

# PURIFICACIÓN GALINDO VILLARDÓN

**Catedrática de Estadística e Investigación Operativa  
Vicerrectora de Postgrado y Planes Especiales en Ciencias de la Salud  
Universidad de Salamanca**

[pgalindo@usal.es](mailto:pgalindo@usal.es)

# Regresión Logística

# En una tabla tetracórica

## Enfermedad

 Factor  
Riesgo

	ENFERMOS	NO ENFERMOS	
PRESENTE	a	b	
AUSENTE	c	d	
			n

		Enfermedad		
		ENFERMOS	NO ENFERMOS	
Factor Riesgo	PRESENTE	a	b	
	AUSENTE	c	d	
				n

## Riesgo Relativo

$$RR = \frac{P(E / +)}{P(E / -)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

*La enfermedad es RR más frecuente entre los casos expuestos al factor de riesgo que en aquellos que no lo están.*

# Riesgo Relativo

El **RR no puede ser utilizado en estudios retrospectivos** ya que no se conocen las probabilidades condicionadas de presentar la enfermedad.

Fijamos de entrada los casos totales con enfermedad y los casos totales sin enfermedad y no es posible conocer la proporción real en los 2 subgrupos estudiados.

# ODDs Ratio :PROSPECTIVOS

$$OR = \frac{\text{razón enfermos en los expuestos}}{\text{razón enfermos en los no expuestos}}$$

		Enfermedad		
		ENFERMOS	NO ENFERMOS	
Factor Riesgo	PRESENTE	a	b	n
	AUSENTE	c	d	

$$OR = \frac{P(E / +)}{P(E / -)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a \cdot d}{b \cdot c}$$

*Se fija el número de individuos expuestos al riesgo y el número de no expuestos.*

Observamos cómo se desarrolla la enfermedad.

# ODDs Ratio :RETROSPECTIVOS

		Enfermedad		
		ENFERMOS	NO ENFERMOS	
Factor Riesgo	PRESENTE	a	b	n
	AUSENTE	c	d	

$$OR = \frac{\text{razón de expuestos al riesgo en enfermos}}{\text{razón de expuestos al riesgo en sanos}}$$

*Se fija el número de individuos con la enfermedad y el número de individuos sin ella.*

**Observamos en cuantos está presente el síntoma.**

$$OR = \frac{\frac{P(+ / E)}{P(+ / noE)}}{\frac{P(- / E)}{P(- / noE)}} = \frac{\frac{a / a + c}{c / a + c}}{\frac{b / b + d}{d / b + d}} = \frac{a.d}{b.c}$$

## PROSPECTIVOS

$$OR = \frac{\frac{P(E / +)}{P(noE / +)}}{\frac{P(E / -)}{P(noE / -)}} = \frac{\frac{a / a + b}{b / a + b}}{\frac{c / c + d}{d / c + d}} = \frac{a.d}{b.c}$$

## RETROSPECTIVO

$$OR = \frac{\frac{P(+ / E)}{P(- / E)}}{\frac{P(+ / noE)}{P(- / noE)}} = \frac{\frac{a / a + c}{c / a + c}}{\frac{b / b + d}{d / b + d}} = \frac{a.d}{b.c}$$

$$RR = \frac{P(E / +)}{P(E / -)} = \frac{\frac{a}{a + b}}{\frac{c}{c + d}}$$



# Introducción

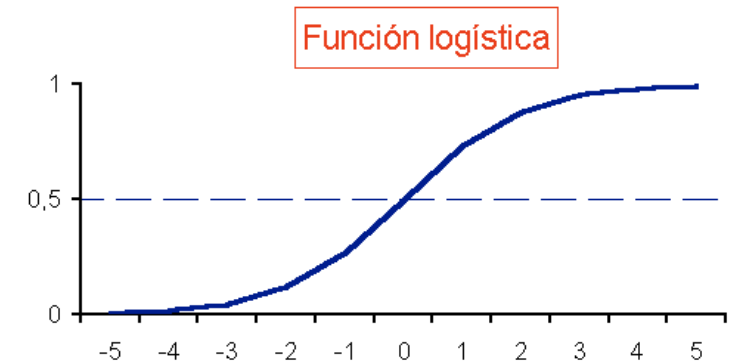
Herramienta para modelizar la relación entre una variable dicotómica de respuesta y una o más variables EXPLICATIVAS.

Es posible incluir regresores de tipo cualitativo, mediante la utilización de variables indicadoras, de manera análoga a como se hace en regresión lineal.

# REGRESIÓN LOGÍSTICA

$$\cancel{Y = \beta_0 + \beta_1 X_1}$$

$$P(Y=1) = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1)}}$$



Con una sencilla transformación (**logit**) puede convertirse en lineal

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1$$

# REGRESIÓN LOGÍSTICA

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$$


$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots)}}$$

  
Función logística

  
Parámetros

Individuos mayores de 40 años

**Y** = Desarrollo enfermedad coronaria (**EC**) en 10 años de seguimiento

$$P(Y=1) = \frac{1}{1 + e^{-(-6.614 + 0.075X_1 + 0.312X_2 + 0.018X_3)}} = 0.68$$

*Se estima que:*

**El 68% de los sujetos con ese perfil desarrollarán un EC en el siguiente decenio**

$X_1$  ,

Edad

**58 años**

$X_2$  ,

Hábito fumar

**Fumador**

$X_3$

Tensión Sistólica

**150**

# Odds Ratio

$$OR(E) = \frac{\frac{P_F(E)}{1 - P_F(E)}}{\frac{P_{\bar{F}}(E)}{1 - P_{\bar{F}}(E)}}$$

$$OR(E) = \frac{\frac{0.68}{1 - 0.68}}{\frac{0.61}{1 - 0.61}} = \frac{2.125}{1.564} = 1.36$$

$$P(E)_{\text{Fumador}} = 0.68$$

$$P(E)_{\text{No Fumador}} = 0.61$$

$$OR(E) = e^{\beta_2} = 1.36$$

## Interpretación de los parámetros

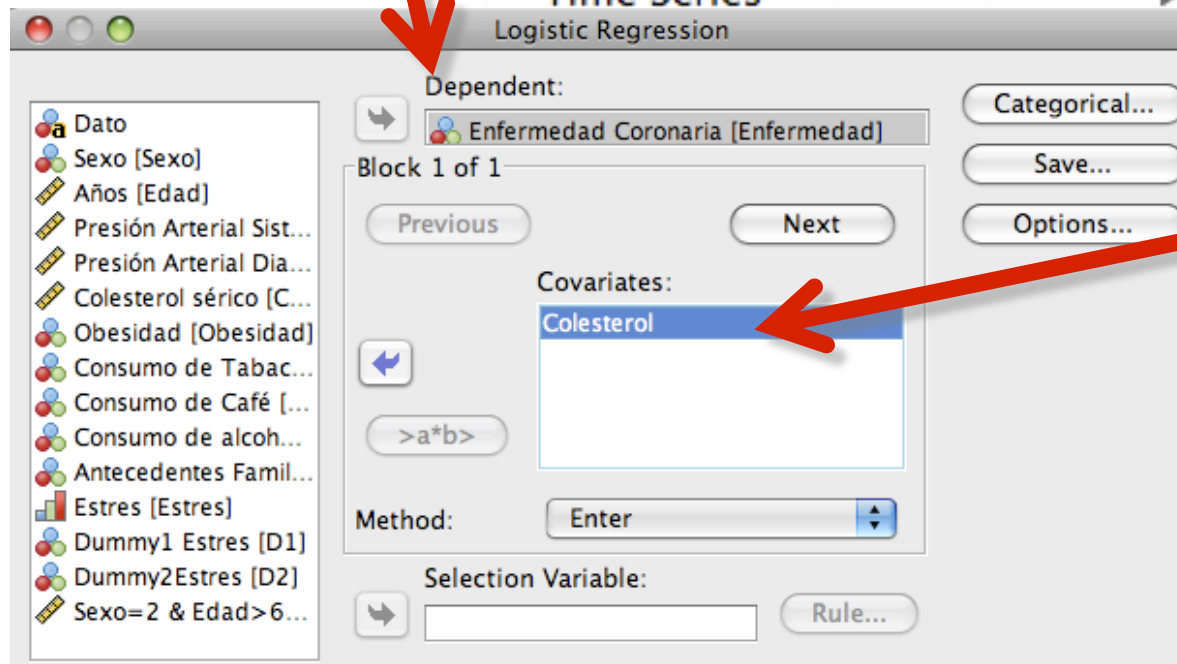
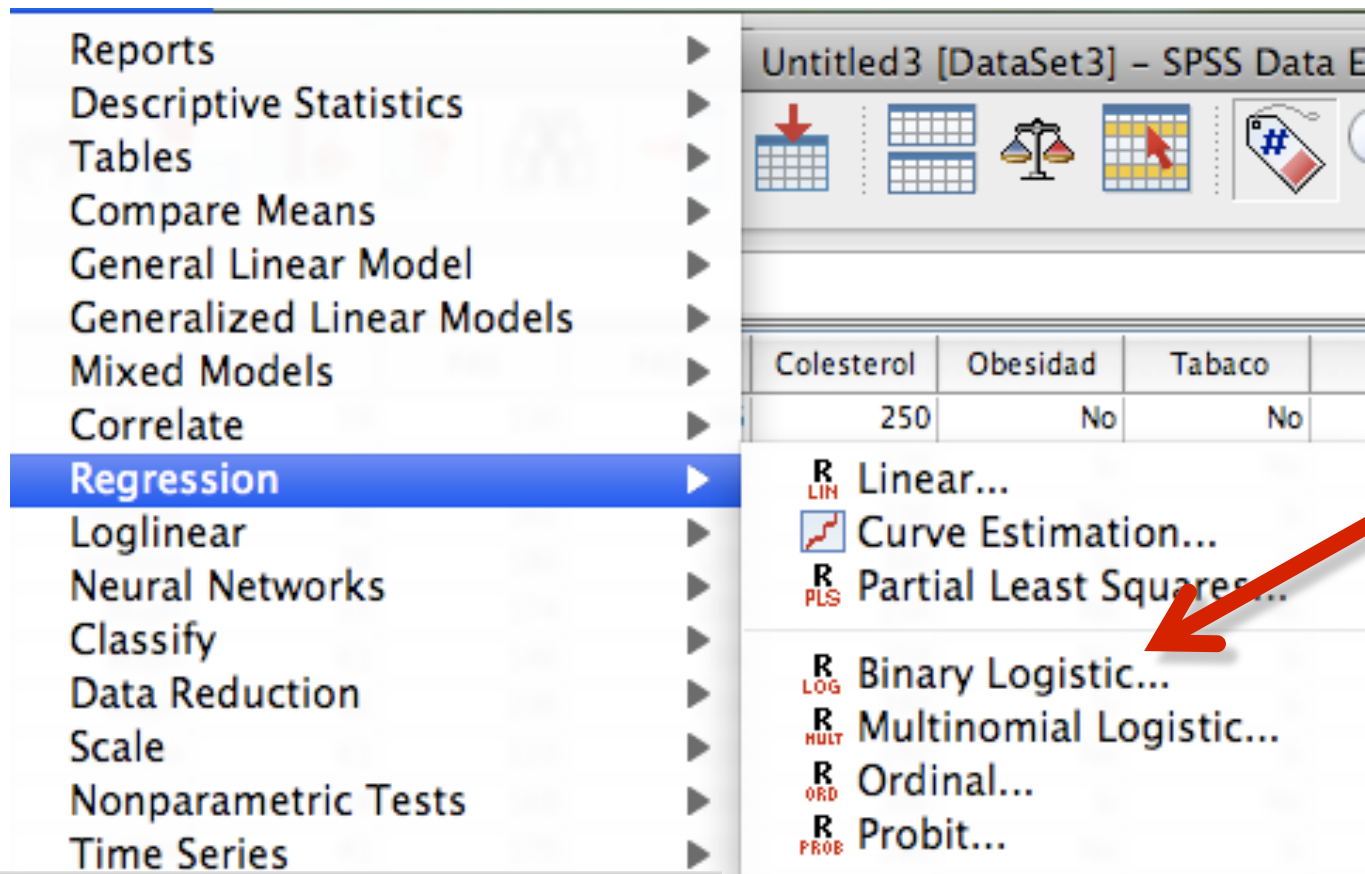
$$OR(E) = \frac{\frac{P_F(E)}{1 - P_F(E)}}{\frac{P_{\bar{F}}(E)}{1 - P_{\bar{F}}(E)}} = e^{\beta}$$

$$e^{\beta} = \frac{\frac{P(E=1/x+1)}{P(E=0/x+1)}}{\frac{P(E=1/x)}{P(E=0/x)}} = \text{Odds Ratio}$$

