

# Ensaio Clínico de Contraceptivos

*José Luiz Padilha da Silva*

*24 de outubro de 2018*

## Exemplo: Ensaio Clínico de Contraceptivos

### Introdução

Este exemplo é de um estudo longitudinal de uso de contraceptivos em mulheres (Machine et al., 1988). Neste estudo, as mulheres receberam uma injeção de 100 mg ou 150 mg de acetato de medroxiprogesterona (DMPA) no dia da aleatorização e três injeções adicionais no intervalo de 90 dias. Houve uma visita final de acompanhamento aos 90 dias para uma quarta injeção, isto é, um ano após a primeira injeção. Durante todo o estudo cada mulher completou um diário menstrual que registrava qualquer distúrbio de padrão de sangramento vaginal. Os dados dos diários foram usados para determinar se cada mulher apresentou amenorreia – ausência de fluxo menstrual – por um número específico de dias.

Um total de 1151 mulheres completaram os diários e os dados desses diários foram usados para gerar uma sequência binária para cada mulher, correspondendo a se ela experimentou amenorreia em quatro intervalos sucessivos de três meses. Uma característica deste ensaio clínico é que houve perda de acompanhamento (*dropout*) considerável. Mais que um terço das mulheres abandonaram o estudo antes de sua conclusão.

- 17% abandonaram após receber apenas uma injeção de DMPA;
- 13% abandonaram após receber apenas duas injeções, e
- 7% abandonaram após receber três injeções.

Para as mulheres que saíram do estudo antes do fim do período de injeção de 90 dias, foi determinado se elas experimentaram amenorreia, usando os dados disponíveis do diário de menstruação para aquele período. O impacto dos dados ausentes nas análises será discutido posteriormente na disciplina.

O objetivo desta análise é determinar mudanças individuais no risco de amenorreia ao longo do estudo (12 meses), e a influência da dose de DMPA nessas mudanças. Vale notar que a covariável de tratamento (dose alta versus dose baixa de DMPA) é fixa no tempo.

### Modelo Misto

Seja  $Y_{ij} = 1$  se a  $i$ -ésima mulher experimentou amenorreia no  $j$ -ésimo intervalo de injeção ( $j = 1, \dots, 4$ ), e  $Y_{ij} = 0$  caso contrário. O seguinte modelo logístico de efeitos mistos para  $Y_{ij}$  foi ajustado aos dados:

$$\text{logito}\{E(Y_{ij}|b_i)\} = \beta_1 + \beta_2 \text{tempo}_{ij} + \beta_3 \text{tempo}_{ij}^2 + \beta_4 \text{dose}_i \times \text{tempo}_{ij} + \beta_5 \text{dose}_i \times \text{tempo}_{ij}^2 + b_{1i},$$

em que  $\text{tempo} = 1, 2, 3, 4$  para os quatro intervalos consecutivos de injeção, e  $\text{dose} = 1$  se a mulher foi aleatorizada em 150 mg de DMPA, e  $\text{dose} = 0$  caso contrário. Note que não há medida *baseline* de amenorreia antes de receber a primeira injeção contraceptiva. Contudo, devido à aleatorização, assumimos que o risco *baseline* (em  $\text{tempo} = 0$ ) é o mesmo em ambos os grupos tratamento e omitimos o efeito principal da dose no modelo.

Dado  $b_i$ , assumimos que os  $Y_{ij}$  são independentes e têm distribuição Bernoulli, com  $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)[1 - E(Y_{ij}|b_i)]$  e  $\phi = 1$ . Finalmente, assumimos que o único efeito aleatório  $b_i$  tem distribuição normal univariada, com média zero e variância  $\sigma_b^2$ ,  $b_{1i} \sim N(0, \sigma_b^2)$ . Este modelo assume que há heterogeneidade natural na propensão ou risco de amenorreia ao longo do estudo e este risco persiste em todas as respostas binárias obtidas ao longo do estudo. A seguir os resultados do ajuste.

```
library(foreign); library(lme4)
```

```
## Loading required package: Matrix
```

```
dados=read.dta("amenorrhea.dta")
head(dados)
```

```
##   id trt time  y
## 1  1  0    1  0
## 2  1  0    2 NA
## 3  1  0    3 NA
## 4  1  0    4 NA
## 5  2  0    1  0
## 6  2  0    2 NA
```

```
#
attach(dados)
time2 <- time^2
trt.time <- trt*time
trt.time2 <- trt*time2
```

