

---

## **Relacionamento dos registros de uma base de dados nacional em terapia renal substitutiva e o Sistema de Informações de Mortalidade através da técnica probabilística.**

### **Autores:**

**Odilon Vanni de Queiroz: Mestrando em Saúde Pública – UFMG.**

**Augusto Afonso Guerra Júnior: Doutorando em Saúde Pública – UFMG**

**Carla Jorge Machado: Departamento de Demografia – Faculdade de Ciências Econômicas – UFMG.**

**Eli Iola Gurgel Andrade: Departamento de Medicina Preventiva e Social -Faculdade de Medicina – UFMG.**

**Wagner Meira Júnior: Departamento de Ciência da Computação – Instituto de Ciências Exatas – UFMG.**

**Francisco de Assis Acúrcio: Departamento de Farmácia Social – Faculdade de Farmácia – UFMG.**

**Walter dos Santos Filho: Mestrando em Ciência da Computação – UFMG.**

**Mariângela Leal Cherchiglia: Departamento de Medicina Preventiva e Social – Faculdade de Medicina – UFMG.**

---

Belo Horizonte, 23/09/2008

## Introdução

---

- Informação é fundamental no planejamento das políticas de saúde.
  - Os Sistemas de Informação do SUS têm grande potencial como fonte de informações estratégicas.
  - No entanto, esse potencial é sub-utilizado em decorrência dos fatores históricos e políticos que determinaram que esses sistemas fossem desarticulados e orientados a uma visão operacional, mais voltada para contagem de procedimentos ou eventos.
-

## Relacionamento de Registros

---

- O relacionamento de registros (record linkage) vem sendo usado para integrar dados dos Sistemas de Informação em Saúde. É um procedimento que permite encontrar registros de uma mesma entidade em bancos de dados distintos.
  - **Relacionamento Determinístico**
    - Considera equivalentes registros que forem iguais em uma determinada chave
    - Apropriado para situações em que exista um identificador único, como CPF
  - **Relacionamento Probabilístico**
    - Registros são comparados aos pares, sendo gerado um escore de similaridade.
    - Com base nesse escore, os pares são divididos em improváveis, prováveis ou duvidosos
    - Leva em consideração as probabilidades de ocorrência de erros de preenchimento, ocorrência de homônimos etc.
-

## O projeto TRS

---

- Projeto de pesquisa realizado pelo Grupo de Pesquisa em Economia da Saúde – GPES/Faculdade de Medicina/UFMG.
  - Avaliar a situação e o desenvolvimento das TRS no Brasil, nos aspectos econômicos e epidemiológicos.
  - Construir uma base nacional de dados em TRS, centrada no indivíduo, que permita uma avaliação multidimensional dos pacientes submetidos a TRS pelo SUS.
  - A partir de dados do subsistema APAC/SIA, construiu-se um cadastro dos pacientes em TRS, através da técnica de relacionamento probabilístico. (CHERCHIGLIA et al, 2007)
-

## Objetivos

---

- Relacionar os registros da Base TRS e o Sistema de Informações de Mortalidade (SIM), de 2000 a 2004
  - Validar este relacionamento, por meio de uma inspeção manual dos registros relacionados.
  - Tal procedimento permite complementar a informação de óbito originária do sub-sistema APAC/SIA, presente na base TRS.
-

## Origem dos Dados

---

- Os dados com os campos de identificação dos indivíduos foram disponibilizados pelo DATASUS através do DES/SCTIE do Ministério da Saúde, conforme termo de compromisso e responsabilidade firmado entre o DES e o GPES/UFMG.
  - APAC/SIA 11/1999 a 05/2005
  - SIM 01/2000 a 12/2004
-

## Desafios

- Grande volume de dados (e.g.:176.773 x 4.636.197 registros)
- Qualidade relativamente ruim dos dados originais (ausência de informação - ex: São Paulo e Minas Gerais, 2000)

Ano	Estado		
	SP	MG	SC
<b>2000</b>	573	7.074	2.737
<b>2001</b>	957	96.284	27.814
<b>2002</b>	235.221	96.908	28.358
<b>2003</b>	238.039	104.234	29.330
<b>2004</b>	234.214	102.887	29.378

## Desafios

Existência de diversos valores para representar a ausência de informações

'NAO INFORMADO', 'IGNORADO', 'O MESMO', 'FALAR C RENATA PAULA NELSON', 'INEXT', 'A MESMA', 'NAO INFORMADA', 'IDEM', 'NAO CONSTA', 'NAO ESPECIFICADO', 'INEX', 'INEXST', 'NAO IDENTIFICADO', 'MESMO', 'XXXX', 'N CONSTA', 'O PROPRIO', 'NAO INFOEMADO', 'NINI', 'DESCONHECIDA', 'RECEM NASCIDO', 'HOMEM', 'DESCONHECIDO', 'NATIMORTO', 'IDENTIDADE DESCONHECIDA', 'UM HOMEM', 'RECEMNASCIDO', 'RECEN NASCIDO', 'ID DESCONHECIDA', 'MULHER', 'DESCONHECIDO FF ', 'NATI MORTO', 'NAO IDENTIFICADO', 'IGNORADA', 'O MEMSO', 'NAO INFORMOU', 'NAO IDENTIFICADA', 'FALECIDA', 'DESC', 'NAO INFROMADO', 'NAO INF', 'NO CONSTA', 'INEXIST', 'INGNORADO', 'XXXXXX', 'AAAA', 'NAO TEM', 'XXXXXXX', 'O PROPIO', 'A PROPRIA', 'INEXIS', 'NADA CONSTA', 'ILEGIVEL', 'FALECIDO', 'IGNORADOS', 'ATUALIZAR', 'INLEGIVEL', 'ESPOSA', 'ANEXO', 'IGN', 'INX', 'INEXST', 'NADA', 'NAO CONSTA NO RG', 'NO CONSTA', 'NTEM', 'NCN', 'SNOME'

