

# Análisis de datos cualitativos

---

TABLAS DE CONTINGENCIA



# Objetivos

---

## Tablas de frecuencias (tablas de contingencia)

- Diseño de cohortes y casos-contróles
- Obtención de los porcentajes correctos según el diseño

## Análisis de Tablas 2x2: Riesgo relativo y odds ratio

- Concepto y uso
- Cálculo con el paquete *epitools*

## Tablas RxC

- Conceptos básicos
- El paquete *CrossTable*
- Interpretación del test de chi-cuadrado
- Interpretación de los residuales ajustados

# Tipos de diseños

---

## Diseño de cohortes

- Seleccionamos muestras de las condiciones que consideramos que pueden influir (causar) una determinada respuesta (p.e. evolución de los pacientes, gravedad, aparición de alergias, etc.).
- Las cohortes pueden estar definidas en base a una intervención. En este caso estaríamos, formalmente, en una situación de estudio experimental (p.e. comparar la evolución de pacientes en respuesta a tratamientos).

## Diseño de casos y controles

- Seleccionamos muestras en función de un problema de salud o de una condición que queremos relacionar con sus posibles causa. Por ejemplo, grupos de personas que han sufrido un ataque cardíaco (casos) y personas que han sufrido episodios de problemas cardíacos leves (controles). Queremos ver qué factores (p.e. fumar, hipertensión, etc.) se asocian a la aparición de ataques cardíacos.

# Tabla de contingencia

## Diseño de cohortes

---

Disponemos de tres métodos para obtener un producto y verificamos la calidad del resultado:

	Método 1	Método 2	Método 3
Baja	12	10	12
Media	22	33	15
Alta	35	15	22
Total	69	58	49

¿Podemos aceptar que los tres métodos dan resultados similares? En caso contrario, ¿qué método es mejor?

# Datos

A partir de una tabla de frecuencias

	Método 1	Método 2	Método 3
Baja	12	10	12
Media	22	33	15
Alta	35	15	22
Total	69	58	49

```
datos <- matrix(c(12,10,12,22,33,15,35,15,22),nrow=3,byrow=T)
colnames(datos) <- c("Método 1","Método 2","Método 3")
rownames(datos) <- c("Baja","Media","Alta")
datos %>% addmargins(1)
```

```
      Método 1 Método 2 Método 3
Baja      12      10      12
Media     22      33      15
Alta     35      15      22
Sum       69      58      49
```

Porcentajes por columnas

```
> prop.table(datos,2) %>% addmargins(1) %>% round(2)
      Método 1 Método 2 Método 3
Baja     0.17     0.17     0.24
Media    0.32     0.57     0.31
Alta     0.51     0.26     0.45
Sum      1.00     1.00     1.00
```

Podemos aceptar que los tres métodos dan resultados similares? En caso contrario, ¿qué método es mejor?

Parece que el método 2 da un porcentaje menor de calidad alta.

Los métodos 1 y 3 dan resultados parecidos.

# Datos

## A partir de un *data.frame*

```
t <- with(regicor, (table(death, sex)))  
t  
prop.table(t, 2) %>% round(3)
```

	sex			sex	
death	Male	Female	death	Male	Female
No	943	1032	No	0.916	0.923
Yes	87	86	Yes	0.084	0.077

Hay que usar la función **table()** para generar resultados que se puedan utilizar en otros procedimientos.

```
data(regicor)  
head(regicor)  
createTable(compareGroups(sex~death, data=regicor))
```

Con el paquete **compareGroups** obtenemos tablas con formato más elegante.

```
-----Summary descriptives table by 'sex'-----
```

	Male N=1030	Female N=1118	p.overall
overall death:			0.574
No	943 (91.6%)	1032 (92.3%)	
Yes	87 (8.45%)	86 (7.69%)	

# Datos

## A partir de un *data.frame*

```
createTable(compareGroups(sex~death,data=regicor))
```

```
-----Summary descriptives table by 'sex'-----
```

	Male N=1030	Female N=1118	p.overall
overall death:			0.574
No	943 (91.6%)	1032 (92.3%)	
Yes	87 (8.45%)	86 (7.69%)	

```
createTable(compareGroups(sex~death,data=regicor),  
            type=1)
```

Indicando `type=1` en `createTable` obtenemos una tabla de porcentajes respecto a la variable de grupo indicada en la fórmula de `compareGroups`.

```
-----Summary descriptives table by 'sex'-----
```

	Male N=1030	Female N=1118	p.overall
overall death:			0.574
No	91.6%	92.3%	
Yes	8.45%	7.69%	

