Rev Saúde Pública 2007;41(Supl. 1):67-76

Ana Luiza Bierrenbach[I]

Antony Peter Stevens[I]

Adriana Bacelar Ferreira Gomes[I]

Elza Ferreira Noronha[II]

Ruth Glatt[I]

Carolina Novaes Carvalho[I]

João Gregório de Oliveira Junior[I]

Maria de Fátima Marinho de Souza[I]

Efeito da remoção de notificações repetidas sobre a incidência da tuberculose no Brasil

Impact on tuberculosis incidence rates of removal of repeat notification records

RESUMO

OBJETIVO: Avaliar o impacto nas taxas de incidência de tuberculose com a exclusão de registros indevidamente repetidos no sistema de notificação.

MÉTODOS: Foram analisados dados do Sistema de Informação de Agravos de Notificação do Ministério da Saúde, referentes ao período de 2000 a 2004. Os registros repetidos foram identificados por pareamento probabilístico e classificados em seis categorias excludentes que determinaram suas remoções, vinculações ou permanências na base.

RESULTADOS: Verificou-se que 73,7% das notificações eram únicas, 18,9% formavam duplas, 4,7% triplas e 2,7% grupos de quatro ou mais registros. Dentre os registros repetidos, 47,3% foram classificados como transferência entre unidades de saúde, 23,6% reingresso, 16,4% duplicidade verdadeira, 10% recidiva, 2,5% foram inconclusivos e 0,2% tinham dados incompletos. Essas percentagens variaram entre estados. A exclusão de registros indevidamente repetidos resultou em redução na taxa de incidência por 100.000 habitantes de 6,1% em 2000 (de 44 para 41,3), 8,3% em 2001 (de 44,5 para 40,8), 9,4% em 2002 (de 45,8 para 41,5), 9,2% em 2003 (de 46,9 para 42,6) e 8,4% em 2004 (de 45,4 para 41,6).

CONCLUSÕES: Os resultados sugerem que as taxas observadas de incidência de tuberculose representem estimativas mais próximas do que seriam os valores reais do que as obtidas com a base em seu estado bruto, tanto em nível nacional como estadual. A prática de pareamento de registros de notificação de tuberculose deve ser estimulada e mantida para melhoria da qualidade dos dados de notificação.

DESCRITORES: Tuberculose, epidemiologia. Notificação de doenças. Registros de doenças. Fontes de dados. Sistemas de Informação Brasil.

[I] Secretaria de Vigilância em Saúde. Ministério da Saúde. Brasília, DF, Brasil

[II] Faculdade de Medicina. Universidade de Brasília. Brasília, DF, Brasil

Correspondência | Correspondence:
Ana L Bierrenbach
Esplanada dos Ministérios, Bloco G
Edifício Sede, 1º andar, sala 150
70058-900 Brasília, DF, Brasil
Telefone: 061-33153496.
E-mail: ana.bierrenbach@saude.gov.br

Recebido: 30/1/2007
Aprovado: 27/3/2007

ABSTRACT

OBJECTIVE: To evaluate the impact on tuberculosis (TB) incidence rates of removal of improper duplicate records from the notification system.

METHODS: Data from the Sistema de Informação de Agravos de Notificação (Brazilian Information System for Tuberculosis Notification) from 2000 to 2004 were analyzed. Repeat records were identified through probabilistic record linkage and classified into six mutually exclusive categories and then kept, combined or removed from database.

RESULTS: Of all TB records, 73.7% had no duplicate, 18.9% were duplicate, 4.7% were triplicate, and 2.7% were quadruplicate or more. Of all repeat records, 47.3% were classified as transfer in/out; 23.6% return after default, 16.4% true duplicates, 10% relapse, 2.5% inconclusive and 0.2% had missing data. These proportions were different in Brazilian states. Removal of improper duplicate records reduced TB incidence rate per 100.000 inhabitants by 6.1% in the year 2000 (from 44 to 41.3), 8.3% in 2001 (from 44.5 to 40.8), 9.4% in 2002 (from 45.8 to 41.5), 9.2% in 2003 (from 46.9 to 42.6) and 8.4% in 2004 (from 45.4 to 41.6).

CONCLUSIONS: The study results indicate that the observed tuberculosis incidence rates represent estimates that would be closer to the actual rates than those obtained from the raw database at state and country level. The use of record linkage approach should be promoted for better quality of notification system data.

KEY WORDS: Tuberculosis, epidemiology. Disease Notification. Diseases registries. Data sources. Information Systems. Brazil.

INTRODUÇÃO

No Brasil, o Sistema de Informação de Agravos de Notificação (Sinan) é usado para coletar e processar dados sobre doenças de notificação compulsória em todo o território nacional.* A presença de registros indevidamente repetidos em um sistema de informação de saúde prejudica a correta interpretação dos dados de vigilância epidemiológica.

Para doenças crônicas como a tuberculose (TB), a geração de notificações repetidas pode decorrer de erros na entrada ou no processamento dos dados. Também, um paciente pode ser notificado repetidas vezes por unidades de saúde diferentes devido a transferências oficiais ou espontâneas entre elas durante o tratamento, ou em tratamentos distintos por recidiva após cura ou reingresso após abandono.** A presença de recidivas e reingressos é considerada legítima nessa base de dados porque são episódios novos de TB, embora correspondam à mesma pessoa. Porém, os demais registros repetidos devem ser removidos.

O objetivo do presente trabalho foi avaliar o impacto nas taxas de incidência de tuberculose, com a exclusão de registros indevidamente repetidos em sistema de notificação.

