

Método de relacionamento de bancos de dados do Sistema de Informações sobre Mortalidade (SIM) e das autorizações de internação hospitalar (BDAIH) no Sistema Único de Saúde (SUS), na investigação de óbitos de causa mal-definida no Estado do Rio de Janeiro, Brasil, 1998

Database Linkage Method Using the Mortality Information System (SIM) and the Hospitalization Authorization Form System (BDAIH) in Brazil's Unified Health System (SUS), in the Investigation of Ill-Defined Causes of Death in Rio de Janeiro State, Brazil, 1998

Cláudio Luiz dos Santos Teixeira

Núcleo de Estudos de Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ

Carlos Henrique Klein

Escola de Saúde Pública, Fundação Instituto Oswaldo Cruz

Kátia Vergetti Bloch

Núcleo de Estudos de Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ

Cláudia Medina Coeli

Núcleo de Estudos de Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ

Resumo

O artigo descreve a metodologia utilizada para realizar o relacionamento do banco de dados do Sistema de Informações sobre Mortalidade (SIM) e das autorizações de internação hospitalar (BDAIH) no Sistema Único de Saúde (SUS) como estratégia de identificação das prováveis causas dos óbitos classificados como de causa mal-definida (OCMD), no Estado do Rio de Janeiro, Brasil, em 1998. Analisaram-se todas as declarações de óbito (DO) do SIM com causa mal-definida e que ocorreram no Rio de Janeiro, em 1998; e todas as AIH pagas nos anos de 1997 e 1998. Outro artigo tratará, exclusivamente, dos resultados obtidos nas reclassificações. A técnica foi o relacionamento probabilístico de registros, utilizando as variáveis nome, sexo e data de nascimento e o programa RecLink II. Foi possível relacionar 20% dos OCMD e 30% de uma amostra dos óbitos de causa definida (OCD), com as AIH. A confiabilidade do método de relacionamento foi elevada ($Kappa=0,93$).

Palavras-chave: causa básica de morte; comparabilidade dos dados; sistemas de gerenciamento de base de dados.

Summary

The article describes the methodology used to link the Mortality Information System (SIM) database to the Hospitalization Authorization Form System (known in Brazil by the Portuguese acronym, AIH) in Brazil's Unified Health System (SUS) as a strategy to investigate deaths classified as due to ill-defined cause in Rio de Janeiro State, Brazil, during 1998. All death certificates classified as ill-defined were analyzed, as well as AIH paid during 1997 and 1998. Another article will deal exclusively with the results obtained in the reclassification. The technique was a probabilistic linkage of records, based on the variables name, sex and date of birth information, using RecLink II software. It was possible to relate 20% of ill-defined, and 30% of a sample of deaths of defined cause, with AIH. The reliability of the relationship method was high ($Kappa=0.93$).

Key Words: *underlying cause of death; data comparability; database management systems.*

Endereço para correspondência:

Rua Anfilóquio de Lima, 78, Campos dos Goytacazes-RJ. CEP: 28051-050
E-mail: cteixeira@fmc.br

Introdução

Óbito de causa mal-definida (OCMD) ou de causa ignorada é aquele em que as causas de morte não foram registradas na declaração de óbito (DO) porque o falecido não teve nenhum tipo de assistência médica, ou por incapacidade do médico atestante em diagnosticar ou registrar, adequadamente, a causa do óbito.

A Classificação Internacional de Doenças, na sua 10ª revisão (CID-10),¹ vigente no Estado do Rio de Janeiro a partir de 1996, para o Sistema de Informações sobre Mortalidade (SIM), e a partir de 1998, para o banco de dados das autorizações de internação hospitalar (BDAIH), no seu Capítulo XVIII – Sintomas, Sinais e Achados Anormais de Exames Clínicos e de Laboratório não Classificados em Outra Parte (R00-R99), apresenta a seqüência de códigos R00-R68 (R69, só para morbidade). Esses códigos remetem a sintomas e sinais relativos aos aparelhos e a outros sintomas e sinais gerais que deveriam ser declarados como *causa mortis* apenas quando a doença desencadeadora do processo que levou à morte e/ou suas intercorrências fosse desconhecida. A seqüência de códigos R70-R94 só será utilizada para morbidade, já que não se refere a “causas” e sim a achados anormais de exames; os demais códigos, R95-R99, referem-se a causas mal-definidas e desconhecidas de mortalidade. A revisão anterior da Classificação Internacional de Doenças (CID-9) já trazia essa mesma lógica.

A mortalidade proporcional por OCMD é um indicador mais simples e direto para avaliação da qualidade da informação do sistema de registro de óbitos.^{2,3} Se a sua frequência é elevada, deduz-se que a qualidade do sistema de mortalidade está comprometida e a mortalidade por outras causas está subestimada, além de essa magnitude ser de difícil estimação.⁴

A ausência de diagnóstico para causa de morte pode ser determinada por vários fatores; entre eles, falta de assistência médica, de recursos para diagnóstico de causa básica e laudos incorretos.³

Entre os casos de OCMD, a proporção de óbitos sem assistência é interpretada como indicador da cobertura e da qualidade da assistência médica prestada à população. No Brasil, em 1995, entre os 95.446 óbitos sem causa definida, 65% ocorreram sem assistência médica.⁵ Essa assistência refere-se ao período que antecede o óbito e não ao momento em que ocorre.