# Gráfico de barras

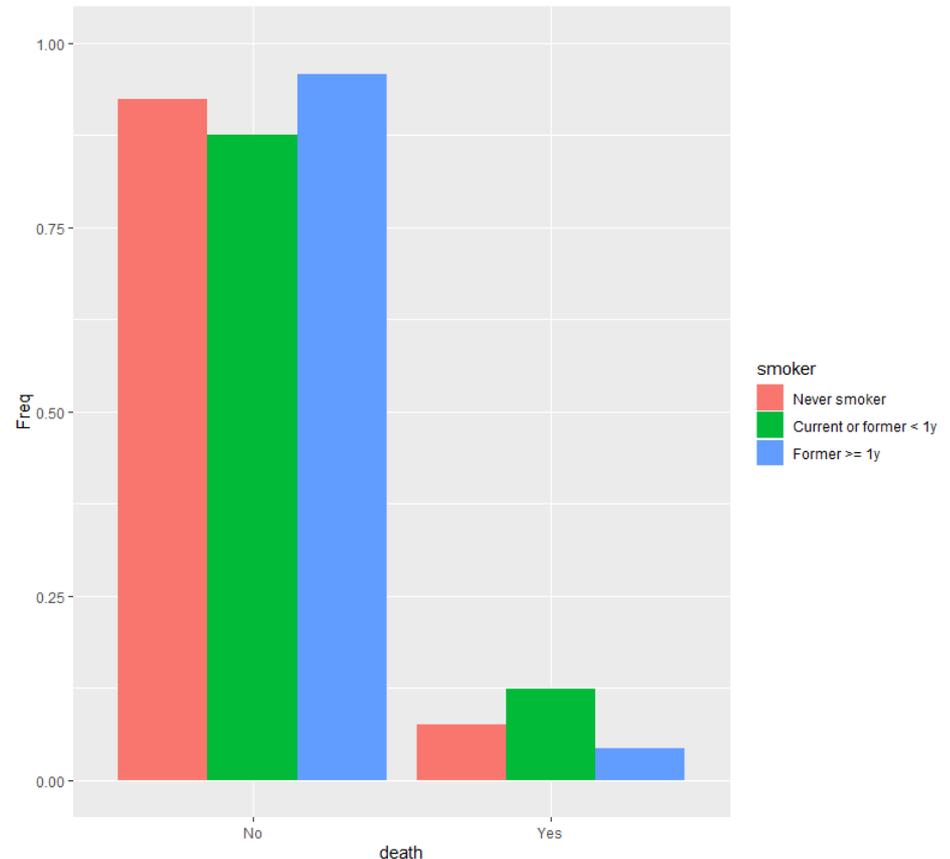
1. Primero creamos la tabla de frecuencias (t)
2. Luego calculamos los porcentajes (tp) y transformamos la tabla en un data.frame para poder usar **ggplot**

```
t <- with(regicor, (table(death, smoker)))
tp <- prop.table(t, 2) %>% round(3) %>%
  as.data.frame()
tp
```

	death	smoker	Freq
1	No	Never smoker	0.924
2	Yes	Never smoker	0.076
3	No	Current or former < 1y	0.876
4	Yes	Current or former < 1y	0.124
5	No	Former >= 1y	0.957
6	Yes	Former >= 1y	0.043

A partir del data.frame, usamos **ggplot** con **geom\_bar**.

```
ggplot(tp, aes(x=death,
               y=Freq,
               fill=smoker))+
  geom_bar(stat='identity',
           position=position_dodge())+
  ylim(0,1)
```



# Comparación de dos grupos

## Riesgo relativo en el diseño de cohortes

---

En un diseño de cohortes seleccionamos a los individuos en función de los grupos (causa).

	No A	A	Total
Grupo de referencia (G0)	f01	f02	f0.
Grupo que se compara (G1)	f11	f12	f1.

### Riesgo relativo

- $RR = (f12/f1.) / (f02/f0.)$

**Un RR=1 indica que el suceso es igualmente probable en los dos grupos.**

**Un RR>1 indica que el suceso es más probable en el grupo que se compara con el de referencia.**

# Comparación de dos grupos

## Diseño de cohortes

---

```
head(dt)
with(dt, table(Status, Exposicio))
createTable(compareGroups(Exposicio~Status, data=dt))
```

```
id Exposicio Status      Exposicio
1      Alta   Viu      Status Baixa Alta
2      Alta   Mort     Viu      36  28
3      Alta   Viu      Mort     17  19
4      Baixa  Viu
5      Baixa  Mort
6      Baixa  Viu
```

```
-----Summary descriptives table by 'Exposicio'--
```

	Baixa N=53	Alta N=47	p. overall
Status:			0.510
Viu	36 (67.9%)	28 (59.6%)	
Mort	17 (32.1%)	19 (40.4%)	

Dos grupos experimentales  
(Exposición: Baja/Alta).

Evento: Vivo/Muerto

Los porcentajes que  
podemos calcular son  
respecto a las columnas.

$$RR=40.4/32.1=1.16$$

# Tabla 2x2

## Diseño de cohortes: Riesgo relativo

-----Summary descriptives table by 'Exposicio'--

	Baixa N=53	Alta N=47	p.overall
Status:			0.510
Viu	36 (67.9%)	28 (59.6%)	
Mort	17 (32.1%)	19 (40.4%)	

$$RR=40.4/32.1=1.26$$

```
dt$status <- relevel(factor(dt$status),ref='viu')
dt$Exposicio <- relevel(factor(dt$Exposicio),ref='Baixa')
t <- with(dt,table(Exposicio,status))
riskratio(t)
```

\$data

Exposicio	Status		Total
	Viu	Mort	
Baixa	36	17	53
Alta	28	19	47
Total	64	36	100

\$measure

Exposicio	risk ratio with 95% C.I.		
	estimate	lower	upper
Baixa	1.000000	NA	NA
Alta	1.260325	0.7467538	2.1271

\$p.value

Exposicio	two-sided			chi.square
	midp.exact	fisher.exact		
Baixa	NA	NA	NA	
Alta	0.3954822	0.4109383	0.3852679	

\$correction

[1] FALSE

```
riskratio(t)$measure %>%
round(2)
```

```
risk ratio with 95% C.I.
Exposicio estimate lower upper
Baixa 1.00 NA NA
Alta 1.26 0.75 2.13
```

RR: 1.26 IC (0.75, 2.13)

Los datos no permiten concluir que la exposición alta conlleve un aumento de riesgo de muerte.

# Tabla Rx2

## Diseño de cohortes: Riesgo relativo

-----Summary descriptives table by 'smoker'-----

	Never smoker N=1138	Current or former N=557	< 1y Former N=415	p.overall
death:				<0.001
No	1052 (92.4%)	488 (87.6%)	397 (95.7%)	
Yes	86 (7.56%)	69 (12.4%)	18 (4.34%)	

```
dt <- regicor
dt$death <- relevel(factor(dt$death),ref='No')
dt$smoker <- relevel(factor(dt$smoker),ref='Never smoker')
t <- with(dt,table(smoker,death))
riskratio(t)$data
prop.table(t,1) %>% round(3)
riskratio(t)$measure %>% round(2)
```

smoker	death		
	No	Yes	Total
Never smoker	1052	86	1138
Current or former < 1y	488	69	557
Former >= 1y	397	18	415
Total	1937	173	2110

Las cohortes deben estar en las filas para poder calcular el RR con riskratio. Por lo tanto, calculamos porcentajes de fila.

smoker	death	
	No	Yes
Never smoker	0.924	0.076
Current or former < 1y	0.876	0.124
Former >= 1y	0.957	0.043

smoker	risk ratio with 95% C.I.		
	estimate	lower	upper
Never smoker	1.00	NA	NA
Current or former < 1y	1.64	1.21	2.21
Former >= 1y	0.57	0.35	0.94

Los RR en este resultado son respecto al grupo de referencia (Never smoker).

