

# **Modelo Logístico para estimar o desenvolvimento de cardiopatia em populações infectadas com *Trypanossoma cruzi* no município de Barcelos, no estado do Amazonas.**

Cecília Lorenzo de Siqueira

cecilialorenzo@gmail.com

Orientada por:

Inez Costa  
Alexandre Federici  
Elaine Machtyngier

Este projeto foi motivado pela tese de mestrado realizada pela médica Lúcia Brum, que visava avaliar a soro prevalência da infecção chagásica e a morbidade da doença de Chagas no município de Barcelos, Amazônia.

Essa enfermidade foi descoberta pelo pesquisador do Instituto Oswaldo Cruz (IOC), Carlos Chagas, mais precisamente em 1909. Foi constatado que a doença de Chagas tem como agente causal o protozoário parasito hemoflagelado denominado *Trypanossoma cruzi* e como seu maior transmissor um inseto triatomíneo conhecido como “barbeiro”.

A complexidade dessa doença dá-se pelas inúmeras formas de transmissão. Dentre as várias maneiras do contágio da doença, temos a transmissão vetorial que acontece a partir do contato do homem suscetível com as excretas contaminadas do vetor, que ao se alojar na corrente sanguínea ocasiona assim o desenvolvimento da enfermidade. Essa é a forma mais clássica de se contrair esse tipo de doença. Outras possíveis formas de contaminação são as transmissões transfusional, congênita e oral.

Ao se contrair a doença de Chagas, ela pode evoluir para as seguintes fases: aguda ou crônica. A primeira pode ser tanto sintomática como assintomática, já a outra é caracterizada por ser indeterminada, cardíaca, digestiva ou mista.

Resolveu, então, estudar a região da Amazônia, visto que tal área era considerada não endêmica por não apresentar o tipo de transmissão mais comum (domiciliar vetorial) para a doença em questão. Então, na década de 80, foi instaurado um Inquérito Sorológico, já que havia suspeitas que essa área apresentava a soro prevalência da doença de Chagas, porém através de uma transmissão menos comum: transmissão vetorial nos piaçabais. As regiões localizadas no Rio Negro e Rio Japurá são as que apresentavam os maiores índices de prevalência, 6,3% e 6,9%, respectivamente.

A partir das informações coletadas, desencadeou-se, na década de 90, uma série de investigações liderada pelo Departamento de Medicina Tropical do Instituto Oswaldo Cruz.

Neste mesmo período, foram abertos novos três Inquéritos Sorológicos na sede do município de Barcelos. Foi utilizada amostragem aleatória sistemática. Selecionava-se uma a cada quatro casas habitadas e coletava o sangue de toda a família. Os resultados encontrados estão esboçados no quadro abaixo.

Ano	Total de pessoas	Prevalência
1991	710	12,5%
1993	658	13,7%
1997	886	13,2%

Os dados utilizados no projeto levaram em consideração as pessoas que haviam sido abordadas nos três Inquéritos. A mestrandia, Lúcia Brum, juntamente com a equipe da FIOCRUZ voltaram à região em questão para reavaliar essas mesmas pessoas nos aspectos epidemiológicos, clínicos e laboratoriais.

Foram entrevistados e submetidos ao questionário epidemiológico **214** indivíduos, porém apenas **172** compareceram à consulta para a avaliação clínica e laboratorial. Após a unificação das informações epidemiológicas, clínicas e laboratoriais em um banco de dados através do programa EpiData, verificamos uma perda de **20** indivíduos gerada em parte pela não complementação de todas as etapas do estudo clínico ou pela falha na coleta do material biológico, fundamental para o diagnóstico, restando um total de **152** indivíduos que conformam à população estudada.

Os 152 indivíduos foram avaliados, após consentimento informado, quanto aos aspectos: clínicos (anamnese, exame físico), eletrocardiográficos e sorológicos. A classificação das 152 amostras de sangue revelou que 25% das amostras eram positivas nos três testes aplicados, constituindo um grupo de 38 indivíduos; 55% apresentaram todas as reações negativas conformando o grupo de 83 indivíduos considerados negativos para a infecção chagásica. Em 20% dos resultados, representado por um grupo de 31 indivíduos, houve divergências sorológicas, (pelo menos um resultado negativo).

Depois de toda análise médica, foi proposto trabalhar com o banco de dados com as seguintes variáveis:

Variável Resposta

Alteração Cardíaca
--------------------

## Variáveis Independentes

Idade	Sorologia	Gênero	Extrativismo
Extrativismo Piaçaba	Indígena	Diabete	Pressão Arterial
Fumar	Beber	Palpitação	Dor Precordial
Perda da Consciência	Disfagia	Odinofagia	Soluço

Após o conhecimento do banco, o trabalho posterior seria tentar criar um modelo que melhor estime a probabilidade de ocorrência de alteração cardíaca na população do município de Barcelos, devido à doença de Chagas. Então, para tentar explicar a ocorrência ou não de alteração cardíaca foram selecionadas as 16 variáveis citadas no quadro acima, de acordo com o parecer médico.