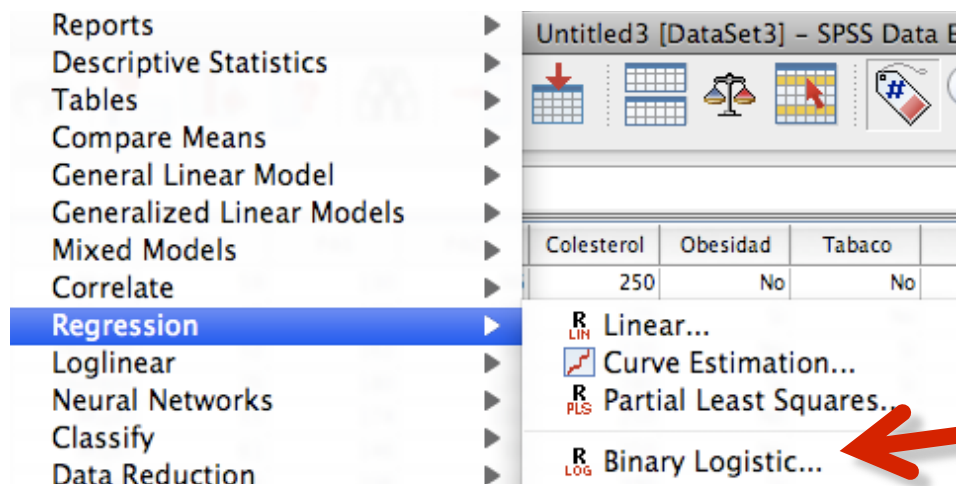
### Para variables continuas:

mide el odds-ratio entre un individuo con un valor  $x+1$  y un individuo con valor  $x$  en la variable independiente, para cualquier valor  $x$

# Menú SPSS



# Menú SPSS



Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	110,104	,186	,255

Bondad del ajuste

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

## Tabla de clasificación

Classification Table

Observed			Predicted		
			Enfermedad Coronaria		
			No	Si	Percentage Correct
Step 1	Enfermedad Coronaria	No	63	1	98,4
		Si	20	16	44,4
	Overall Percentage				79,0

a. The cut value is ,500

MODELO

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Colesterol	,020	,005	15,567	1	,000	1,020
	Constant	-6,406	1,498	18,280	1	,000	,002

OR



$$\text{Prob}[Enferm = SI / X = x_i] = p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

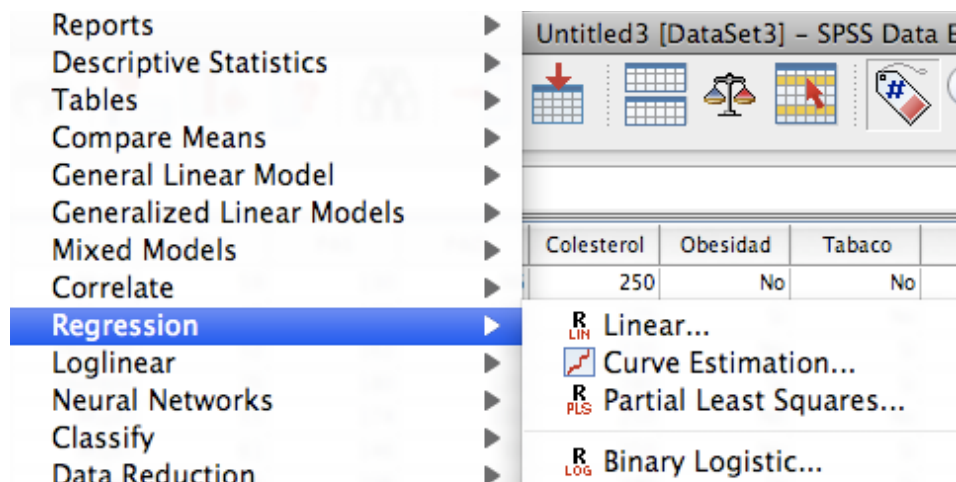
$e^{\beta_1}$

		Estimadores	Errores estándar			Significaciones	Odds-Ratio
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Colesterol	,020	,005	15,567	1	,000	1,020
	Constant	-6,406	1,498	18,280	1	,000	,002

$\beta_1$  (points to Coolesterol row, B column)  
 $\beta_0$  (points to Constant row, B column)

Por cada incremento de una unidad en el **colesterol** el logit se incrementa en 0.02, por tanto el odds (ventaja) de “SI” estar enfermo frente a “NO” estar enfermo se multiplica por 1.020 al incrementarse una unidad el colesterol.

# Menú SPSS



**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	110,104	,186	,255

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

**Classification Table**

de clasificación

Observed			Predicted		
			Enfermedad Coronaria		
			No	Si	Percentage Correct
Step 1	Enfermedad Coronaria	No	63	1	98,4
		Si	20	16	44,4
		Overall Percentage			79,0

El 98,4% de los que no tienen enfermedad coronaria se han clasificado correctamente utilizando el modelo de regresión logística, mientras que de los que "SI" tiene la enfermedad se clasifican correctamente el 44.4%. Conociendo los valores del Colesterol se predice correctamente la presencia/ausencia de enfermedad coronaria en el 79% de los sujetos.

Obsérvese que el modelo es mucho más adecuado para descartar que para confirmar.







TESIS DOCTORAL

VNiVERSiDAD  
D SALAMANCA

# Epidemiología de las sibilancias en el primer año de vida. Estudio multicéntrico

Javier Pellegrini Belinchón

## Objetivos.

- Detectar **factores de riesgo sobre los que fuera posible actuar** para variar la evolución de las sibilancias recurrentes en los primeros meses de vida.

# Probabilidad de presentar Sibilancias Recurrentes.

## Modelo Pellegrini.

$$p(x) = \frac{1}{1 + e^{(-2.312 + 0.9086 \text{ Guardería} + 1.251 \text{ Eczema} + 0.806 \text{ Asma madre} + 0.715 \text{ Fuma madre 3er Trim} - 0.831 \text{ Com Mediter})}}$$

	B	E.T.	Wald	gl	Sig.	Exp(B)	Inf.	I.C. 95% para EXP(B)	
								Inferior	Superior
<b>Guardería</b>	,908	,197	21,287	1	,000	2,480		1,686	3,647
<b>Eczema</b>	1,251	,224	31,105	1	,000	3,494		2,251	5,424
<b>Asma madre</b>	,806	,338	5,678	1	,017	2,293		1,154	4,344
<b>Fuma madre 3er Trim.</b>	,715	,281	6,479	1	,011	2,044		1,179	3,545
<b>Com. Mediterranea</b>	-,831	,196	17,927	1	,000	,436		,297	,640
<b>Constante</b>	-2,312	,170	185,604	1	,000	,099			

**Factores incluidos en el modelo matemático**



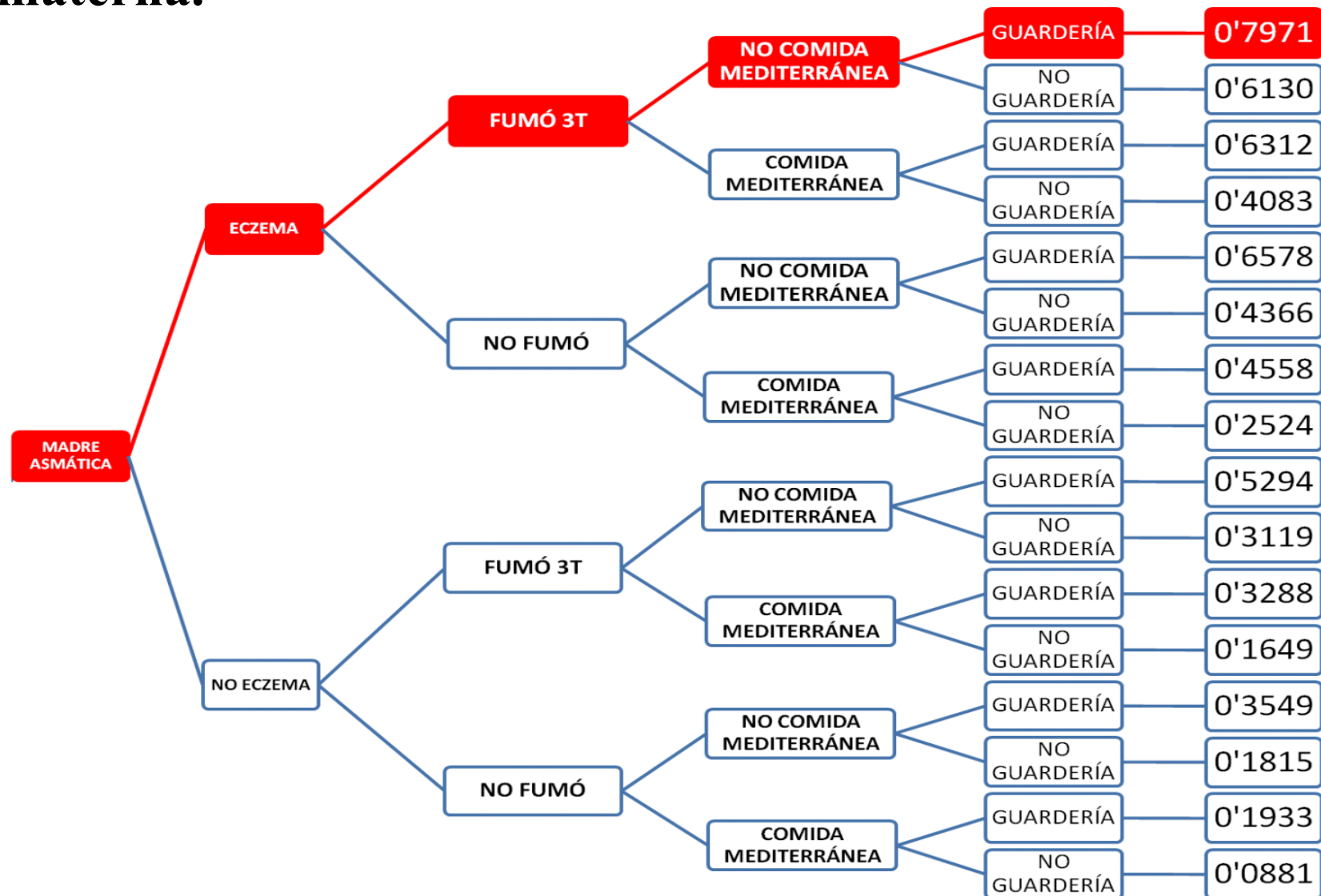
## Modelo Pellegrini para estimar la Probabilidad de Sibilancias Recurrentes.