MÉTODOS

Foram utilizados os registros de notificação de TB de todo o território nacional, ocorridos de 2000 a 2004. Esses dados foram disponibilizados em fevereiro de 2006 pela Gerência Nacional do Sinan-TB, a partir da consolidação dos dados enviados pelas Secretarias Estaduais de Saúde.

A identificação de registros repetidos seguiu as etapas: 1) pré-processamento da base de dados; 2) identificação de registros pareados (*matches*) utilizando o programa de pareamento Link-Plus; 3) verificação de quais registros pareados se referiam ao mesmo indivíduo (*links*); 4) pós-processamento, com reagrupamento dos registros pertencentes ao mesmo indivíduo. Os registros pareados pertencentes ao mesmo indivíduo foram considerados registros repetidos.

* Ministério da Saúde. Secretaria de Vigilância em Saúde. Sistema de Informação de Agravos de Notificação. Normas e rotinas. Brasília; 2004. (Série A: normas e manuais técnicos).
** Ministério da Saúde. Fundação Nacional de Saúde. Tuberculose - Guia de vigilância epidemiológica. Brasília; 2002.

Durante o pré-processamento da base de dados foram feitas correções e depurações no conteúdo das variáveis "nome do paciente" e "nome da mãe do paciente", visando aumentar a chance de descobrir registros pareados. Os procedimentos adotados incluíram: 1) correção de erros óbvios de digitação; 2) eliminação ou substituição de caracteres especiais (%, /); 3) alteração da fonte dos nomes para letra maiúscula; 4) remoção de qualquer letra que estivesse isolada e de preposições dos nomes, 5) remoção de termos que indicavam a falta de conhecimento sobre o nome do paciente ou da mãe do paciente (e.g. ignorado, desconhecido).

Para a identificação de registros pareados utilizou-se o programa Link-Plus (CDC, Atlanta, Georgia, EUA),* por meio de método probabilístico para procurar registros repetidos. Por este método, desenvolvido por Fellegi & Sunter,[2] foi calculada a probabilidade de concordância e discordância das variáveis selecionadas para parear os registros (variáveis de pareamento).

Para que o programa encontre registros repetidos é necessário preparar sua configuração. As variáveis "nome do paciente", "nome da mãe" e "data de nascimento" foram adotadas como variáveis de pareamento. A variável "sexo" foi escolhida para ser a variável de blocagem, ou seja, a variável usada para repartir o arquivo em blocos menores, visando a aumentar a rapidez do processo de pareamento.

As probabilidades usadas no processo de pareamento foram extraídas pelo método indireto. Isso significa que a variedade dos registros da base Sinan-TB submetidos ao pareamento determinou o cálculo de tais probabilidades; não foram usadas as probabilidades-padrão sugeridas pelo programa ou predefinidas pelos pesquisadores.

O Link-Plus calcula uma pontuação para cada dupla de registros pareados. Quanto maior a pontuação, maior a probabilidade de a dupla ser referente ao mesmo indivíduo. Valores acima de um determinado ponto de corte dessa pontuação são considerados registros repetidos e valores abaixo do ponto de corte são considerados registros únicos. O valor seis foi escolhido como ponto de corte. Ao final do processo de pareamento, o programa emite relatórios contendo as listas de duplas de registros pareados e de registros únicos.

Três sucessivas depurações manuais foram realizadas com o objetivo de considerar como duplas de registros pareados as que tivessem ambos os registros pertencentes ao mesmo indivíduo. O desmembramento da dupla cujos registros não se referiam ao mesmo indivíduo baseou-se em várias informações e critérios. Por exemplo, a data de nascimento é frequentemente mal preenchida, pois havia inconsistências entre a data de nascimento e a idade do paciente. A presença de registros com datas de nascimento diferentes tem baixo valor preditivo negativo no reconhecimento de uma dupla de registros do mesmo indivíduo, enquanto a presença de datas iguais tem alto valor preditivo positivo. O conhecimento dos pesquisadores a respeito da formação dos nomes próprios no Brasil também foi levado em consideração. Por exemplo, o fato de ser comum famílias darem nomes parecidos aos seus filhos permitiu, pelo programa Link-Plus, o reconhecimento de registros de possíveis irmãos como pertencente a um mesmo indivíduo, e que essa dupla indevida fosse desmembrada durante a depuração manual. Quando em dúvida, os pesquisadores optaram pela alternativa conservadora de não considerar os registros pareados como registros repetidos.

As duas primeiras depurações foram realizadas utilizando somente as variáveis de pareamento e a pontuação atribuída pelo programa. A terceira depuração aconteceu após o reagrupamento dos registros repetidos; foram comparadas outras variáveis de pareamento, como município e unidade de saúde de notificação e município e logradouro de residência. Em todas essas etapas, a pontuação atribuída pelo programa serviu para determinar quais registros mereciam maior atenção na depuração dos registros pareados.

Embora o Link-Plus forneça seus resultados no formato de duplas de registros, existem duplas relacionadas entre si de maneira transitiva. Pela lógica transitiva, se o registro A está relacionado com o registro B e com o C, então os registros B e C também estão necessariamente relacionados. Conseqüentemente, A, B e C foram reagrupados como uma tripla de registros pertencente a um mesmo indivíduo, mesmo que o programa de pareamento não tivesse identificado A e C como uma dupla.

No pós-processamento das duplas de registros repetidos, foram obtidos grupos de três, quatro ou mais registros considerados como um indivíduo. O grupo com maior número de registros repetidos relacionados possuía 15 registros.

