

---

# Plano Analítico para Associação entre os marcadores CK19, EpCAM e CD44 e óbito por carcinoma hepatocelular: coorte retrospectiva

DOCUMENTO: SAP-2023-028-AN-v01

De: Felipe Figueiredo Para: Aline Nascimento

2023-11-01

## SUMÁRIO

1	LISTA DE ABREVIATURAS.....	2
2	CONTEXTO.....	2
2.1	Objetivos.....	2
2.2	Hipóteses.....	2
3	DADOS.....	2
3.1	Dados brutos.....	2
3.2	Tabela de dados analíticos.....	2
4	PARÂMETROS DO ESTUDO.....	4
4.1	Desenho do estudo.....	4
4.2	Critérios de inclusão e exclusão.....	4
4.3	Exposições.....	4
4.4	Desfechos.....	4
4.5	Covariáveis.....	4
5	MÉTODOS ESTATÍSTICOS.....	5
5.1	Análises estatísticas.....	5
5.1.1	Análise descritiva.....	5
5.1.2	Análise inferencial.....	5
5.1.3	Modelagem estatística.....	6
5.1.4	Dados faltantes.....	6
5.2	Significância e Intervalos de Confiança.....	6
5.3	Tamanho da amostra e Poder.....	7
5.4	Softwares utilizados.....	7
6	OBSERVAÇÕES E LIMITAÇÕES.....	7
7	REFERÊNCIAS.....	7
8	APÊNDICE.....	8
8.1	Análises associadas.....	8
8.2	Disponibilidade.....	8

---

# Plano Analítico para Associação entre os marcadores CK19, EpCAM e CD44 e óbito por carcinoma hepatocelular: coorte retrospectiva

## Histórico do documento

Versão	Alterações
01	Versão inicial

## 1 LISTA DE ABREVIATURAS

- CHC: carcinoma hepatocelular
- DP: desvio padrão
- HR: taxa de risco (*hazards ratio*)
- IC: intervalo de confiança

## 2 CONTEXTO

### 2.1 Objetivos

Identificar fatores de risco que alteram a associação entre marcadores de carcinoma hepatocelular e óbito em pacientes do INCA.

### 2.2 Hipóteses

Os marcadores CK19, EpCAM e CD44 são associados ao óbito em pacientes com CHC.

## 3 DADOS

### 3.1 Dados brutos

A base de dados original possuía 50 variáveis coletadas de 31 observações.

### 3.2 Tabela de dados analíticos

O tempo de observação, em anos, foi calculado entre a data da primeira consulta e a data da última consulta disponível na base de dados, ou data de óbito quando ocorreu. O desfecho foi classificado em eventos competitivos, como óbito por CHC, óbito por outras causas e censurado quando o paciente estava vivo. Variáveis dicotômicas foram codificadas como indicadores com valores 0 e 1. Variáveis categóricas foram codificadas

Plano Analítico (SAP)

considerando todos os níveis possíveis, mesmo que não tenham sido observados na amostra. A unidade de análise é o participante e indivíduos em duplicata serão removidas para ter uma única observação por unidade.

Duas variáveis contínuas foram cortadas: a idade foi cortada em 60 anos, e o maior tamanho de tumor foi cortado em 5 centímetros, ambos inclusivos no nível superior. Estas novas variáveis serão apresentadas ao método *stepwise* junto com suas versões contínuas para selecionar a versão da mensuração que apresente maior bondade de ajuste no modelo (ver seção 5.1.3).

A variável “Alfa-feto proteína” é contínua mas possui valores censurados (“>4000”) assim como valores numéricos superiores ao corte (eg 12.000 e 30.000). É comum métodos analíticos quantitativos censurarem valores fora dos limites de quantificação do teste, mas a presença de valores superiores ao indicado indica que o limite superior de quantificação não é 4000, o que inviabiliza o tratamento do dado. Esta variável não pode ser tratada nem como categórica nem como contínua, e foi removida da análise.

As variáveis com dados faltantes farão parte da análise descritiva, mas serão excluídas do universo da análise inferencial (ver seção 5.1.3).

Depois dos procedimentos de limpeza e seleção 44 variáveis foram incluídas na análise com 30 observações.