Observa-se uma tendência de diminuição da proporção de OCMD, tanto em países desenvolvidos como em desenvolvimento. Nos Estados Unidos da América, em 1978, a proporção foi de 1,6%, e em 1991, de 1,1%. No Japão, em 1980, a proporção foi de 7,1%, passando para 3,7% em 1993. Suécia e Inglaterra tiveram proporções de 0,5% em 1980 e de 0,4% em 1981. Em países com grau de desenvolvimento semelhante ao do Brasil, observam-se proporções de OCMD em patamares bem inferiores ao do nosso país, com tendências de queda. O México, que, em 1978, tinha uma proporção de OCMD de 8,9%, em 1992, estava com 2,0%; e a Argentina, que contava uma proporção de 4,3% em 1978, em 1991, apresentava 2,8% de OCMD.⁶

O Brasil, no início da década de 80, apresentou proporções bem mais elevadas: em 1980, de 21,5%; e em 1994, de 17%.⁶ Mesmo entre as regiões do País, observam-se razoáveis variações; porém, sempre com tendências de queda. Nas áreas menos desenvolvidas, observam-se quedas mais acentuadas; a Região Nordeste, por exemplo, apresentou redução de 24%, ao passar de 48,6% em 1980 para 37,0% em 1994. Na Região Sul, no mesmo período, a redução foi de 37,1%, ao passar de 15,1% para 9,5%.⁶

No Brasil, em 2000, 14,3% dos óbitos foram classificados como OCMD. Apesar desse valor ser alto, houve decréscimo em relação ao ano anterior, quando a proporção foi de 15,1%.⁷

Na Região Metropolitana do Rio de Janeiro, contudo, no ano de 1980, a proporção de OCMD foi de 2,12%, passando para 4,95% em 1990 e chegando a 10,4% em 1995.⁸ A partir de então, as proporções se estabilizaram nesse patamar, sempre acima de 10%.

Oliveira, estudando a evolução da mortalidade por doenças do aparelho circulatório (DAC), doenças isquêmicas do coração (DIC) e cerebrovasculares (DCBV), no Rio de Janeiro, em São Paulo e no Rio Grande do Sul, no período de 1980 a 2000, observou declínio das mortalidades compensadas e ajustadas por DAC, DIC e DCBV, nesses três Estados. As taxas de mortalidade por DCBV apresentaram declínio constante, durante todo o período, naqueles Estados; enquanto isso, as taxas por DIC exibiram queda relevante no RJ e em quase todas as suas regiões de saúde, a partir de 1990, com aumento simultâneo das taxas por OCMD. É preciso considerar, entretanto, que as manobras de compensação das taxas de mortalidade

pelas causas específicas (DIC e DCBC) foram baseadas na inclusão de parte dos óbitos por OCMD entre os das causas definidas de interesse, de acordo com as proporções de cada grupo delas no número total de óbitos por causa conhecida.³ A suposição era de que as distribuições das verdadeiras causas de óbito entre os mal-definidos eram semelhantes às observadas entre os óbitos de causas definidas. Contudo, não havia qualquer evidência de que, entre os OCMD, a taxa de mortalidade proporcional, tanto para doença cardiovascular como para qualquer outro grupo, fosse a mesma encontrada entre os OCD. É possível que parte desse declínio, assim como da elevação dos OCMD, tenha ocorrido em razão da mudança na qualidade do registro da causa do óbito.

No Rio de Janeiro, uma medida administrativa – a Resolução nº 550, publicada pela Secretaria de Estado de Saúde (SES/RJ) em 23 de janeiro de 1990 – pode ter levado ao aumento na classificação de óbitos como OCMD, e, conseqüentemente, à redução de outras causas de óbitos. Essa resolução, ainda em vigor, determina em seu Artigo 2º:

*“Esgotadas todas as tentativas de se determinar a causa básica da morte e não havendo suspeita de óbito por causa violenta, deverá ser declarado, na parte I do atestado médico, Causa Indeterminada”.*⁹

Segundo a coordenação de dados da SES/RJ, essa medida tinha como um de seus objetivos reduzir a emissão de atestados falsificados. É possível que os óbitos classificados a partir de então como OCMD correspondam a óbitos anteriormente classificados como de causas cardiovasculares em adultos e respiratórias em crianças.

Neste trabalho, para relacionamento de DO e AIH dos casos de OCMD e de uma amostra de OCD, utilizou-se o método probabilístico de relacionamento de registros (MPRR). A principal dificuldade no relacionamento é que não existe um identificador único que permita vincular uma AIH de um indivíduo à DO que registra o seu óbito. Ainda assim, mesmo sem um identificador comum e utilizando-se outras informações – nome, sexo e data de nascimento –, que irão compor a chave de identificação, estimaram-se as possibilidades, por meio de escores, de determinados registros dos dois bancos de dados pertencerem às mesmas pessoas.

Embora o preenchimento da AIH vise à remuneração pelo atendimento prestado e o preenchimento da

declaração de óbito esteja voltado para as avaliações epidemiológicas e para as questões legais, espera-se um grau de coerência entre os dois documentos, pelo menos para alguns grupos de doenças, maior do que o atribuído ao acaso, meramente.

Metodologia

Foram utilizados os seguintes programas para organização e edição, e para aplicação do método probabilístico de relacionamento de registros: Epi Info 6.04,¹⁰ Foxbase 2.1¹¹ e Reclink II.¹²

Os dois primeiros prestaram-se a modificações das estruturas dos bancos de dados e recodificações de variáveis selecionadas. O Reclink II, por se tratar de um programa desenvolvido, especificamente, para a aplicação do método de relacionamento, foi usado nas fases de padronização, blocagem e reconstrução dos bancos após o pareamento. A mescla de alguns bancos de dados também foi feita mediante esse programa.

A mortalidade proporcional por óbitos de causa mal-definida (OCMD) é um indicador mais simples e direto para avaliação da qualidade da informação do sistema de registro de óbitos.