$$p(x) = \frac{1}{1 + e^{(-2.312 + 0.9086 \text{ Guarderia} + 1.251 \text{ Ezcema} + 0.806 \text{ Asma madre} + 0.715 \text{ Fuma mdre3er Trim} - 0.831 \text{ Com Mediter})}}$$

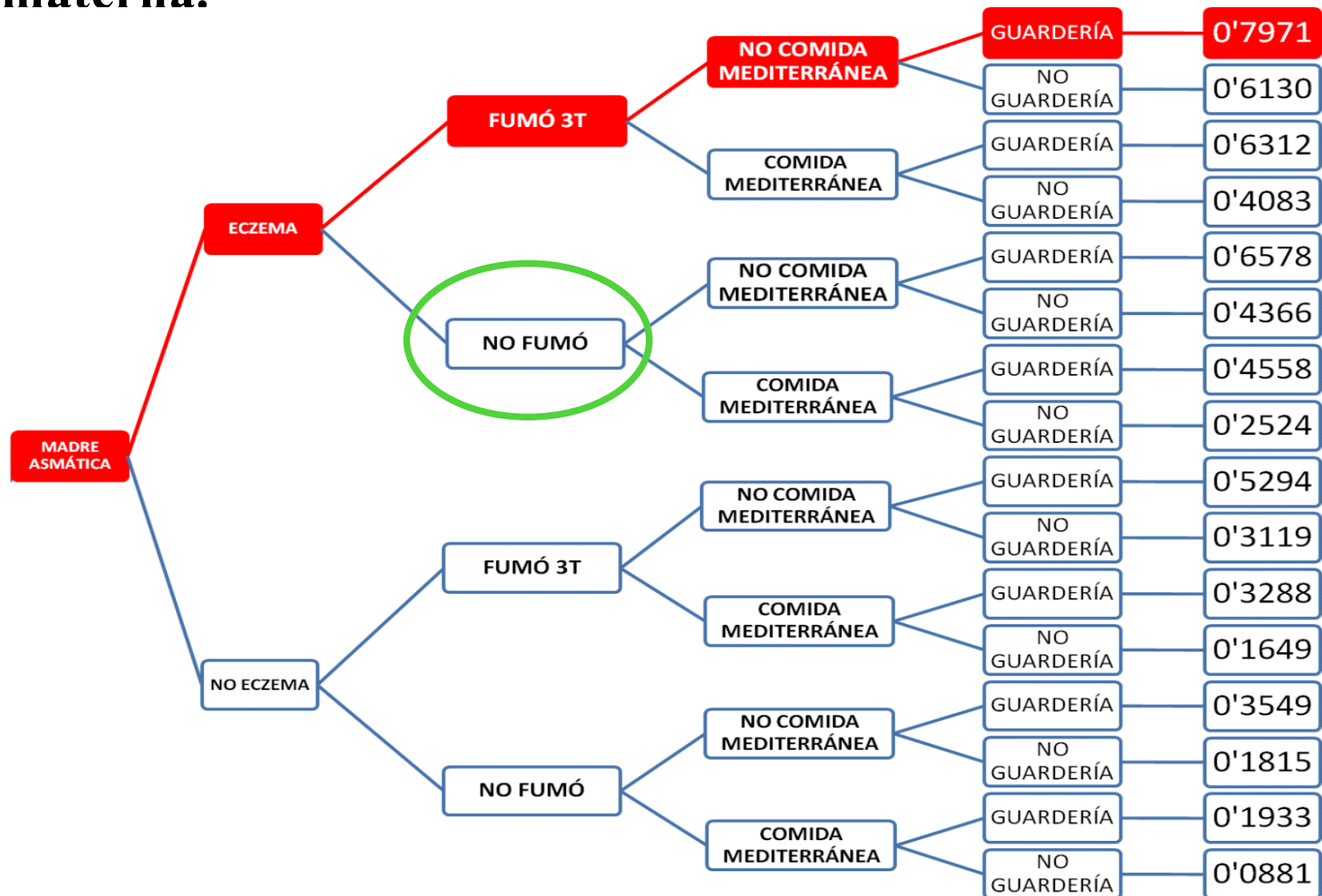


# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.

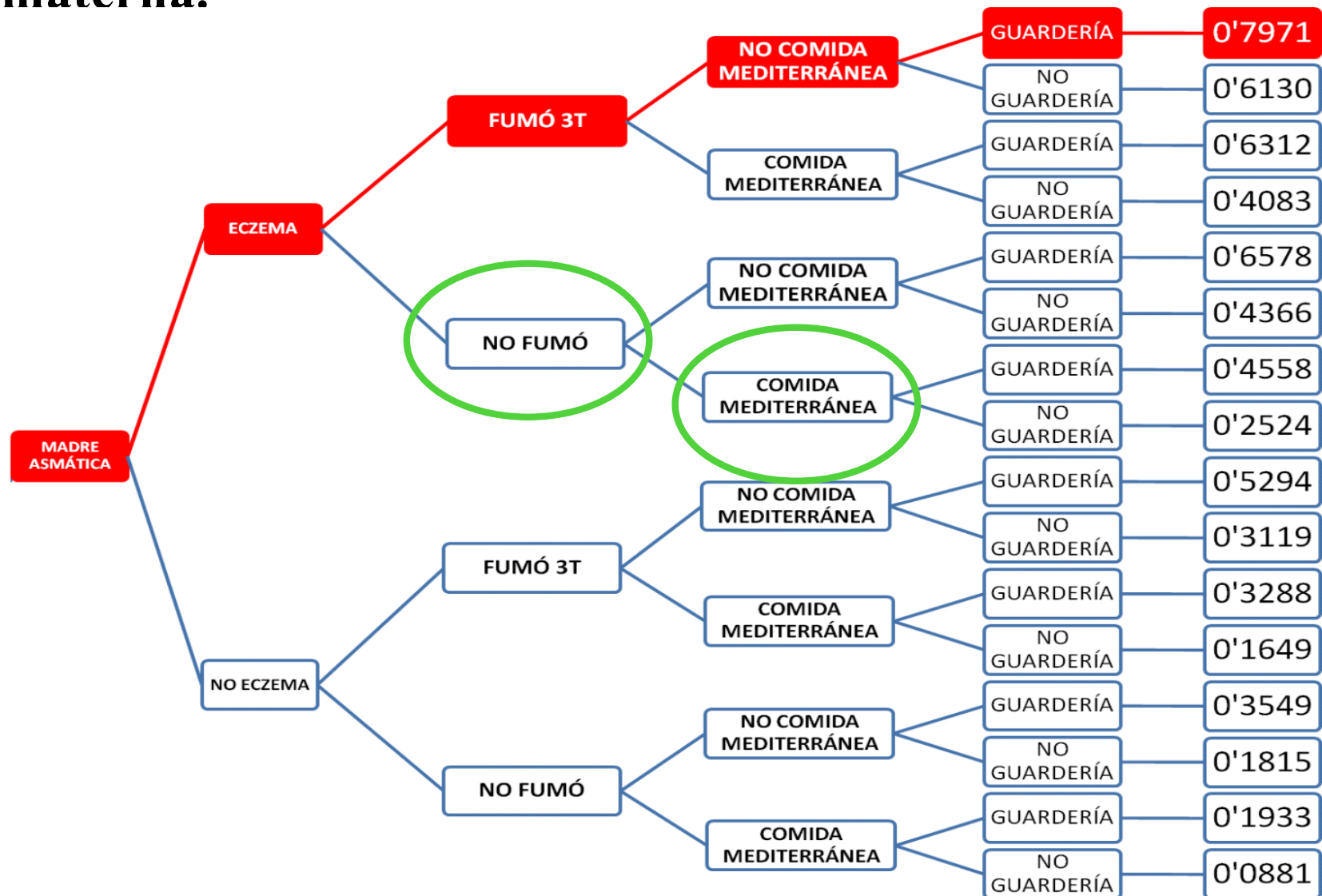




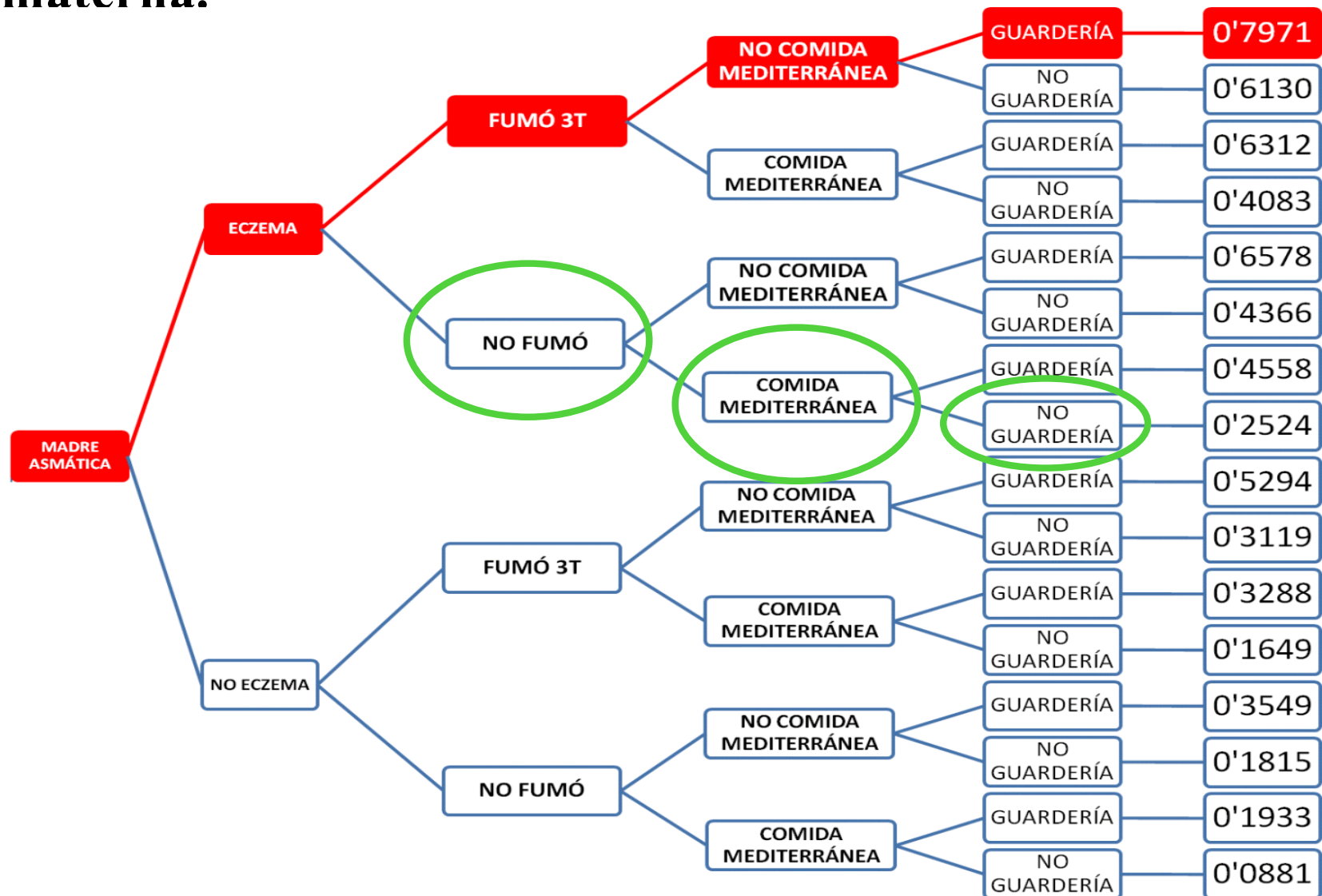
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



