Acurácia do relacionamento probabilístico de registros na identificação de óbitos em uma coorte de pacientes com insuficiência cardíaca descompensada

Accuracy of probabilistic record linkage for identifying deaths in a cohort of patients with decompensated heart failure

Exactitud de vinculación probabilística de registros para identificar muertes en una cohorte de pacientes con insuficiencia cardiaca descompensada

Pedro Pimenta de Mello Spineti [1,2]
Andrea Silvestre de Souza [2,3]
Luiz Augusto Feijó [2]
Marcelo Iorio Garcia [2,4]
Sergio Salles Xavier [2,3]

Resumo

O relacionamento probabilístico de registros vem sendo cada vez mais empregado na identificação de desfechos em estudos de coorte. O objetivo deste trabalho foi avaliar a acurácia deste método na identificação de óbitos em uma coorte de 450 pacientes admitidos em um hospital universitário por insuficiência cardíaca descompensada, em um período de seis anos. O estado vital dos membros da coorte foi determinado a partir dos registros no prontuário eletrônico dos pacientes (padrão-ouro). O software OpenRecLink foi utilizado para relacionar os registros da coorte com aqueles da base do Sistema de Informações de Mortalidade, visando à identificação de óbitos. Apenas 53,6% pacientes apresentavam estado vital conhecido ao final do seguimento e destes 59,3% haviam falecido. O método apresentou sensibilidade de 97,9%, especificidade de 100%, valor preditivo positivo de 100%, valor preditivo negativo de 97% e acurácia de 98,8%. Esses resultados sugerem que o relacionamento probabilístico de registros é uma valiosa ferramenta na identificação de óbitos para estudos de coorte.

Registro Médico Coordenado; Registros de Mortalidade; Base de Dados

[1] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.*
[2] *Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.*
[3] *Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*
[4] *Hospital Pró-Cardíaco, Rio de Janeiro, Brasil.*

Correspondência
P. P. M. Spineti
Rua Henrique Stamile Coutinho 430, apto. 302, Rio de Janeiro, RJ 22795-200, Brasil.
pedrospineti@yahoo.com.br

Introdução

A insuficiência cardíaca é a via final comum da maioria das cardiopatias, sendo um importante desafio clínico na área da saúde. Trata-se de um problema em progressão no Brasil e no mundo [1]. Estudos de coorte têm sido conduzidos em todo o mundo para melhor compreender sua epidemiologia [2,3,4].

Perdas de seguimento são uma importante fonte de viés para estudos de coorte. Pacientes não localizados ao longo do estudo podem ter desenvolvido a doença investigada, não ter sido diagnosticados, ou até mesmo falecido. Diversas alternativas têm sido empregadas na tentativa de reduzir esta perda de seguimento como: uma melhor seleção dos indivíduos incluídos na pesquisa, contatos frequentes através de consultas de revisão, telefonemas, cartas ou telegramas [5]. A possibilidade da identificação de desfechos como morte ou internação através do relacionamento entre bancos de dados administrativos e epidemiológicos vem sendo cada vez mais utilizada por pesquisadores na área da saúde [6].

Esta tarefa é muito simples quando existe uma chave de identificação comum entre os diferentes bancos de dados de interesse como o número do prontuário, número do registro no cadastro de pessoas físicas (CPF) ou o número do cartão nacional de saúde (CNS). Embora o número da CNS seja um dos campos da Declaração de Óbito (DO), este dado é pouco preenchido, não podendo ser utilizado como chave de identificação no relacionamento com a base de dados do Sistema de Informações sobre Mortalidade (SIM). Na ausência deste identificador comum o relacionamento probabilístico de dados é uma alternativa [7].

O principal objetivo do relacionamento probabilístico de registros é encontrar pares de registros que se referem a uma mesma pessoa. Isto é possível através do pareamento das bases de dados utilizando-se uma combinação de campos com dados pessoais e calculando-se, para cada um deles, razões de verossimilhança positivas ou negativas, nas situações em que, respectivamente, concordam ou discordam [6].

Normalmente, são usadas para a identificação de indivíduos variáveis como: nome, nome da mãe, sexo, endereço e data de nascimento. Informações adicionais como estado civil, escolaridade, município de residência, entre outras, podem ser utilizadas, dependendo da qualidade do seu preenchimento.

