

# Preparo da base de dados de estudo retrospectivo de complicações após artroplastia total de quadril primária em idosos

DOCUMENTO: SAR-2021-023-LP-v01

De: Felipe Figueiredo Para: Lourenço Peixoto

2022-02-02

## SUMÁRIO

1	LISTA DE ABREVIATURAS.....	2
2	CONTEXTO.....	2
2.1	Objetivos.....	2
2.2	Recepção e tratamento dos dados.....	2
3	METODOLOGIA.....	3
3.1	Variáveis.....	3
3.2	Expressões regulares.....	3
3.3	Software utilizado.....	3
4	ESTRATÉGIA DE BUSCA.....	4
4.1	Complicações individuais.....	4
4.2	Qualquer complicação.....	6
5	OBSERVAÇÕES E LIMITAÇÕES.....	6
6	REFERÊNCIAS.....	7
7	APÊNDICE.....	7
7.1	Análise exploratória de dados.....	7
7.2	Disponibilidade.....	7
7.3	Dados utilizados.....	7

# Preparo da base de dados de estudo retrospectivo de complicações após artroplastia total de quadril primária em idosos

## Histórico do documento

Versão	Alterações
01	Versão inicial

## 1 LISTA DE ABREVIATURAS

- ATQ: artroplastia total de quadril
- FX:
- TEV: tromboembolismo venoso
- TEP: tromboembolia pulmonar
- TVP: trombose venosa profunda
- ITU:

## 2 CONTEXTO

O projeto “Artroplastia total do Quadril em septuagenários, complicações e resultados a curto prazo” identificará ocorrências de complicações após ATQ em pacientes septuagenários quando comparados a pacientes de faixas etárias inferiores. As complicações foram descritas de maneira qualitativa em campo de texto aberto, onde havia a possibilidade de múltiplas observações para um mesmo paciente. Para a análise das ocorrências das complicações, elas precisam ser codificadas como variáveis dicotômicas.

Este relatório técnico descreve como as complicações foram identificadas na base de dados original, em campo de texto aberto, usando expressões regulares.

### 2.1 Objetivos

Limpeza da base de dados e identificação de cada tipo de complicação pós-cirúrgica dos participantes incluídos na base com chaves de busca por expressões regulares.

## 2.2 Recepção e tratamento dos dados

A base de dados original possuía 21 variáveis coletadas de 926 observações. Depois dos procedimentos de limpeza e seleção a base final contava com 38 variáveis de 926 observações. Nenhuma variável foi removida no preparo da base, e a diferença é composta apenas pelas novas variáveis indicadoras criadas para classificar todas as possíveis complicações em cada participante incluído na base. Nenhuma observação foi excluída como parte dos processos de limpeza de dados.

## 3 METODOLOGIA

### 3.1 Variáveis

#### 3.1.1 Complicações listadas no projeto

- tromboembolismo venoso
- embolia pulmonar
- anemia com necessidade de transfusão sanguínea
- distúrbios hidroeletrolíticos com repercussão clínica
- infecção pós operatória
- deiscência de ferida
- instabilidade (luxação aguda ou crônica)

#### 3.1.2 Outras complicações

- TVP
- óbito
- delirium
- ITU
- sepse
- pneumonia
- FX
- parestesia

### 3.2 Expressões regulares

Expressões regulares permitem fazer buscas em texto por palavras-chave de forma flexível, permitindo considerar variações de grafia, maiúsculas e minúsculas entre outras coisas. Ao definir chaves de busca específicas para cada complicação, é possível identificar a presença da chave de busca no texto original.

Esta estratégia além de flexível também é expansível e permite ampliar a estratégia de busca que objetiva identificar cada complicação alvo. Na seção 4 serão descritas as

---

chaves de busca usadas para identificação de cada uma das complicações listadas na seção 3.1.

### 3.3 Software utilizado

Esta análise foi realizada utilizando-se o software R versão 4.1.2.

## 4 ESTRATÉGIA DE BUSCA

Todas as expressões regulares descritas abaixo foram mapeadas ignorando o caso de grafia, isto é, todas funcionam da mesma forma independente se o texto base foi escrito em maiúsculas ou minúsculas.

