

**DETERMINACIÓN DE UN MODELO PARA EL CÁLCULO DEL RIESGO DE  
MORTALIDAD PERINATAL Y DE COMPLICACIONES NEONATALES  
TEMPRANAS NO FATALES ENTRE LAS GESTANTES AFILIADAS A LA  
EPS - S EMSSANAR EN LA CIUDAD DE PASTO**

**CARLOS ALBERTO HIDALGO PATIÑO**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
SAN JUAN DE PASTO  
2008**

**DETERMINACIÓN DE UN MODELO PARA EL CÁLCULO DEL RIESGO DE  
MORTALIDAD PERINATAL Y DE COMPLICACIONES NEONATALES  
TEMPRANAS NO FATALES ENTRE LAS GESTANTES AFILIADAS A LA  
EPS - S EMSSANAR EN LA CIUDAD DE PASTO**

**CARLOS ALBERTO HIDALGO PATIÑO**

**Trabajo de GRADO**

**DIRECTOR  
JORGE HUMBERTO MAYORGA**

**UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
SAN JUAN DE PASTO  
2008**

## TABLA DE CONTENIDO

	Pág.
RESUMEN.....	7
INTRODUCCIÓN.....	8
1. OBJETIVO GENERAL.....	9
2. MARCO TEÓRICO.....	10
2.1. DEFINICION DE TERMINOS.....	10
2.2. COMENTARIOS SOBRE EL RIESGO EN SALUD.....	11
2.3. DETERMINACION DE RIESGOS EN SALUD PERINATAL.....	12
2.4. METODO ESTADÍSTICO.....	12
2.4.1. La Regresión Logística.....	13
2.4.1.1. El Modelo de Regresión Logística con respuesta Dicotómica.....	13
2.4.1.2. El modelo univariado.....	13
2.4.1.3. El Modelo Multivariado.....	15
2.4.1.4. Pruebas para la significancia de los coeficientes.....	16
2.4.1.5. Interpretación de los coeficientes del modelo.....	17
2.4.1.6. Métodos para selección de variables.....	19
2.4.1.7. Pruebas de bondad de ajuste.....	20
• El Chi-cuadrado de Pearson.....	20
• El deviance.....	21
• El Test de Hosmer – Lemeshow.....	21
• Las tablas de clasificación.....	22
• El área bajo la curva ROC.....	22
• Coeficiente de determinación.....	23
• Pruebas de correlación de Pearson.....	23
2.5. ESTUDIOS DE CASOS Y CONTROLES Y APLICACIÓN DE REGRESION LOGISTICA.....	24
3. METODOLOGÍA.....	26

<b>3.1. TIPO DE ESTUDIO .....</b>	<b>26</b>
<b>3.2. RECOLECCION DE INFORMACION .....</b>	<b>26</b>
<b>3.3. VARIABLES .....</b>	<b>27</b>
<b>4. RESULTADOS .....</b>	<b>28</b>
<b>4.1. MODELO 1.....</b>	<b>28</b>
<b>4.2. MODELO 2.....</b>	<b>33</b>
<b>5. DISCUSION.....</b>	<b>39</b>
<b>6. CONCLUSIONES.....</b>	<b>41</b>
<b>BIBLIOGRAFÍA .....</b>	<b>42</b>

## LISTA DE CUADROS

	Pág.
<b>Cuadro 1. Valores del Modelo de Regresión Logística cuando la Variable Independiente es Dicotómica .....</b>	<b>18</b>
<b>Cuadro 2. Tabla de Clasificación para un modelo de Regresión Logística.....</b>	<b>22</b>
<b>Cuadro 3. Descripción de las variables .....</b>	<b>27</b>
<b>Cuadro 4. Modelos Logísticos Univariados para la MPN .....</b>	<b>28</b>
<b>Cuadro 5. Modelo Logístico Multivariado la MPN .....</b>	<b>29</b>
<b>Cuadro 6. Modelo Logístico Multivariado seleccionado por el método Step – Wise hacia delante para la MPN.....</b>	<b>29</b>
<b>Cuadro 7. Modelo Logístico Multivariado preliminar final para la MPN.....</b>	<b>30</b>
<b>Cuadro 8. Prueba de Chi cuadrado del modelo preliminar final para MPN ....</b>	<b>31</b>
<b>Cuadro 9. Coeficientes de determinación del modelo preliminar final para MPN.....</b>	<b>31</b>
<b>Cuadro 10. Estadístico de Prueba y Tabla de Contingencias para el Test de Hosmer y Lemeshow para MPN.....</b>	<b>31</b>
<b>Cuadro 11. Cuadro de clasificación del modelo preliminar final para MPN ...</b>	<b>32</b>
<b>Cuadro 12. Área bajo la curva ROC para el modelo ajustado de MPN.....</b>	<b>33</b>
<b>Cuadro 13. Modelos Logísticos Univariados para UCI.....</b>	<b>34</b>
<b>Cuadro 14. Modelo Logístico Multivariado para la UCI .....</b>	<b>34</b>
<b>Cuadro 15. Modelo Logístico Multivariado seleccionado por el método Step – Wise hacia atrás para UCI.....</b>	<b>35</b>
<b>Cuadro 16. Modelo Logístico Multivariado preliminar final para UCI.....</b>	<b>35</b>
<b>Cuadro 17. Prueba de Chi cuadrado del modelo preliminar final para UCI ....</b>	<b>36</b>
<b>Cuadro 18. Coeficientes de determinación del modelo preliminar final para UCI .....</b>	<b>36</b>

