

Resultados do projeto “Eventos adversos da quimioterapia anticâncer em pacientes pediátricos com tumores do sistema nervoso central: estudo retrospectivo”

Um caderno aberto de pesquisa

Francisco Hélder Cavalcante Félix, Centro Pediátrico do Câncer - Hospital Infantil Albert Sabin

Abstract: Um banco de dados de pacientes pediátricos com tumores do sistema nervoso central diagnosticados em um grande hospital estadual foi criado pelos autores. Os resultados e análises foram colocados num repositório do serviço GitHub. O código das análises foi checado com o serviço de integração contínua em nuvem Travis Ci e o resultado final publicado neste arquivo. Este arquivo traz tabelas, gráficos e texto mostrando os resultados do projeto. Foi elaborado em Rmarkdown, utilizando a linguagem de marcação simplificada Markdown com “pedaços” de código da linguagem estatística R entremeados. O arquivo foi avaliado pelo pacote rmarkdown e compilado para o formato pdf neste texto. Devido à integração contínua, pode ser atualizado em tempo real, enquanto dados novos são acrescentados.

Keywords: tumores do sistema nervoso central, cancerologia pediátrica, estatística escrita, rmarkdown, integração contínua, ciência aberta

January 30, 2020

Introdução:

A ciência aberta baseia-se principalmente na capacidade de divulgar (compartilhar) eletronicamente as informações coletadas (dados brutos) e produzidas (análises e seus resultados) de um projeto de pesquisa através da internet. Dessa forma, 2 consequências advêm imediatamente: 1 - Transparência da informação e do processo científicos, inclusive para públicos não técnicos. 2 - Capacidade irrestrita de comentários, tanto por especialistas (análogo à *revisão por pares*), quanto por não especialistas (que poderíamos chamar de *revisão cidadã*).

O observador arguto já pode levantar a questão de que o controle de comentários numa plataforma é dos controladores daquele serviço, ou seja, é possível criar um canal de comunicação tipo “ciência aberta”, porém unidirecional, sem recepção de comentários (ou pior, é possível censurá-los). Independentemente disso, os comentários a uma publicação livremente disponível na internet podem ser publicados em qualquer canal sem relação com o canal original e referenciado ao primeiro. Ou seja, não há como verdadeiramente censurar comentários a uma publicação livre na internet.

Outra característica importante da ciência aberta é a capacidade de *reuso* de informações, o que pode ser entendido como a principal utilidade social da ciência aberta. Esta característica simples tem o potencial de otimizar a produção científica a nível global. Bastaria isso para justificar a implementação em larga escala da ciência aberta. Outros benefícios podem ser descritos de forma ilimitada.

Esta é a implementação de ciência aberta que criei, baseando-me largamente em projetos já existentes. Trata-se de um caderno de pesquisa aberto, armazenado num repositório remoto para o programa Git (existem vários), gerado através de um serviço de integração contínua (CI) em nuvem (vários idem) e com a estrutura de um pacote da linguagem estatística R, usada para as análises. *Não se trata de um pacote verdadeiro*, apesar de ter um diretório de código R e um arquivo de definições DESCRIPTION. O objetivo desse mimetismo é facilitar as análises numa plataforma de CI. Um pacote de R é um programa com funções utilizáveis. Não é isso que este(s) caderno(s) é(são). Assim, propositadamente deixo de fora partes imprescindíveis de um pacote, como o NAMESPACE e os manuais.

Na seção a seguir, são mostrados resultados de análises estatísticas concernentes a este caderno aberto de pesquisa em particular. Todos os dados pertinentes a seres humanos são adequadamente desidentificados.

Análises:

```
require(pander)
```

```
## Loading required package: pander
```

```
require(survival)
```

```
## Loading required package: survival
```

```
snc<-read.csv('../data/EA2007-2010.csv')
```

```
attach(snc)
```

```
require(DescTools)
```

```
## Loading required package: DescTools
```

```
options(OutDec= ",")
```

```
barplot(summary(as.factor(sex)),names.arg=c("masculino","feminino"),  
xlab="Sexo",width=0.5,xlim=c(0,1.7),space=0.5,col=0)
```