Na sequência, ajustamos o modelo de efeitos mistos. Foi preciso alterar o otimizador pois a opção padrão retornou um aviso de que o ajuste não convergiu. Vamos considerar 50 pontos de quadratura.

```
m1 <- glmer(y ~ time + time2 + trt.time + trt.time2 + (1 | id), family=binomial,
            na.action=na.omit, control=glmerControl(optimizer="bobyqa"), nAGQ=50)
summary(m1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
## Gauss-Hermite Quadrature, nAGQ = 50) [glmerMod]
## Family: binomial ( logit )
## Formula: y ~ time + time2 + trt.time + trt.time2 + (1 | id)
## Control: glmerControl(optimizer = "bobyqa")
##
```

```
##      AIC      BIC   logLik deviance df.resid
## 3880.9   3918.1 -1934.5   3868.9     3610
##
```

```
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.9194 -0.4295 -0.2273  0.4711  4.0102
##
```

```
## Random effects:
## Groups Name      Variance Std.Dev.
## id      (Intercept) 5.065    2.25
## Number of obs: 3616, groups: id, 1151
##
```

```
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.80568    0.30495 -12.480 < 2e-16 ***
## time         1.13321    0.26819  4.225 2.39e-05 ***
## time2        -0.04192    0.05480 -0.765 0.44428
## trt.time      0.56444    0.19223  2.936 0.00332 **
## trt.time2     -0.10955    0.04961 -2.208 0.02722 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Correlation of Fixed Effects:
##           (Intr) time  time2  trt.tm
## time      -0.831
## time2      0.710 -0.967
## trt.time  -0.053 -0.345  0.434
## trt.time2  0.041  0.333 -0.456 -0.953
```

Estes resultados fornecem evidência de que as log-odds de amenorreia específicas de indivíduo aumentam ao longo dos 12 meses de estudo, e que as mudanças específicas de indivíduo no risco de amenorreia dependem da dose. Por exemplo para uma mulher alocada à dose baixa de DMPA, as log odds de amenorreia aumentam aproximadamente de forma linear, com um aumento na log odds de

- 1.09 (ou  $1.1332 - 0.0419$ ) em 3 meses,
- 2.10 (ou  $2 \times 1.1332 - 4 \times 0.0419$ ) em 6 meses,
- 3.02 (ou  $3 \times 1.1332 - 9 \times 0.0419$ ) em 9 meses, e
- 3.86 (ou  $4 \times 1.1332 - 16 \times 0.0419$ ) em 12 meses.

Estes aumentos no risco correspondem a odds ratio de 3.0 (ou  $e^{1.09}$ ), 8.2 (ou  $e^{2.10}$ ), 20.5 (ou  $e^{3.02}$ ), e 47.5 (ou  $e^{3.86}$ ) em 3, 6, 9, e 12 meses, respectivamente. Por outro lado, para uma mulher alocada à alta dose de DMPA, as log odds de amenorreia aumenta quadraticamente, com um aumento de

- 1.55 em 3 meses,
- 2.79 em 6 meses,
- 3.73 em 9 meses, e
- 4.37 em 12 meses.

Estes aumentos no risco correspondem a odds ratios de 4.7 (ou  $e^{1.55}$ ), 16.3 (ou  $e^{2.79}$ ), 41.7 (ou  $e^{3.73}$ ), e 79.0 (ou  $e^{4.37}$ ) em 3, 6, 9, e 12 meses, respectivamente. Para ambos os grupos, todos esses aumentos nas odds ratio são significativos ao nível 0.05.