## Resultados

O relacionamento foi feito em três etapas uma determinística e duas probabilísticas

Etapa	Número de registros		Pares gerados
	Base TRS	SIM	
<b>ETAPA 1</b>	176.773	4.636.197	16.347
<b>ETAPA 2</b>	160.426	4.636.197	523.077.601
<b>ETAPA 3</b>	160.426	4.636.197	1.022.137

- O relacionamento foi feito por meio de uma rotina em linguagem SQL, no Sistema Gerenciador de Bancos de Dados MySQL. O tempo de processamento foi cerca de 6 horas.
- Após análise, foram encontradas 52.048 declarações de óbito pelo relacionamento.

## Validação

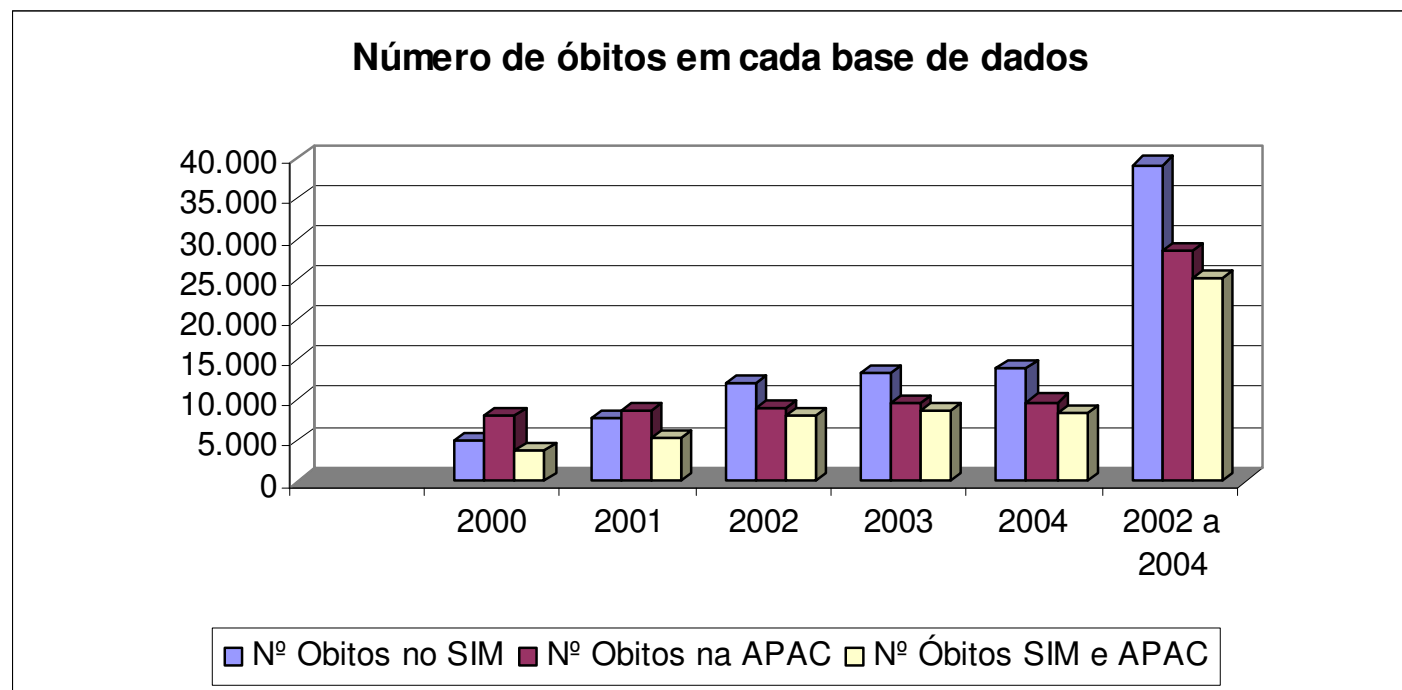
**Número de pares considerados falsos pela validação manual, por decil do peso de concordância, no relacionamento da base nacional de dados em TRS com o SIM**

Decil	Menor Escore	Maior Escore	Pares Falsos	% Pares Falsos
1	29,95	34,1	5	16,67
2	34,11	38,08	3	10,00
3	38,09	41,41	0	0,00
4	41,42	44,63	0	0,00
5	44,64	47,78	0	0,00
6	47,79	51,02	0	0,00
7	51,03	54,55	0	0,00
8	54,56	58,75	0	0,00
9	58,76	64,83	0	0,00
10	64,84	103,76	0	0,00

No total 97.3% dos pares inspecionados foram considerados verdadeiros

## Validação

Havia, na base TRS, 45.203 óbitos identificados por informações do sub-sistema APAC/SIA. Desses, 75,6% foram encontrados no SIM. O percentual aumenta para 87,9% para os anos de 2002 a 2004, devido à maior completude dos dados do SIM nesse período.



## Conclusão

---

Ainda são raros estudos envolvendo o sub-sistema APAC/SIA e relacionamento de registros, ainda mais envolvendo um número tão grande de registros. Porém este estudo conseguiu relacionar os registros das duas bases de dados com precisão satisfatória.

---

## Conclusão

---

- **CNPq**  
Conselho Nacional de Desenvolvimento Científico e Tecnológico
  - **FNS**  
Fundo Nacional de Saúde
  - **FAPEMIG**  
Fundação de Amparo à Pesquisa do Estado de Minas Gerais
-