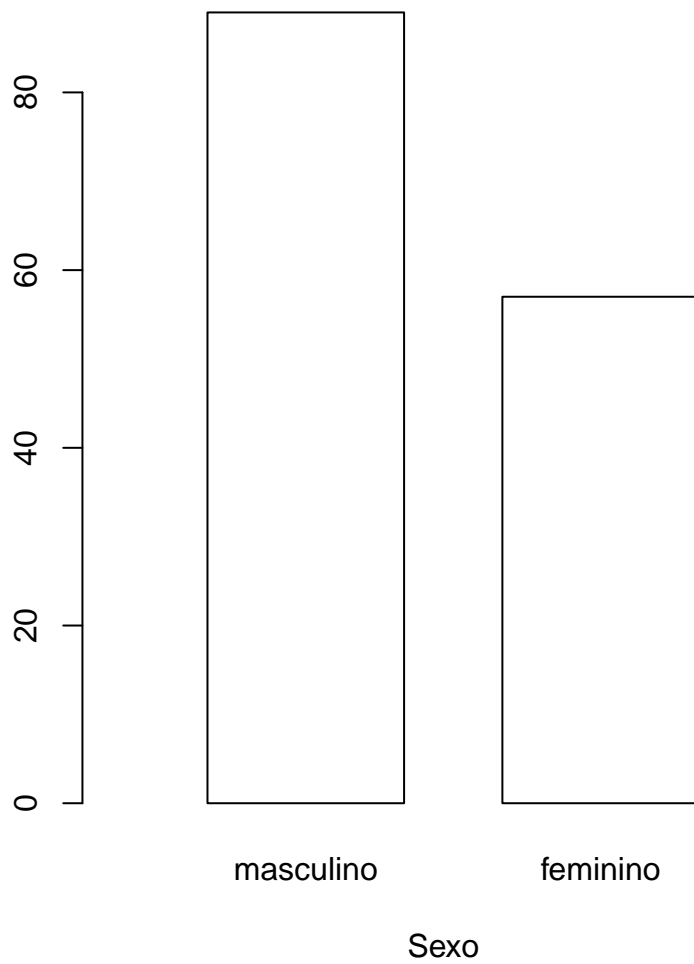


Figura 1: número de pacientes segundo o sexo.

```
boxplot(age/365.25,xlab="Idade (anos)",boxwex=0.6,staplewex=0.4,  
frame.plot=F)
```

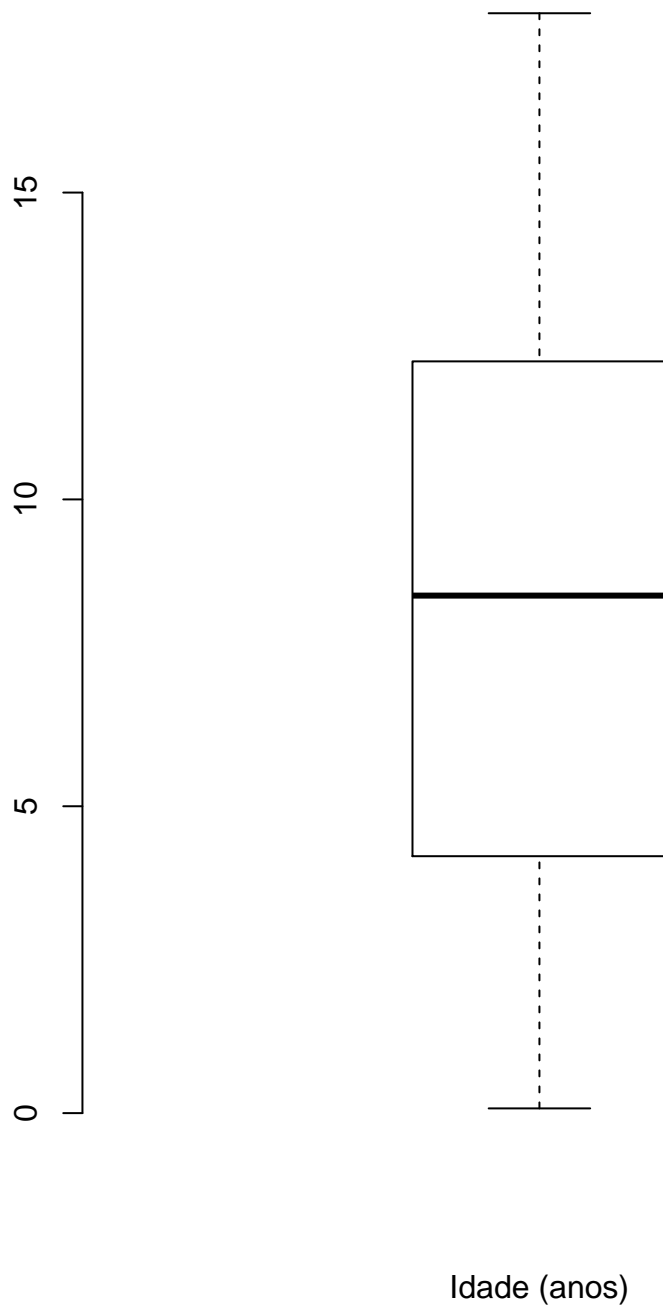


Figura 2: idade dos pacientes ao diagnóstico.

```
panderOptions('table.split.table', Inf)
set.caption("Tratamentos utilizados como primeira linha")
pander(Freq(ct1,ord='desc')[-4], style = 'rmarkdown')
```

