



## PROPOSTA DE LINKAGE ENTRE OS SISTEMAS DE INFORMAÇÃO EM SAÚDE SIM E SINASC: UMA ANÁLISE SOBRE A MORTALIDADE INFANTIL E CARACTERÍSTICAS MATERNAS

**Ricardo Morsoleto<sup>1</sup>; Maria C. Batista<sup>2</sup>; Simone M. F. Miranda<sup>3</sup>; Juliano Caliarí<sup>4</sup>; Hiran N. M. Ferreira<sup>5</sup>; Vinícius A. Silva<sup>6</sup>**

### RESUMO

Este trabalho apresenta a proposta de linkage probabilístico entre Sistemas de Informação a Saúde (SIS) - Sistemas de Informação sobre Mortalidade (SIM) e Nascidos Vivos (Sinasc) - a partir do uso das bibliotecas RecordLinkage e Pandas, disponíveis em linguagem de programação Python. Os bancos de Dados SIM e o SINASC foram agrupados e, conseqüentemente, foi feita uma análise do perfil da mãe presente no banco agrupado. Os resultados parciais obtidos já demonstram o poder que a integração desses SIS pode trazer para futuras pesquisas.

**Palavras-chave:** Linkage, Sistemas de Informação e Mortalidade Infantil.

### 1. INTRODUÇÃO

Segundo a Organização Mundial de Saúde (OMS), os Sistemas de Informação em Saúde (SIS) são parte crucial no âmbito das políticas públicas e do cuidado em saúde, uma vez que envolvem mecanismos de coleta, processamento, análise e transmissão da informação (Indicators, 2010). As informações obtidas norteiam o planejamento, organização, operacionalização e avaliação dos serviços de saúde. No Brasil, os SIS foram desenvolvidos e implantados pelo Ministério da Saúde (MS), sendo que cada um atende a diferentes finalidades (Drumond, 2009).

Quanto mais integrados e interoperáveis forem os SIS, mais informações poderão ser disponibilizadas e utilizadas para orientar a tomada de decisões, aumentando a eficiência do sistema de saúde como um todo (Galvão, 2024). Porém, os sistemas foram desenvolvidos de maneira independente, criando problemas como: falta na padronização das informações, impossibilidade de vincular atendimentos para um mesmo indivíduo e descontinuidade na coleta.

Um SIS deve garantir a avaliação constante da situação de saúde da população, bem como dos efeitos das ações de saúde implementadas, fornecendo subsídios para adequações de acordo com as políticas de saúde vigentes. Dentre os indicadores utilizados pelo Sistema Único de Saúde (SUS), temos a taxa de mortalidade infantil, que abrange as mortes de crianças que não

<sup>1</sup>Bolsista Decit/SECTICS/MS - CNPq, IFSULDEMINAS – *Campus Passos*. E-mail: ricardo.morsoleto.developer@gmail.com.

<sup>2</sup>Bolsista Decit/SECTICS/MS - CNPq, IFSULDEMINAS – *Campus Passos*. E-mail: maria.batista@alunos.ifsuldeminas.edu.br

<sup>3</sup>Bolsista Decit/SECTICS/MS - CNPq. E-mail: simone.mara@batistanet.com.br.

<sup>4</sup>Docente, IFSULDEMINAS – *Campus Passos*. E-mail: juliano.caliari@ifsuldeminas.edu.br.

<sup>5</sup>Docente, IFSULDEMINAS – *Campus Passos*. E-mail: hiran.ferreira@ifsuldeminas.edu.br.

<sup>6</sup>Docente, IFSULDEMINAS – *Campus Passos*. E-mail: vinicius.silva@ifsuldeminas.edu.br.

completaram um ano de vida (França, 2017). A mortalidade infantil é um importante indicador de saúde pois reflete as condições de vida de uma população, em especial as condições da gestação, pré-natal e parto. No SUS, as informações a respeito dessas mortes são disponibilizadas em dois SIS distintos, o Sistema de Informação sobre Mortalidade (SIM) e o sobre Nascidos Vivos (SINASC), em formato anonimizado através do site do Departamento de Informática do SUS, o DATASUS. As informações produzidas, por meio das análises dos dados gerados, são utilizadas para avaliar as ações e os resultados destas sobre a situação de saúde e para propor intervenções a fim de impactar nas taxas de óbitos infantis (Brasil, 2009).