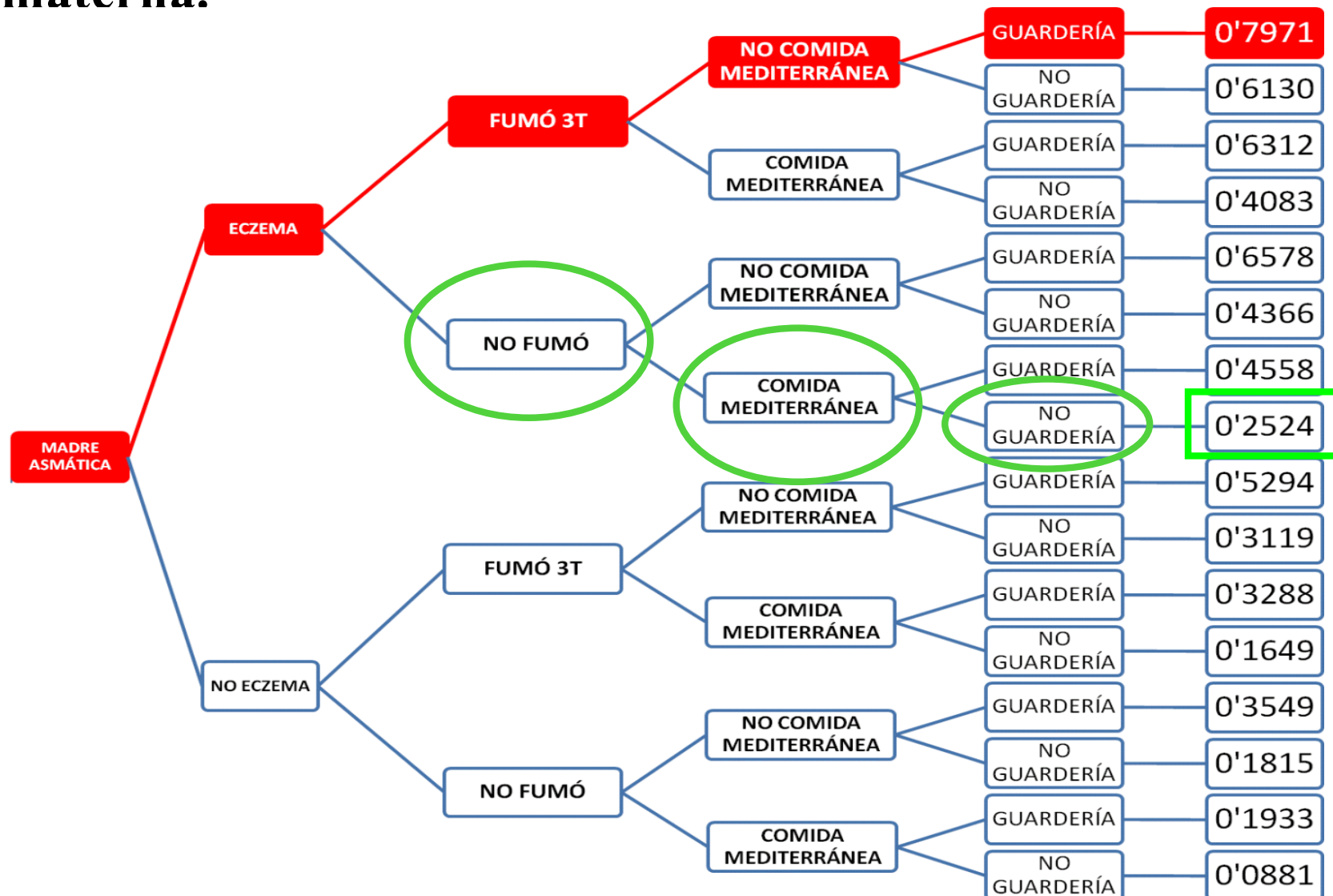
## Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



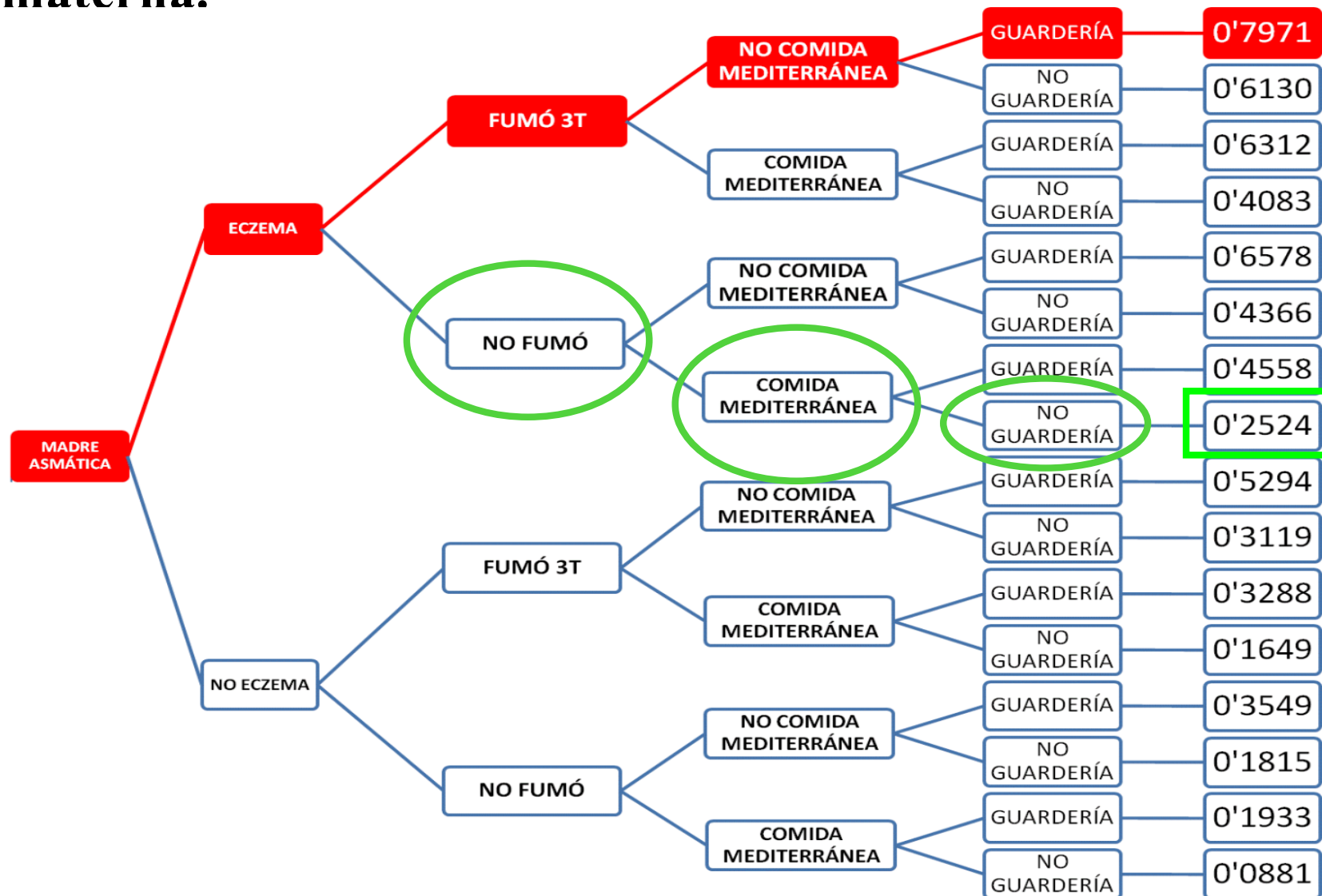
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



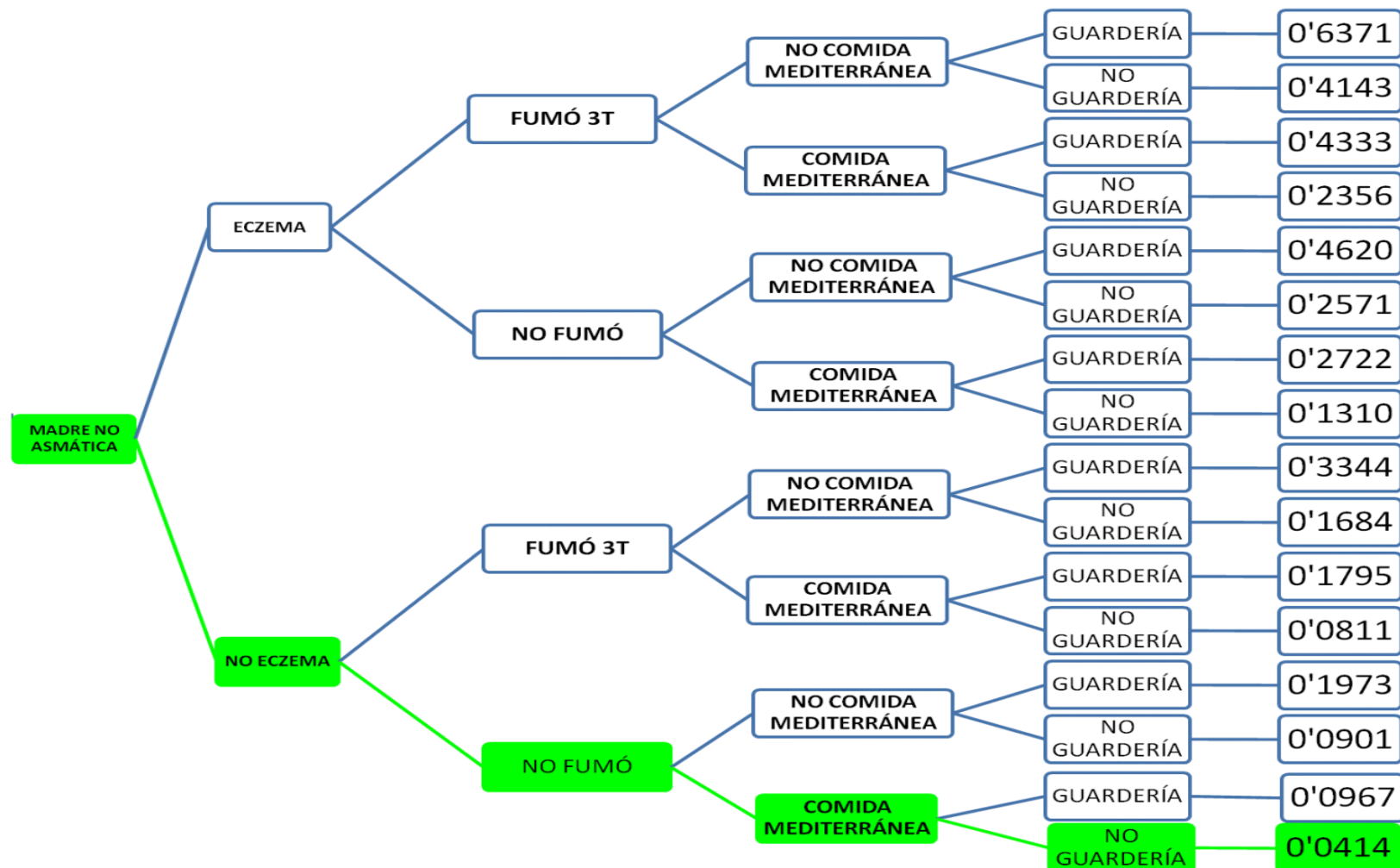
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.

