

# Linkage entre registros do Sihsus e Sinasc: possíveis vieses decorrentes do não-pareamento

Eliane de Freitas Drumond\*  
Carla Jorge Machado\*\*

## Introdução

O *linkage* de bancos de dados possibilita grande incremento nas informações, com relativo baixo custo, agregando dados de mais de uma fonte. Há dois tipos de *linkages*: o determinístico, baseado na concordância exata, e o probabilístico, no qual mensura-se o grau de concordância entre dois registros em bancos de dados distintos (JARO, 1995). Espera-se que registros pertencentes ao mesmo indivíduo tenham grau de concordância superior em um conjunto de variáveis, quando comparados a registros que pertençam a diferentes indivíduos.

A acurácia dos *linkages* probabilísticos está relacionada a diversos fatores, como qualidade do preenchimento das variáveis identificadoras, selecionadas de acordo com sua capacidade discriminatória para formação dos pares (JARO, 1995), e problemas de especificação correta dos escores de pareamento. Por exemplo, o *linkage* de registros de nomes muito comuns corrobora menos para se declarar um par correto do que aquele de registros que contenham nomes raros. Logo, o escore de pareamento do primeiro par deve ser menor do que o do segundo, com base apenas no identificador nome. Mas, por diversas razões, podem ocorrer tanto o *linkage* incorreto de registros de indivíduos diferentes (falsos-positivos) quanto a não-identificação de pares verdadeiros (falsos-negativos).

Os *linkages* probabilísticos são realizados em etapas, visando minimizar a ocorrência dos erros relacionados à técnica. Após a padronização das variáveis, utiliza-se a blocagem, que é a criação de blocos lógicos de registros nos arquivos que serão relacionados. A técnica de blocagem aumenta a eficiência do *linkage* ao permitir comparar registros dos dois bancos de dados apenas quando concordam em uma variável altamente discriminatória. Em seguida, dentro dos blocos, procede-se à comparação dos registros e a cada par é dado um escore.

No Brasil, nos anos recentes, os *linkages* probabilísticos de dados têm sido cada vez mais utilizados para integração de informações provenientes dos Sistemas de Informações do Ministério da Saúde, de acordo com técnicas internacionalmente utilizadas.

Esta nota de pesquisa, que tem origem no estudo das autoras, objetiva avaliar as características da assistência ao parto na rede SUS e a provável subnotificação de nascidos vivos no Sistema de Informações sobre Nascido Vivo – Sinasc (DRUMOND; MACHADO; FRANÇA, 2006; DRUMOND; MACHADO; FRANÇA, no prelo). Para tanto, realizou-se o *linkage* dos registros das internações para partos de nascidos vivos na rede SUS (Sihsus – Sistema de Informações Hospitalares do Sistema Único de Saúde) aos registros de nascidos vivos do Sinasc de Minas Gerais, em 2001. Por meio de cinco variáveis (nome da mãe, data de internação, data de nascimento, estabelecimento de saúde e município de ocorrência), buscou-se encontrar no Sinasc pelo menos uma notificação (Declaração de Nascido Vivo – DN) correspondente a cada nascido vivo do Sihsus (poderia ser mais de uma, em caso de nascidos vivos de partos gemelares). Contudo, nem todas as notificações do Sihsus foram pareadas com sucesso (sendo este considerado o parâmetro mínimo de subnotificações no Sinasc), restando algumas AIH não pareadas a registros do Sinasc.

\* Técnica da Gerência de Epidemiologia e Informação da Secretaria Municipal de Saúde de Belo Horizonte.

\*\* Professora adjunta do Departamento de Demografia da Faculdade de Ciências Econômicas da UFMG.

Daí, a motivação desta nota de pesquisa, na qual se propõe identificar diferenças nas proporções de algumas variáveis selecionadas em dois grupos de registros – pares e não-pares. Pretende-se discutir possíveis vieses ligados ao não-pareamento de registros do Sihsus. Os vieses relacionar-se-iam ao processo e desenho do procedimento de *linkage* utilizado no estudo ou a problemas sistemáticos ligados à informação obtida nos registros, sendo, portanto, resultado de um erro sistemático que comprometeria análises realizadas a partir dos resultados do *linkage*. Este tópico – a interferência do viés nos resultados de *linkage* – tem sido pouco explorado no Brasil, o que foi constatado pela escassez de estudos na literatura nacional sobre o assunto.