**Tabela 1** Estrutura da tabela de dados analíticos após seleção e limpeza das variáveis. (continua abaixo)

id	ck19	ep_cam	c644	outcome	time	sexo	idade	idade_cat	tx	etiologia_principal	etiologia_zaria	terapia_loccorreg	metodo	momento	recidiva	epoca_recidiva_em_meses	sorafenibe	numero_de_lesoes	maior_tam_mambo	maior_tam_mambo_cat	classif_mscro	cirrose	padrao_histologico_preadominante	padrao_histologico_secundario	
1																									
2																									
3																									
...																									
N																									

tipo_histologico	dif_edmonson_predominante	dif_edmonson_secundaria	dif_oms	dif_oms_secundaria	dif_oms_terciaria	inv_vascular_micro	inv_vas_macro	margin	inf_inflamatorio_peri_intra_tumoral	linfocitos	neutrofilos	esteatose_asociada	necrose_tu	necrose_qt	prolif_ductular	mtx	local_mtx	epoca_mtx_anos

Todas as variáveis da tabela de dados analíticos foram identificadas de acordo com as descrições das variáveis, e os valores foram identificados de acordo com o dicionário de dados providenciado. Estas identificações possibilitarão a criação de tabelas de resultados com qualidade de produção final.

## 4 PARÂMETROS DO ESTUDO

### 4.1 Desenho do estudo

Coorte retrospectiva, baseada em dados de prontuários.

### 4.2 Critérios de inclusão e exclusão

N/A

### 4.3 Exposições

- CK19
- EpCAM
- CD44

### 4.4 Desfechos

**Especificação das medidas de desfecho** (Zarin, 2011):

1. (Domínio) Câncer
2. (Mensuração específica) Óbito por CHC
3. (Métrica específica) Tempo até evento
4. (Método de agregação) Risco (*hazard*)

#### **Desfecho primário**

Risco (*hazard*) de óbito por CHC (causa específica).

### 4.5 Covariáveis

1. Sexo
2. Idade
3. Idade (Maior ou igual a 60 anos)
4. Transplante
5. Etiologia primária
6. Etiologia secundária
7. Terapia localregional
8. Método da terapia localregional
9. Momento da terapia localregional
10. Recidiva
11. Quanto tempo apresentou recidiva (meses)
12. Tratamento com Sorafenibe
13. Número de lesões
14. Tamanho da maior lesão
15. Tamanho da maior lesão (maior ou igual a 5 cm)
16. Classificação macroscópica

17. Cirrose
18. Padrão histológico predominante
19. Padrão histológico secundário
20. Tipo histológico
21. Grau de diferenciação segundo Edmonson predominante
22. Grau de diferenciação segundo Edmonson secundário
23. Grau de diferenciação segundo OMS predominante
24. Grau de diferenciação segundo OMS secundário
25. Grau de diferenciação segundo OMS terciário
26. Invasão vascular microscópica
27. Invasão vascular macroscópica
28. Margem cirúrgica
29. Infiltrado inflamatório peri/intra tumoral
30. Linfócitos
31. Neutrófilos
32. Esteatose associada ao tumor
33. Necrose tumoral
34. Necrose pós terapia locorregional
35. Proliferação ductular
36. Metástase
37. Local da metástase
38. Quanto tempo depois apresentou metástase (anos)

## 5 MÉTODOS ESTATÍSTICOS

### 5.1 Análises estatísticas

#### 5.1.1 Análise descritiva

O perfil epidemiológico dos participantes do estudo será descrito na baseline. As características demográficas (sexo e idade) e clínicas serão descritas como média (DP), mediana (amplitude) ou frequência e proporção (%), conforme apropriado. As distribuições das características dos participantes serão resumidas em tabelas e visualizadas em gráficos exploratórios

#### 5.1.2 Análise inferencial

Todas as análises inferenciais serão realizadas com base nos modelos estatísticos (descritos na próxima seção).

### 5.1.3 Modelagem estatística

A incidência de óbito por CHC será estimada com o modelo de Cox de riscos proporcionais, testando sua associação com os marcadores (seção 4.3) e considerando as covariáveis como candidatas para inclusão (seção 4.5). Para estimar a sobrevida de causa específica os óbitos de outras causas serão censurados na data do óbito. Nesta análise não serão avaliadas as subdistribuições das várias causas de óbito, mas apenas a mortalidade de causa específica. Devido ao grande número de covariáveis relativo ao baixo tamanho amostral, as covariáveis que contenham dados faltantes serão excluídas do universo de candidatas a inclusão no modelo.