Os arquivos utilizados foram fornecidos no formato *data base file* (DBF) e contêm informações de identificação dos pacientes, médicas e sobre os hospitais.

Os bancos de dados, assim como as documentações pertinentes, encontram-se disponíveis em CD-ROM e em bibliotecas especializadas em saúde; ou na Internet, no *site* do Departamento de Informática do SUS (Datusus).¹³

Três arquivos deram origem a todos os outros: aquele com as informações das DO de ocorrência no Estado do Rio de Janeiro, em 1998; e os dos bancos com os registros de todas as AIH referentes aos períodos de 1998 e de 1997.

Tais bases de dados, disponíveis na Internet, por razões éticas, não são completas; elas não contêm as informações que permitem identificar os pacientes, como nome, endereço e CPF.

O banco das DO de 1998 tem 119.325 registros, de 397 caracteres para cada registro, em 77 variáveis. O banco de dados das AIH de 1998 reúne 1.002.890 registros, de 585 caracteres cada, em 71 variáveis. O banco das AIH de 1997 tem 1.061.472 registros, de 524 caracteres cada, em 64 variáveis.

A partir do banco de dados com as declarações de óbito de 1998, foram selecionados dois conjuntos de registros: um com todos os 10.692 casos de OCMD; e outro composto por uma amostra aleatória simples, de mesmo tamanho, de óbitos de causa definida, para utilização no processo de validação do método, ou seja, para avaliar a concordância entre a classificação do diagnóstico na DO e na AIH.

Os registros sem nome haviam sido excluídos previamente, de todos os bancos, por não serem passíveis de relacionamento.

O banco de dados das AIH de 1998 foi desmembrado em duas partes: uma apenas com as internações que terminaram em óbito; e outra com os pacientes que tiveram alta ou foram transferidos.

No relacionamento probabilístico, foram executados os seguintes processos: (1) cumprimento de rotinas para padronização do formato das variáveis dos bancos; (2) blocagem, que consistiu na criação de conjuntos comuns de registros, de acordo com a chave de identificação; (3) aplicação de algoritmos para comparações entre campos (por exemplo: comparação aproximada de cadeias de caracteres); (4) cálculo de escores, que resumiram o grau de concordância global entre registros de um mesmo par; (5) definição de limiares para o relacionamento dos pares de registros classificados como verdadeiros, duvidosos e não-pares; e (6) revisão manual dos pares duvidosos, visando à sua reclassificação como pares verdadeiros ou não-pares.¹⁴

Na aplicação do método probabilístico de recuperação de registros, a hierarquia de identificação utilizou-se das seguintes variáveis-chave: nome, sexo e data de nascimento, nesta ordem.

A padronização foi usada para a preparação das variáveis e diminuição de erros na fase de pareamento. De campos do tipo caractere, foram retirados acentos, cedilhas, espaços, Algarismos e outros símbolos; e feita a conversão de letras minúsculas em maiúsculas.

Nem todos os campos dos arquivos em questão (DO e AIH) seriam úteis no processo de relacionamento. Assim, promoveu-se uma seleção dos campos que,

de alguma forma, colaborassem na identificação dos pares verdadeiros. Essa decisão reduziu o tempo de processamento e o espaço de memória a ser ocupado no computador.

A blocagem consistiu na criação de blocos lógicos de registros dentro dos arquivos relacionados. O objetivo dessa etapa foi o de permitir que o processo de pareamento se fizesse de forma otimizada. A blocagem permitiu que as bases de dados fossem divididas logicamente, em blocos mutuamente exclusivos; as comparações foram limitadas aos registros pertencentes a um mesmo bloco. Os blocos foram constituídos de forma a aumentar a probabilidade de que os registros, neles contidos, representassem pares verdadeiros. O processo consistiu na indexação dos arquivos a serem relacionados, segundo uma chave formada pela combinação dos campos – nome, sexo e data de nascimento. Os registros de um determinado bloco apresentaram o mesmo valor para a chave escolhida.¹⁴

Inicialmente, foi planejada uma estratégia de blocagem constituída de cinco passos. O primeiro passo, o mais seletivo, considerou os códigos Soundex do primeiro e último nome, e o sexo. O código Soundex é um código fonético, onde pequenas diferenças, tanto na grafia como na pronúncia, geram o mesmo código; foi utilizada uma versão adaptada para o idioma Português. Seu resultado é expresso em um conjunto de quatro caracteres: o primeiro é, sempre, uma letra que corresponde à letra inicial do nome; os subsequentes são algarismos. Por exemplo: o código de João é "J000".¹⁴

Apenas nessa etapa, foram consumidas mais de 23 horas de processamento, utilizando-se um Microcomputador XP 2400MHz, com 512Mb de memória RAM, tendo sido gerado um banco com 48.968 registros. O segundo passo examinou o código Soundex do primeiro nome e o sexo; o terceiro observou o código Soundex do último nome e o sexo; o quarto, os códigos Soundex do primeiro e do último nome; e o quinto passo, o menos seletivo, considerou o ano de nascimento e o sexo.

A aplicação de algoritmos para a comparação aproximada de cadeias de caracteres consistiu na criação dos blocos, em que foram utilizados os códigos fonéticos Soundex. Para a comparação entre cadeias de caracteres, empregou-se um algoritmo baseado na distância de Levenstein, que identifica o número de operações necessárias (por exemplo: inserções,

deleções, trocas) para transformar uma cadeia de caractere na outra que se encontra em comparação.¹⁴

Os escores foram calculados considerando a probabilidade dos pares serem verdadeiros, isto é, pertencerem à mesma pessoa; ou a probabilidade de os registros pertencerem a pessoas diferentes. O escore final de cada par foi construído a partir da soma dos escores ponderados de cada campo – nome, último nome, sexo e data de nascimento –, para permitir que cada campo contribuísse, de forma diferenciada, no escore total do par. A contribuição diferenciada foi necessária, uma vez que os campos apresentavam poder discriminatório distinto e, ao mesmo tempo, probabilidades variadas de terem seus conteúdos registrados incorretamente.