## Metodologia

As fontes de dados foram obtidas junto à Secretaria de Estado da Saúde de Minas Gerais e ao Datasus, para 2001. Foi efetuado o *linkage* e, em seguida, separaram-se os registros em pareados (Sinasc-Sihsus pares) e não-pareados (Sinasc-Sihsus não-pares).<sup>1</sup> Nos grupos de registros pares e nos não-pares calculou-se a proporção de algumas variáveis, a saber: se a mãe havia tido pelo menos um óbito fetal concomitantemente ao nascido vivo de referência; se a saída do neonato do hospital havia sido por óbito ou por alta; se o número de dias de permanência no hospital havia sido superior a três; se a especialidade do médico assistente era obstetrícia ou não; o ano da apresentação da AIH para pagamento (2001 ou 2002); o caráter da internação (eletivo ou de urgência); e a natureza jurídica do hospital (contratado/conveniado, próprio, filantrópico ou universitário).

Por último, calculou-se a diferença percentual entre as proporções. Como referência, considerou-se que diferenças percentuais entre as proporções superiores a 10% seriam significativas, seguindo estudo de Ford, Roberts e Taylor (2006). Utilizaram-se, para processamento e análise dos da-

dos, os *softwares* Stata 8.0 e Excel 2000. O estudo foi aprovado pelo Comitê de Ética da Faculdade de Medicina da UFMG.

## Resultados

Foram obtidas informações de 223.443 AIH para partos de nascidos vivos. Destas, encontrou-se par para 196.404 AIH (193.259 AIH de mães de não-gemelares e 3.145 AIH de mães de gemelares). O percentual de sucesso do pareamento foi, então, de 87,9%. Estas 196.404 AIH formaram par com 199.575 declarações de nascido vivo (DN) provenientes do Sinasc (193.259 DN de não-gemelares e 6.136 DN de gemelares). Assim, estes 199.575 pares formados foram chamados de “Sinasc-Sihsus pares”. Contudo, restaram 27.039 AIH de mães de nascidos vivos cujos pares formados obtiveram baixos escores, inferiores ao ponto de corte a partir do qual foram admitidos os pares verdadeiros (12,1%). Estes foram chamados de “Sinasc-Sihsus não-pares”.

A Tabela 1 indica os números absolutos, a distribuição percentual e a diferença das proporções entre os pares formados, nos grupos “Sinasc-Sihsus pares” e “Sinasc-Sihsus não-pares”. No segundo grupo, observaram-se maiores proporções de: pelo menos um óbito fetal; saída do neonato do hospital com óbito; tempo de permanência superior a três dias; especialidade diferente de obstetrícia; internações pagas em 2002; caráter eletivo da internação; e hospital cuja natureza jurídica era contratado/conveniado ao SUS.

## Discussão

O não-pareamento de alguns registros de forma sistemática é um aspecto importante e deve ser considerado após a realização de *linkages*. Esta questão é algumas vezes negligenciada, quando da análise dos resultados do *linkage* (FORD; ROBERTS; TAYLOR, 2006).

Neste estudo, observou-se que algumas variáveis assistenciais apresentaram

<sup>1</sup> Maiores detalhamentos sobre o estudo encontram-se em Drumond, Machado, França (2006) e em Drumond, Machado, França (no prelo).

**TABELA 1**  
**Números absolutos, proporções e diferença entre proporções no Sinasc-Sihsus, segundo variáveis**  
**Minas Gerais – 2001**

Variáveis	Números absolutos		Proporções		Diferença entre proporções (4) - (3) / (4)
	SIH-SUS pares (1)	SIH-SUS não pares (2)	SIH-SUS pares (3)	SIH-SUS não pares (4)	
<b>Óbito fetal</b>					
0	199,460	26,976	99,94	99,77	-0,18
1+	115	63	0,06	0,23	75,27
<b>Saída do hospital</b>					
Transferência/alta	198,781	26,568	99,60	98,26	-1,37
Com óbito	794	471	0,40	1,74	77,16
<b>Dias de permanência</b>					
0 a 3	193,498	25,740	96,96	95,20	-1,85
4 e mais	6,077	1,299	3,04	4,80	36,62
<b>Especialidade</b>					
Obstetrícia	199,344	26,970	99,88	99,74	-0,14
Outra	231	69	0,12	0,26	54,64
<b>Ano de apresentação</b>					
2001	174,056	22,074	87,21	81,64	-6,83
2002	25,519	4,965	12,79	18,36	30,36
<b>Caráter da internação</b>					
Eletivo	19,767	3,521	9,90	13,02	23,94
Urgência	179,808	23,518	90,10	86,98	-3,58
<b>Natureza Jurídica</b>					
Contratado	48,708	7,433	24,41	27,49	11,22
Próprio	25,194	3,752	12,62	13,88	9,03
Filantropico	112,150	14,206	56,19	52,54	-6,96
Universitário	13,523	1,648	6,78	6,09	-11,17

Fonte: Sihsus/Datusus e Sinasc/SESMG.