# Ejemplo

En un estudio sobre un tipo de alergia, se comparan dos grupos de personas según su lugar de residencia. ¿Podemos concluir que las alergias son más frecuentes en la ciudad?

	Alergias	Sin alergias	Total
Ciudad	34	30	64
Campo	43	27	70

$$RR = \frac{34/64}{43/70} = 0.86$$

```
datos <- matrix(c(27,43,30,34),ncol = 2,byrow=T)
rownames(datos) <- c('Campo','Ciudad')
colnames(datos) <- c('Sin alergias','Alergias')
datos
riskratio(datos)$measure %>% round(2)
```

	Sin alergias	Alergias	Total
Campo	27	43	70
Ciudad	30	34	64
Total	57	77	134

```
risk ratio with 95% C.I. estimate lower upper
      Campo      1.00      NA      NA
      Ciudad      0.86    0.64    1.16
```

IC RR: (0.64,1.16)

No podemos concluir que en la ciudad se produzca un aumento de la probabilidad de alergias

# Ejemplos de interpretación

-----Summary descriptives table by 'Genotipo'-----

	AA N=328	Aa N=334	aa N=338	p.overall
Status:				0.332
Normal	240 (73.2%)	252 (75.4%)	264 (78.1%)	
Patològic	88 (26.8%)	82 (24.6%)	74 (21.9%)	

risk ratio with 95% C.I.			
Genotipo	estimate	lower	upper
AA	1.00	NA	NA
Aa	0.92	0.71	1.19
aa	0.82	0.62	1.07

$$\text{RR Aa vs AA: } 24.6/26.8 = 0.92$$

$$\text{RR aa vs AA: } 21.9/26.8 = 0.82$$

Los genotipos de este carácter que tienen un alelo mutante no se asocian a un aumento de riesgo de esta patología (referencia genotipo AA). Los IC del RR incluyen en ambos casos el valor 1.

# Ejemplos de interpretación

-----Summary descriptives table by 'Genotipo'-----

	AA N=328	Aa N=334	aa N=338	p.overall
Status:				0.002
Normal	262 (79.9%)	252 (75.4%)	230 (68.0%)	
Patològic	66 (20.1%)	82 (24.6%)	108 (32.0%)	

risk ratio with 95% C.I.			
Genotipo	estimate	lower	upper
AA	1.00	NA	NA
Aa	1.22	0.92	1.62
aa	1.59	1.22	2.07

RR Aa vs AA:  $24.6/20.1 = 1.22$

RR aa vs AA:  $32.0/20.1 = 1.59$

El genotipo aa se asocia a un aumento de riesgo de esta patología respecto al genotipo AA (IC: 1.22, 2.07). El genotipo Aa no presenta un aumento de la probabilidad de patología respecto a AA (IC: 0.92, 1.62).

# Tablas 2x2

## Diseño de casos y controles

---

Seleccionamos muestras en función de un problema de salud o de una condición que queremos relacionar con sus posibles causa. Por ejemplo, grupos de personas que han sufrido un ataque cardiaco (casos) y personas que han sufrido episodios de problemas cardiacos leves (controles). Queremos ver qué factores (p.e. fumar, hipertensión, etc.) se asocian a la aparición de ataques cardiacos. Por ejemplo:

En este ejemplo se han seleccionado 50 personas que presentan una patología (casos) y 50 personas sanas (controles). Se verifica si un individuo es hipertenso o si tiene tensión normal.

Los porcentajes se calculan para hipertensión y normal en cada grupo (casos y controles).

	Controles N=50	Casos N=50	p. overall
Hiper:			0.317
Normal	28 (56.0%)	22 (44.0%)	
Hipertenso	22 (44.0%)	28 (56.0%)	

# El concepto de odds

---

Si tenemos un suceso de probabilidad  $p$ , el concepto de **odds** se define como:

$$\frac{p}{1-p}$$



Las **odds** indican cuantas veces es más probable observar un suceso que no observarlo.

- Si la probabilidad de padecer una enfermedad a una cierta edad es de 0.6, las **odds** son  $0.6/0.4=1.5$

# El concepto de odds ratio (OR)

---

El OR es el cociente de las **odds** en dos situaciones distintas

- Si la probabilidad de padecer una enfermedad a los 50 años es de 0.6, y a los 60 es de 0.75, el OR (60 años respecto a 50) es:

$$OR = \frac{0.75/0.25}{0.6/0.4} = 2$$

- Por lo tanto, a los 60 años se ha doblado la relación entre la probabilidad de padecer o no la enfermedad con respecto a los 50 años.

El OR ratio **suele interpretarse** como una estimación de riesgo relativo (en el ejemplo sería el riesgo de padecer la enfermedad a los 60 respecto a los 50), aunque **no es una interpretación totalmente correcta**.

Es más adecuado interpretar el OR como una estimación de la existencia de correlación entre las dos variables (p.e. edad y enfermedad).