Como o tratamento é uma variável que varia entre indivíduos, a interpretação dos efeitos fixos para interações de *dose*  $\times$  *tempo* é mais difícil. Os efeitos de interação têm interpretação em termos de contraste na log odds (ou odds ratio) de amenorreia para duas mulheres que tenham o mesmo risco subjacente de experimentar amenorreia antes da aleatorização, mas que diferem em termos de dose. Assim, a razão das odds de amenorreia em 12 meses para uma mulher alocada ao grupo alta dose, versus uma mulher com o mesmo risco de amenorreia antes da aleatorização e que foi alocada ao grupo dose baixa, é 1.66 (ou  $e^{4.37-3.86}$ ), com intervalo de confiança de 95%: 1.03 a 2.66.

A estimativa de variância dos interceptos aleatórios é relativamente grande,  $\hat{\sigma}_b^2 = 5.065$ . Isto implica que há uma considerável variabilidade na propensão de experimentar amenorreia, já que 95% das mulheres têm um risco *baseline* de amenorreia que varia de

$$\frac{\exp(-3.8057 - 1.96\sqrt{5.0646})}{1 + \exp(-3.8057 - 1.96\sqrt{5.0646})}$$

até

$$\frac{\exp(-3.8057 + 1.96\sqrt{5.0646})}{1 + \exp(-3.8057 + 1.96\sqrt{5.0646})},$$

ou 0.03% até 64.68%.

Alternativamente, podemos interpretar  $\hat{\sigma}_b^2$  recorrendo à noção de distribuição de variável latente. Isto é, podemos assumir um modelo linear de efeitos mistos para a variável latente  $U_{ij}$ ,

$$U_{ij} = \beta_1 + \beta_2 tempo_{ij} + \beta_3 tempo_{ij}^2 + \beta_4 dose_i \times tempo_{ij} + \beta_5 dose_i \times tempo_{ij}^2 + b_{1i} + \varepsilon_{ij},$$

em que  $b_{1i}$  segue uma distribuição normal, com média zero e variância  $\sigma_b^2$ , e os  $\varepsilon_{ij}$  têm uma distribuição logística padrão, com média zero e variância  $\pi^2/3$ . Sem perda de generalidade, podemos assumir que o limiar

para categorizar  $U_{ij}$  é zero, com

$$\begin{aligned} Y_{ij} &= 1, \text{ se } U_{ij} > 0, \\ Y_{ij} &= 0, \text{ se } U_{ij} \leq 0. \end{aligned}$$

Este modelo para a variável latente implica o modelo logístico de efeitos mistos para  $Y_{ij}$ ,

$$\text{logito}\{E(Y_{ij}|b_i)\} = \beta_1 + \beta_2 \text{tempo}_{ij} + \beta_3 \text{tempo}_{ij}^2 + \beta_4 \text{dose}_i \times \text{tempo}_{ij} + \beta_5 \text{dose}_i \times \text{tempo}_{ij}^2 + b_{1i}.$$

Assim, usando a noção de uma distribuição de variável latente, podemos comparar as magnitudes das fontes de variabilidade entre e intra-indivíduos de  $U_{ij}$  em termos da correlação intra-indivíduo (frequentemente chamada de correlação intraclass)

$$\rho = \text{Corr}(U_{ij}, U_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \pi^2/3}.$$

A correlação intra-indivíduo estimada para as respostas latentes é

$$\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \pi^2/3} = \frac{5.065}{5.065 + 3.290} = 0.61,$$

indicando que há heterogeneidade substancial na propensão latente de experimentar amenorreia. Note que  $\rho$  é a correlação marginal (ponderada na distribuição dos efeitos aleatórios) entre os  $U_{ij}$  não observados; não é a correlação marginal entre os  $Y_{ij}$ .

O modelo misto considerado acima inclui apenas um efeito aleatório,  $b_{1i}$ . Com dados binários, e medidas em apenas quatro ocasiões, deve-se tomar muito cuidado na especificação dos efeitos aleatórios já que uma quantidade limitada de dados pode não suportar a estimação de mais de um único componente de variância. A inclusão de interceptos e inclinações aleatórias resultou em problemas de convergência durante a estimação. Intuitivamente, tentativas de ajustar uma série de regressões logísticas aos dados com no máximo 4 observações (o número de medidas repetidas em cada mulher com dados completos; lembre que devido ao *dropout* 37% das mulheres tiveram menos que 4 medidas) provavelmente resultarão em problemas numéricos e/ou produzirão estimativas instáveis.