Ao final dessas etapas, os registros estavam identificados como únicos (uma notificação sem repetição), duplas (uma notificação com uma repetição), triplas (uma notificação com duas repetições) e assim por diante.

Para classificar os registros repetidos, foram comparados os valores das seguintes variáveis: número de notificação, data de notificação, data do diagnóstico, data de notificação atual, data do início do tratamento atual, data de encerramento da notificação, código do

* Centers for Disease Control and Prevention. Link Plus fact sheet. Atlanta: 2004 [Acesso em 2 set 2005]. Disponível em: http://ftp.cdc.gov/pub/Software/RegistryPlus/Link_Plus/Link%20Plus.htm

município de notificação, código de identificação da unidade de saúde de notificação, código de identificação da unidade de saúde responsável pelo acompanhamento do paciente, tipo de entrada no sistema, forma clínica e situação de encerramento.

Os registros repetidos foram classificados em seis categorias mutuamente excludentes, a saber:

- Falta de dados: registros repetidos com valores faltantes nas variáveis referentes à data de notificação e/ou ao tipo de entrada no sistema e/ou ao código da unidade de saúde de notificação.

- Duplicidade verdadeira: registros repetidos com valores idênticos (e não faltantes) na variável referente ao código do município de notificação, que apresentassem a mesma data de notificação ou com intervalo inferior a 60 dias e que fossem provenientes da mesma unidade de saúde de notificação. Havia a possibilidade da concomitância de uso de duas tabelas de códigos de unidades de saúde. Assim, os registros eram considerados provenientes da mesma unidade de saúde caso tivessem códigos iguais ou que o código em uma tabela correspondesse ao código na outra. O mapa de trocas dos códigos de unidades de saúde foi solicitado a todos os estados, mas somente por metade deles foi disponibilizado em tempo de ser incluído no estudo.

- Recidiva: registros repetidos em que as categorias assinaladas nas variáveis relativas ao tipo de entrada no sistema e/ou à situação de encerramento indicassem cura anterior.

- Reingresso: registros repetidos em que as categorias assinaladas nas variáveis relativas ao tipo de entrada no sistema e/ou à situação de encerramento indicassem que abandono anterior.

- Transferência entre unidades de saúde: registros repetidos que tivessem sido notificados por unidades de saúde diferentes e que tivessem valores nas variáveis referentes ao tipo de entrada no sistema e/ou à situação de encerramento indicando que o caso havia sido transferido. Também foram classificados como transferência entre unidades de saúde registros repetidos que, embora possuíssem códigos iguais (ou correspondentes) da unidade de saúde de notificação, tivessem um diferente código da unidade de saúde responsável pelo acompanhamento do paciente.

- Inconclusiva: não foi possível chegar a uma classificação, apesar de as variáveis utilizadas não apresentarem valores faltantes.

Os registros repetidos da categoria transferência entre unidades de saúde foram classificados como: intra-municipais quando as unidades de saúde pertenciam ao mesmo município; inter-municipais se seus registros eram de municípios diferentes do mesmo estado; e inter-estaduais se seus registros eram de estados diferentes.

A comparação dos valores e a classificação foram realizadas utilizando-se uma rotina automática escrita no programa Stata 8.2.

Após a classificação, os registros repetidos foram ou excluídos ou permaneceram na base de dados, seguindo as normas operacionais do Sinan. Assim, foram mantidos os registros classificados como recidivas, reingressos, inconclusivos. Na categoria de duplicidade verdadeira, o registro mais antigo (ou mais completo, se ambos tinham a mesma data de notificação) permaneceu. Na categoria de transferência entre unidades de saúde, os dados da ficha de notificação do registro mais antigo foram vinculados aos dados da ficha de acompanhamento do registro mais atual.* Denominou-se "completa" a base de dados contendo todos os registros notificados, e de "enxuta" aquela contendo somente os registros não excluídos.

De acordo com as orientações sobre o uso do Sinan para ações de vigilância epidemiológica,** foi considerado como caso novo de TB: 1)qualquer notificação em que a variável "entrada no sistema" estivesse preenchida com as categorias de "caso novo" ou "não sabe"; 2) a variável situação de encerramento não estivesse preenchida com a categoria de "mudança de diagnóstico".

As taxas de incidência de TB foram calculadas como número de casos novos de TB residentes em uma área diagnosticados em determinado ano, dividido pela população residente da área no mesmo ano e multiplicado por 100.000. Os dados populacionais foram provenientes do Instituto Brasileiro de Geografia e Estatística (IBGE).***

RESULTADOS

Na base de dados de notificações de TB de 2000 a 2004 havia 482.501 registros, englobando todos os tipos de entrada no sistema e todas as formas clínicas. Desses, mais de 70% eram registros únicos, e a proporção de registros únicos, duplas, triplas e grupos de quatro ou mais não apresentaram tendência nítida (Tabela 1). Para cada região brasileira, a proporção de registros únicos, duplas, triplas e grupos de quatro ou mais também variou ao longo dos anos estudados, mas para alguns estados a variação foi considerada alta.

* Ministério da Saúde. Fundação Nacional de Saúde. Tuberculose - Guia de Vigilância Epidemiológica. Brasília; 2002.
** Ministério da Saúde. Secretaria de Vigilância em Saúde. Sistema de Informações de Agravos de Notificação Normas e Rotinas. Brasília; 2004. (Série A: normas e manuais técnicos).
*** Departamento de Informática do Sistema Único de Saúde. Informações de saúde: demográficas e socioeconômicas. Brasília; 2005. [Acesso em 2 set 2005]. Disponível em: http://w3.datasus.gov.br/datasus/datasus.php?area=359A1B379C6D0E0F359G23HIJd6L26M0N&VInclude=../site/infsaude.php

Tabela 1. Número de registros notificados de cada paciente no Sistema de Informação de Agravos de Notificação-Tuberculose, segundo ano de notificação. Brasil, 2000 a 2004.