# El concepto de *odds ratio (OR)*

## Diseño de casos y controles

En un estudio de casos y controles, donde seleccionamos sanos y enfermos y observamos si presentaban o no hipertensión de manera retrospectivas, el OR puede calcularse también como:

	Controles N=50	Casos N=50	p. overall
Hiper:			0.317
Normal	28 (56.0%)	22 (44.0%)	
Hipertenso	22 (44.0%)	28 (56.0%)	

$$OR = \frac{P(\text{Hiper}|\text{Casos})/P(\text{Normal}|\text{Casos})}{P(\text{Hiper}|\text{Controles})/P(\text{Normal}|\text{Controles})}$$

$$OR = \frac{(28/50)/(22/50)}{(22/50)/(28/50)} \longrightarrow OR = \frac{28 \times 28}{22 \times 22} = 1.61$$

# El concepto de *odds ratio (OR)*

## Diseño de casos y controles

	Controles N=50	Casos N=50	p. overall
Hiper:			0.317
Normal	28 (56.0%)	22 (44.0%)	
Hipertenso	22 (44.0%)	28 (56.0%)	

$$OR = \frac{(28/50)/(22/50)}{(22/50)/(28/50)}$$

$$OR = \frac{28 \times 28}{22 \times 22} = 1.61$$

```
createTable(compareGroups(Grupo~Hiper,
                          data=EjOR),
            show.ratio = T)
```

-----Summary descriptives table by 'Grupo'-----

	Controles N=50	Casos N=50	OR	p.ratio	p. overall
Hiper:					0.317
Normal	28 (56.0%)	22 (44.0%)	Ref.	Ref.	
Hipertenso	22 (44.0%)	28 (56.0%)	1.61 [0.73;3.60]	0.239	

# OR para distintas variables

```
data(regicor)
head(regicor)
regicor$histhtn <- relevel(factor(regicor$histhtn),ref='No')
regicor$sex <- relevel(factor(regicor$sex),ref='Female')
createTable(compareGroups(death~histhtn+sex+smoker,data=regicor),
             show.ratio = T,show.p.overall = F)
```

-----Summary descriptives table by 'Overall death'-----

	No N=1975	Yes N=173	OR	p.ratio
histhtn:				
No	1365 (69.4%)	97 (56.1%)	Ref.	Ref.
Yes	603 (30.6%)	76 (43.9%)	1.77 [1.29;2.43]	<0.001
sex:				
Female	1032 (52.3%)	86 (49.7%)	Ref.	Ref.
Male	943 (47.7%)	87 (50.3%)	1.11 [0.81;1.51]	0.522
Smoking status:				
Never smoker	1052 (54.3%)	86 (49.7%)	Ref.	Ref.
Current or former < 1y	488 (25.2%)	69 (39.9%)	1.73 [1.23;2.42]	0.002
Former >= 1y	397 (20.5%)	18 (10.4%)	0.56 [0.32;0.92]	0.021

## Serum Carotenoids and Risk of Cervical Intraepithelial Neoplasia in Southwestern American Indian Women<sup>1</sup>

Melissa A. Schiff, Ruth E. Patterson,<sup>2</sup>  
Richard N. Baumgartner, Mary Masuk,  
Linda van Asselt-King, Cosette M. Wheeler, and  
Thomas M. Becker

### Abstract

The objective of this research was to evaluate the association between serum carotenoids and cervical intraepithelial neoplasia (CIN) among Southwestern American Indian women. Cases were American Indian women with biopsy-proven CIN II/III cervical lesions ( $n = 81$ ) diagnosed between November 1994 and October 1997. Controls were American Indian women from the same clinics with normal cervical epithelium ( $n = 160$ ). All of the subjects underwent interviews and laboratory evaluations. Interviews evaluated demographic information, sexual history, and cigarette smoking. Serum concentrations of  $\alpha$ -carotene,  $\beta$ -carotene,  $\beta$ -cryptoxanthin, lycopene, and lutein/zeaxanthin were measured by high performance liquid chromatography. Cervical human papillomavirus infection was detected using a PCR-based test. Increasing levels of  $\alpha$ -carotene,  $\beta$ -cryptoxanthin, and lutein/zeaxanthin were associated with decreasing risk of CIN II/III. In addition, the highest tertiles of  $\beta$ -cryptoxanthin (odds ratio = 0.39, 95% confidence interval = 0.17–0.91) and lutein/zeaxanthin (odds ratio = 0.40, 95% confidence interval = 0.17–0.95) were associated with the lowest risk of CIN. In conclusion, specially targeted intervention efforts to increase consumption of fruits and vegetables may protect Southwestern American Indian women from developing CIN.

Table 1 Demographic characteristics of study subjects, American Indian dysplasia study, 1994–1997

	Controls ( $n = 160$ ) No. (%)	Cases ( $n = 81$ ) No. (%)	$P_s$
Age (years)			
18–21	29 (18.1)	17 (21.0)	0.86
22–30	99 (61.9)	49 (60.5)	
31–45	32 (20.0)	15 (18.5)	
Years of education			
12 years or below	47 (29.3)	31 (38.3)	0.35
13–15 years	89 (55.6)	38 (46.9)	
16+ years	24 (15.1)	12 (14.8)	
Annual family income			
<\$10,000	38 (23.8)	32 (39.5)	0.005
\$10,000–\$19,999	41 (25.6)	25 (30.9)	
$\geq$ \$20,000	81 (50.6)	24 (29.6)	
Marital status			
Single, never married	46 (28.8)	26 (32.1)	0.03
Married or with partner	106 (66.2)	45 (55.6)	
Divorced	6 (3.8)	3 (3.7)	
Separated	2 (1.2)	7 (8.6)	
Tribal affiliation			
Southwestern	146 (91.3)	72 (88.9)	0.56
Other	14 (8.7)	9 (11.1)	
Blood quantum <sup>a</sup>			
Full	119 (74.4)	63 (77.8)	0.26
$\frac{3}{4}$	15 (9.4)	3 (3.7)	
$\frac{1}{2}$	17 (10.6)	12 (14.8)	
$\frac{1}{4}$	7 (4.4)	1 (1.2)	
$<\frac{1}{4}$	2 (1.2)	2 (2.5)	
Reason for screening			
Family planning	0	2 (2.5)	<0.001
Annual exam	140 (87.5)	45 (55.6)	
Suspected STD	1 (0.6)	1 (1.2)	
Planned follow-up	1 (0.6)	9 (10.8)	
Other <sup>b</sup>	18 (11.3)	24 (29.9)	

<sup>a</sup> Blood quantum is amount of Indian blood reported by study subjects.