### 4.1 Complicações individuais

#### 4.1.1 Anemia com necessidade de transfusão sanguínea

Chaves de busca:

- anemia

Número de identificações: 18

#### 4.1.2 Embolia pulmonar

Chaves de busca:

- tep
- embol

Número de identificações: 2

#### 4.1.3 Tromboembolismo venoso

Chaves de busca:

- tev

Número de identificações: 0

#### 4.1.4 TVP

Chaves de busca:

- tvp

Número de identificações: 5

#### 4.1.5 Instabilidade (luxação aguda ou crônica)

Chaves de busca:

- instab
- luxa
- soltura

Número de identificações: 10

#### 4.1.6 Deiscência de ferida

Chaves de busca:

- deisc

Número de identificações: 2

#### 4.1.7 Infecção pós operatória

Chaves de busca:

- infec

Número de identificações: 34

#### 4.1.8 Distúrbios hidroeletrólitos com repercussão clínica

Chaves de busca:

- hidro
- eletr[oó]l[íi]t

Número de identificações: 0

#### 4.1.9 Delirium

Chaves de busca:

- delir

Número de identificações: 5

#### 4.1.10 Óbito

Chaves de busca:

- [óó]bito
- morte

Número de identificações: 0

---

#### 4.1.11 ITU

Chaves de busca:

- itu

Número de identificações: 4

#### 4.1.12 Seps

Chaves de busca:

- s[ée]p[st]

Número de identificações: 1

#### 4.1.13 Pneumonia

Chaves de busca:

- pneumonia

Número de identificações: 4

#### 4.1.14 FX

Chaves de busca:

- fx

Número de identificações: 3

#### 4.1.15 Parestesia

Chaves de busca:

- parest

Número de identificações: 3

### 4.2 Qualquer complicação

A fim de criar um desfecho composto que capturasse o risco de qualquer complicação pós-cirúrgica no curto prazo, foi criada uma nova variável que é positiva se qualquer das complicações identificadas na seção 4.1. Ela é negativa apenas caso nenhuma das complicações individuais tenha ocorrido para aquele paciente.

Número de identificações: 84

## 5 OBSERVAÇÕES E LIMITAÇÕES

A estratégia utilizada neste relatório técnico são flexíveis o suficiente para permitir a identificação de informações alvo com diferentes grafias, incluindo acentuação, erros tipográficos típicos entre outros. Adicionalmente, caso novas chaves se mostrem necessárias, elas podem ser facilmente acrescentadas ao algoritmo de busca para ampliar o escopo da identificação.

A principal limitação desta estratégia pode ser identificada na busca de chaves de texto pré-determinadas para identificar informações alvo. A qualidade do resultado depende do estabelecimento prévio de chaves de busca abrangentes, suficientes e adequadas ao texto alvo.

## 6 REFERÊNCIAS

N/A

## 7 APÊNDICE

### 7.1 Análise exploratória de dados

N/A

### 7.2 Disponibilidade

Este documento pode ser obtido no seguinte endereço:

<https://philsf-biostat.github.io/SAR-2021-023-LP/>

### 7.3 Dados utilizados

Os dados utilizados neste relatório não podem ser publicados online por questões de sigilo.

**Tabela A1** Estrutura da tabela de dados analíticos (continued below)

prontuario	nome	sexo	data_de_nascimento	dia_da_cirurgia	dia_de_alta	dias_ate_alta	colunas 4	cid	cor	idade	asa	has	dm
1													
2													
3													
...													
926													

Relatório de Análise Estatística (SAR)

Table continues below

outras_comorbidades	tabagismo	ex_tabagista	complicacoes	lateralidade	lateralidade_obs	coluna1	comp_anemia	comp_tep	comp_tev

Table continues below

comp_tvp	comp_inst	comp_deisc	comp_infec	comp_hidro	comp_delir	comp_obito	comp_itu	comp_sepse	comp_pneumo	comp_fx

comp_parest	comp_qualquer	group