Ano	Número de registros de cada paciente								Total
	Único		Duplo		Triplo		Quádruplo ou maior		
	N	%	N	%	N	%	N	%	
2000	70.151	77,9	14.911	16,6	3.189	3,5	1.795	2,0	90.046
2001	68.975	74,2	17.071	18,3	4.353	4,7	2.620	2,8	93.019
2002	70.491	71,8	19.377	19,7	5.160	5,3	3.116	3,2	98.144
2003	72.468	71,4	20.577	20,3	5.399	5,3	3.054	3,0	101.498
2004	73.259	73,4	19.422	19,5	4.625	4,6	2.488	2,5	99.794
Total	355.344	73,7	91.358	18,9	22.726	4,7	13.073	2,7	482.501

Fonte: Sistema de Informação de Agravos de Notificação/Secretaria de Vigilância Sanitária/Ministério da Saúde (Sinan/SVS/MS)

Na Tabela 2 observa-se que em 2003 os estados com a menor e a maior proporção de registros únicos foram, respectivamente, Goiás (21,1%) e Roraima (86,9%).

A Tabela 3 apresenta a proporção anual das seis categorias de registros repetidos. A categoria de transferências entre unidades de saúde foi a mais prevalente em todos os anos, compreendendo 55,4% dos registros repetidos no primeiro ano da série e estabilizando-se em torno de 47% nos anos seguintes. A proporção de reingressos foi de 12% em 2000 e depois permaneceu estável em torno de 25%. De uma maneira geral, o número de duplicidades verdadeiras diminuiu e de recidivas aumentou ao longo do período estudado.

Do total de 32.341 registros repetidos classificados como transferências entre unidades de saúde, 40,4% correspondiam a transferências intramunicipais, 47,8% intermunicipais e 11,8% interestaduais.

A Tabela 4 apresenta a classificação dos registros repetidos notificados em 2003, por regiões e estados. Houve diferença na proporção de registros repetidos em cada categoria entre os estados, mesmo pertencentes à mesma região. Enquanto Roraima, Amazonas e Amapá apresentaram as maiores proporções de transferências entre unidades de saúde, o Acre apresentou a menor, em que pese o pequeno número de registros repetidos de alguns desses estados. Em Goiás, as duplicidades verdadeiras representavam 74% dos registros repetidos, mais do que o dobro do encontrado na Paraíba, o segundo de maior proporção nessa categoria.

A Tabela 5 compara as taxas anuais de incidência de TB entre as bases de dados completa e enxuta, ou seja, respectivamente, antes e depois da remoção das duplicidades e vinculação dos registros de casos transferidos. Houve diferenças nas taxas anuais de incidência de TB entre as bases para todos os estados ao longo do período estudado, com raras exceções. As diferenças ultrapassaram 10% em pelo menos um ano para os estados de Amapá, Goiás, Paraíba, Piauí, Rio Grande do Norte, São Paulo e Tocantins. Goiás apresentou diferenças acima de 34% em todos os anos estudados. Para o Brasil como um todo, as diferenças observadas entre as taxas de incidência variaram entre as bases, de 6,1% no ano 2000 a 9,4% no ano 2002, sem tendência nítida. Na Tabela 5 também é possível constatar diferenças nas taxas ao longo dos anos e entre regiões e estados que não podem ser explicadas pela presença de registros repetidos no banco de dados e que, portanto, não foram objeto de análise do estudo.

DISCUSSÃO

O Sinan foi desenvolvido no início da década de 90 e tem passado por diversas atualizações no sentido de corrigir suas imperfeições e de continuamente adequá-lo às novas demandas da vigilância epidemiológica. Embora todos os municípios brasileiros enviem suas notificações ao Sinan, a entrada direta de dados informatizados ocorre em cerca de 70%. A atualização das bases dos níveis hierárquicos superiores é realizada rotineiramente por meio de transferências verticais de dados. As normas de operacionalização e a definição das atribuições das três esferas de governo estão regulamentadas em documentos oficiais e estão disponíveis aos usuários.*

Coerente com as normas da vigilância epidemiológica, o Sinan dispõe de rotinas específicas para manejo de registros de pacientes de TB notificados mais de uma vez e de ferramentas próprias que facilitam a identificação de possíveis duplicidades e a realização de procedimentos para solucioná-las. Contudo, pelo

* Ministério da Saúde. Secretaria de Vigilância em Saúde. Sistema de Informação de Agravos de Notificação. Normas e rotinas. Brasília; 2004. (Série A: Normas e Nanuais Técnicos).

Tabela 2. Número de registros notificados de cada paciente no Sistema de Informação de Agravos de Notificação-Tuberculose, segundo regiões e estados. Brasil, 2003.