Embora o preenchimento da AIH e da DO tenham objetivos de avaliação diferentes, espera-se um grau de coerência entre eles mais além do mero acaso.

O tipo de algoritmo e os parâmetros para cálculo dos escores foram os mesmos utilizados em outras pesquisas que relacionavam as mesmas bases de dados.¹⁴

O número total de pares possíveis é o produto dos registros dos dois bancos de dados. Dois dos bancos de dados a parear continham 879.539 e 10.692 registros, individualmente; logo, existiam 9.404.030.988 pares possíveis. Mesmo com a aplicação da rotina de blocagem, que reduz o número de pares formados, se não fosse estabelecido um limite a partir do qual fosse pouco provável encontrar um par verdadeiro, a seleção manual tornar-se-ia mais trabalhosa – e infrutífera, na maior parte do tempo. Os escores menores de -3 não foram aproveitados; esse ponto de corte foi adotado em todas as etapas seguintes, no cruzamento de todos os bancos de dados, e é o mesmo sugerido no tutorial do programa de relacionamento.¹⁴

Nessa fase de revisão “manual” que visava à reclassificação dos pares duvidosos como pares verdadeiros ou não-pares, eles foram analisados um a um, conforme a concordância do conteúdo dos campos.

As revisões “manuais”, não automáticas, foram feitas ao final de cada passo, desprezando-se, inicialmente, os registros com escores inferiores a -3. Os

registros restantes foram submetidos ao julgamento do pesquisador, que decidiu se cada registro correspondia a um par verdadeiro ou falso. O critério era o de descartar como par os registros duvidosos. A variável “local” e a variável “data de óbito” serviram como subsídio a esses julgamentos. Uma vez identificados os pares verdadeiros, estes foram removidos das buscas subsequentes.

O relacionamento dos OCMD com as AIH dos indivíduos com alta (vivos) em 1998 foi realizado duas vezes, em dois momentos diferentes, sem que o pesquisador soubesse que estava relacionando os mesmos pares de bancos. O banco OCMD foi duplicado e identificado com nomes diferentes. Isso possibilitou aos autores medir a concordância do método de relacionamento com o coeficiente Kappa, índice de confiabilidade usado para variáveis dicotômicas.¹⁵

Após o relacionamento, foram excluídos os pares em que o intervalo entre o óbito e a data de alta excedesse 365 dias.

A descrição detalhada de todas as etapas do relacionamento encontra-se disponível na Internet, no endereço <http://www.fmc.br/ocmd>

Considerações éticas

O projeto da pesquisa foi encaminhado ao Comitê de Ética em Pesquisa do Núcleo de Estudos de Saúde Coletiva (NESC/UFRJ), pelo processo nº 029/2003, e aprovado em 26 de junho de 2003.

Resultados

O banco das declarações de óbito de 1998 contém todos os 119.325 óbitos ocorridos no Estado, durante o ano de 1998: 12.633 (10,5%) são OCMD, 17,3% desses registros estão sem nome e 76,6% ocorreram em alguma unidade de saúde.

O banco de dados das AIH de 1998 dispõe de 1.002.890 registros, dos quais 9,0% estão sem nome, 12,7% são do tipo 5 – isto é, de longa permanência –, e, em 3,6% deles, a saída foi por óbito.

O banco de dados das AIH de 1997 conta com 1.061.472 registros, dos quais 15,3% sem nome, 14,5% do tipo 5 e 3,3% concluídos com óbito.

Tais bancos diferiam em sua estrutura, categorias de algumas variáveis e versão da CID utilizada – em 1997, ainda se adotava a CID-9.

A Figura 1 apresenta os números de registros e de excluídos e os desmembramentos de cada banco.