O objetivo do presente estudo foi avaliar a acurácia do relacionamento probabilístico de registros utilizado para identificação de óbitos no SIM de uma coorte de pacientes admitidos com insuficiência cardíaca descompensada em um hospital universitário, empregando-se como padrão-ouro as informações de estado vital registradas no prontuário eletrônico dos pacientes (PEP).

Metodologia

Para este estudo utilizou-se uma coorte retrospectiva de 450 pacientes admitidos com insuficiência cardíaca descompensada em um hospital universitário, na cidade do Rio de Janeiro, Brasil, no período compreendido entre 1º de janeiro de 2006 e 31 de dezembro de 2011. Os pacientes foram seguidos até 31 de dezembro de 2012 de forma a garantir um período de seguimento mínimo de um ano para o último paciente incluído na amostra. Determinou-se o estado vital dos pacientes nesta data através da revisão do prontuário eletrônico, sendo considerado como padrão-ouro. Óbito foi determinado através do registro de óbito no PEP. Foram considerados vivos os pacientes que apresentaram registro de passagens no PEP (consultas de revisão, visitas à emergência e re-hospitalizações) posteriores a data final do seguimento.

Foi realizado o relacionamento probabilístico do banco de dados original da pesquisa com o banco de dados do SIM contendo as informações referentes a 890.898 declarações de óbitos do Estado do Rio de Janeiro entre 1º de janeiro de 2006 e 31 de dezembro de 2012. As bases de dados de mortalidade foram obtidas junto ao Departamento de Dados Vitais, Secretaria de Estado de Saúde do Rio de Janeiro.

Este relacionamento foi realizado utilizando-se o programa OpenRecLink na plataforma Windows 7, 64 *bits* da Microsoft Corp. Foram selecionadas quatro variáveis (nome, nome da mãe, sexo e data de nascimento) para os procedimentos de blocagem e comparação dos dados.

Inicialmente procedeu-se à padronização das variáveis e à quebra em componentes dos campos nome, nome da mãe e data de nascimento. Foi utilizada uma estratégia de relacionamento em 14 passos (Tabela 1), empregando-se como chaves de blocagem a combinação dos seguintes campos: Soundex do primeiro nome (PBLOCO), Soundex do último nome (UBLOCO), Soundex do primeiro nome da mãe (PBLOCOMAE), Soundex do último nome da mãe (UBLOCOMAE), sexo e ano de nascimento [8].

Como regra de classificação, no intuito de minimizar a ocorrência de falsos-positivos e o tempo dispensado na revisão manual, foi utilizado primeiramente o passo com o numero máximo de variáveis (6) e, em seguida, todas as

Tabela 1

Chaves de blocagem utilizadas em cada passo na estratégia de relacionamento probabilístico entre os bancos: prontuário eletrônico do paciente e declarações de óbito.

Passo	Chave de blocagem
1	Sexo + PBLOCO + UBLOCO + PBLOCOMAE + UBLOCOMAE + ano de nascimento
2	Sexo + PBLOCO + UBLOCO + PBLOCOMAE + UBLOCOMAE
3	Sexo + PBLOCO + UBLOCO + PBLOCOMAE + ano de nascimento
4	Sexo + PBLOCO + UBLOCO + UBLOCOMAE + ano de nascimento
5	Sexo + PBLOCO + PBLOCOMAE + UBLOCOMAE + ano de nascimento
6	Sexo + UBLOCO + PBLOCOMAE + UBLOCOMAE + ano de nascimento
7	PBLOCO + UBLOCO + PBLOCOMAE + UBLOCOMAE + ano de nascimento
8	Sexo + PBLOCO + UBLOCO + ano de nascimento
9	Sexo + PBLOCO + PBLOCOMAE + ano de nascimento
10	Sexo + UBLOCO + UBLOCOMAE + ano de nascimento
11	Sexo + PBLOCOMAE + UBLOCOMAE + ano de nascimento
12	PBLOCO + UBLOCO + PBLOCOMAE + UBLOCOMAE
13	Sexo + PBLOCO + UBLOCO
14	PBLOCO + UBLOCO + ano de nascimento

PBLOCO: Soundex do primeiro nome; PBLOCOMAE: Soundex do primeiro nome da mãe;
UBLOCO: Soundex do últim nome; UBLOCOMAE: Soundex do último nome da mãe.

combinações possíveis com cinco variáveis. Para um número menor de variáveis, não foram esgotadas todas as possibilidades (seriam necessárias mais 50 combinações com quatro, três e duas variáveis).