No sentido de compreender as influências nas taxas de mortalidade infantil, torna-se importante analisar o SINASC e o SIM, através do processo de linkage, que possui dois tipos, o determinístico e o probabilístico. O primeiro ocorre por meio da concordância exata entre os registros, com uma variável de identificação compartilhada entre os SIS. Já o segundo ocorre por meio do grau de concordância, que deve ser menor entre registros de uma mesma pessoa e maior entre os registros de pessoas diferentes. Ressalta-se que, como no DATASUS, os dados são disponibilizados publicamente em formato anonimizado, só é possível efetuar o linkage probabilístico.

Considerando os dados disponíveis, o presente estudo propõe realizar um linkage entre os SIS, SINASC e SIM, referentes aos anos de 2012 a 2022, com abrangência nacional, para analisar se determinadas características da mãe possuem relação com casos de mortalidade infantil.

## 2. MATERIAL E MÉTODOS

O linkage foi efetuado por meio de um processo de quatro etapas, conforme representado na Figura 1:

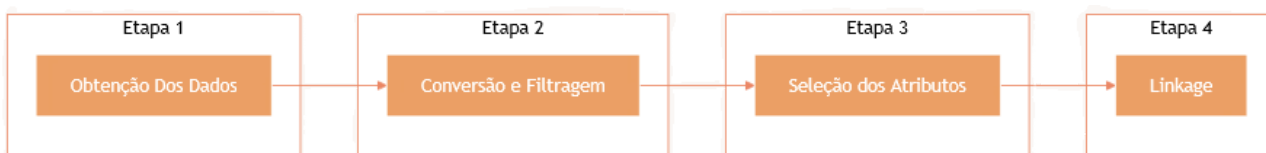


Figura 1: Etapas para a realização do linkage entre os SIS

Na Etapa 1: foi feita a obtenção dos dados de interesse nos sistemas de informação SINASC e SIM através do site do DATASUS<sup>7</sup>. Os dados adquiridos correspondem aos anos de 2012 a 2022 com abrangência de todo o território nacional, totalizando 605 arquivos.

Na Etapa 2: uma das barreiras encontradas para o uso desses dados, foi com a extensão DBC, usada em todos os arquivos de bancos de dados oriundos do DATASUS. O site disponibiliza uma ferramenta para a conversão, o TABWIN, porém, a mesma é ineficiente para a conversão de uma quantidade massiva de arquivos. Foi optado pela implementação de um conversor, escrito em

<sup>7</sup> <https://datasus.saude.gov.br/>

linguagem de programação C, para DBF, e um script, escrito em linguagem Python, para converter para CSV. Ao final da conversão, os arquivos totalizaram mais de 21,5 GB de espaço no disco.

Após a padronização da extensão, foi realizado um processo de filtragem de registros dos bancos do SIM obtidos. No primeiro, permaneceram apenas registros que possuíam idade menor que 1 ano e óbitos fetais. Os bancos do SINASC não foram alterados.

Etapa 3: Consiste na escolha de uma lista de variáveis comuns, e sua padronização, aos bancos e que respondem aos questionamentos e objetivos da pesquisa. A lista para os bancos SIM e SINASC possui variáveis socioeconômicas, referentes a gestação, parto e geográficas. Para o presente estudo foram selecionadas as seguintes variáveis: DTNASC, data de nascimento da criança, IDADEMAE, idade da mãe em anos, ESCMAE2010, nível de escolaridade da mãe, SERIESMAE, séries escolares completas da mãe, OCUPMAE, código de ocupação da mãe, CODMUNNATU, código do município de naturalidade, SEMAGESTAC, número de semanas de gestação, GRAVIDEZ, tipo de gravidez, PARTO, tipo de parto, PESO, peso ao nascer.

Etapa 4: foi feito o uso da linguagem de programação Python, junto às bibliotecas Pandas, para o manuseio dos bancos, e RecordLinkage, para implementação do linkage, através da plataforma para computação Jupyter Notebook. O critério usado para a comparação entre os registros foi a total concordância, pareando apenas registros com o mesmo valor nas variáveis referentes a data de nascimento e o código do município, acrescidos de pelo menos 4 outras.

Ao fim do processo, foi analisada a quantidade de registros pareados com sucesso. Para os bancos referentes a Minas Gerais do ano de 2022, foi possível parear 1660 registros de 28942 presentes no SIM, totalizando 5% dos registros.