diferenças de proporção. É possível que a maior proporção de partos com relato de óbitos fetais e de saída do neonato com óbito entre os registros Sihsus não-pares esteja refletindo a presença de problemas na qualidade da assistência ao parto e ao recém-nascido e/ou a maior vulnerabilidade social das gestantes entre os Sihsus não-pares. Pesquisas com dados brasileiros têm demonstrado que ainda é significativa a ocorrência de óbitos fetais e neonatais evitáveis. O paradoxo na prevalência de recém-nascidos de baixo peso, reconhecido fator de risco para mortalidade infantil, também tem despertado o interesse dos pesquisadores. As possíveis dificuldades de acesso e as falhas na assistência qualificada no pré-natal e no período intraparto recaem mais fortemente sobre o grupo de mulheres socialmente vulneráveis e estão relacionadas às causas de mortes evitáveis e ao excessivo nascimento de crianças de baixo peso. Assim, os resultados aqui obtidos são comparáveis aos descritos em países desenvolvidos (ZINGMOND; ZHISTEN; ETTNER; LIU, 2004), onde se observou maior concentração de registros não pareados na população tradicional-

mente considerada vulnerável socialmente. Isto constitui um problema denominado por Lawn, Cousens e Zupan (2005) de lei da informação invertida, tendo em vista que, além da maior vulnerabilidade social dos neonatos e de suas mães, suas notificações – fundamentais para adoção de políticas de saúde adequadas e oportunas – também não possuem informações que possibilitem a correta identificação em outros bancos de dados.

Por outro lado, a maior proporção de óbitos entre Sinasc-Sihsus não-pares, verificada neste estudo, poderia também estar ligada aos já reconhecidos problemas na classificação de abortos, óbitos fetais e óbitos infantis. É possível que casos de óbitos infantis tenham sido corretamente registrados no Sihsus e incorretamente considerados abortos para fins de registro em estatísticas vitais, o que tornaria desnecessária a emissão das declarações de nascidos vivos e de óbito. Neste caso, a DN, indevidamente, não constaria do banco de dados do Sinasc.

Em Minas Gerais, como em outros Estados do Brasil, partos são muitas vezes atendidos em estruturas hospitalares com

recursos técnicos e infra-estruturais inadequados. Os reflexos dessas inadequações são maiores para partos e recém-nascidos com alto perfil de gravidade. As maiores proporções de partos não assistidos por obstetras no grupo Sinasc-Sihsus não-pares pode ser resultado de deficiências de recursos humanos. Por outro lado, as internações com necessidade de permanência superior a sete dias no grupo Sinasc-Sihsus não-pares podem ser resultado de deficiências do funcionamento regionalizado e hierarquizado da rede de atendimento ao pré-natal e ao parto.

Tendo em vista o caráter jurídico do hospital na rede SUS e no Sihsus, deve ser sempre considerada a hipótese de utilização de estratégias para aumento do faturamento. Nesse estudo, tais estratégias poderiam estar relacionadas às maiores proporções de internações com AIH pagas somente

no ano seguinte e à significativa proporção de registros de internações eletivas, que deveriam ser incomuns em internações para partos.

Também deve-se ressaltar a maior proporção do grupo Sinasc-Sihsus não-pares nos hospitais contratados/conveniados. Reconhece-se, também, que nas maternidades contratadas/conveniadas são observadas as piores condições de acolhimento e assistência à gestante e de assistência ao parto.

Especialmente no Brasil, onde são reconhecidos os problemas ligados à qualidade da informação em saúde, esses resultados ressaltam a importância da avaliação de possíveis diferenças entre pares e não-pares após *linkage* de dados. A exclusão sistemática de registros pode comprometer seriamente as análises de risco.

### Referências bibliográficas

DRUMOND, E. F.; MACHADO, C. J.; FRANÇA, E. B. SIHSUS e SINASC: utilização do método probabilístico para relacionamento de dados. **Cadernos Saúde Coletiva**, 14, p. 251-264, 2006.

\_\_\_\_\_. Subnotificação de nascidos vivos: procedimentos de mensuração a partir do Sistema de Informações Hospitalares. **Revista de Saúde Pública** (no prelo).

FORD, J. B.; ROBERTS, C. L.; TAYLOR, L. K. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge. **Pediatric and Perinatal Epidemiology**, 20, p.329-337, 2006.

JARO, M. A. Probabilistic linkage of large public health. **Statistics in Medicine**, 14, p. 491-498, 1995.

ZINGMOND, D. S.; ZHISHEN, Y.; ETTNER, S. L.; LIU, H. Linking hospital discharge and death records – accuracy and source of bias. **Journal of Clinical Epidemiology**, 57, p. 21-19, 2004.

LAWN J. E.; COUSENS, S.; ZUPAN, J. 4 million neonatal deaths: when? where? why? **Lancet**, 365, p. 891-900, 2005.

Recebido para publicação em 20/12/2007.

Aceito para publicação em 25/04/2008.