Região / Estado	Número de registros de cada paciente								
	Único		Dupla		Tripla		Quádrupla ou maior		Total
	N	%	N	%	N	%	N	%	
Centro-Oeste	2.864	56,2	1.414	27,8	451	8,8	365	7,2	5.094
Distrito Federal (DF)	511	75,8	126	18,7	27	4,0	10	1,5	674
Goiás (GO)	438	21,1	952	45,9	357	17,2	327	15,8	2.074
Mato Grosso do Sul (MS)	859	81,2	155	14,7	30	2,8	14	1,3	1.058
Nordeste	22.090	73,7	5.812	19,4	1.370	4,6	701	2,3	29.973
Alagoas (AL)	1.118	78,2	243	17,0	48	3,4	20	1,4	1.429
Bahia (BA)	6.500	73,1	1.792	20,2	365	4,1	232	2,6	8.889
Ceará (CE)	4.670	76,2	1.026	16,7	283	4,6	153	2,5	6.132
Maranhão (MA)	2.405	76,1	566	17,9	142	4,5	48	1,5	3.161
Paraíba (PB)	1.030	66,0	425	27,2	73	4,7	32	2,1	1.560
Pernambuco (PE)	3.881	72,6	1.029	19,2	297	5,6	142	2,6	5.349
Piauí (PI)	966	66,8	371	25,6	82	5,7	27	1,9	1.446
Rio Grande do Norte (RN)	1.001	73,8	248	18,3	67	4,9	40	3,0	1.356
Sergipe (SE)	519	79,7	112	17,2	13	2,0	7	1,1	651
Norte	6.415	76,7	1.592	19,0	251	3,0	108	1,3	8.366
Acre (AC)	290	83,6	44	12,7	12	3,4	1	0,3	347
Amazonas (AM)	1.823	73,2	563	22,6	73	2,9	31	1,3	2.490
Amapá (AP)	203	71,2	73	25,6	6	2,1	3	1,1	285
Pará (PA)	3.199	78,4	716	17,5	117	2,9	50	1,2	4.082
Rondônia (RO)	537	79,6	105	15,5	21	3,1	12	1,8	675
Roraima (RR)	172	86,9	21	10,6	3	1,5	2	1,0	198
Tocantins (TO)	191	66,1	70	24,2	19	6,6	9	3,1	289
Sudeste	32.629	70,0	9.704	20,8	2.751	5,9	1.543	3,3	46.627
Espírito Santo (ES)	1.276	84,8	176	11,7	33	2,2	20	1,3	1.505
Minas Gerais (MG)	5.173	80,3	1.001	15,5	206	3,2	65	1,0	6.445
Rio de Janeiro (RJ)	12.164	75,2	2.713	16,8	796	4,9	498	3,1	16.171
São Paulo (SP)	14.016	62,3	5.814	25,8	1.716	7,6	960	4,3	22.506
Sul	8.470	74,1	2.055	18,0	576	5,0	337	2,9	11.438
Paraná (PR)	2.617	76,3	585	17,1	135	3,9	91	2,7	3.428
Rio Grande do Sul (RS)	4.385	72,7	1.109	18,4	345	5,7	196	3,2	6.035
Santa Catarina (SC)	1.468	74,3	361	18,3	96	4,9	50	2,5	1.975
Brasil	72.468	71,4	20.577	20,3	5.399	5,3	3.054	3,0	101.498

Fonte: Sinan/SVS/MS

montante de registros repetidos existentes na base nacional do Sinan-TB, essas rotinas não devem ser executadas com a devida freqüência e/ou o devido cuidado pelos usuários do sistema, especialmente nos municípios. A execução das rotinas é prioritariamente de responsabilidade dos responsáveis pela vigilância do agravo nas esferas administrativas existentes, em colaboração com os responsáveis pela gerência do sistema de informação.[3,*]

Os resultados mostram problemas na qualidade dos dados do Sinan-TB em todos os estados brasileiros. As reduções nas taxas anuais de incidência de tuberculose

* Glatt R. Análise da qualidade da base de dados de Aids do Sistema de Informação de Agravos de Notificação (Sinan) [dissertação de mestrado]. Rio de Janeiro: Escola Nacional de Saúde Pública da FIOCRUZ; 2004.

conseqüentes ao processo de pareamento, classificação e exclusão de registros, indevidamente repetidos da base de dados do Sinan-TB, poderiam ter sido ainda maiores se não houvesse registros repetidos não classificados e se os mapas de trocas dos códigos de unidades de saúde estivessem disponíveis para todos os

Tabela 3. Classificação dos registros repetidos presentes no Sistema de Informação de Agravos de Notificação-Tuberculose, segundo ano de notificação. Brasil, 2000 a 2004.

Ano	Classificação dos registros repetidos												
	Transferência		Reingresso		Duplicidade		Recidiva		Inconclusiva		Falta de dados		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N
2000	3.985	55,4	884	12,3	2.119	29,4	102	1,4	76	1,1	28	0,4	7.194
2001	5.654	47,5	2.917	24,5	2.159	18,2	831	7,0	270	2,3	59	0,5	11.890
2002	6.903	45,0	3.670	23,9	2.831	18,5	1.510	9,9	385	2,5	37	0,2	15.336
2003	7.925	46,4	4.141	24,3	2.487	14,6	1.991	11,7	493	2,9	18	0,1	17.055
2004	7.874	46,4	4.555	26,9	1.606	9,5	2.418	14,3	478	2,8	9	0,1	16.940
Total	32.341	47,3	16.167	23,6	11.202	16,4	6.852	10,0	1.702	2,5	151	0,2	68.415

Fonte: Sinan/SVS/MS

Tabela 4. Classificação dos registros repetidos presentes no Sistema de Informação de Agravos de Notificação-Tuberculose após remoção das duplicidades e vinculação dos registros de casos transferidos, segundo região e estado. Brasil, 2003.