### 3. RESULTADOS PRELIMINARES

Os registros pareados foram submetidos a uma análise a fim de buscar características relacionadas ao perfil das mães cujas crianças vieram a óbito. A primeira a ser discutida é a idade da mãe, cuja distribuição pode ser vista no gráfico da Figura 2-A.

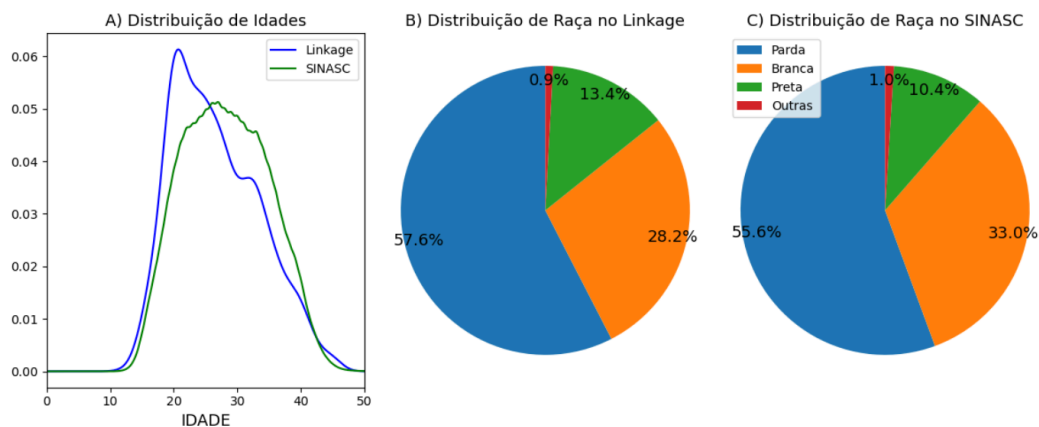


Figura 2: Distribuições dos bancos.

A distribuição dos registros pareados (em azul), ao contrário da distribuição do SINASC, possui maior número de mães adolescentes, com até 20 anos, e menor número de mães jovens, entre 20 e 34 anos. Outra característica interessante é a distribuição de raça disponível nos gráficos da Figura 2-B e 2-C, no qual 58% dos registros pertencem a mulheres pardas, 28% brancas e 13% negras, sendo muito próxima das existentes no banco do SINASC. Porém, as raças consideradas em situação de risco (Parda e Negra) tiveram aumento nas porcentagens.

#### 4. CONCLUSÃO

Com base nos resultados preliminares obtidos do linkage, pode-se afirmar que é possível executar o pareamento entre os bancos SINASC e SIM. Além disso, o mesmo pode ser usado para a análise do perfil da mãe. Porém, esse pareamento deve ser feito com certa atenção em algumas áreas como, ausência de registros e falta de padronização entre os bancos. Erros no pareamento podem gerar diminuição no grau de concordância dos registros, gerando duplas de registros que não pertencem a mesma pessoa.

Diferentes análises poderão ser feitas com esses dados, sendo possível a continuidade da pesquisa em trabalhos futuros. Uma das barreiras encontradas foi a quantidade massiva de dados, totalizando 605 arquivos.

#### REFERÊNCIAS

- BRASIL; MINISTÉRIO DA SAÚDE. Manual de vigilância do óbito infantil e fetal e do Comitê de Prevenção do Óbito Infantil e Fetal. 2009.
- DRUMOND, Eliane de Freitas et al. Utilização de dados secundários do SIM, Sinasc e SIH na produção científica brasileira de 1990 a 2006. **Revista Brasileira de Estudos de População**, v. 26, p. 7-19, 2009.
- FRANÇA, Elisabeth Barboza et al. Principais causas da mortalidade na infância no Brasil, em 1990 e 2015: estimativas do estudo de Carga Global de Doença. *Revista brasileira de epidemiologia*, v. 20, p. 46-60, 2017.
- GALVÃO, Pauliana Valéria Machado et al. Percurso metodológico para implementação do linkage nos sistemas de informação em saúde brasileiros. **Peer Review**, v. 6, n. 13, p. 158-166, 2024.
- INDICATORS, A. H. O. Monitoring the building blocks of health systems. **Geneva, Switzerland: WHO Document Production Services**, 2010.