Para o cálculo dos escores empregaram-se os campos: nome completo, nome completo da mãe e data de nascimento, que foram comparados utilizando-se algoritmos baseados na distância de Levenshtein para nome e nome da mãe e na comparação caractere a caractere para data de nascimento [9]. Foram utilizados valores de parâmetros de pareamento sugeridos por Camargo Jr. & Coeli [9]. Para os passos 8, 13 e 14 não foi utilizado o nome completo da mãe para o cálculo dos escores, porque tinham por objetivo identificar pares em que o nome da mãe estivesse ausente em uma ou ambas as bases.

Todos os pares obtidos no primeiro passo de blocagem foram revisados manualmente por um único avaliador cego para o estado vital dos pacientes. Nos passos seguintes, para reduzir o tempo gasto na revisão manual, o mesmo avaliador revisou somente os pares com escore acima de nove, pois não foram identificados pares verdadeiros ou duvidosos abaixo deste valor no primeiro passo. Os pares com escore abaixo de nove foram descartados.

Um conjunto de critérios foi previamente estabelecido para classificação dos pares como verdadeiros ou falsos. Estes critérios consideraram não somente a concordância dos nomes, nomes das mães e datas de nascimento como a raridade de nomes e sobrenomes. Os campos endereço (logradouro e bairro) e estado civil foram utilizados adicionalmente no processo de revisão manual para confirmar ou excluir pares duvidosos. O campo data da alta hospitalar, presente apenas no banco da coorte, foi utilizado para excluir pares duvidosos quando a data do óbito era inferior à data da alta hospitalar.

Utilizando-se o estado vital avaliado através dos registros no PEP como padrão-ouro, estimaram-se a sensibilidade, a especificidade e os valores preditivos positivo (VPP) e negativo (VPN) para os dados obtidos por meio do relacionamento dos bancos.

O presente trabalho é um subestudo do projeto *Insuficiência Cardíaca Descompensada (ICD): Análise do Perfil Etiológico, Preditores Prognósticos e Impacto da Clínica de Insuficiência Cardíaca na Qualidade e Abordagem Diagnóstica e Terapêutica*, tendo sido registrado e aprovado pelo Comitê de Ética em Pesquisa do Hospital Universitário Clementino Fraga Filho, Universidade Federal do Rio de Janeiro, sob o número 065/09.

Resultados

Dos 450 pacientes incluídos na coorte, apenas 241 (53,56%) apresentavam seu estado vital conhecido em 31 de dezembro de 2012, sendo considerados para análise. Destes, 143 (59,34%) faleceram até 31 de dezembro de 2012. O relacionamento com a base de dados do SIM identificou 140 óbitos.

Na Tabela 2 encontram-se os dados de ocorrência de óbito, obtidos pelo prontuário eletrônico (padrão-ouro) e pelo relacionamento probabilístico. O relacionamento de dados não identificou três óbitos (falsos-negativos) e não classificou nenhum paciente incorretamente como tendo falecido (falso-positivo). A sensibilidade foi de 97,9% e a especificidade foi de 100%. O VPP e VPN foram iguais a 100% e 97%, respectivamente. A acurácia foi de 98,8%.

Na Tabela 3 é apresentado o número total de pares formados e de pares identificados como verdadeiros em cada passo da estratégia de relacionamento. Os passos 1, 2, 4, 8 e 13 identificaram, em conjunto, 123 pares verdadeiros (87,85%). Os passos 10, 11 e 12 não identificaram nenhum par. Na Tabela 4 são descritos a sensibilidade, especificidade, o VPP, o VPN e a acurácia acumulativos a cada etapa.

Os três casos falsos negativos foram identificados através de busca retrospectiva no banco do SIM pela data conhecida do óbito. No primeiro caso, houve a troca de uma consoante no primeiro nome do paciente fazendo com que o Soundex do primeiro nome nas duas bases fosse diferente, além disto, o nome da mãe estava ausente no registro do PEP. Como todos os passos continham o Soundex do primeiro nome ou um dos dois componentes do nome da mãe, este par não foi gerado.

No segundo caso faltava um sobrenome no nome da paciente no banco do registro, o que fez com que o Soundex do último nome fosse distinto entre os bancos. Além disto, houve uma troca no primeiro nome da mãe, alterando o Soundex do primeiro nome da mãe e, como todos os passos continham Soundex do último nome ou o Soundex do primeiro nome da mãe, este par não foi gerado.