<sup>b</sup> Other reasons for screening included postpartum visits, pregnancy testing, and vaginitis.

high serum carotenoids appear to be associated with a decreased risk of cervical neoplasia, although the association with individual carotenoids may differ because of adjustment for different confounders and different dietary habits among diverse populations.

Table 2 Serum carotenoid concentrations as associated with cervical intraepithelial neoplasia; American Indian dysplasia study, 1994–1997

Carotenoids	Controls <i>n</i> = 160 No. (%)	Cases <i>n</i> = 81 No. (%)	Crude		Adjusted**	
			OR	95% CI	OR	95% CI
<i>α</i> -Carotene ( $\mu\text{g}/\text{dl}$ )						
6.5–12.7	53 (33.1)	42 (51.8)	1.0		1.0	
12.8–18.0	53 (33.1)	21 (25.9)	0.50	0.26–0.96	0.52	0.24–1.12
18.1–87.3	54 (33.8)	18 (22.2)	0.42	0.23–0.86	0.46	0.21–1.00
Test for trend			<i>P</i> = 0.01		<i>P</i> = 0.04	
<i>β</i> -Carotene ( $\mu\text{g}/\text{dl}$ )						
11–24	56 (35.0)	32 (39.5)	1.0		1.0	
25–38	51 (31.9)	33 (40.7)	1.13	0.61–2.1	0.96	0.45–2.03
39–174	53 (33.1)	16 (19.8)	0.53	0.26–1.1	0.46	0.20–1.07
Test for trend			<i>P</i> = 0.10		<i>P</i> = 0.08	
<i>β</i> -Cryptoxanthin ( $\mu\text{g}/\text{dl}$ )						
2.2–6.4	55 (34.4)	34 (42.0)	1.0		1.0	
6.5–10.3	52 (32.5)	31 (38.3)	0.96	0.52–1.79	0.80	0.38–1.69
10.4–30.3	53 (33.1)	16 (19.7)	0.49	0.24–0.99	0.39	0.17–0.91
Test for trend			<i>P</i> = 0.06		<i>P</i> = 0.03	
Lycopene ( $\mu\text{g}/\text{dl}$ )						
14.8–30.7	53 (33.1)	37 (45.7)	1.0		1.0	
30.8–40.7	54 (33.8)	25 (30.9)	0.66	0.35–1.25	0.88	0.41–1.89
40.8–69.8	53 (33.1)	19 (23.4)	0.51	0.26–1.00	0.85	0.38–1.91
Test for trend			<i>P</i> = 0.05		<i>P</i> = 0.69	
Lutein/Zeaxanthin ( $\mu\text{g}/\text{dl}$ )						
12–21	49 (30.6)	34 (42.0)	1.0		1.0	
22–28	55 (34.4)	31 (38.3)	0.81	0.44–1.51	0.60	0.28–1.28
29–76	56 (35.0)	16 (19.7)	0.41	0.20–0.84	0.40	0.17–0.95
Test for trend			<i>P</i> = 0.02		<i>P</i> = 0.03	

\* Adjusted for age, cervical infection with any type of HPV, income, and urban vs rural residence.

RESEARCH ARTICLE

Open Access

# Brief episodes of rapid irregular atrial activity (micro-AF) are a risk marker for atrial fibrillation: a prospective cohort study



Tove Fredriksson<sup>\*</sup>, Katrin Kemp Gudmundsdottir, Viveka Frykman, Leif Friberg, Faris Al-Khalili, Johan Engdahl and Emma Svennberg

## Abstract

**Background:** Short supraventricular tachycardias with atrial fibrillation (AF) characteristics are associated with an increased risk of developing AF over time. The aim of this study is to determine if presence of very short-lasting episodes of AF-like activity (micro-AF) can also be used as a marker of undiagnosed silent atrial fibrillation.

**Methods:** In the STROKESTOP II study, a Swedish mass screening study for AF among 75- and 76-year-olds, participants with NT-proBNP  $\geq 125$  ng/L performed intermittent ECG recordings 30 s, four times daily for 2 weeks. Participants with micro-AF (sudden onset of irregular tachycardia with episodes of  $\geq 5$  consecutive supraventricular beats and total absence of p-waves, lasting less than 30 s) were invited to undergo extended AF screening using continuous event recording for 2 weeks. A control group of individuals without micro-AF was examined using the same ECG modalities.

**Results:** Out of 3763 participants in STROKESTOP II who had elevated NT-proBNP levels and were free of AF,  $n = 221$  (6%) had micro-AF. The majority of participants with micro-AF ( $n = 196$ ) accepted further investigation with continuous ECG monitoring which showed presence of AF in 26 of them. In the control group ( $n = 250$ ), continuous monitoring detected 7 new AF cases. Thus, AF was significantly more common in the micro AF group (13%) compared to the control group (3%),  $p < 0.001$ .

**Conclusions:** Presence of short-lasting episodes of AF-like activity (micro-AF) indicates increased likelihood for undetected AF. Continuous screening therefore seems recommendable if a finding of AF would change clinical management.

**Trail registration:** ClinicalTrials.gov, identifier: NCT02743416, registered April 19, 2016.

**Keywords:** Atrial fibrillation, Electrocardiogram, Micro-AF, Supraventricular ectopic beats, Screening, Supraventricular tachycardia

