



Universidade Federal do Rio de Janeiro

Instituto de Estudos de Saúde Coletiva

**Comparação dos Parâmetros de Relacionamento entre as bases das  
Autorizações de Internação Hospitalar (AIH) e das Declarações de Óbitos (DO)**

---

**Flávia dos Santos Barbosa**

Co-autores:

Cláudia Medina Coeli

Katia Vergetti Bloch

Rejane Sobrino Pinheiro

Alexandre dos Santos Brito

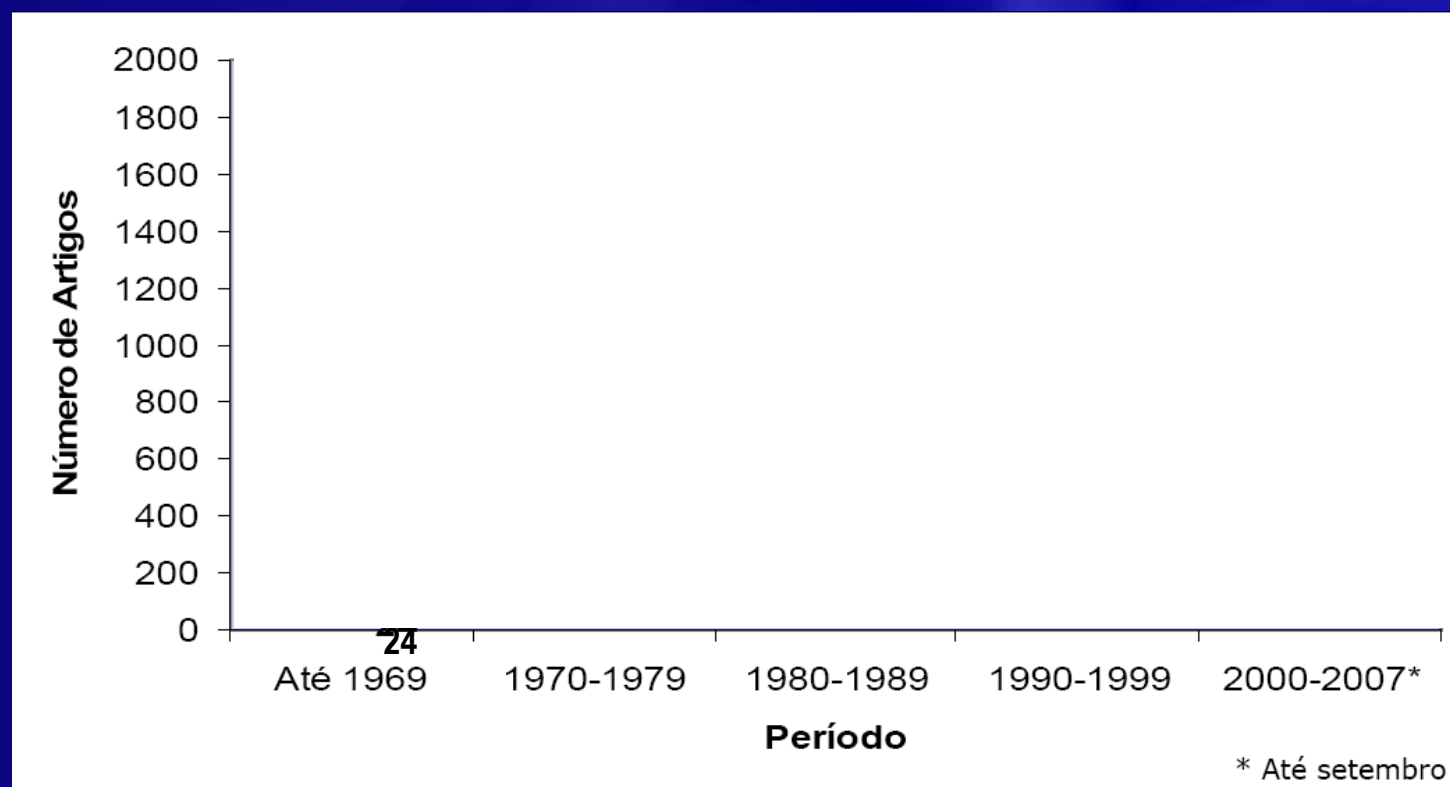
# Introdução

# Introdução

O interesse em relacionar registros em diferentes bases de dados aumentou nas últimas décadas juntamente com a crescente disponibilidade de grandes bases de dados em saúde informatizadas (Coeli, 2004; Bopp & Minder, 2003; Camargo Jr & Coeli, 2000; Newcombe et al., 1983).