Planeja-se uma análise de sensibilidade onde vários modelos serão testados, incrementando em complexidade, para avaliar a robustez das estimativas obtidas. Pretende-se ajustar três modelos aos dados. O modelo mais simples incluirá apenas as exposições. Um modelo intermediário acrescentará as interações de segunda ordem entre as exposições. A premissa de riscos proporcionais será avaliada em ambos os modelos e os termos que a violarem serão removidos, assim como termos que resultem em falhas de convergência.

Um terceiro modelo será especificado com o maior número de covariáveis possível na amostra do estudo para uma estimativa ajustada. O modelo intermediário será submetido ao processo de seleção de variáveis *stepwise* para tentar acrescentar covariáveis e ajustar as estimativas de efeito entre os marcadores e o óbito. Estipula-se tentar o método *stepwise* em duas direções (*forward* e *backward*) e o modelo resultante com maior número de covariáveis será selecionado como melhor modelo. Termos que violem a premissa de riscos proporcionais serão removidos, assim como termos que induzam o modelo a estimativas excessivamente imprecisas (relativas às obtidas nos dois modelos anteriores).

Caso as tentativas com a técnica *stepwise* não resultem em modelos interpretáveis, a modelagem será refeita incluindo covariáveis manualmente, seguindo a ordem de prioridade: Sexo, Idade, Idade (Maior ou igual a 60 anos), Etiologia primária, Recidiva, Cirrose, Grau de diferenciação segundo OMS predominante e Invasão vascular microscópica.

### 5.1.4 Dados faltantes

Não serão realizadas imputações de dados faltantes. Todas as avaliações serão realizadas como análises de casos completos. As frequências e proporções de dados faltantes serão reportadas em tabelas.

## 5.2 Significância e Intervalos de Confiança

Todas as análises serão realizadas ao nível de significância de 5%. Todos os testes de hipóteses e intervalos de confiança calculados serão bicaudais.

### 5.3 Tamanho da amostra e Poder

N/A

### 5.4 Softwares utilizados

Esta análise será realizada utilizando-se o software R versão 4.3.0.

## 6 OBSERVAÇÕES E LIMITAÇÕES

### Guideline recomendado de relato completo

A adoção dos guidelines de relato da rede EQUATOR (<http://www.equator-network.org/>) têm sido ampliada em revistas científicas. Recomenda-se que todos os estudos observacionais sigam o guideline STROBE (von Elm et al, 2014).

Em particular quando um estudo retrospectivo é conduzido usando registros hospitalares de rotina, incluindo informações de prontuários, é recomendado que a extensão RECORD do guideline STROBE seja considerado (Benchimol et al, 2015).

## 7 REFERÊNCIAS

- **SAR-2023-028-AN-v01** – Associação entre os marcadores CK19, EpCAM e CD44 e óbito por carcinoma hepatocelular: coorte retrospectiva
- **SAR-2023-031-AN** – Associação entre fatores diagnósticos e marcadores de CHC: estudo caso-controle
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (<https://doi.org/10.1056/NEJMsa1012065>).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (<https://doi.org/10.1001/jama.2017.18556>).
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9 (<https://doi.org/10.1016/j.ijsu.2014.07.013>).
- Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015 Oct 6;12(10):e1001885 (<https://doi.org/10.1371/journal.pmed.1001885>).

## 8 APÊNDICE

Este documento foi elaborado seguindo recomendações de estrutura para Planos de Análise Estatística (Gamble, 2017) para maior transparência e clareza.

### 8.1 Análises associadas

Esta análise é parte de um projeto maior e é suportada por outras análises, disponíveis abaixo.

**Associação entre fatores diagnósticos e marcadores de CHC: estudo caso-controle**

<https://philsf-biostat.github.io/SAR-2023-031-AN/>

### 8.2 Disponibilidade

Todos os documentos gerados nessa consultoria foram incluídos no portfólio do consultor.

O portfólio pode ser visto em:

<https://philsf-biostat.github.io/SAR-2023-028-AN/>