**Table 3** Multivariable analysis for the development of atrial fibrillation

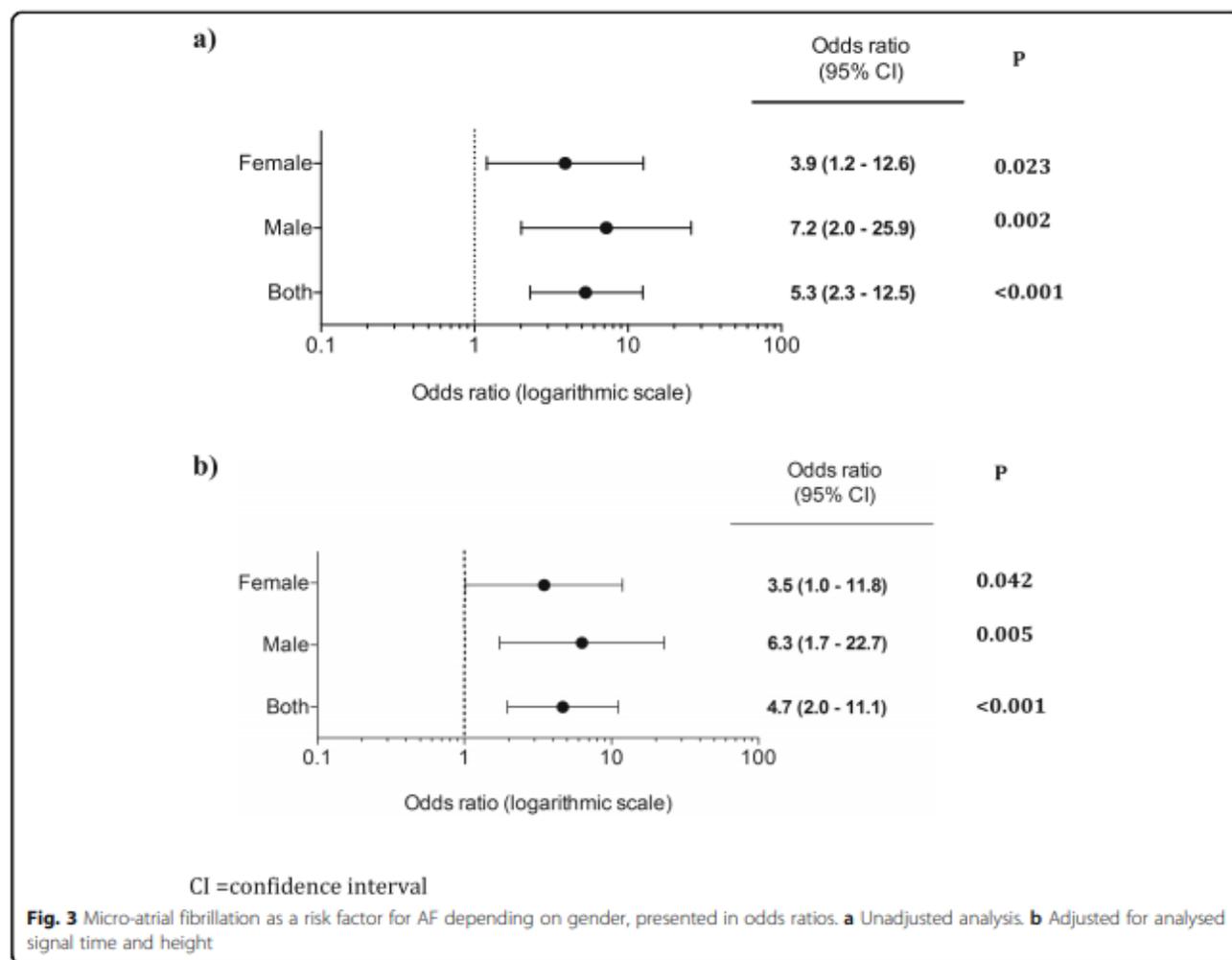
Variable	Unadjusted OR (CI 95%)	P	Model 1		Model 2		Model 3	
			Adjusted OR (CI 95%)	P	Adjusted OR (CI 95%)	P	Adjusted OR (CI 95%)	P
<b>Micro-AF</b>	5.3 (2.3–12.5)	< 0.001	4.7 (2.0–11.1)	< 0.001	5.1 (2.1–12.8)	< 0.001	4.6 (1.8–11.5)	< 0.001

Model 1: Adjusted for height and analysed signal time

Model 2: Adjusted for age, hypertension, heart failure and stroke

Model 3: Adjusted for height, analysed signal time, age, hypertension, heart failure and stroke

AF atrial fibrillation, CI confidence interval, Micro-AF short episodes of irregular supraventricular tachycardia, always refers to when it is seen during intermittent ECG recordings; OR odds ratio, SVEBs supraventricular ectopic beats



# Análisis de tablas RxC

---

# Tablas RxC

---

El análisis anterior no es aplicable cuando tenemos más de dos grupos (p.e. tres cohortes y tres grados de gravedad en la evolución de los pacientes). Estas tablas las denominaremos RxC.

En este tipo de situaciones, nos interesa:

- ¿Hay diferencias entre las probabilidades de evolución para las distintas cohortes?
- ¿En caso afirmativo, en qué cohorte se observa la mejor evolución?

# Tablas RxC (librería *gmodels*)

---

La función **CrossTable** permite representar tablas de frecuencias y proporciona una manera de analizar casos con más de dos cohortes.

Las opciones de **CrossTable** especifican qué porcentajes se calculan, p.e. **prop.r=F** indica que no se calculen los porcentajes de filas, **prop.t** calcularía los porcentajes con respecto al total de la table (sólo tiene sentido en un diseño transversal).

Por defecto se calculan todos los porcentajes. Por lo tanto, deberemos poner a **FALSE** los que no interesen.

**CrossTable** proporciona una evaluación global sobre si podemos aceptar que las proporciones de las respuestas son iguales para todas las cohortes (**p-valor de la prueba chi-cuadrado**).

Además, veremos que los residuales ajustados (**asresid**) permiten evaluar las diferencias entre cohortes.