Dentre os modelos existentes, foi escolhido o logístico para modelar essa situação, já que a variável resposta tem caráter binário, ou seja, o indivíduo é positivo ou negativo para alteração cardíaca, logo, só tem duas respostas plausíveis para a variável em questão.

Definido o modelo a se trabalhar, a próxima etapa foi constituída de uma análise mais detalhada das variáveis independentes. A partir de uma análise exploratória, foi verificado que tanto as variáveis *diabete*, *disfagia*, *odinofagia* e *soluço* deveriam ser excluídas. A primeira por não ter um número expressivo dentro das pessoas coletadas e, também, por ser uma variável de auto-declaração, ou seja, não foi necessário fazer exame de sangue para constar se o pesquisado era ou não diabético. Como o objetivo é tentar explicar a alteração cardíaca e as outras variáveis caracterizam a alteração no sistema digestivo, em relação à doença de Chagas, e nenhuma das pessoas pesquisadas apresentaram alteração à não ser a cardíaca, logo não é cabível continuar com elas no banco de dados.

Em um segundo momento, tanto a variável *fumar* quanto a *beber* foram também descartadas. Acredita-se que ocorreu uma distorção nos números, visto que a área pesquisada é predominante indígena, e sabe-se que culturalmente eles têm o hábito em fumar e beber. Além disso, essa variável também foi de auto-declaração. Possivelmente, a maioria dos indivíduos pesquisados ficaram constrangidos em se declarar fumantes ou que tinham o hábito de ingerir bebida alcoólica.

Conhecendo bem as características da área de estudo, verifica-se uma predominância em habitantes indígenas, mão-de-obra extrativista em geral e extrativista nos piaçabais, com isso, essas variáveis não dariam nenhuma resposta significativa, visto que são características peculiares locais. Logo, excluiu-se as variáveis *indígena*, *extrativismo* e *extrativismo piaçaba*.

Por fim, depois de todas essas análises, restaram sete variáveis para tentar explicar a existência ou não de alteração cardíaca.

Foram feitos uma quantidade considerada de experimentos para conseguir um modelo que melhor se ajustasse à situação proposta. Foram feitas combinações entre as variáveis, porém o melhor resultado foi colocando-as separadamente, sem nenhum tipo de interação. Abaixo segue um experimento realizado.

Classification Table <sup>a</sup>

Observado			Predito		
			ECGAlterado		Porcentagem Correta
			Não	Sim	
Step 1	ECGAlterado	Não	93	11	89,4
		Sim	29	19	39,6
	Porcentagem Completa				73,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	PosDiv(1)	1,256	,427	8,635	1	,003	3,511
	Idade	,046	,014	10,609	1	,001	1,047
	Genero(1)	,865	,436	3,941	1	,047	2,375
	PressaoArterial(1)	,384	,474	,655	1	,418	1,468
	Palpitação(1)	-,343	,455	,569	1	,451	,709
	DorPrecordial(1)	,413	,456	,824	1	,364	1,512
	PerdadaConsciência(1)	-,326	,571	,326	1	,568	,722
	Constant	-4,125	,796	26,821	1	,000	,016

a. Variable(s) entered on step 1: PosDiv, Idade, Genero, PressaoArterial, Palpitação, DorPrecordial, PerdadaConsciência.

Depois de escolher realizar um modelo sem nenhuma interação entre as variáveis independentes, verifica-se no exemplo acima que a *pressão arterial*, *palpitação*, *dor precordial* e *perda da consciência* são variáveis não significativas a um nível de 5%. Analisando por uma ótica médica, nenhuma das variáveis citadas poderiam tentar explicar o fato de um indivíduo apresentar alteração cardíaca, visto que são consequência desse tipo de alteração. Como ambas as variáveis não apresentaram nenhuma significância para o modelo, então foram retiradas.

Assim, após a exclusão das últimas variáveis, o modelo proposto terá *alteração cardíaca* como sendo variável resposta e o conjunto das variáveis *idade*, *gênero* e *sorologia* como sendo as predictoras. Vale ressaltar que foi criado uma nova variável a partir da variável sorologia, visto que a mesma pode ser subdividida em *positiva*, *divergente* ou *negativa*, e como tomaremos como base ser soro-negativo, foi criado à variável *PosDiv*.