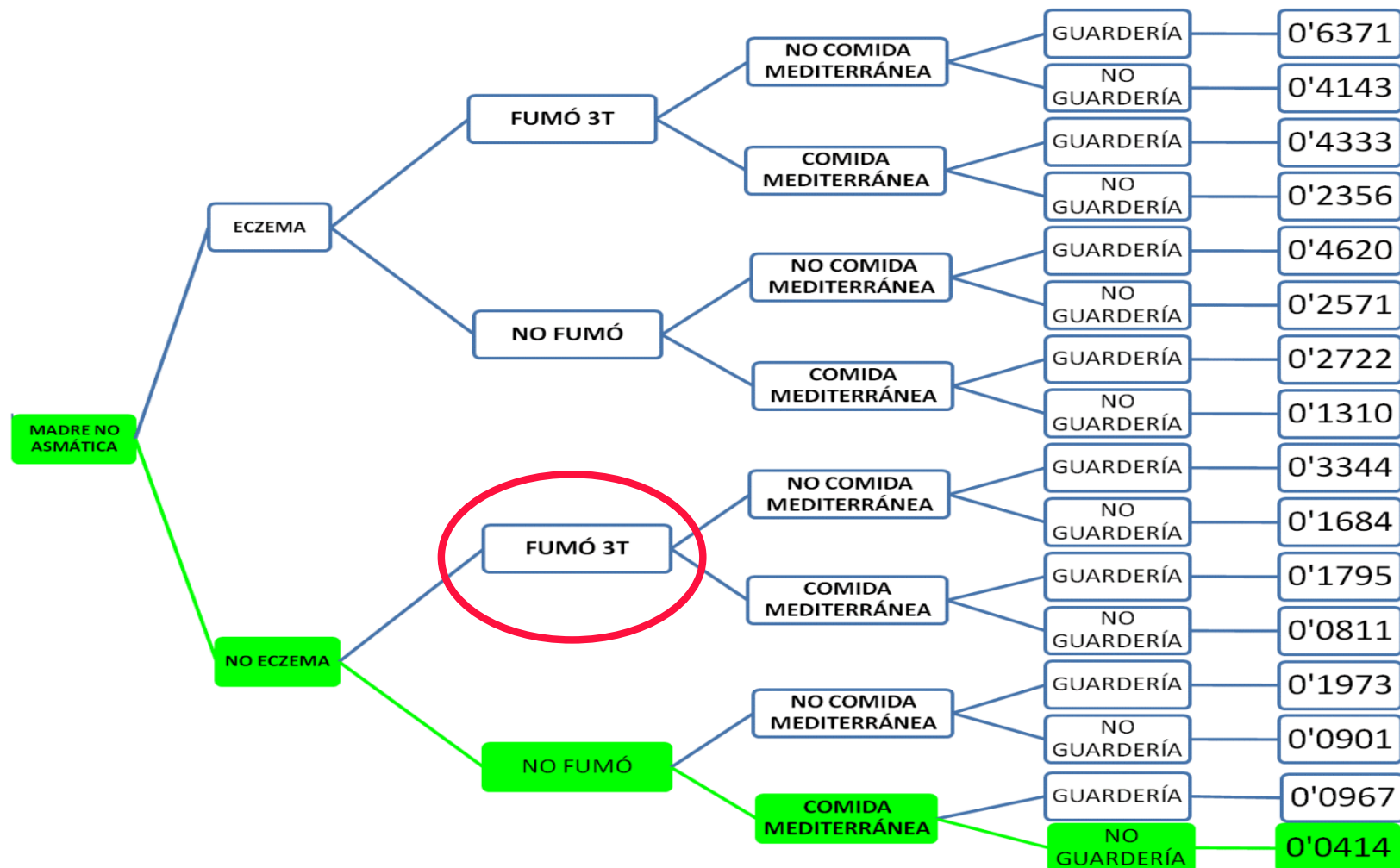


# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.

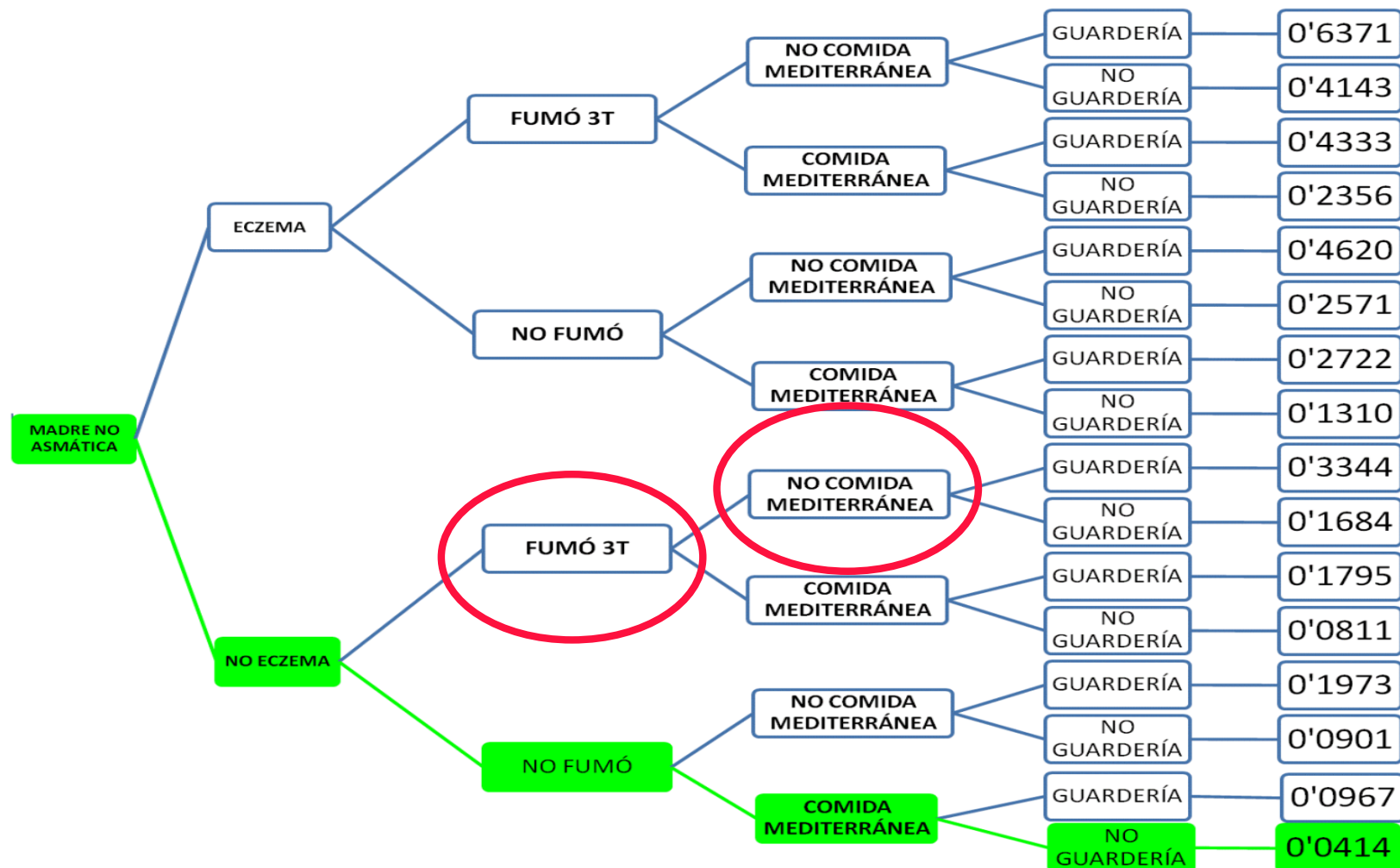




# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.

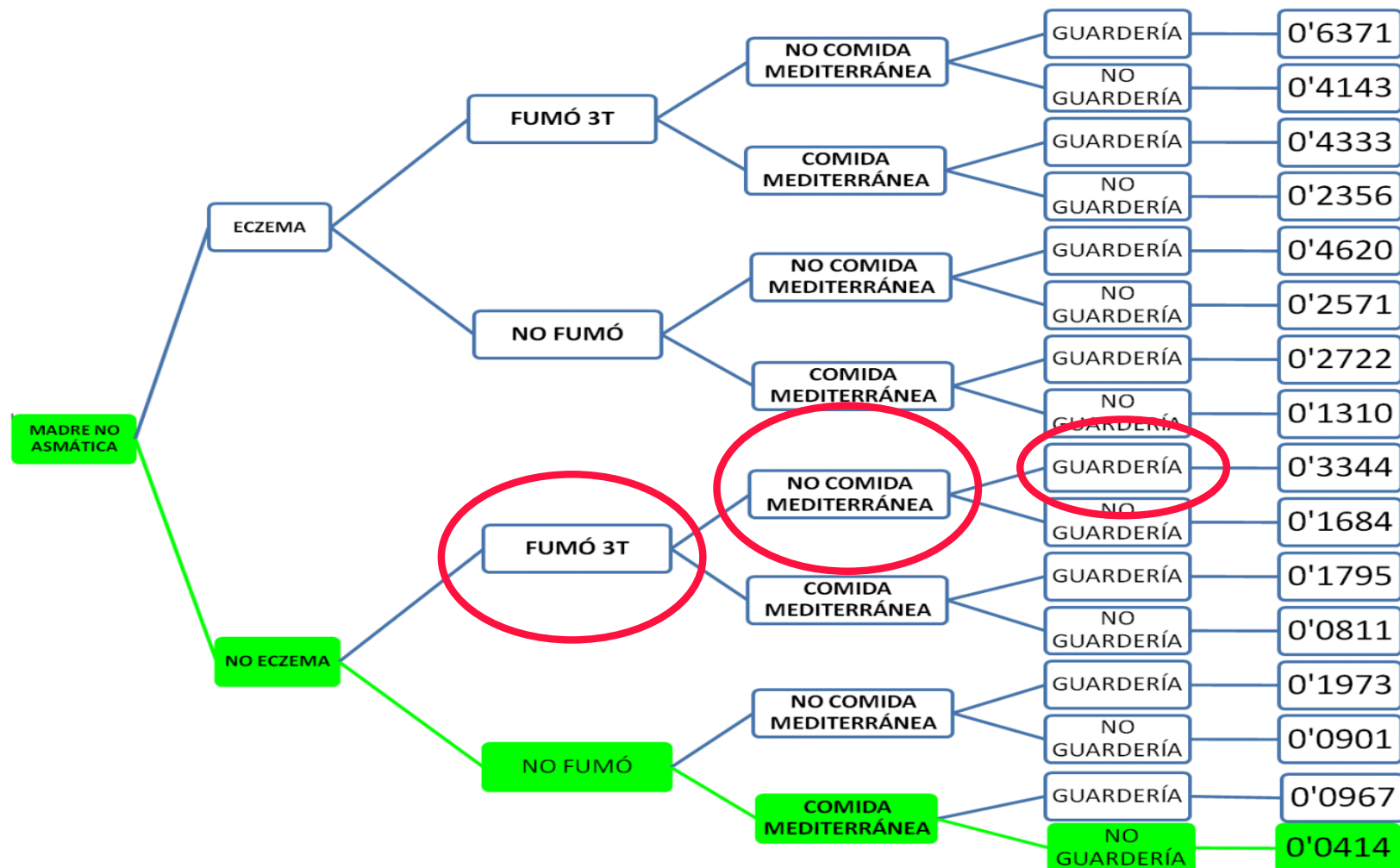


# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.