Região/ Estado	Classificação dos registros repetidos												
	Transferência		Reingresso		Duplicidade		Recidiva		Inconclusiva		Falta de dados		Total
	N	%	N	%	N	%	N	%	N	%	N	%	N
Centro-Oeste	298	23,1	137	10,6	753	58,4	85	6,6	15	1,2	1	0,1	1.289
DF	39	47,0	16	19,3	8	9,6	16	19,3	3	3,6	1	1,2	83
GO	164	16,8	63	6,4	725	74,0	23	2,4	4	0,4	0	0,0	979
MS	46	46,0	24	24,0	14	14,0	14	14,0	2	2,0	0	0,0	100
MT	49	38,6	34	26,8	6	4,7	32	25,2	6	4,7	0	0,0	127
Nordeste	2.187	47,7	1.021	22,3	602	13,1	607	13,3	155	3,4	10	0,2	4.582
AL	90	51,4	51	29,2	4	2,3	27	15,4	3	1,7	0	0,0	175
BA	780	56,1	291	20,9	110	7,9	137	9,9	65	4,7	7	0,5	1.39
CE	324	36,6	199	22,4	207	23,3	129	14,6	27	3,1	0	0,0	886
MA	276	58,7	90	19,2	23	4,9	72	15,3	9	1,9	0	0,0	470
PB	125	43,7	40	14,0	100	35,0	19	6,6	2	0,7	0	0,0	286
PE	365	42,1	227	26,2	104	12,0	136	15,7	32	3,7	3	0,3	867
PI	132	59,7	29	13,1	11	5,0	43	19,5	6	2,7	0	0,0	221
RN	71	32,7	70	32,3	39	18,0	27	12,4	10	4,6	0	0,0	217
SE	24	34,3	24	34,3	4	5,7	17	24,3	1	1,4	0	0,0	70
Norte	629	59,6	240	22,7	59	5,6	111	10,5	17	1,6	0	0,0	1.056
AC	2	8,0	10	40,0	0	0,0	12	48,0	1	4,0	0	0,0	25
AM	228	68,9	46	13,9	15	4,5	40	12,1	2	0,6	0	0,0	331
AP	30	63,8	13	27,7	1	2,1	2	4,3	1	2,1	0	0,0	47
PA	268	54,3	141	28,5	32	6,5	46	9,3	7	1,4	0	0,0	494
RO	38	46,9	23	28,4	11	13,6	6	7,4	3	3,7	0	0,0	81
RR	13	76,4	2	11,8	0	0,0	2	11,8	0	0,0	0	0,0	17
TO	50	82,0	5	8,2	0	0,0	3	4,9	3	4,9	0	0,0	61
Sudeste	4.051	48,4	2.242	26,7	962	11,5	880	10,5	233	2,8	4	0,1	8.372
ES	51	39,2	43	33,1	6	4,6	27	20,8	3	2,3	0	0,0	130
MG	315	48,2	158	24,1	104	15,9	51	7,8	22	3,4	4	0,6	654
RJ	954	38,4	937	37,8	228	9,2	258	10,4	105	4,2	0	0,0	2.482
SP	2.731	53,5	1.104	21,6	624	12,2	544	10,7	103	2,0	0	0,0	5.106
Sul	760	43,3	501	28,5	111	6,3	308	17,5	73	4,2	3	0,2	1.756
PR	183	37,4	140	28,6	56	11,5	85	17,4	22	4,5	3	0,6	489
RS	433	44,3	294	30,0	38	3,9	177	18,1	36	3,7	0	0,0	978
SC	144	49,8	67	23,2	17	5,9	46	15,9	15	5,2	0	0,0	289
Brasil	7.925	46,4	4.141	24,3	2.487	14,6	1.991	11,7	493	2,9	18	0,1	17.055

Fonte: Sinan/SVS/MS

Tabela 5. Taxas de incidência* de tuberculose por estado e ano de notificação nas bases de dados completa e enxuta e diferença percentual das taxas de ambas as bases. Brasil, 2000 a 2004.