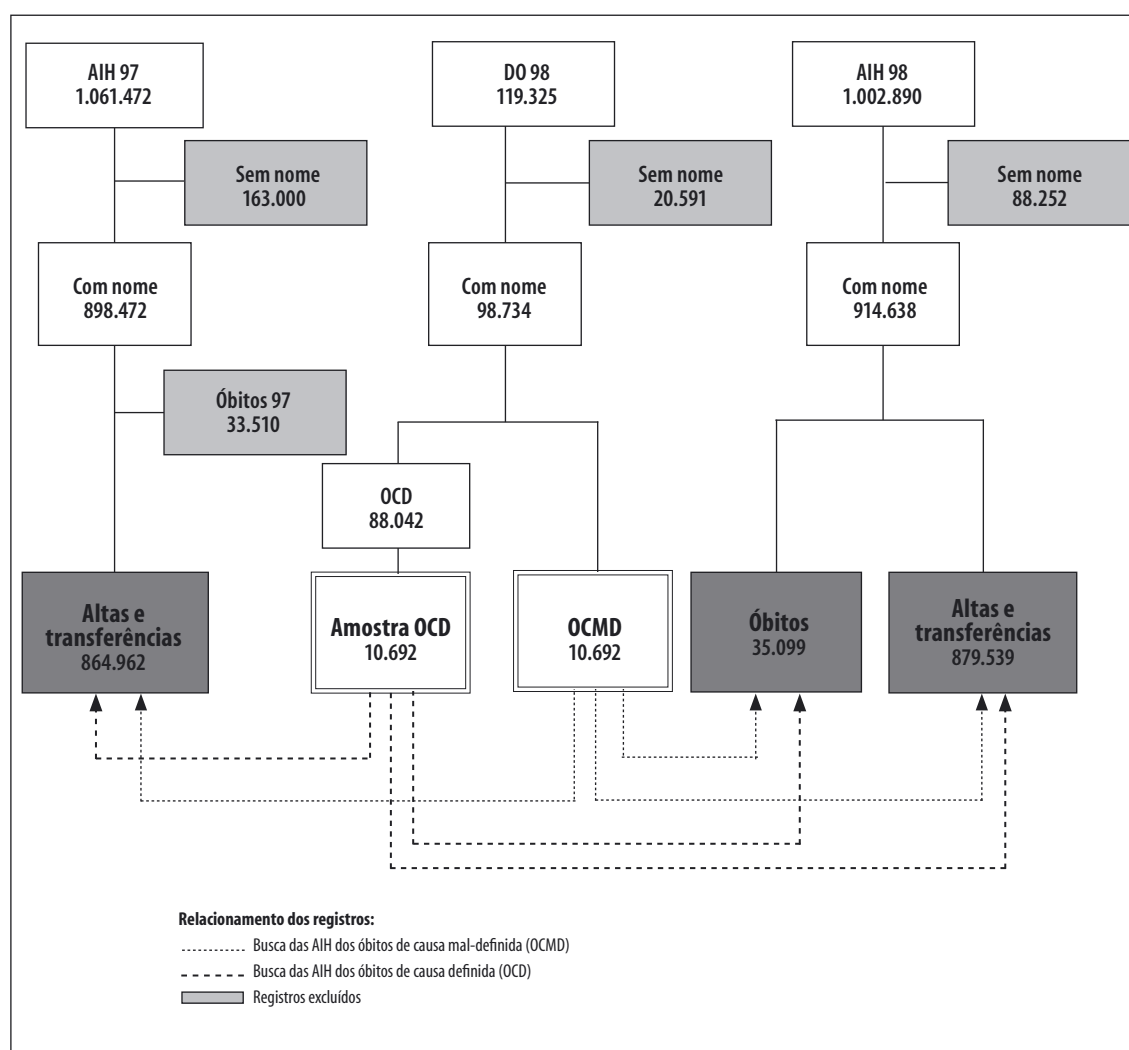
As 3.377 interações dos OCMD identificadas no banco das AIH, em realidade, referem-se a 2.133 óbitos (1,6 interações por indivíduo), que correspondem a 20% dos OCMD passíveis de relacionamento. Com respeito aos OCD, foram encontrados 3.252 óbitos em 6.030 interações (1,9 interações por indivíduo), ou 30% do total.

O número de pares identificados em cada etapa do processo de relacionamento encontra-se na Tabela 1;

excluído o quinto passo, que não identificou qualquer par. Os escores dos pares relacionados variaram de -354 a 17,22, com média de 8,4 e desvio-padrão de 3,2.

Na Tabela 2, encontram-se os valores obtidos nos relacionamentos – processos 1 e 2 – de pares idênticos de arquivos.

O grau de concordância obtido entre os dois relacionamentos, medido pelo índice Kappa, foi de 0,93. Esse valor é bastante elevado, tendo-se em conta que o valor de Kappa ajustado (por prevalência e discordantes) foi de 0,96.^{16,17}



Fonte: Ministério da Saúde, Departamento de Informática do SUS – Sistema de Informações sobre Mortalidade e Banco de Dados de Autorizações de Internação Hospitalar.

Figura 1 - Distribuição dos arquivos e desmembramento dos registros dos bancos de dados de declarações de óbito (DO) e de autorizações de internação hospitalar (AIH) ocorridas no Estado do Rio de Janeiro, Brasil, 1997 e 1998

Tabela 1 - Número de pares obtidos nos vários passos do processo de relacionamento dos arquivos das declarações de óbito (DO) e das autorizações de internação hospitalar (AIH) ocorridas no Estado do Rio de Janeiro. Brasil, 1997 e 1998

Passos	Óbitos de causa mal-definida (OCMD)		Óbitos de causa definida (OCD)	
	AIH/mortos 98	AIH/vivos 97 e 98	AIH/mortos 98	AIH/vivos 97 e 98
Passo 1	457	3.201	1.955	4.801
Passo 2	42	219	84	210
Passo 3	22	176	102	153
Passo 4	7	73	89	64
Subtotal	528	3.669	2.230	5.228
Repetições ^a	7	45	157	366
Maior que um ano ^b	0	768	0	905
Total de retirados	7	813	157	1.271
Total de pares verdadeiros	521	2.856	2.073	3.957

Retirados:

a) Repetição do número da AIH

b) Mais de um ano entre a data do óbito e a última internação

Fonte: Ministério da Saúde, Departamento de Informática do SUS – Sistema de Informações sobre Mortalidade; Banco de Dados de Autorizações de Internação Hospitalar.

Tabela 2 - Número de pares e não-pares obtidos em dois processos de relacionamento dos bancos de dados de declarações de óbito [óbitos de causas mal-definidas (OCMD)] e de autorizações de internação hospitalar (AIH98) ocorridas no Estado do Rio de Janeiro. Brasil, 1998

Relacionamentos	Processo 1		Total	
	Par	Não-par		
Processo 2	Par	1.786	93	1.879
	Não-par	121	8.692	8.813
TOTAL	1.907	8.785	10.692	

Fonte: Ministério da Saúde, Departamento de Informática do SUS – Sistema de Informações sobre Mortalidade; Banco de Dados de Autorizações de Internação Hospitalar.