No terceiro caso, houve a troca de uma consoante do primeiro nome, o que alterou o Soundex do primeiro nome e um erro no último dígito do ano de nascimento. Como todos os passos continham o Soundex do primeiro nome ou ano de nascimento, este par também não foi gerado.

Discussão

Os resultados demonstram que a estratégia de relacionamento probabilístico de dados adotada demostrou excelente acurácia, apresentando especificidade de 100% e sensibilidade acima de 90%. Os resultados estão entre os melhores descritos na literatura [6] (sensibilidade 74 a 98% e especificidade 99% a 100%) com sensibilidade superior a encontrada em outros estudos nacionais [7,10,11,12,13].

Dois dos três falsos negativos foram atribuídos a erros de digitação em pelo menos dois campos utilizados na estratégia de blocagem. O terceiro falso negativo foi fruto da associação de um erro de digitação com a ausência de um dos campos utilizados para blocagem.

Silveira & Artman [6] sugerem que acurácia em estudos de relacionamento probabilístico é altamente dependente da quantidade de campos e qualidade dos registros utilizados para o relacionamento. Sousa et al. [14], Capuani et al. [10] e Coutinho et al. [13] utilizaram estratégias de pareamento de três campos em três passos. Coutinho & Coeli [7] e Migowski et al. [12] relacionaram quatro campos em cinco passos. Fonseca et al. [11], por sua vez, relacionou cinco campos em três passos. No presente estudo, foram relacionados seis campos em 14 passos de pareamento, o que pode ter contribuído para melhor sensibilidade da estratégia.

Além dos seis campos utilizados no pareamento, outros quatro campos (logradouro e bairro de residência, estado civil e data da alta hospitalar) utilizados para identificar pares verdadeiros entre os pares considerados duvidosos, após a comparação dos campos de blocagem, também podem ter contribuído para melhor acurácia da estratégia empregada.

A qualidade da informação entre os bancos de estudo utilizados para comparação com o SIM também pode ter contribuído para as diferenças observadas entre o presente e os demais estudos nacionais. Coutinho et al. [13] atribui a baixa sensibilidade encontrada em seu estudo, em parte, ao elevado percentual de registros sem informação no Sistema de Informações sobre Nascidos Vivos (SINASC). O presente estudo também difere dos estudos anteriores por ter empregado o software OpenRecLink, enquanto os demais artigos citados utilizaram o RecLink II [7,13,14] e III [10,11,12].

Apesar de poder contribuir para uma melhor sensibilidade, o grande número de passos utilizados na estratégia de relacionamento utilizada pode inviabilizar sua aplicação em bancos de dados maiores, em virtude do tempo e custo de processamento necessários. Cinco dos 14 passos foram responsáveis por identificar mais de 85% dos óbitos e três passos não identificaram

Tabela 2

Acurácia na identificação de óbitos através do relacionamento probabilístico com a Declaração de Óbito (DO) em relação à informação do prontuário eletrônico dos pacientes (PEP) (padrão-ouro).

Relacionamento probabilístico (DO)	Padrão-ouro (PEP)		
	Óbito	Vivo	Total
Óbito	140	0	140
Vivo	3	98	101
Total	143	98	241

Nota: Sensibilidade: 97,9%; Especificidade: 100%; Valor preditivo positivo: 100%; Valor preditivo negativo: 97%; Acurácia: 98,8%.

Tabela 3

Número de pares identificados conforme o passo do relacionamento probabilístico entre os bancos: prontuário eletrônico do paciente e Declarações de Óbito.

Passo	Intervalo de escores		Número de pares		
	Total	Pares selecionados	Totais	Escore > 9	Verdadeiros
1	17,21 a 7,19	17,21 a 14,74	239	237	85
2	16,69 a 2,4	16,69 a 15,94	395	171	8
3	16,87 a 7,11	16,87 a 13,81	94	41	5
4	16,95 a 6,9	16,95 a 14,47	179	143	10
5	16,85 a 4,16	16,85 a 15,65	137	30	1
6	16,96 a 3,27	16,96 a 13,92	174	70	3
7	17,21 a 7,16	17,21 a 15,96	10	9	5
8	10,69 a 4,04	10,69 a 10,17	1.365	61	12
9	16,27 a 4,16	16,27 a 11,62	2.674	380	2
10	15,15 a 3,21	-	4.443	1.565	0
11	8,59 a 2,07	-	1.726	0	0
12	14,02 a 2,34	-	234	99	0
13	10,69 a 1,77	10,69 a 9,93	52.836	169	8
14	10,69 a 4,04	10,69 a 10,64	1.149	17	1
Total			**65.655**	**2.992**	**140**