Um aspecto importante de dados longitudinais binários é que geralmente não há muita informação disponível sobre os efeitos aleatórios, além de um efeito aleatório em nível de indivíduo (ou intercepto aleatório), quando o número de medidas repetidas é pequeno. Assim, é comum encontrar problemas de convergência em modelos mistos para dados binários.

As estimativas anteriores de efeitos fixos e componentes de variância foram obtidas maximizando uma verossimilhança integrada, em que a integração sobre a distribuição dos efeitos aleatórios foi realizada usando métodos numéricos de quadratura. A escolha dos pontos de quadratura determina o grau de precisão da estimativa. É recomendado avaliar a análise, com um número crescente de pontos de quadratura, até que todas as estimativas e seus erros padrões se tornem estáveis.

## Modelo Marginal

Note que estimativas dos efeitos fixos em modelos logísticos de efeitos mistos são maiores que aqueles obtidos em análises similares usando um modelo marginal. Considere o seguinte modelo marginal ajustado aos dados de amenorreia

$$\text{logito}\{E(Y_{ij})\} = \beta_1 + \beta_2 \text{tempo}_{ij} + \beta_3 \text{tempo}_{ij}^2 + \beta_4 \text{dose}_i \times \text{tempo}_{ij} + \beta_5 \text{dose}_i \times \text{tempo}_{ij}^2$$

e assumo uma estrutura do tipo simetria composta.

O resultado do ajuste, obtido usando o enfoque GEE, é dado a seguir.

```
library(geepack)
m2 <- geeglm(y ~ time + time2 + trt.time + trt.time2, id=id, corstr="exchangeable",
             family=binomial)
summary(m2)
```

```
##
## Call:
## geeglm(formula = y ~ time + time2 + trt.time + trt.time2, family = binomial,
##       id = id, corstr = "exchangeable")
##
## Coefficients:
##              Estimate Std.err    Wald Pr(>|W|)
## (Intercept) -2.23696   0.17649  160.643 < 2e-16 ***
## time         0.69674   0.15855   19.310 1.11e-05 ***
## time2        -0.03275   0.03197    1.050 0.30551
## trt.time      0.32836   0.11000    8.910 0.00284 **
## trt.time2     -0.06366   0.02857    4.966 0.02585 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Estimated Scale Parameters:
##              Estimate Std.err
## (Intercept)    1.001 0.02871
##
## Correlation: Structure = exchangeable Link = identity
##
## Estimated Correlation Parameters:
##              Estimate Std.err
## alpha    0.3629 0.02429
## Number of clusters: 1151 Maximum cluster size: 4
```

Note que os coeficientes estimados do modelo marginal (em valores absolutos) são menores que aqueles estimados do modelo misto. Contudo, diferenças nos coeficientes estimados e nas odds ratio se devem a diferentes interpretações de  $\beta$  nestas duas classes de modelos, ou seja, eles respondem questões distintas:

- As estimativas dos efeitos fixos de dose no modelo de efeitos mistos descreve o efeito da dose no risco de amenorreia para uma mulher específica.
- O efeito correspondente no modelo marginal descreve os efeitos da dose na prevalência de amenorreia na população de mulheres alocadas à dose alta versus dose baixa de DMPA.

Utilizando a aproximação de Zeger (1988), chegamos nas seguintes valores para os efeitos marginais:

```
beta.lme=summary(m1)$coef[,1]
var_bi=as.numeric(VarCorr(m1)); var_bi
```

```
## [1] 5.065
```

```
k=16*sqrt(3)/(15*pi)
fator=sqrt(1+k^2*var_bi)
fator
```

```
## [1] 1.659
```

```
beta.mar.est=beta.lme/fator
beta.mar.est
```

```
## (Intercept)      time      time2    trt.time    trt.time2
##      -2.29427      0.68316     -0.02527      0.34027     -0.06604
```

Note a concordância entre os efeito marginal obtido e aquele predito pela aproximação usando o modelo misto.