Discussão

Os bancos de dados das AIH, apesar de terem a finalidade básica de propiciar a remuneração de atividades hospitalares, constituem a única fonte abrangente de informações sobre a distribuição das causas de internação no SUS.

Já haviam sido feitos estudos de confiabilidade das informações contidas em base de dados de AIH. Em es-

tudo sobre a confiabilidade dos dados nos formulários das AIH referentes ao Município do Rio de Janeiro, em 1986, observou-se alto grau de concordância, tanto das variáveis socioeconômicas como das variáveis clínicas, em que o diagnóstico (com três dígitos) apresentava um Kappa de 0,81.¹⁸

Em uma análise de 1.936 formulários de AIH, registrados com o diagnóstico principal de infarto agudo do miocárdio, no Município do Rio de Janeiro, em 1997,

observou-se concordância absoluta dos registros de prontuário com os procedimentos declarados como realizados, de 95% dos casos. Entre as variáveis demográficas e administrativas, a concordância absoluta para o Município de residência foi de 71,6%; para a data de internação, de 90,4%; para a data da alta, de 96,4%; e para a data de nascimento, de 69,8%.¹⁹

Em 1992, foram estudadas 1.595 internações referentes a uma amostra representativa das internações ocorridas nos oito hospitais gerais do Município de Maringá, Estado do Paraná, em que os diagnósticos registrados nos prontuários médicos foram comparados aos contidos nas AIH correspondentes; o Kappa variou de 0,79, para doenças do aparelho geniturinário, a 0,98, para complicações da gravidez, parto e puerpério.²⁰ No presente estudo, mesmo que 76,6% das OCMD tenham ocorrido em unidades de saúde, os óbitos que ocorreram em atendimentos não pagos pelo SUS (tanto em unidades conveniadas como não conveniadas), não foram considerados, pois não geraram AIH. Não é possível estimar o percentual de relacionamentos obtidos quando se consideram apenas as internações do SUS, porque o modelo de DO não especifica se a internação aconteceu pelo sistema SUS ou pelo sistema privado.

Óbitos de pessoas que não foram internadas pelo sistema SUS não podem ser relacionados utilizando-se as bases de dados da AIH. As bases de dados das internações no sistema exclusivamente privado, que não geraram AIH, são fragmentadas por instituição e não se encontram disponíveis para pesquisa. Ademais, este trabalho utilizou apenas as bases de AIH pagas, por estarem disponíveis e serem mais fidedignas; porém, mesmo nas instituições prestadoras de serviços ao SUS, podem ocorrer internações atribuídas ao sistema que não constam das bases analisadas (por glosa ou não-apresentação da AIH para faturamento); assim como os atendimentos nas emergências dos hospitais, que não são registrados nas bases dos formulários de AIH. Também não se dispõe de estimativas sobre a cobertura das internações pagas pelo SUS e a sua relação com o total de internações, que inclui as glosadas, no ano de 1998, no Estado do Rio de Janeiro.

Neste estudo, nem todos os campos contribuíram, da mesma maneira, para a seleção dos pares; houve uma grande preocupação quanto à uniformidade no critério de seleção dos pares verdadeiros.

Entre os OCMD, alguns nomes e sobrenomes apresentaram uma capacidade discriminatória muito baixa: contando-se os sobrenomes, constatou-se que 17% eram "SILVA"; e entre os nomes, 12% eram "MARIA". Entre os OCD, os percentuais foram semelhantes.

A complexidade do processo cresce à medida que o número de registros a serem relacionados aumenta, tornando necessária a utilização de rotinas automatizadas para a sua execução.

Entre as dificuldades encontradas, destacamos o aparecimento de pares repetidos, atribuídos à presença de 135 registros com o mesmo identificador no banco da DO. Esses registros apresentam-se em seqüência, no final desse banco.

Com a crescente disponibilidade de grandes bases de dados em saúde informatizadas, além da maior capacidade de processamento dos computadores, o interesse pelo relacionamento de registros em diferentes bases vem aumentando, nas últimas décadas. Essas bases são empregadas, muitas vezes, para monitorar a ocorrência de eventos de interesse, como os óbitos em estudos de coorte,^{21,22} ou com o objetivo de combinar bases distintas, com informações complementares.^{23,24}

Nas últimas duas décadas, encontramos dezenas de exemplos da aplicação desse método,^{20,23,25-28} usado para vários tipos de população, tanto de adultos^{23,25,28-31} como de crianças.^{26,32-35}

No processamento de todos os conjuntos de arquivos, no primeiro passo, já citado, foram relacionados quase 90% dos pares; estes autores sugerem, portanto, que, na rotina de serviço, o trabalho possa-se restringir ao primeiro passo, para economia de tempo e demais recursos.

Encontramos quase quatro vezes mais pares relacionados entre os OCD, comparativamente aos OCMD, quando nos restringimos às AIH com saídas de óbitos; ao considerar as AIH de vivos, essa relação caiu para cerca de 1,4 vezes maior nas OCD. Esse desequilíbrio já era esperado; explica-se, em parte, pela omissão das causas de óbito entre os OCMD. Igualmente, poder-se-ia esperar que os 5% [521 pares relacionados no primeiro passo, divididos por todos os 10.633 óbitos mal-definidos, multiplicados por 100 (Tabela 1)] dos OCMD com AIH gerada na ocasião imediatamente anterior à morte tivessem causa definida ou, pelo menos, algum indício do motivo de óbito. Outrossim, como se observou que o índice de internações dos OCD

foi maior que o dos OCMD (1,9 e 1,6 por indivíduo, respectivamente), é possível especular que aqueles que morreram sem causa de óbito definida tiveram menor cobertura de atenção médico-hospitalar, e, por essa razão, menor probabilidade de causa de óbito identificada.