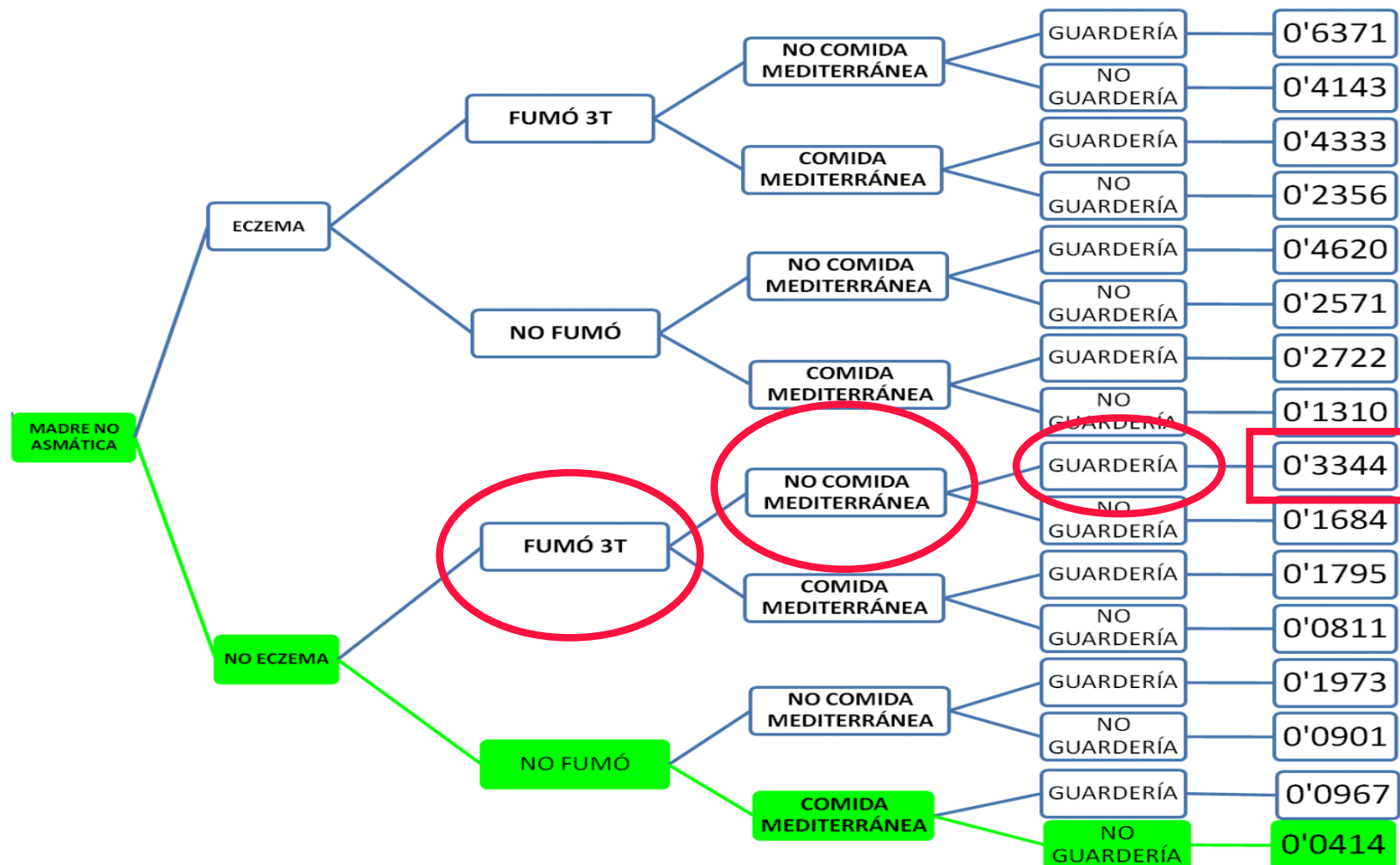




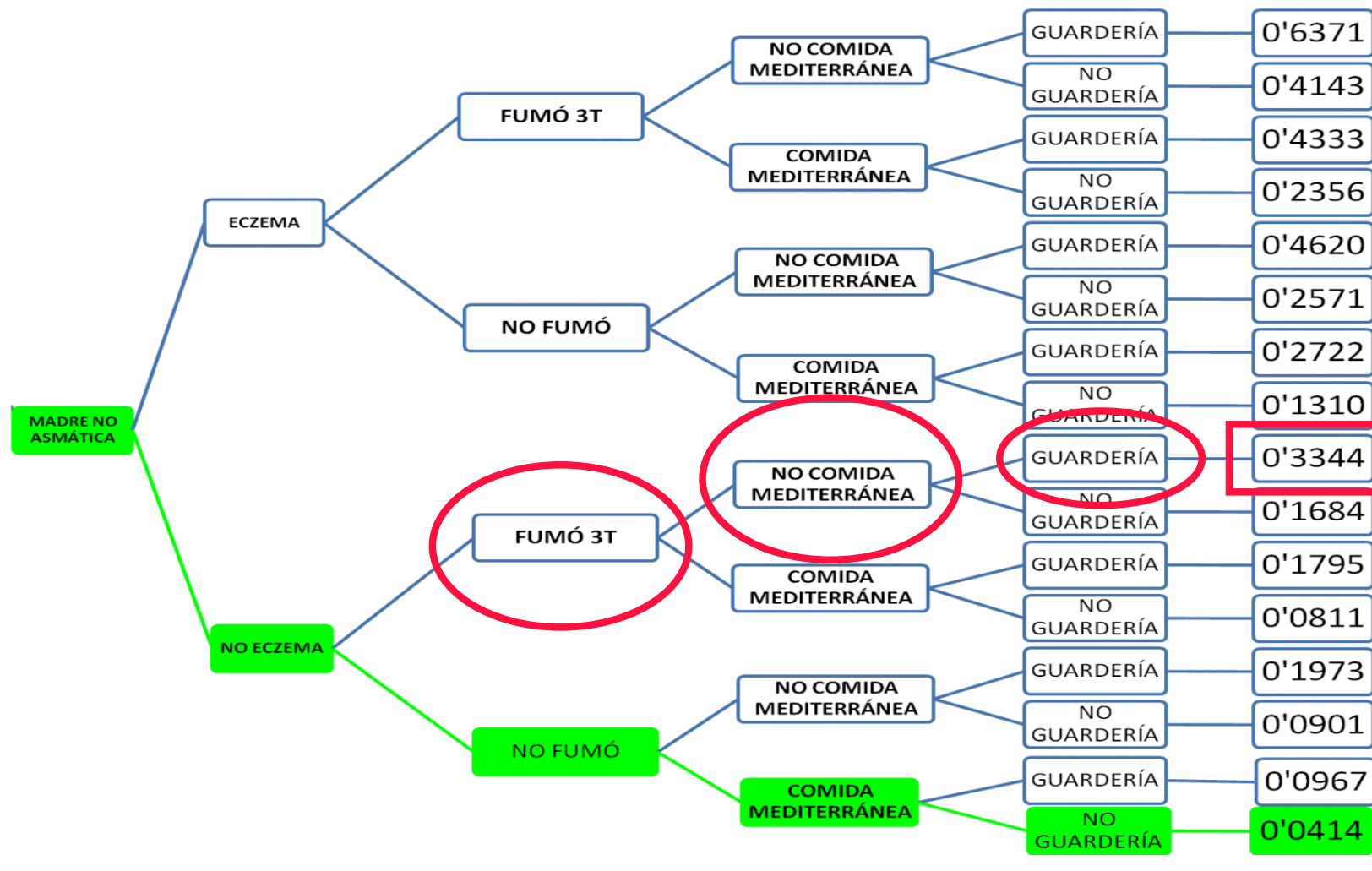
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



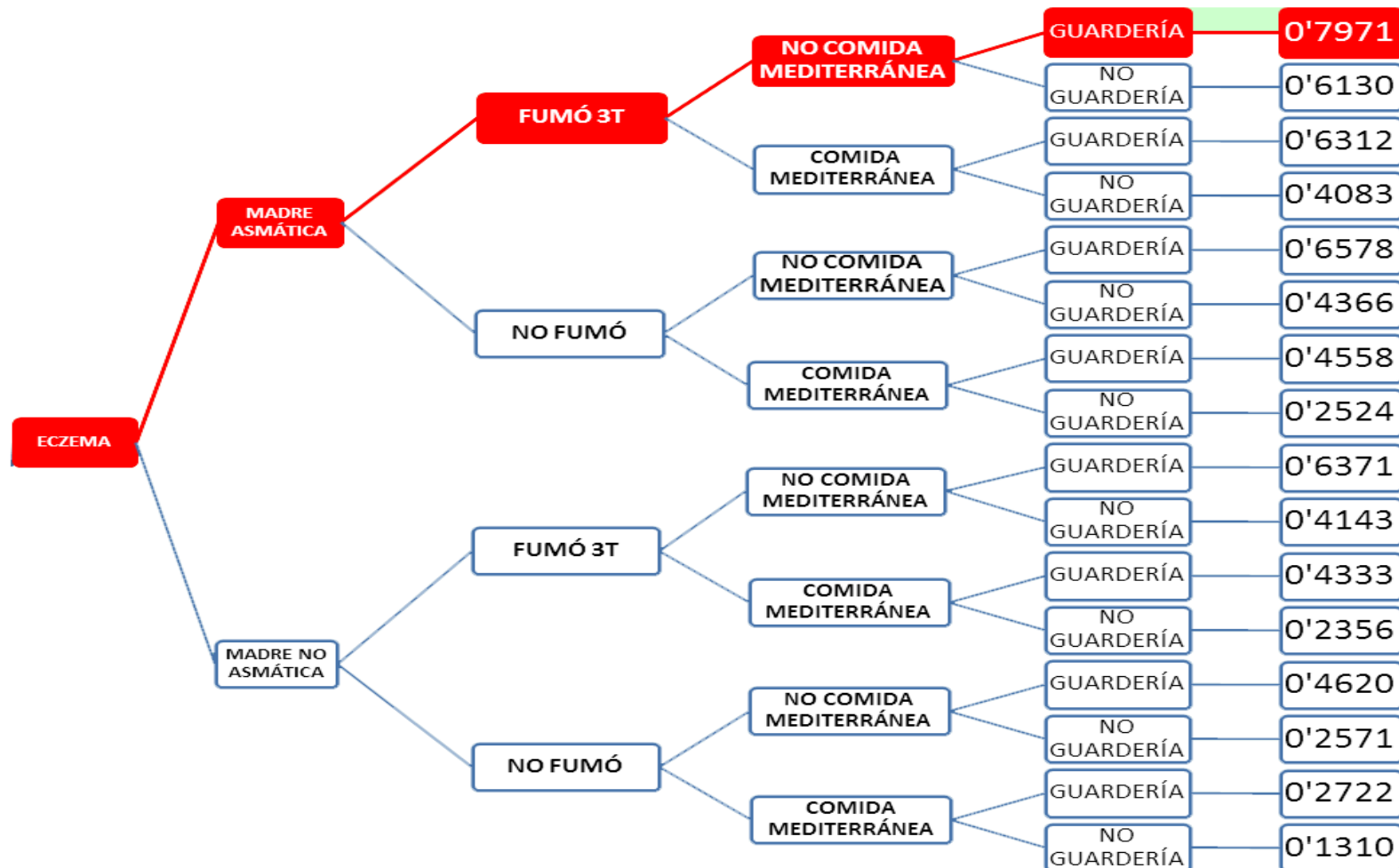
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