Rodando o modelo proposto, com ajuda do software SPSS, obteve-se o seguinte resultado:

**Classification Table<sup>a</sup>**

Observado		Predito		
		ECGAlterado		Porcentagem Correta
		Não	Sim	
Step 1:	ECGAlterado Não	94	10	90,4
	Sim	29	19	39,6
Porcentagem Completa				74,3

a. The cut value is ,500

**Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Idade	,048	,013	13,945	1	,000	1,049
PosDiv(1)	1,197	,407	8,646	1	,003	3,310
Genero(1)	,896	,427	4,412	1	,036	2,451
Constant	-4,190	,786	28,412	1	,000	,015

a. Variable(s) entered on step 1: Idade, PosDiv, Genero.

Com o fim do trabalho proposto, pode-se tirar algumas conclusões. Primeiramente, que o modelo encontrado tem caráter exclusivamente peculiar da região em estudo, ou seja, não se teria êxito utilizar esse modelo para talvez tentar explicar a alteração cardíaca, devido à doença de Chagas, de um estado da região Sul do país.

Ao analisar o quadro imediatamente acima, verifica-se que, a cada incremento de idade, a chance de uma pessoa vir a ter alteração aumenta em 4,9%. Já, se uma pessoa pertence ao grupo PosDiv, ela terá 3, 310 mais chance de ter alteração cardíaca do que uma pessoa que pertence o grupo negativo. Pode ser citado também o fato de que uma pessoa do sexo masculino tem 2, 451 mais chance de apresentar qualquer tipo de alteração cardíaca do que uma pessoa do sexo feminino, porém essa característica é bem local, visto que tem uma maior quantidade de homem exposto nos piaçabais, assim uma maior probabilidade de ser picado pelo barbeiro, já que esse é o tipo de transmissão mais comum na Amazônia.

Analisando o modelo, de modo geral, verifica-se uma maior eficácia em relação ao grupo de pessoas que não apresentam alteração cardíaca. Possivelmente, a não eficácia em relação ao grupo que apresentam alteração cardíaca dá-se pela omissão de variáveis importantes no desenvolvimento do trabalho, mas não havia possibilidade de não retirá-las porque, assim poderia ocorrer uma distorção nos resultados a serem obtidos.

Outro empecilho encontrado no decorrer do projeto foi o tamanho da amostra coletada, considerada pequena. Para uma possível continuação do trabalho, necessita-se de um banco de dados mais completo porque só assim se poderá suprir as limitações para o modelo construído.

Como o modelo acima supõe a variável Idade de forma linear, o que não ocorre normalmente no modelo logístico, procurou-se trabalhar com esta variável em intervalos de

classe. As faixas de idade estão subdivididas da seguinte maneira: Faixa (1) é composta por indivíduos que apresentam idade inferior a 25 anos. Faixa (2) é composta por indivíduos entre 25 e 35 anos, exclusive. Faixa (3) é composta por indivíduos entre 35 e 55 anos, exclusive. Faixa (4) é composta por indivíduos entre 55 a 70 anos, exclusive. Faixa (5) é composta por indivíduos de 70 anos ou mais.

**Classification Table<sup>a</sup>**

Observed			Predicted		
			ECGAlterado		Percentage Correct
			Não	Sim	
Step 1	ECGAlterado	Não	95	9	91,3
		Sim	34	14	29,2
		Overall Percentage			71,7

a. The cut value is ,500

**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	PosDiv(1)	1,171	,409	8,214	1	,004	3,225
	ClasseEtaria			14,610	4	,006	
	ClasseEtaria(1)	1,364	,907	2,264	1	,132	3,913
	ClasseEtaria(2)	1,526	,821	3,454	1	,063	4,601
	ClasseEtaria(3)	2,542	,850	8,938	1	,003	12,702
	ClasseEtaria(4)	3,253	1,066	9,318	1	,002	25,880
	Genero(1)	,869	,432	4,055	1	,044	2,385
	Constant	-3,680	,870	17,875	1	,000	,025

a. Variable(s) entered on step 1: PosDiv, ClasseEtaria, Genero.

Analisando o modelo acima, percebe-se que a tabela de análise da variância responde de forma significativa ao usarmos as faixas etárias, o modelo foi menos sensível na estimação dos pacientes que apresentam ECG alterado.

Como trabalho futuro, seria interessante estudar este modelo em áreas conhecidamente endêmicas para Chagas onde se teria um maior banco de dados para o teste da consistência do modelo.

### **Bibliografia**

David W. Hosmer, Jr. e Stanley Lemeshow. *Applied Logistic Regression*. Wiley-Interscience Publication; 2 edition (September 15, 2000).

John Neter e William Wasserman. *Applied Linear Statistical Models*. McGraw-Hill/Irwin; 4 edition (February 1, 1996).