# Introdução

## Medline 1960-2007



Fonte: [www.nates.ufjf.br/novo/encontro/xaps/pdf/palestras/usointegrado.pdf](http://www.nates.ufjf.br/novo/encontro/xaps/pdf/palestras/usointegrado.pdf)

# Introdução

O Brasil possui grandes bases de dados de abrangência nacional, comparáveis às que existem em diversos países. Anualmente, é produzido um grande volume de dados, amplamente disponíveis via internet pelo Datasus/MS, cujo percentual de aproveitamento deixa ainda a desejar (Pinheiro et al. 2006).

# Introdução

Campo comum que estabelece uma relação direta entre os diversos sistemas de informação em saúde

Na ausência de um **campo identificador**, o relacionamento probabilístico representa uma alternativa metodológica eficiente para preencher esta lacuna oferecendo como vantagens a rapidez e o aumento da acurácia do processo de relacionamento (Coutinho & Coeli, 2006; Camargo Jr & Coeli, 2000).

# Introdução

Países como Canadá e a Austrália vêm desenvolvendo experiências bem sucedidas de criação de registros de saúde amplos, integrando várias bases de dados epidemiológicos e administrativos, que são utilizados para a condução de estudos epidemiológicos e a avaliação de políticas públicas, por meio do método probabilístico (Roos et al., 2004; Holman et al., 1999).

# Objetivo

# Objetivo

Comparar os valores dos parâmetros de relacionamento entre as bases das Autorizações de Internação Hospitalar (AIH) e das Declarações de Óbitos (DO) considerando o subconjunto de óbitos de causas mal-definidas (OCMD) e o óbitos de causas definidas (OCD).

# Método

Método

# Método

## Fonte de dados

**DO\_2001  
(RJ)**

**OCMD**

**Amostra aleatória simples de  
mesmo tamanho de OCD**

**AIH 2000\_2001  
(RJ)**

**Internações que  
terminaram em óbito**

# Método

## Método Probabilístico

Utilização conjunta de campos comuns presentes em ambos os bancos de dados com o objetivo de identificar o quanto é provável que um par de registros se refira a um mesmo indivíduo (Coeli & Camargo Jr, 2002).

# Método

## Software: RecLink III



<http://paginas.terra.com.br/educacao/kencamargo/RecLink.html>

# Método

**Etapas para o relacionamento probabilístico**

# Método

## Etapas para o relacionamento probabilístico

### 1ª - Padronização



# Método

## Etapas para o relacionamento probabilístico

1ª - Padronização

2ª - Blocagem

# Método

## Etapas para o relacionamento probabilístico

1ª - Padronização

2ª - Blocagem

3ª - Aplicação de algoritmos para comparação entre campos

<b>Campo</b>	<b>Algoritmo</b>	<b>Sensibilidade (mi)</b>	<b>1-especificidade (ui)</b>	<b>Proporção mínima de concordância</b>
Nome	Aproximado	92%	1%	85%
Data de nascimento	character	90%	5%	65%

Fonte Camargo Jr & Coeli, 2002

# Método

## Etapas para o relacionamento probabilístico

1ª - Padronização

2ª - Blocagem

3ª - Aplicação de algoritmos para comparação entre campos

4ª - Cálculo de escores

LINK	PAR	
	Verdadeiro	Falso
SIM	Verdadeiro positivo ( $m_i$ )	Falso positivo ( $u_i$ )
NÃO	Falso Negativo ( $1-m_i$ )	Verdadeiro Negativo ( $1-u_i$ )