# Uso de CrossTable

En el data.frame dt tenemos:

```
> head(dt)
  Status Genotipo
1:    ++      aa
2:   +++      AA
3:    +      Aa
4:    +      Aa
5:    +      aa
6:   +++      aa
```

```
with(dt,
      CrossTable(Genotipo, Status,
                 prop.c=F,
                 prop.t=F,
                 prop.chisq = F,
                 asresid = T,
                 format='SPSS',
                 chisq = T))
```

## Cell Contents

```
Count
Row Percent
Adj Std Resid
```

Total Observations in Table: 100

Genotipo	Status			Row Total
	+	++	+++	
AA	13 44.828% 1.175	8 27.586% -0.993	8 27.586% -0.199	29 29.000%
Aa	9 32.143% -0.501	9 32.143% -0.374	10 35.714% 0.923	28 28.000%
aa	14 32.558% -0.623	18 41.860% 1.249	11 25.581% -0.654	43 43.000%
Column Total	36	35	29	100

Statistics for All Table Factors

Pearson's Chi-squared test

```
Chi^2 = 2.612463    d.f. = 4    p = 0.6246169
```

Minimum expected frequency: 8.12

# Interpretación de CrossTable

1. El p-valor=0.0005 nos indica que las proporciones observadas para cada cohorte (genotipo) no permiten concluir que entre genotipos sea igual.
2. Los residuales ajustados negativos indican que la estimación de la probabilidad para esa respuesta y ese genotipo está por debajo de los esperado si no existieran diferencias entre cohortes.
3. Los valores de un residual ajustado superiores (en valor absoluto) a 1.96 indican que en esa casilla la estimación de la probabilidad es superior (valor positivo) o inferior (negativo) al resto para ese nivel de respuesta .

## Cell Contents

Count
Row Percent
Adj Std Resid

Total Observations in Table: 200

Genotipo	Status			Row Total
	+	++	+++	
AA	25 45.455% 1.909	18 32.727% 0.917	12 21.818% -2.739	55 27.500%
Aa	33 44.595% 2.180	15 20.270% -1.866	26 35.135% -0.419	74 37.000%
aa	12 16.901% -3.981	23 32.394% 1.027	36 50.704% 2.978	71 35.500%
Column Total	70	56	74	200

## Statistics for All Table Factors

### Pearson's Chi-squared test

Chi^2 = 19.91589      d.f. = 4      p = 0.0005188585

Minimum expected frequency: 15.4

# Uso de CrossTable

Podemos concluir que existen diferencias en la evolución de los pacientes según el genotipo ( $p < 0.0001$ ).

¿Qué genotipo proporciona un mejor resultado de evolución de los pacientes?

Cell Contents

	Count	Row Percent	Adj Std Resid	
Genotipo				Status
				+
				++
				+++
				Row Total
AA	34	51.515%	3.208	66
				33.000%
				0.509
				-3.684
Aa	28	47.458%	2.184	59
				29.500%
				-0.870
				-1.370
aa	10	13.333%	-5.173	75
				37.500%
				0.325
				4.869
Column Total	72			200

Statistics for All Table Factors

Pearson's Chi-squared test

Chi<sup>2</sup> = 33.97393      d.f. = 4      p = 7.544198e-07

Minimum expected frequency: 16.52

# Interpretación de CrossTable

Si miramos la proporción de resultados (+), vemos que es mayor para el genotipo AA. El valor del residual ajustado (*asresid*) es 3.2. Para Aa es de 2.18 y para aa es de -5.17. Esto permite concluir que los genotipos AA y Aa muestran resultados (+) superiores a lo que cabría esperar en caso de que no hubiera diferencias entre los genotipos. El genotipo aa muestra resultados inferiores.

Cell Contents

	Count	Row Percent	Adj Std Resid	
Genotipo				Row Total
AA	34	51.515%	3.208	66
Aa	28	47.458%	2.184	59
aa	10	13.333%	-5.173	75
Column Total	72			200

Total Observations in Table: 200

Genotipo	Status +	++	+++	Row Total
AA	34 51.515% 3.208	20 30.303% 0.509	12 18.182% -3.684	66 33.000%
Aa	28 47.458% 2.184	14 23.729% -0.870	17 28.814% -1.370	59 29.500%
aa	10 13.333% -5.173	22 29.333% 0.325	43 57.333% 4.869	75 37.500%
Column Total	72	56	72	200

## Statistics for All Table Factors

### Pearson's Chi-squared test

Chi<sup>2</sup> = 33.97393      d.f. = 4      p = 7.544198e-07

Minimum expected frequency: 16.52

# Interpretación de CrossTable

Si miramos la proporción de resultados (++) , vemos que es menor para el genotipo AA. El valor del residual ajustado (asresid) es 3.6. Para Aa es de 1.3 y para aa es de 4.86. Esto permite concluir que el genotipos AA muestra resultados (+++) inferiores a lo que cabría esperar en caso de que no hubiera diferencias entre los genotipos. El genotipo aa muestra resultados más altos.

Cell Contents

	Count	Row Percent	Adj Std Resid	
Genotipo				Row Total
AA	34	51.515%	3.208	66
Aa	28	47.458%	2.184	59
aa	10	13.333%	-5.173	75
Column Total	72			200

Total Observations in Table: 200

Genotipo	Status			Row Total
	+	++	+++	
AA	34	20	12	66
Aa	28	14	17	59
aa	10	22	43	75
Column Total	72	56	72	200

## Statistics for All Table Factors

### Pearson's Chi-squared test

Chi^2 = 33.97393      d.f. = 4      p = 7.544198e-07

Minimum expected frequency: 16.52

# Interpretación de CrossTable

1. El p-valor=0.624 nos indica que las proporciones observadas para cada cohorte (genotipo) no permiten concluir que existan diferencias reales entre la evolución entre genotipos.
2. Los residuales ajustados son todos (en valor absoluto) inferiores a 1.96.
3. En este caso, las diferencias observadas no permiten concluir que la evolución de los pacientes dependa del genotipo.

