) possui elevado número de notificações, são sempre adotadas regras para minimizar o número de pares falso-positivo. Esse

procedimento pode, por outro lado, induzir ao aumento do número de resultados falso-negativo. Parece razoável supor, portanto, 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. Dessa 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>14</sup> No entanto, 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. Suas principais desvantagens são: impossibilidade de afirmar quando trata-se do mesmo indivíduo registrado em cada um dos sistemas de informação e 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 desse 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 compatível com o estimado pelo Ministério da Saúde (13,7%)<sup>b</sup> para Minas Gerais e corrobora a adequação dos resultados obtidos pelo relacionamento probabilístico. Também corrobora essa 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 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 era de 18,9%, sendo de 45,5% na capital e

<sup>a</sup> Organização Mundial de Saúde. Objectifs du Millénaire pour le développement [acesso em 26/6/2007]. Disponível em: <http://www.un.org/french/millenniumgoals/index.html>

<sup>b</sup> Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde. Saúde Brasil 2005: uma análise de saúde no Brasil. [acesso em 27/11/2007]. Disponível em: [http://portal.saude.gov.br/portal/arquivos/pdf/saude\\_brasil\\_2005.pdf](http://portal.saude.gov.br/portal/arquivos/pdf/saude_brasil_2005.pdf)

de 31,6% na região metropolitana.<sup>a</sup> Embora estejam excluídos do SIH os partos ocorridos em domicílio ou em trânsito, os quais são mais frequentes 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 completude 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 completude.<sup>7</sup> Porém, optou-se pela sua utilização devido à chamada “peregrinação das parturientes”,<sup>13</sup> que por receio de não serem atendidas em município diferente daquele onde residem, não declaram corretamente 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 bloqueio. A estratégia de relacionamento de revisão manual de todos os pares formados teve como um dos objetivos a redução de prováveis falso-negativos. No entanto, esse procedimento requer critérios claramente definidos e 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 do presente trabalho. No entanto, entre os aspectos relacionados à subnotificação poderiam ser citados os problemas de entendimento de conceitos como aborto, óbito fetal e nascido vivo. Mello-Jorge et al (1993)<sup>10</sup> observaram aumento da subnotificação de nascidos vivos em casos considerados como pouco viáveis (de baixo peso, por exemplo) e de neomortos que vêm a óbito nos primeiros minutos pós-parto (asfixiados, por exemplo). Outro aspecto refere-se às evasões de partos, que podem resultar em falhas na retroalimentação de dados entre os municípios.<sup>7</sup>

Finalmente, deve ser considerado em estudos baseados em dados secundários a possibilidade de viés de informação. No presente 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 completude, disponíveis desde o início dos anos 90,<sup>10</sup> mostram sua progressiva qualificação.<sup>2,3,12</sup> Veras & Martins<sup>16</sup> verificaram que a confiabilidade dos dados do SIH era superior àquela que lhe era imputada – inclusive para a variável nascido vivo. Novas análises de validade e confiabilidade dos dados do SIH contribuirão para que esse sistema possa ser cada vez mais utilizado como fonte de dados confiável e oportuna para pesquisa, planejamento e avaliação em saúde.

<sup>a</sup> Ministério da Saúde. Agência Nacional de Saúde Suplementar. Caderno de Informação de Saúde Suplementar: beneficiários, operadoras e planos. [acesso em 16/6/2007]. Disponível em: [http://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)



## REFERÊNCIAS

1. Almeida MF, Alencar GP. Informações em saúde: necessidade de introdução de mecanismos de gerenciamento dos sistemas. *Inf Epidemiol Sus.* 2000;9(4):241-9.
2. 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.
3. Andrade CL, Szwarcwald CL. Desigualdades sócio-espaciais da adequação das informações de nascimentos e óbitos do Ministério da Saúde, Brasil, 2000-2002. *Cad Saude Publica.* 2007;23(5):1207-16.
4. Bittencourt SA, Camacho LAB, Leal MC. O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva. *Cad Saude Publica.* 2006;22(1):19-30.
5. Camargo Jr KR, Coeli CM. *Reclink*: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. *Cad Saude Publica.* 2000;16(2):439-47.
6. Coutinho ESF, Coeli, CM. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevivência. *Cad Saude Publica.* 2006;22(10):2249-52.
7. Drumond EF, Machado CJ, França E. SIHSUS e Sinasc: utilização do método probabilístico para relacionamento de dados. *Cad Saude Coletiva.* 2006;14(2):251-64.
8. Machado CJ. Procedimentos para relacionamento de registros: revisão bibliográfica com enfoque na saúde infantil. *Cad Saude Publica.* 2004;20(2):362-71.
9. Machado CJ, Hill K. Relacionamento probabilístico de dados e um procedimento automático para minimizar o problema da incerteza no pareamento de registros. *Cad Saude Publica.* 2004;20(4):915-25.
10. 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 Saude Publica.* 1993;27(Supl ):1-46.
11. Mello-Jorge MHP, Gotlieb SLD. O Sistema de Informação de Atenção Básica como fonte de dados para os Sistemas de Informações sobre Mortalidade e sobre Nascidos Vivos. *Inf Epidemiol Sus.* 2001;10(10):7-18.
12. Mello-Jorge MHP, Laurenti R, Gottlieb SLD. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do Sinasc. *Cienc Saude Coletiva.* 2007;12(3):643-54.
13. 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 Saude Publica.* 2006;22(3):553-9.
14. Perpétuo IHO, França-Mendonça E. Avaliação das estatísticas de nascimentos em Belo Horizonte 1974-1994. In: Anais do Encontro Nacional de Estudos Populacionais; 1996; Belo Horizonte, Brasil. Belo Horizonte: ABEP, 1996.
15. Schramm JMA, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e natimortalidade. *Rev Saude Publica.* 2000;34(3):272-9.
16. Veras CMT, Martins MS. A confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH), Rio de Janeiro, Brasil. *Cad Saude Publica.* 1994;10(3):339-55.