Estado	Taxa de incidência - 2000			Taxa de incidência – 2001			Taxa de incidência - 2002			Taxa de incidência - 2003			Taxa de incidência - 2004		
	Completa	Enxuta	Diferença %	Completa	Enxuta	Diferença %	Completa	Enxuta	Diferença %	Completa	Enxuta	Diferença %	Completa	Enxuta	Diferença %
AC	59,4	57,8	2,7	56,6	55,9	1,2	54,3	53,5	1,5	50,1	49,1	2,0	46,2	44,8	3,0
AL	39,6	38,2	3,5	39,7	37,7	5,0	40,7	37,9	6,9	40,9	38,9	4,9	41,3	39,0	5,6
AM	73,0	72,8	0,3	81,0	79,8	1,5	73,4	72,0	1,9	68,8	66,6	3,2	72,7	69,0	5,1
AP	10,1	9,0	10,9	38,3	37,1	3,1	49,4	45,5	7,9	40,8	38,1	6,6	40,9	37,1	9,3
BA	52,8	51,2	3,0	56,5	52,3	7,4	48,0	43,9	8,5	52,9	49,4	6,6	50,2	47,2	6,0
CE	45,4	43,8	3,5	43,8	41,7	4,8	45,0	41,9	6,9	67,5	61,3	9,2	48,7	45,5	6,6
DF	17,9	17,1	4,5	16,4	15,6	4,9	15,9	15,4	3,1	17,1	16,3	4,7	15,7	14,8	5,7
ES	41,6	40,8	1,9	42,0	40,4	3,8	42,6	41,5	2,6	40,2	39,5	1,7	38,5	37,8	1,8
GO	31,1	20,5	34,1	29,2	19,1	34,6	30,5	19,2	37,0	30,0	19,1	36,3	25,5	16,7	34,5
MA	49,7	47,2	5,0	46,7	43,7	6,4	47,7	44,7	6,3	46,0	43,8	4,8	46,6	43,2	7,3
MG	0,3	0,3	0,0	7,1	6,8	4,2	29,4	28,0	4,8	29,5	27,8	5,8	29,1	27,2	6,5
MS	41,2	39,9	3,2	39,2	38,0	3,1	35,7	34,1	4,5	39,8	38,3	3,8	41,8	39,1	6,5
MT	46,8	45,0	3,8	48,4	46,8	3,3	42,2	40,5	4,0	39,8	38,4	3,5	37,1	35,3	4,9
PA	46,4	44,5	4,1	49,4	46,2	6,5	52,2	49,2	5,7	53,6	50,9	5,0	53,6	51,2	4,5
PB	37,8	34,0	10,1	33,3	30,9	7,2	33,3	31,0	6,9	33,3	31,4	5,7	33,7	31,1	7,7
PE	46,3	43,3	6,5	47,5	43,5	8,4	51,6	47,3	8,3	53,6	49,0	8,6	55,3	51,0	7,8
PI	41,1	35,8	12,9	41,6	36,8	11,5	37,9	33,3	12,1	34,6	32,0	7,5	39,2	35,5	9,4
PR	25,3	24,4	3,6	26,3	25,1	4,6	26,9	25,5	5,2	27,8	26,3	5,4	26,1	24,6	5,7
RJ	95,9	90,7	5,4	95,1	87,9	7,6	94,6	87,2	7,8	89,1	82,1	7,9	85,8	79,7	7,1
RN	40,1	39,4	1,7	38,8	37,3	3,9	41,7	38,4	7,9	39,9	36,3	9,0	41,7	37,4	10,3
RO	38,3	37,5	2,1	40,5	38,9	4,0	37,6	36,5	2,9	37,8	36,7	2,9	36,2	35,4	2,2
RR	56,1	55,8	0,5	50,1	49,8	0,6	43,0	42,4	1,4	47,3	45,6	3,6	53,3	51,9	2,6
RS	45,3	43,2	4,6	42,6	39,2	8,0	44,9	42,4	5,6	46,5	43,8	5,8	47,0	44,1	6,2
SC	24,5	23,2	5,3	25,8	23,8	7,8	28,5	26,5	7,0	28,1	26,6	5,3	27,2	26,2	3,7
SE	28,5	26,5	7,0	23,6	23,1	2,1	25,5	25,0	2,0	29,0	28,1	3,1	26,6	25,7	3,4
SP	49,9	45,8	8,2	48,0	42,6	11,3	44,1	36,8	16,6	44,4	37,5	15,5	43,8	38,2	12,8
TO	21,2	18,2	14,2	23,0	20,7	10,0	23,2	21,3	8,2	19,0	16,4	13,7	19,9	18,0	9,5
Brasil	44,0	41,3	6,1	44,5	40,8	8,3	45,8	41,5	9,4	46,9	42,6	9,2	45,4	41,6	8,4

Fonte: Sinan/SVS/MS
* Por 100 mil habitantes.

estados. Também é possível que registros repetidos não tenham sido detectados pelo programa de pareamento utilizado. Não existe um padrão-ouro que permita a averiguação da sensibilidade do programa Link-Plus. Em estudos preliminares utilizando a base de dados do Sinan (dados não publicados), a sensibilidade alcançada pelo programa Link-Plus foi comparável ao uso da metodologia de distância editorial de Levenshtein, aplicada ao nome do paciente, nome da mãe do paciente e data de nascimento.*

Por outro lado, as reduções nas taxas anuais de incidência de TB podem ter sido superestimadas se registros pareados pertencentes a indivíduos diferentes foram erroneamente considerados como repetidos. Outro fator de superestimação dessa redução seria a classificação errônea de registros repetidos como casos de duplicidade verdadeira ou transferência entre unidades de saúde. Embora possíveis, essas situações são improváveis, devido a metodologia utilizada ser conservadora.

A metodologia probabilística não exige concordância exata entre os valores das variáveis de pareamento para o pareamento de dois registros. Mas, esse aspecto não aumentou indevidamente o número de registros repetidos encontrados, visto que os registros pareados foram subseqüentemente avaliados pelo pesquisador. A rigorosa depuração manual dos registros pareados contribuiu para incrementar a especificidade sem grande prejuízo da sensibilidade de encontrar registros repetidos na base do Sinan-TB.

Quanto à classificação dos registros repetidos, apenas as recidivas, os reingressos após abandono e as transferências entre unidades de saúde localizadas em estados diferentes seriam legitimamente esperados na base de dados de nível nacional. As demais categorias

* Black PE. Levenshtein distance. In: Black PE, editor. Dictionary of algorithms and data structures. Gaithersburg: National Institute of Standards and Technology; 2005. [Acesso em 3 nov 2006]. Disponível em: http://www.nist.gov/dads/HTML/Levenshtein.html

representam falhas na operacionalização e gerenciamento do sistema de informação nas diversas instâncias responsáveis pela vigilância e controle da TB.

Foram encontrados valores faltantes nas variáveis "data de notificação", "tipo de entrada no sistema" e "código de identificação da unidade de saúde", apesar de essas variáveis serem de preenchimento obrigatório no Sinan. Isso pode ocorrer por problemas no sistema, gerando arquivos corrompidos pela utilização inadequada de outras ferramentas que acessam a base de dados original (Sinanw.GDB) e acabam por danificá-lo, ou pela utilização de sistemas paralelos por alguns estados. Os dados gerados por esses outros sistemas são enviados ao Sinan, e muitas vezes não possuem campos de preenchimento obrigatório, gerando bases incompletas.