O tempo gasto no processamento global de relacionamento depende, fundamentalmente, do número de registros dos bancos, da complexidade da chave de identificação – ou seja, da quantidade de variáveis usadas para pareamento – e da necessidade de julgamentos não automáticos.

Só poderão constituir pares verdadeiros as DO e as AIH referentes a indivíduos com o seu nome registrado corretamente, nas duas bases, e que tiveram as suas internações custeadas pelo SUS. Porém, ainda que essas condições sejam satisfeitas, restará uma limitação do método, decorrente das coincidências de identidades de indivíduos: mesmo nome, sexo e data de nascimento. O número único de saúde deverá dirimir, parcialmente, essa limitação, mas ainda ficarão fora do âmbito de relacionamento todos os indivíduos que tiverem morrido sem AIH da internação que resultou em óbito, que venha a ser paga, ou que não tenham sido internados pelo SUS anteriormente (no período de até um ano). Observa-se que apenas 1/5 dos OCD foi relacionado pelo método com AIH de internações que terminaram em óbito. Isso significa que os demais – a grande maioria – tiveram assistência médica na circunstância do óbito, sem geração de AIH. Uma parte deles, todavia, pode ter sido relacionada por ter sido gerada AIH em internação anterior ao óbito, de até um ano.

A identificação de todos os casos de OCMD deve ser, portanto, uma meta inatingível para o atual estágio do sistema de informações sobre morbidade e mortalidade. Por fim, o método heurístico utilizado, que envolveu pelo menos um certo grau de decisão arbitrária por parte do pesquisador, na sua determinação de formar pares, pode ser considerado bastante satisfatório no aspecto relacionado à confiabilidade, uma vez que o Kappa foi muito elevado e próximo do padrão ideal.

É possível especular que aqueles que morreram sem causa de óbito definida tiveram menor cobertura de atenção médico-hospitalar, e, por essa razão, menor probabilidade de causa de óbito identificada.

A validação do método de identificação das causas dos OCMD será objeto de um próximo artigo, no qual apresentar-se-ão resultados obtidos na reclassificação das causas dos OCD, inclusivamente.

Apesar de o método adotado ser trabalhoso e exigir computadores de grande velocidade e capacidade de armazenamento, foi possível relacionar 20% dos OCMD com AIH e 30% de OCD.

Entre os obstáculos à utilização dessa fonte de informação, destacam-se:

- a má qualidade das informações registradas, como por exemplo, nomes em branco, sexo mal codificado e número de DO repetidas; a SES-RJ relata que “até 1999, o sistema DO tinha um arquivo principal e um de complemento (onde eram digitados os nomes), alguns Municípios com o SIM descentralizado não encaminhavam à Secretaria Estadual de Saúde o arquivo complemento”;
- a grande variação na estrutura dos bancos, que impede comparações diretas; e
- o tamanho considerável dos bancos, que implica grande tempo de processamento.

A qualidade da informação pode ser aprimorada por meio do ensino nas escolas médicas e da implementação de estratégias de educação continuada nos serviços de saúde.

A padronização dos bancos e a inclusão de uma variável identificadora, comum às diversas bases de dados, são fundamentais para a expansão dessa metodologia de relacionamento de bases de dados, que permita, de forma relativamente rápida e barata, o acesso às informações disponíveis, obtidas em diferentes passagens do usuário pelo sistema de saúde.