## Cell Contents

Count
Row Percent
Adj Std Resid

Total Observations in Table: 100

Genotipo	Status			Row Total
	+	++	+++	
AA	13 44.828% 1.175	8 27.586% -0.993	8 27.586% -0.199	29 29.000%
Aa	9 32.143% -0.501	9 32.143% -0.374	10 35.714% 0.923	28 28.000%
aa	14 32.558% -0.623	18 41.860% 1.249	11 25.581% -0.654	43 43.000%
Column Total	36	35	29	100

## Statistics for All Table Factors

### Pearson's Chi-squared test

Chi<sup>2</sup> = 2.612463      d.f. = 4      p = 0.6246169

Minimum expected frequency: 8.12

	Método 1	Método 2	Método 3
Baja	12	10	12
Media	22	33	15
Alta	35	15	22
Total	69	58	49

#### Cell Contents

	Count	Column Percent	Adj Std Resid
Baja	12	17.391%	-0.520
Media	22	31.884%	-1.717
Alta	35	50.725%	2.127
Column Total	69	39.205%	

## Ejemplo

Total Observations in Table: 176

El p-valor es inferior a 0.05. Por lo tanto, debemos concluir que existen diferencias entre métodos.

En este caso, el porcentaje de resultados **altos** es superior en el método 1 y menor en el 2.

	Método 1	Método 2	Método 3	Row Total
Baja	12 17.391% -0.520	10 17.241% -0.489	12 24.490% 1.079	34
Media	22 31.884% -1.717	33 56.897% 3.254	15 30.612% -1.542	70
Alta	35 50.725% 2.127	15 25.862% -2.846	22 44.898% 0.669	72
Column Total	69 39.205%	58 32.955%	49 27.841%	176

#### Statistics for All Table Factors

#### Pearson's Chi-squared test

Chi^2 = 12.35564      d.f. = 4      p = 0.01489358

**Nota:** Ahora hemos puesto los grupos en columnas, por lo tanto los porcentajes que calculamos son los de columnas.

# Ejemplos de interpretación

---

Cell Contents

Count
Column Percent
Adj Std Resid

Total Observations in Table: 168

	Control	Trata 1	Trata 2	Row Total
+	30 54.545% 3.554	20 34.483% -0.242	10 18.182% -3.309	60
++	15 27.273% -1.053	20 34.483% 0.350	20 36.364% 0.699	55
+++	10 18.182% -2.601	18 31.034% -0.104	25 45.455% 2.706	53
Column Total	55 32.738%	58 34.524%	55 32.738%	168

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 17.48711      d.f. = 4      p = 0.001553961

- El p-valor indica que los porcentajes observados no son compatibles con la hipótesis de igualdad entre tratamientos.
- El tratamiento 2 tiene un residual ajustado de 2.706 en la mejora +++. Esto indica que este tratamiento es superior al tratamiento 1.
- El grupo control presenta un porcentaje muy inferior de este nivel de mejora (asresid=-2.6) y un mayor porcentaje de poca mejora (asresid=3.55)

- El p-valor indica que los porcentajes observados son compatibles con la hipótesis de igualdad entre tratamientos.
- Todos los residuales ajustados son menores, en valor absoluto, que 1.96.

```

Cell Contents
-----|
|                Count |
|      Column Percent |
|      Adj Std Resid  |
|-----|

```

Total Observations in Table: 215

	Control	Trata 1	Trata 2	Row Total
+	30 29.412% 0.467	20 34.483% 1.307	10 18.182% -1.864	60
++	40 39.216% 0.578	20 34.483% -0.503	20 36.364% -0.150	80
+++	32 31.373% -1.026	18 31.034% -0.720	25 45.455% 1.907	75
Column Total	102 47.442%	58 26.977%	55 25.581%	215

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 5.450878      d.f. = 4      p = 0.2440812

Cell Contents

	Count
	Column Percent
	Adj Std Resid

Total Observations in Table: 209

	AA	Aa	aa	Row Total
+	30 44.776% 3.285	17 27.419% -0.462	15 18.750% -2.720	62
++	22 32.836% -1.016	25 40.323% 0.489	32 40.000% 0.517	79
+++	15 22.388% -2.151	20 32.258% -0.056	33 41.250% 2.118	68
Column Total	67 32.057%	62 29.665%	80 38.278%	209

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 13.10838      d.f. = 4      p = 0.01075821

- Queremos valorar si una mutación en un gen se asocia a la gravedad de una enfermedad.
- El p-valor < 0.05 indica que los porcentajes observados no son compatibles con la hipótesis de homogeneidad de gravedad entre genotipos.
- Los residuales ajustados indican que la probabilidad de gravedad +++ es más alta en el genotipo aa.
- Los residuales ajustados indican que la probabilidad de menor gravedad es mayor en el genotipo AA.
- Por lo tanto, podemos concluir que una mayor gravedad se asocia a la presencia de alelos a.

Cell Contents

Count
Column Percent
Adj Std Resid

Total Observations in Table: 161

	AA	Aa	aa	Row Total
+	7 15.909% 1.992	5 10.000% 0.394	2 2.985% -2.171	14
++	22 50.000% 0.145	25 50.000% 0.159	32 47.761% -0.280	79
+++	15 34.091% -1.283	20 40.000% -0.386	33 49.254% 1.522	68
Column Total	44 27.329%	50 31.056%	67 41.615%	161

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 6.814906      d.f. = 4      p = 0.1459989

Minimum expected frequency: 3.826087

Cells with Expected Frequency < 5: 2 of 9 (22.22222%)

Mensajes de aviso perdidos

In chisq.test(t, correct = FALSE, ...) :

Chi-squared approximation may be incorrect

- El resultado del análisis no es válido cuando se dispone de pocos datos.
- En particular, si las frecuencias esperadas son inferiores a 5.
- Cuando esto sucede, debemos agrupar categorías (en este caso agruparemos + y ++ en Baja gravedad, y +++ en Alta gravedad).

Cell Contents

Count
Column Percent
Adj Std Resid

Total Observations in Table: 161

	AA	Aa	aa	Row Total
Baja	29 65.909% 1.283	30 60.000% 0.386	34 50.746% -1.522	93
Alta	15 34.091% -1.283	20 40.000% -0.386	33 49.254% 1.522	68
Column Total	44 27.329%	50 31.056%	67 41.615%	161

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 2.651413      d.f. = 2      p = 0.2656152

Minimum expected frequency: 18.58385

- Ahora el análisis es válido.
- Concluimos que no hay diferencia de gravedad entre genotipos.