Table 1: Tratamentos utilizados como primeira linha

level	freq	perc	cumperc
[0,2]	116	0.7945	0.7945
(10,12]	19	0.1301	0.9247
(2,4]	4	0.0274	0.9521
(8,10]	4	0.0274	0.9795
(12,14]	2	0.0137	0.9932
(22,24]	1	0.006849	1
(4,6]	0	0	1
(6,8]	0	0	1
(14,16]	0	0	1
(16,18]	0	0	1
(18,20]	0	0	1
(20,22]	0	0	1

```
require(VennDiagram)

## Loading required package: VennDiagram
## Loading required package: grid
## Loading required package: futile.logger

v<-draw.triple.venn (area1 = length(subset(surg1,surg1!=0)),
  area2 = length(subset(rt1,rt1!=0)),
  area3 = length(subset(ct1,ct1!="Não realizou quimioterapia")),
  n12 = length(subset(surg1,surg1!=0&rt1!=0)),
  n23 = length(subset(surg1,rt1!=0&ct1!="Não realizou quimioterapia")),
  n13 = length(subset(surg1,surg1!=0&ct1!="Não realizou quimioterapia")),
  n123 = length(subset(surg1,surg1!=0&rt1!=0&ct1!="Não realizou quimioterapia")),
  category = c('Cirurgia', 'Radioterapia', 'Quimioterapia'),
  fill = c('gray90', 'gray70', 'gray50'),
  )
grid.draw(v)
```

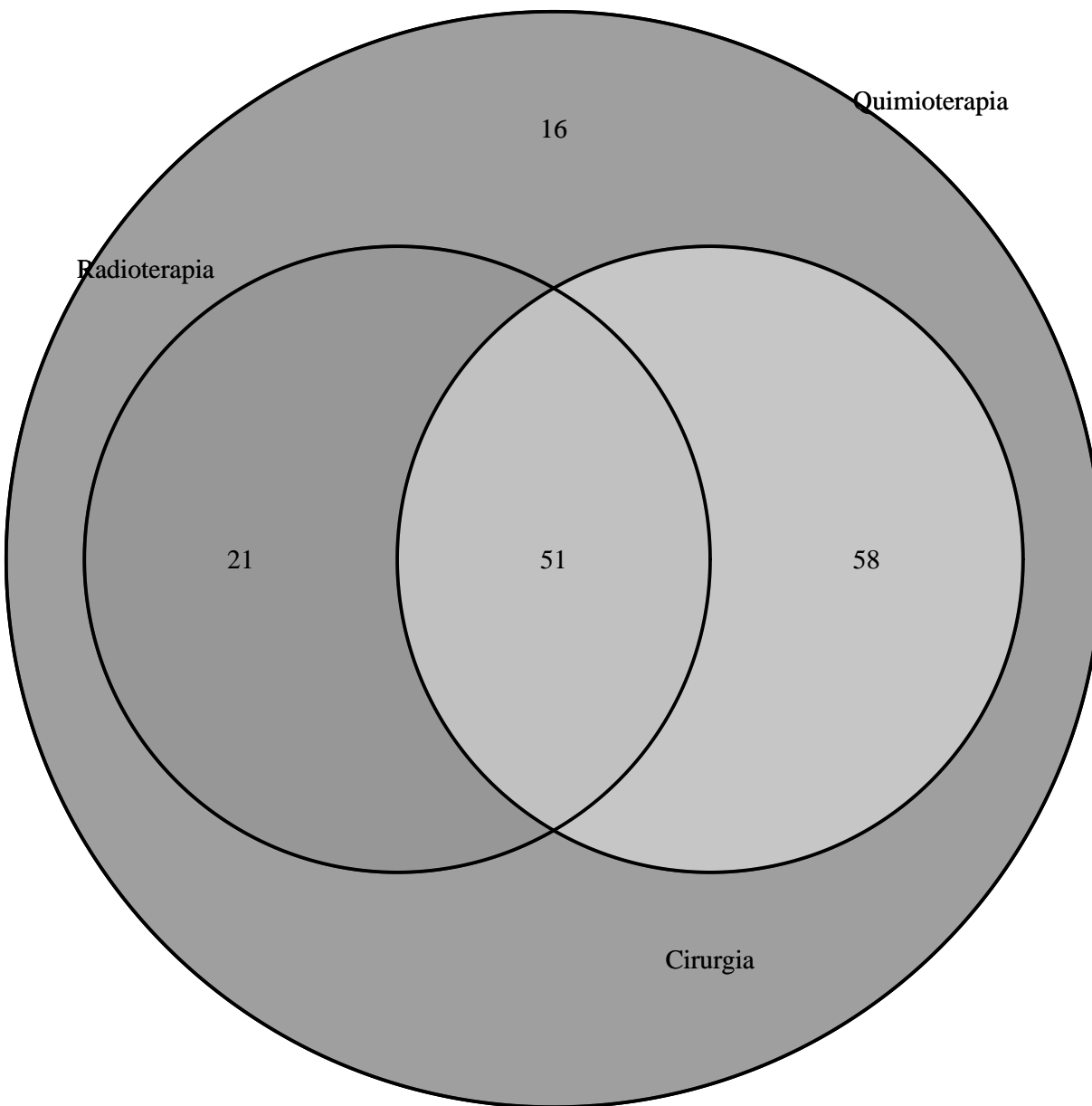


Figura 8: número de pacientes que foram tratados com cirurgia, radioterapia e/ou quimioterapia.

Dados e código para replicação estão disponíveis no repositório do GitHub do projeto