nenhum par verdadeiro. Ao se comparar a sensibilidade cumulativa dos cinco primeiros passos da estratégia utilizada (76,2%) com os resultados de outros estudos nacionais que empregaram estratégias de relacionamento em até cinco passos [7,10,11,12,13], observa-se que somente o estudo de Coutinho et al. [13] apresentou sensibilidade inferior (72,8%). Isto sugere que o maior determinante da sensibilidade do relacionamento probabilístico não seria o número de passos de blocagem, mas sim, a combinação de campos utilizada em cada passo e que uma melhor seleção dos passos de blocagem poderia minimizar o tempo empregado no relacionamento, sem apresentar grande impacto no resultado final.

Conclusão

O relacionamento probabilístico de dados realizado através do programa OpenRecLink apresentou uma boa acurácia na identificação de óbitos, em um estudo de coorte de pacientes portadores de insuficiência cardíaca com informações provenientes de prontuário eletrônico do paciente.

Tabela 4

Sensibilidade, especificidade, valor preditivo positivo (VPP), valor preditivo negativo (VPN) e acurácia conforme o passo do relacionamento probabilístico entre os bancos: prontuário eletrônico do paciente e declarações de óbito.

Passo	Sensibilidade (%)	Especificidade (%)	VPP (%)	VPN (%)	Acurácia (%)
1	59,4	100,0	100,0	62,8	75,9
2	65,0	100,0	100,0	66,2	79,2
3	68,5	100,0	100,0	68,5	81,3
4	75,5	100,0	100,0	73,7	85,5
5	76,2	100,0	100,0	74,2	85,9
6	78,3	100,0	100,0	76,0	87,1
7	81,8	100,0	100,0	79,0	89,2
8	90,2	100,0	100,0	87,5	94,2
9	91,6	100,0	100,0	89,1	95,0
10	91,6	100,0	100,0	89,1	95,0
11	91,6	100,0	100,0	89,1	95,0
12	91,6	100,0	100,0	89,1	95,0
13	97,2	100,0	100,0	96,1	98,3
14	97,9	100,0	100,0	97,0	98,8

Contribuíram para os resultados encontrados a boa qualidade dos registros no PEP, a utilização de seis variáveis para blocagem, uma estratégia de relacionamento em múltiplos passos e o uso de quatro variáveis auxiliares para exclusão de pares duvidosos.

Os resultados sugerem, ainda, que o maior determinante da sensibilidade encontrada foi a combinação de campos utilizada em cada passo de blocagem. Uma nova estratégia de relacionamento empregando apenas os passos que identificaram um maior número de pares poderia minimizar o tempo de processamento, sem apresentar grande impacto no resultado final.

Colaboradores

P. P. M. Spineti participou da concepção e projeto do estudo, interpretação dos dados e redação do artigo. A. S. Souza participou da interpretação dos dados, revisão crítica do conteúdo intelectual e aprovação final da versão a ser publicada. L. A. Feijó e M. I. Garcia participou da concepção e projeto do estudo, interpretação dos dados e revisão crítica do conteúdo intelectual. S. S. Xavier participou da concepção e projeto do estudo, interpretação dos dados, revisão crítica do conteúdo intelectual e aprovação final da versão a ser publicada.

Agradecimentos

A Paula Dias Maia, Luiza Lapolla Perruso, Eliene Ferreira Salles e Patricia Ferreira pelo trabalho de coleta de dados e a Flavia Carneiro da Cunha Oliveira pela colaboração na revisão e editoração do artigo.