Referências bibliográficas

1. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. Geneva: WHO; 1993.
2. Paula AMC, Filho DE, Pereira IPA, Albano AHBL, Fernandes RM. Avaliação dos dados de mortalidade, Brasil – 1979 a 1989. Informe Epidemiológico do SUS 1994;3(1):21-31.
3. Oliveira GMM. Mortalidade cardiovascular no Estado do Rio de Janeiro no período de 1980 a 2000 [tese de Doutorado]. Rio de Janeiro (RJ): Universidade Federal do Rio de Janeiro; 2003.
4. Vermelho L, Costa AJL, Kale PL. Indicadores de Saúde. In: Medronho RA. Epidemiologia. 1a ed. São Paulo: Ateneu; 2002. p.33-55.
5. Vasconcelos AMN. Estatísticas de mortalidade por causas: uma avaliação da qualidade da informação. Anais do 10o Encontro de Estudos Populacionais; 1996; Caxambu, Brasil. Rio de Janeiro: Associação Brasileira de Estudos Populacionais; 1996.
6. Laurenti, R. As condições de saúde no Brasil – 1997. Rio de Janeiro: Editora Fiocruz; 2000.
7. Laurenti R. Mortalidade Brasil 2000: Sistema de Informação Sobre Mortalidade (SIM) 1996 a 2000 [CD-ROM]. Brasília: Datasus; 2000.
8. Reis ACGV. Mortalidade por causas mal-definidas na Região Metropolitana do Rio de Janeiro, de 1980 a 1995 [dissertação de Mestrado]. Rio de Janeiro (RJ): Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz; 1998.
9. Estado do Rio de Janeiro. Secretaria de Estado de Saúde. Resolução no 550, de 23 de janeiro de 1990. Alteração no preenchimento da Declaração de Óbito. Diário Oficial do Estado do Rio de Janeiro, Rio de Janeiro, 26 de jan. 1990.
10. Dean AG, Dean JA, Burton AH, Dicker RC. Epi Info, version 6.04d: A word processing database, and statistics program for epidemiology on microcomputers. Atlanta: Centers for Disease Control and Prevention; 1990.
11. Bordland Visual dBASE V 5.5, 1995. California: Borland Software Corporation, Scotts Valley; 1995.
12. Camargo JRK, Coeli CM. RecLink: aplicativo para relacionamento de base de dados, implementando o método probabilistic record linkage. Cadernos de Saúde Pública 2000;16(2):439-47.
13. Ministério da Saúde. Informações de saúde [dados na Internet]. Brasília: Ministério da Saúde [acesso 2003 jul. 2]. Disponível em: <http://www.saude.gov.br>
14. Camargo JRK, Coeli CM. ReclinkII: manual do usuário [dados na Internet] [acesso 2003 jul. 6]. Disponível em: <http://planeta.terra.com.br/educacao/kencamargo/ReclinkII.html>
15. Fleiss JL. Statistical methods for rates and proportions. New York: Wiley; 1981.
16. Byrt BJ, Carlin JB. Bias, prevalence and kappa. Journal Clinical of Epidemiology 1993;45(5):423-9.
17. Lantz CA, Nebenzall E. Behavior an interpretation of the statistic: resolution of the two paradoxes. Journal Clinical of Epidemiology 1996;49(4):431-4.
18. Veras CMT, Martins MAS. Confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH) – Rio de Janeiro, Brasil. Cadernos de Saúde Pública 1994;10(3):339-55.
19. Escosteguy CC, Portela MC, Medronho RA, Vasconcellos MTL. O Sistema de Informações Hospitalares e a assistência ao infarto agudo do miocárdio. Revista de Saúde Pública 2002;36(4):491-9.
20. Mathias TAF, Soboll MLMS. Confiabilidade de diagnósticos nos formulários de autorização de internação hospitalar. Revista de Saúde Pública 1998;32:526-32.
21. Rogot E, Sorlie P, Johnson NJ. Probabilistic methods in matching census samples to the National Death Index. Journal of Chronic Diseases 1986;39:719-34.
22. Van Den Brabdt PA, Schouten L, Goldbohm RA, Dorant E, Hunen PMH. Development of a record linkage protocol for use in the Dutch Cancer Registry for epidemiological research. International Journal of Epidemiology 1990;19:553-8.
23. Newcombe HB, Smith ME, Howe GR, Mingay J, Strugnell A, Abbatt JD. Reliability of computerized versus manual death searches in a study of the health of Eldorado uranium workers. Comput Biological Medicine 1983;13(3):157-69.
24. Dean JM, Vernon DD, Cook L, Nechodom P, Reading J, Suruda A. Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: a potential tool for evaluation of emergency medical services. Annals of Emergency Medical 2001;37(6):616-26.

25. Bopp M, Minder CE, Swiss National Cohort. Mortality by education in German speaking Switzerland, 1990-1997: results from the Swiss National Cohort. *International Journal of Epidemiology* 2003;32(3):346-54.
26. Herman AA, Mccarthy BJ, Bakewell JM, Ward RH, Mueller BA, Maconochie NE, Read AW, Zadka P, Skjaerven R. Data linkage methods used in maternally-linked birth and infant death surveillance data sets from the United States (Georgia, Missouri, Utah and Washington State), Israel, Norway, Scotland and Western Australia. *Paediatric and Perinatal Epidemiology* 1997;11 Suppl 1:5-22.
27. Smith R, Cook LJ, Olson LM, Reading JC, Dean JM. Trends of behavioral risk factors in motor vehicle crashes in Utah, 1992-1997. *Accident Analysis and Prevention* 2004;36(2): 249-55.
28. Coeli CM. Vigilância do diabetes mellitus em uma população idosa: aplicação da metodologia de captura-recaptura [tese de Doutorado]. Rio de Janeiro (RJ): Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro; 1998.
29. Melo ECP. Infarto agudo do miocárdio no Município do Rio de Janeiro: qualidade dos dados, sobrevivência e distribuição espacial [tese de Doutorado]. Rio de Janeiro (RJ): Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz; 2004.
30. Goldberg MS, Carpenter M, Theriault G, Fair M. The accuracy of ascertaining vital status in a historical cohort study of synthetic textiles workers using computerized record linkage to the Canadian. *Canadian Journal of Public Health* 1993; 84(3):201-4.
31. Brown MH, Weinberg M, Chong N, Levine R, Holowaty E. A cohort study of breast cancer risk in breast reduction patients. *Plastic and Reconstructive Surgery* 1999;3(6):1674-81.
32. Wen SW, Joseph KS, Kramer MS, Demissie K, Oppenheimer L, Liston R, Allen A. Fetal and infant mortality study group, Canadian Perinatal Surveillance System. Recent trends in fetal and infant outcomes following post-term pregnancies. *Chronic Diseases Canadian* 2001;22(1):1-5.
33. Fair M, Cyr M, Allen AC, Wen SW, Guyon G, Macdonald RC. An assessment of the validity of a computer system for probabilistic record linkage of birth and infant death records in Canada. The fetal and infant health study group. *Chronic Diseases Canadian* 2000;21(1):8-13.
34. Soares, EP. Associação entre peso ao nascer e mortalidade infantil no Município de Campos dos Goytacazes, RJ [dissertação de Mestrado]. Rio de Janeiro (RJ): Universidade Federal do Rio de Janeiro, Núcleo de Estudos em Saúde Coletiva; 2003.
35. Machado, CJ. Procedimentos para relacionamento de registros: revisão bibliográfica com enfoque na saúde infantil. *Cadernos de Saúde Pública* 2004;20(2):362-71.