# Método

## Etapas para o relacionamento probabilístico

**1ª - Padronização**

**2ª - Blocagem**


**3ª - Aplicação de algoritmos para comparação entre campos**

**4ª - Cálculo de escores**

**5ª - Definição dos limiares para o relacionamento dos pares verdadeiros classificados como verdadeiros, duvidosos e não-pares**

**6ª - Revisão manual dos pares duvidosos, visando à sua reclassificação como pares verdadeiros ou não-pares.**

# Resultados



Resultados

# Resultados

**Tabela.** Probabilidades  $m_i$  e  $u_i$ , escores individuais para concordância e discordância para cada variável (identificador)

Variáveis (identificadores)	Probabilidade ( $m_i$ )		Probabilidade ( $u_i$ )		Escore para concordância $\text{Log}_2 (m/u)$		Escore para discordância $\text{Log}_2 \{(1-m)/(1-u)\}$	
	OCD	OCMD	OCD	OCMD	OCD	OCMD	OCD	OCMD
Nome completo	0.84	0.54	0.0002	0.0002	15.34	14.88	-1.1	-2.6
Data de nascimento	0.85	0.58	0.02	0.02	5.61	4.99	-1.2	-2.7

**OCD:** Óbito com causa definida

**OCMD:** Óbito com causa mal-definidas

# Resultados

Tabela. Probabilidades  $m_i$  e  $u_i$ , escores individuais para concordância e discordância para cada variável (identificador)

Variáveis (identificadores)	Probabilidade ( $m_i$ )		Escore para concordância $\text{Log}_2 (m/u)$		Escore para discordância $\text{Log}_2 \{(1-m)/(1-u)\}$	
	OCD	OCMD	OCD	OCMD	OCD	OCMD
	Nome completo	0.84	0.54	15.34	14.88	-1.1
Data de nascimento	0.85	0.58	5.61	4.99	-1.2	-2.7

# Resultados

**Tabela.** Probabilidades  $m_i$  e  $u_i$ , escores individuais para concordância e discordância para cada variável (identificador)

Variáveis (identificadores)	Probabilidade ( $m_i$ )		Probabilidade ( $u_i$ )		Escore para concordância $\text{Log}_2 (m/u)$		Escore para discordância $\text{Log}_2 \{(1-m)/(1-u)\}$	
	OCD	OCMD	OCD	OCMD	OCD	OCMD	OCD	OCMD
	Nome completo	0.84	0.54	0.0002	0.0002	15.34	14.88	-1.1
Data de nascimento	0.85	0.58	0.02	0.02	5.61	4.99	-1.2	-2.7

**OCD:** Óbito com causa definida

**OCMD:** Óbito com causa mal-definidas

# Resultados

Tabela. Probabilidades  $m_i$  e  $u_i$ , escores individuais para concordância e discordância para cada variável (identificador)

Variáveis (identificadores)	ade	Pr	Escore para concordância $\text{Log}_2 (m/u)$		Escore para discordância $\text{Log}_2 \{(1-m)/(1-u)\}$	
	OCMD	OCD	OCD	OCMD	OCD	OCMD
	0.54	0.00			-1.1	-2.6
	0.58	0.0			-1.2	-2.7
Nome completo			15.34	14.88		
Data de nascimento			5.61	4.99		

# Resultados

A pior performance observada no relacionamento da base dos OCMD indica uma menor qualidade do preenchimento dos campos de identificação nesses registros.

# Resultados

Teixeira et al. (2006) realizaram o relacionamento de DO e AIH dos casos de OCMD e de uma amostra de OCD, utilizando o método probabilístico e encontraram quase quatro vezes mais pares relacionados entre os OCD, comparativamente aos OCMD, quando eles restringiram às AIH com saídas de órbitos. Segundo os autores esse desequilíbrio explica-se, em parte, pela omissão das causas de órbitos entre os OCMD.

# Conclusão


Conclusão

# Conclusão

A redução de alguns obstáculos, como:

- a má qualidade das informações registradas;
- a grande variação na estrutura dos dados;

A inclusão de uma variável identificadora comum às diversas bases de dados.



São fundamentais para o relacionamento de bases de dados, permitindo, de forma relativamente rápida e barata, o acesso às informações.