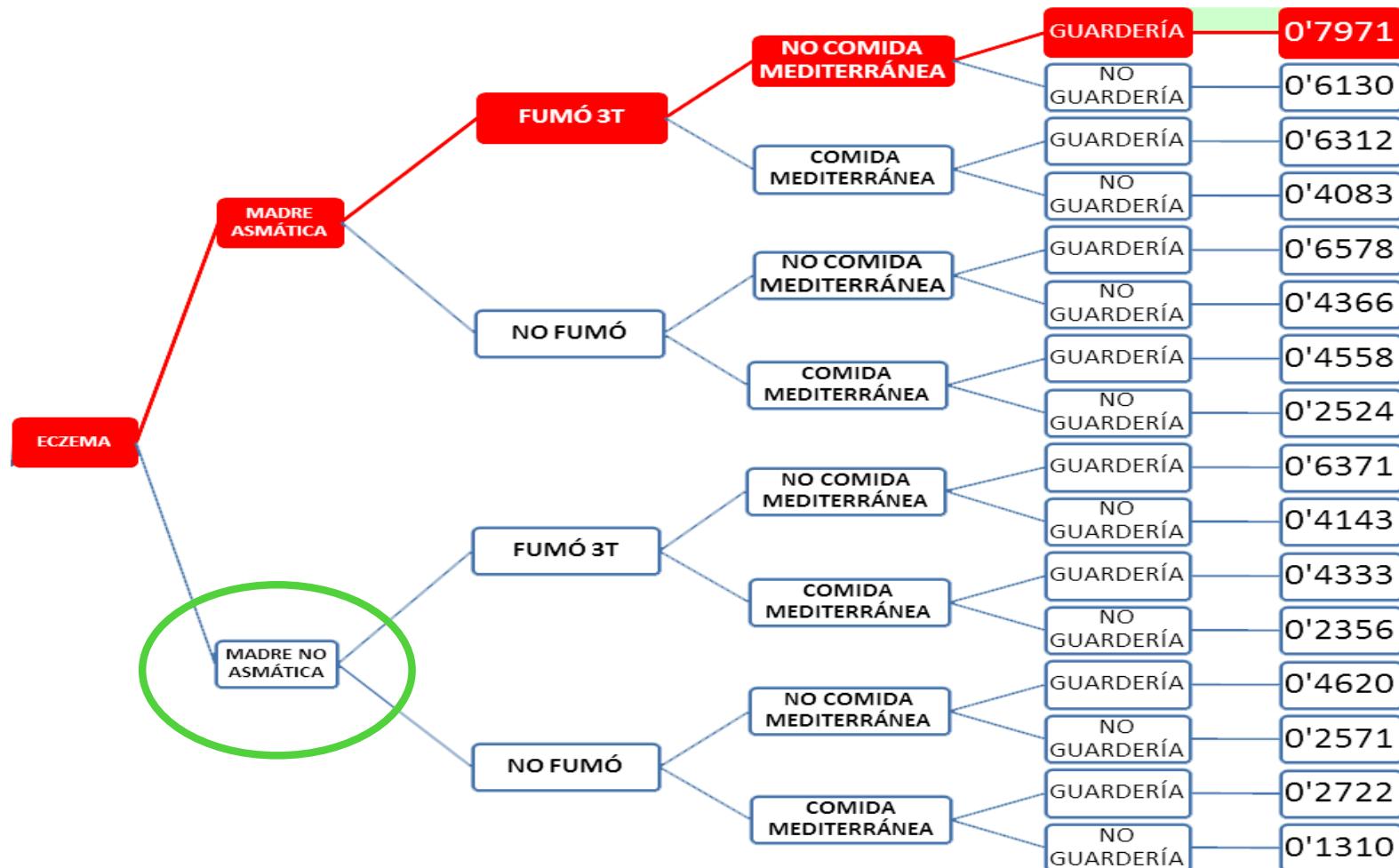
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Asma materna.



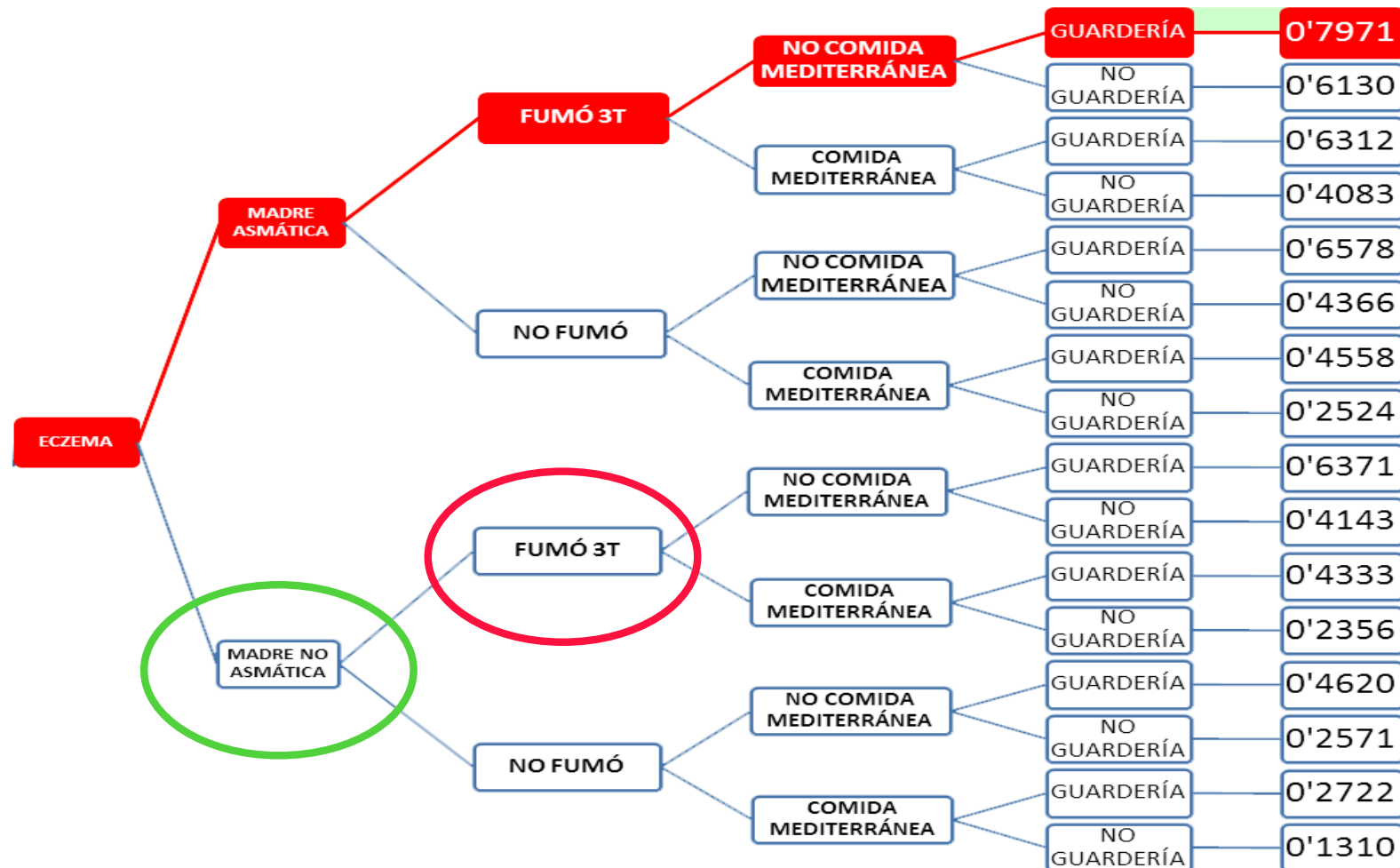
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema



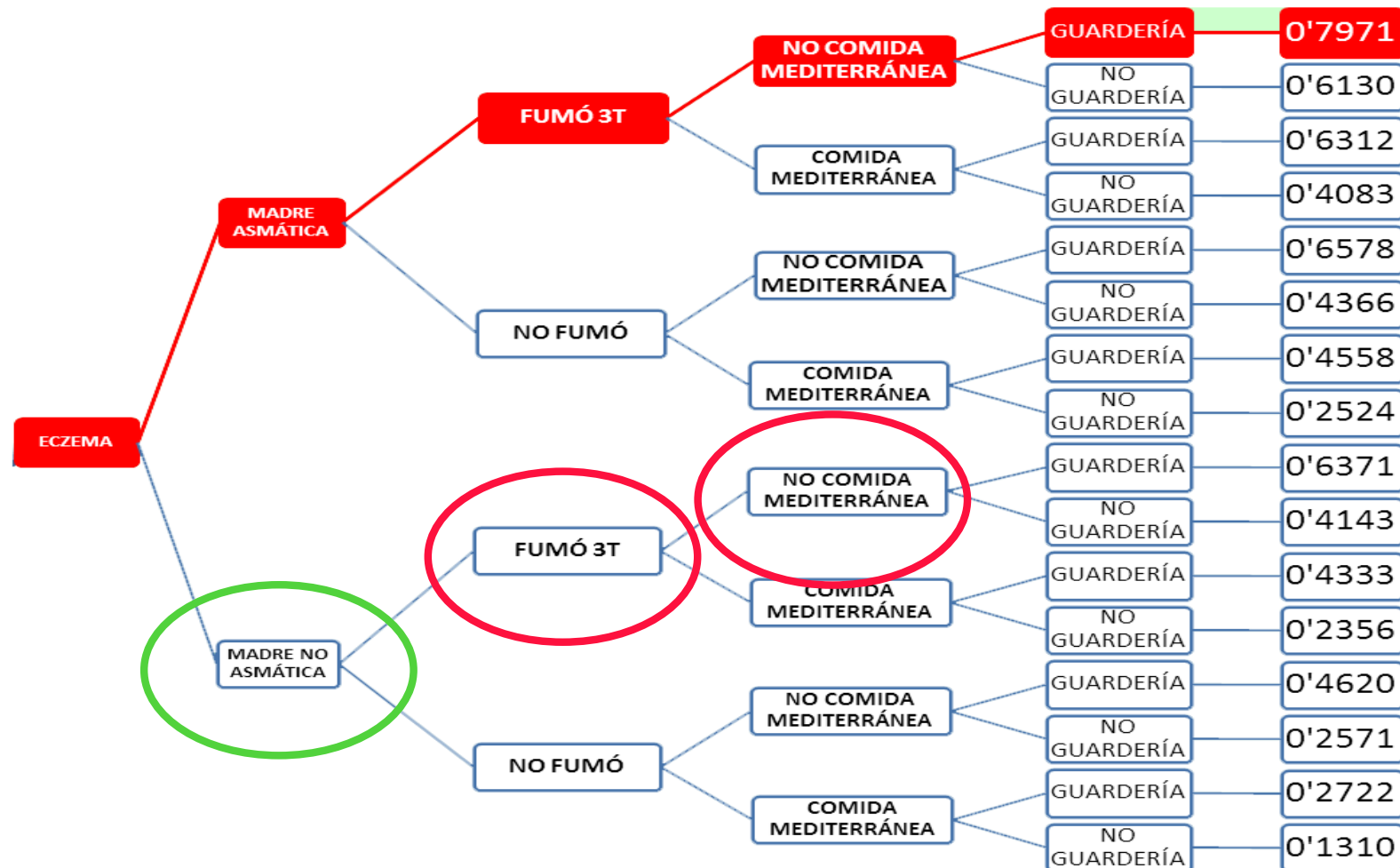
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema



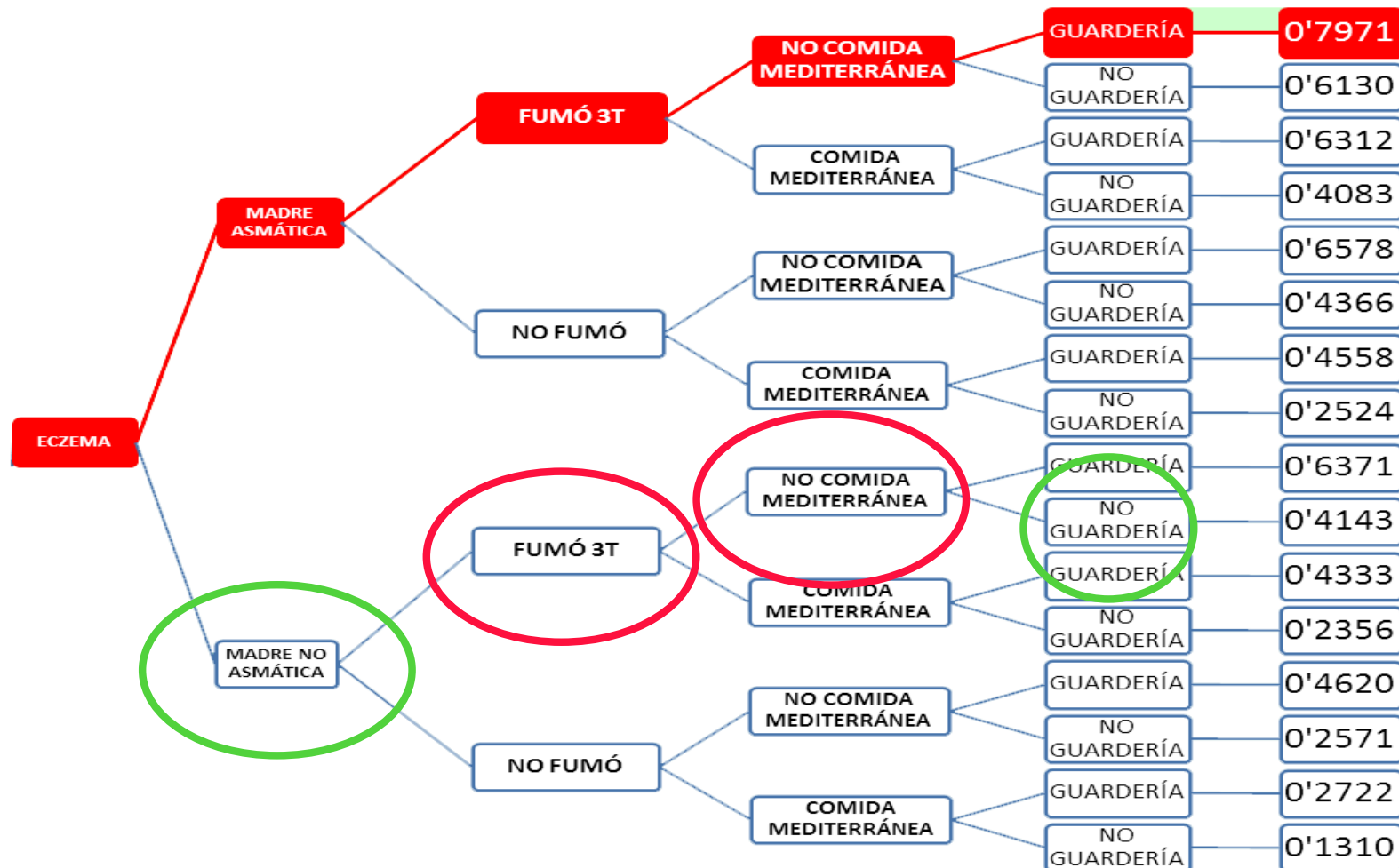
# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema



# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema

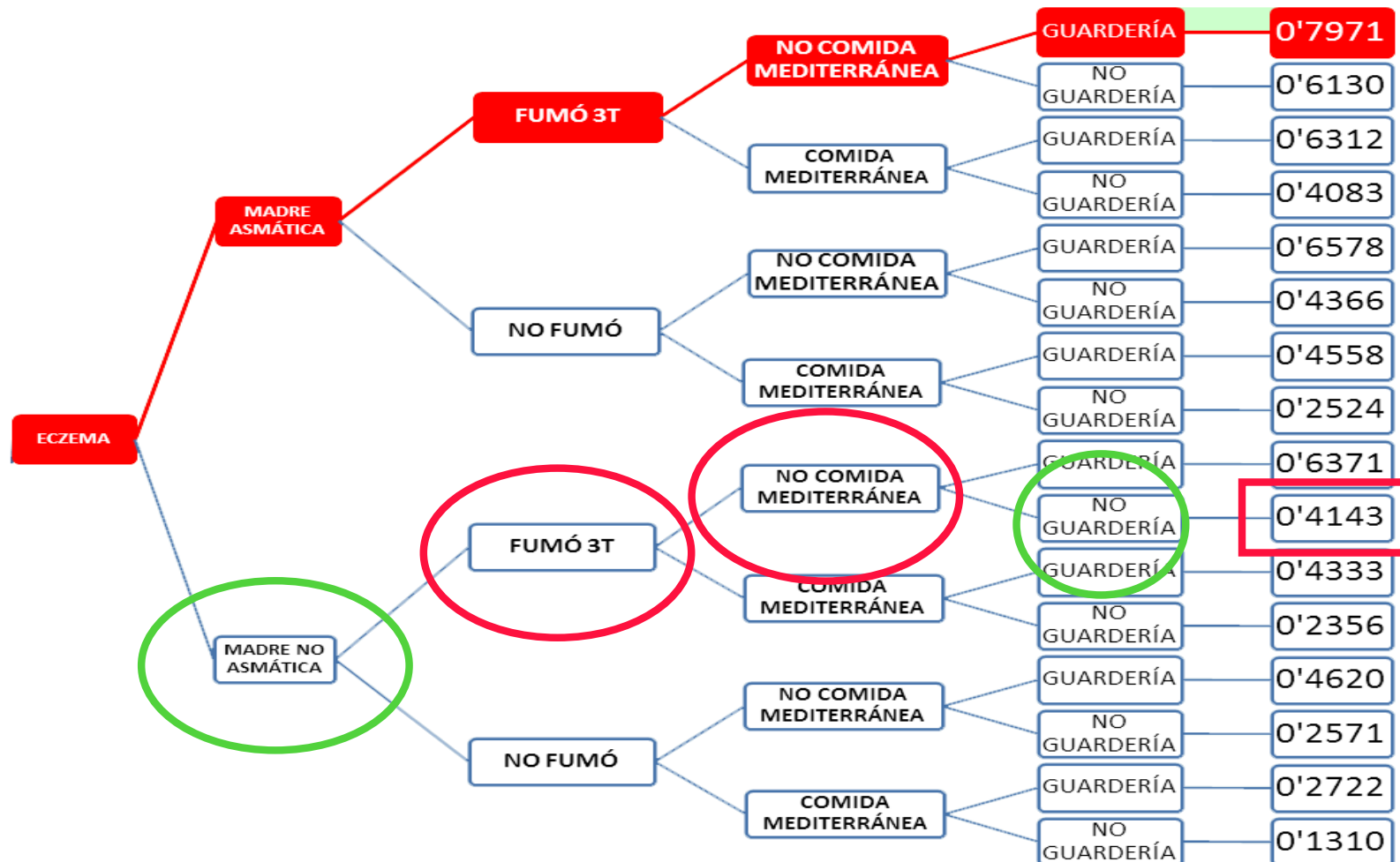


# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema

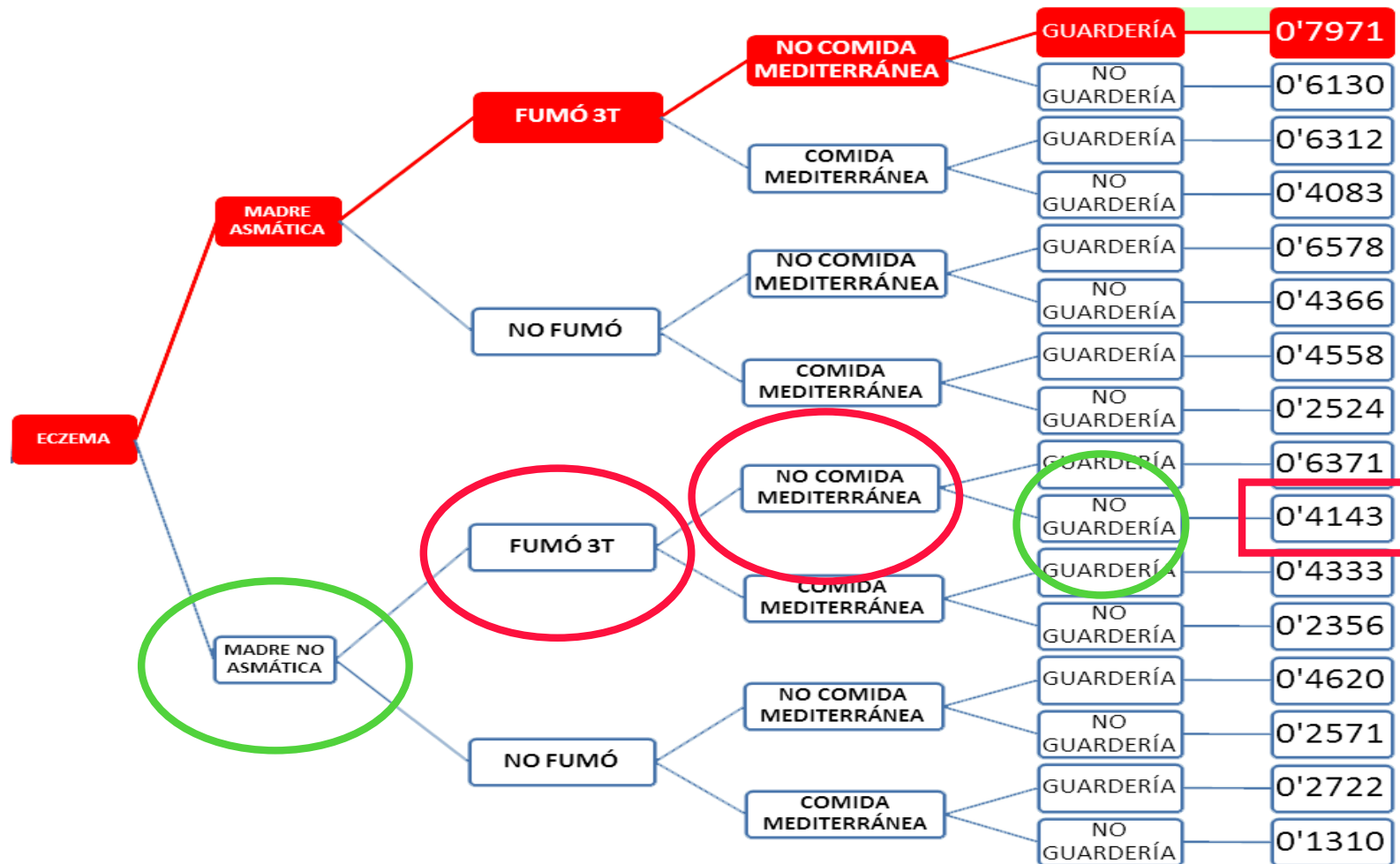




# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema



# Modelo para estimar la Probabilidad de Sibilancias Recurrentes. Niño con Eczema



# BIBLIOGRAFIA

**Luis Carlos Silva Aycaguer**

**Excursión a la regresión logística en ciencias de la salud.**

**Ed. Díaz de Santos Madrid 1995**

**David W. Hosmer y Stanley Lemeshow**

**Applied Logistic Regression**

**Ed. John Wiley New York 1989**