A geração de registros da categoria duplicidade verdadeira pode ser devida à passagem do paciente por diversos profissionais numa mesma unidade de saúde após a consulta que gerou a primeira notificação, no momento da entrega da amostra para o exame de escarro ou para obtenção do medicamento. Nesses momentos, o profissional de saúde gera uma nova notificação por segurança e ambos os registros são encaminhados para digitação. Entretanto, se há qualquer diferença no valor das variáveis que compõe os campos-chave (número de notificação, data de notificação, município de notificação e unidade notificadora), o sistema não reconhece que os registros são do mesmo paciente, gerando a duplicidade.

A presença de possíveis duplicidades na base de dados pode ser verificada de duas maneiras no Sinan. A primeira, por meio de listas de notificações com os nomes dos pacientes ou de suas mães ordenados alfabeticamente. A segunda maneira é por meio de listas de possíveis duplicidades identificadas por possuírem valores idênticos em uma variável criada automaticamente pelo programa. Estas variável automática é composta pela junção do primeiro e último nome do paciente, do sexo e de sua data de nascimento. Cabe ao profissional de saúde responsável pela vigilância do agravo analisar essas listas, investigar as possíveis duplicidades contatando as unidades de saúde notificadoras, de modo a decidir seu encaminhamento adequado. Quando esses procedimentos não são realizados regularmente, as duplicidades se acumulam na base de dados em todos os níveis do sistema.

A existência de registros contendo códigos de unidade de saúde diferentes e valores idênticos nas demais variáveis analisadas foi conseqüente à introdução de uma nova tabela de códigos de unidades de saúde e a uma falha na padronização dos códigos da tabela nova. Isso fez com que registros contendo os códigos antigos não fossem substituídos por registros contendo os códigos novos quando da transferência vertical dos dados, gerando duplicidades. Após a identificação dessa falha de programação, a Gerência Nacional do Sinan encaminhou aos estados uma nota técnica explicativa e um aplicativo corretivo. Atualmente, o número de duplicidades geradas por essa falha e ainda não removidas da base de dados é pequeno. Desse modo, no presente trabalho optou-se por apresentar essa informação com os demais registros repetidos da categoria duplicidade verdadeira. Entretanto, esse aplicativo ainda não havia sido largamente utilizado em Goiás à época da extração da base de dados, resultando em 97,6% das duplicidades verdadeiras apresentadas pelo estado e influenciando suas taxas de incidência.

Em relação às transferências entre unidades de saúde entre os registros repetidos, quase 90% delas eram intramunicipais ou intraestaduais e deveriam ter tido seus registros vinculados pelo nível municipal ou estadual, respectivamente. As rotinas disponíveis no Sinan para identificação e vinculação de registros de pacientes transferidos não são executadas automaticamente. Elas exigem familiaridade com conceitos específicos relativos à vigilância do agravo e necessitam, portanto, da atuação dos responsáveis pela vigilância na gerência dos dados. Os motivos pelos quais as rotinas de vinculação não tem sido executadas devem ser investigados para que se intervenha com propriedade.

É também possível que na categoria de registros repetidos inconclusivos existam transferências entre unidades de saúde ou abandonos não reconhecidos pelo sistema de saúde e consequentemente registrados de forma adequada no Sinan. Isso implica que os técnicos responsáveis pela vigilância da tuberculose deveriam aprimorar o acompanhamento de seus pacientes e informar as unidades de saúde de origem o recebimento de um caso de transferência ou reingresso após abandono.

A comparação da qualidade dos dados da base do Sinan-TB de 2003 entre os estados deve ser criteriosa, pois a responsabilidade pela geração dos registros repetidos é compartilhada entre os níveis de gestão dos dados. Além disso, a interpretação dos dados aqui apresentados se limita à comparação da qualidade desses dados quanto à presença de registros repetidos. A análise da subnotificação de registros, da falta de completitude dos campos e inconsistência de dados e do atraso na remessa de informações não foram objeto de investigação no presente estudo, mas seriam necessárias para se completar o estudo da qualidade dos dados do Sinan-TB.

Não obstante as considerações sobre a metodologia empregada, acredita-se que as taxas anuais de incidência de TB obtidas no presente trabalho representem estimativas mais próximas do que seriam os valores reais do que as obtidas com a base em seu estado bruto, tanto em nível nacional como estadual. A prática de pareamento de registros de notificação de TB por meio da utilização das ferramentas intrínsecas do Sinan ou do uso acoplado de outros aplicativos de pareamento

deve, portanto, ser estimulada e mantida para melhoria da qualidade dos dados de notificação.[1]

O presente estudo faz parte de uma pesquisa de avaliação do Programa Nacional de Controle de Tuberculose coordenada pelo Departamento de Análise de Situação de Saúde da Secretaria de Vigilância do Ministério da Saúde. O pareamento de dados por meio da metodologia utilizada permitiu a obtenção do diagnóstico da linha de base da qualidade dos dados do Sinan-TB de 2000 a 2004 e a elaboração de uma estratégia de intervenção implementada no segundo semestre de 2005.

REFERÊNCIAS

1. 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.
2. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc*.1969; 64(328):1183-210.
3. Laguardia J, Domingues CMA, Carvalho C, Lauerman CR, Macário E, Glatt R. Sistema de Informação de Agravos de Notificação (Sinan): desafios no desenvolvimento de um sistema de informação em saúde. *Epidemiol Serv Saude*. 2004;13(3):135-46.

Nota: Ver Carta ao Editor neste Suplemento.