Referências

1. Bocchi EA, Braga FGM, Ferreira SMA, Rohde LEP, Oliveira WA, Almeida DR, et al. III diretriz brasileira de insuficiência cardíaca crônica. Arq Bras Cardiol 2009; 93(1 Suppl 1):1-71.
2. Fonarow GC, Adams Jr. KF, Abraham WT, Yancy CW, Boscardin WJ; ADHERE Scientific Advisory Committee, Study Group, and Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 2005; 293:572-80.
3. O'Connor CM, Abraham WT, Albert NM, Clare R, Gattis Stough W, Gheorghiade M, et al. Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). Am Heart J 2008; 156:662-73.
4. Albuquerque DC, Souza Neto JD, Bacal F, Rohde LEP, Bernardez-Pereira S, Berwanger O, et al. I registro brasileiro de insuficiência cardíaca – aspectos clínicos, qualidade assistencial e desfechos hospitalares. Arq Bras Cardiol 2015; 104:433-42.
5. Hunt JR, White E. Retaining and tracking cohort study members. Epidemiol Rev 1998; 20:57-70.
6. Silveira DP, Artmann E. Acurácia em métodos de relacionamento probabilístico de bases de dados em saúde: revisão sistemática. Rev Saúde Pública 2009; 43:875-82.
7. Coutinho ESF, Coeli CM. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevida. Cad Saúde Pública 2006; 22:2249-52.
8. Coeli CM, Camargo Jr. KR Avaliação de diferentes estratégias de blocagem. Rev Bras Epidemiol 2002; 5:185-96.
9. Camargo Jr. KR, Coeli CM. OpenRecLink: guia do usuário. http://sourceforge.net/projects/reclink/files/guiausuario.pdf/download (acessado em 10/Jul/2014).
10. Sousa MH, Cecatti JG, Hardy E, Serruya SJ. Relacionamento probabilístico de registros: uma aplicação na área de morbidade materna grave (*near miss*) e mortalidade materna. Cad Saúde Pública 2008; 24:653-62.
11. Capuani L, Bierrenbach AL, Abreu F, Takecian PL, Ferreira JE, Sabino EC. Accuracy of a probabilistic record-linkage methodology used to track blood donors in the Mortality Information System database. Cad Saúde Pública 2014; 30:1623-32.
12. Fonseca MGP, Coeli CM, Lucena FFA, Veloso VG, Carvalho MS. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. Cad Saúde Pública 2010; 26:1431-8.
13. Migowski A, Chaves RB, Coeli CM, Ribeiro AL, Tura BR, Kuschnir MC, et al. Accuracy of probabilistic record linkage in the assessment of high complexity cardiology procedures. Rev Saúde Pública 2011; 45:269-75.
14. Coutinho RGM, Coeli CM, Faerstein E, Chor D. Sensitivity of probabilistic record linkage for reported birth idetification: Pró-Saúde Study. Rev Saúde Pública 2008; 42:1097-100.

Abstract

Probabilistic record linkage has been used increasingly to identify outcomes in cohort studies. This study aimed to assess the method's accuracy for identifying deaths in a cohort of 450 patients admitted to a university hospital for decompensated heart failure over a six-year period. Vital status of cohort members was determined from electronic patient file data (gold standard). OpenRecLink software was used to link cohort records with those from the Mortality Information System, aimed at identifying deaths. Only 53.6% of patients had vital status known at the end of follow-up, and 59.3% of these had died. The method showed 97.9% sensitivity, 100% specificity, 100% positive predictive value, 97% negative predictive value, and 98.8% accuracy. The results suggest probabilistic record linkage as a valuable tool for identifying deaths in cohort studies.

Medical Record Linkage; Mortality Registries; Database

Resumen

La vinculación probabilística de registros es cada vez más empleada para identificar los resultados de los estudios de cohortes. El objetivo de este estudio fue evaluar la exactitud de este método en la identificación de las muertes en una cohorte de 450 pacientes ingresados en un hospital universitario por insuficiencia cardíaca descompensada en un período de seis años. El estado vital de integrantes de la cohorte se determinó a partir de los datos en los registros médicos electrónicos de pacientes (patrón oro). El software OpenRecLink fue utilizado para relacionar los registros de la cohorte con los de la base del Sistema de Información sobre Mortalidad, dirigido a la identificación de las muertes. Sólo el 53,6% de los pacientes presentaban estado vital conocido al final del seguimiento y de éstos el 59,3% habían muerto. El método tuvo una sensibilidad de un 97,9%, una especificidad de un 100%, valor predictivo positivo de un 100%, valor predictivo negativo de un 97% y exactitud de un 98,8%. Estos resultados sugieren que la vinculación probabilística de registros es una herramienta valiosa para la identificación de las muertes en los estudios de cohortes.

Registro Médico Coordinado; Registros de Mortalidad; Base de Datos

Recebido em 21/Jun/2015
Versão final reapresentada em 01/Nov/2015
Aprovado em 16/Nov/2015