<b>Cuadro 19. Estadístico de Prueba y Tabla de Contingencias para el Test de Hosmer y Lemeshow para UCI .....</b>	<b>36</b>
<b>Cuadro 20. Cuadro de clasificación del modelo preliminar final para UCI .....</b>	<b>37</b>
<b>Cuadro 21. Área bajo la curva ROC para el modelo ajustado de UCI .....</b>	<b>38</b>
<b>Cuadro 22. Valores para las variables ESCOL_DIC y ABOR_PREV en los Modelos 1 y 2 .....</b>	<b>39</b>

## RESUMEN

Se pretende determinar un modelo que permita cuantificar el riesgo de mortalidad perinatal y/o de complicaciones neonatales tempranas no fatales entre las gestantes afiliadas a la EPS-S EMSSANAR residentes en la ciudad de Pasto, para lo cual se desarrolla un estudio analítico con base en el diseño epidemiológico de casos y controles teniendo para cada modelo estudiado la definición de caso y realizando una selección aleatoria, sin pareamiento, de 4 controles por cada caso. Se tomó información del año 2006 de los casos de mortalidad perinatal y los casos de recién nacidos que padecieron complicaciones neonatales tempranas. Los controles se seleccionaron aleatoriamente del grupo de nacidos vivos procedentes de la base de datos de Estadísticas Vitales del municipio de Pasto. Se aplicó regresión logística con respuesta dicotómica para obtener 2 modelos posibles: el primero, tomando como casos las muertes perinatales Vs. controles de nacidos vivos, y el segundo modelo tomando como casos las complicaciones neonatales tempranas Vs. controles de nacidos vivos. Luego de realizar la selección de variables, las pruebas de bondad de ajuste para ambos modelos y un ajuste propuesto para validar la aplicación de la regresión logística con base en un diseño muestral de casos y controles, se concluye que el primer modelo es el que mejor ajuste y significancia presenta. Dicho modelo permite evaluar únicamente el riesgo (probabilidad) de que una gestante afiliada a la EPS-S EMSSANAR, residente en la ciudad de Pasto, pueda presentar una muerte perinatal. Las variables Escolaridad e Historia de Abortos Previos, son las de mayor significancia estadística en el modelo y que, a su vez, determinan un riesgo relativo muy alto frente al evento de una Muerte Perinatal. Se requiere realizar otros estudios que incluyan variables de tipo clínico y paraclínico que han demostrado ser factores de riesgo para eventos adversos relacionados con la mobimortalidad perinatal para generar modelos logísticos alternos.

## INTRODUCCIÓN

Una muerte perinatal es la ocurrida en el periodo comprendido entre las 22 semanas completas (154 días) de gestación y los siete días completos después del nacimiento o mayor de 500 gramos de peso.

La mortalidad perinatal es un evento objeto de Vigilancia en Salud Pública en Colombia debido a su alto impacto social, a su peso relativo en el análisis general de la mortalidad infantil y a que tiene múltiples factores que determinan su ocurrencia, no solo desde el punto de vista del sector salud sino también de tipo, familiar, demográfico, socioeconómico, cultural y tecnológico.

Dada la relevancia de este evento como indicador de condiciones de salud y condiciones de vida de una población, así como su alta incidencia en el país (18,6 por 1.000 nacidos vivos para Colombia en el 2002) y en las distintas regiones del mismo (12,3 por cada 1.000 nacidos vivos para Nariño en el 2006) frente a la incidencia en países desarrollados, se considera necesario generar un abordaje de la mortalidad perinatal con un enfoque de riesgo, con miras a prevenir la ocurrencia del evento con base en variables ya reconocidas como determinantes del evento.

Así mismo se introduce el evento de una “complicación neonatal extrema no fatal” que se considera como un caso intermedio entre el hecho de la mortalidad perinatal y un nacimiento sin mención de complicaciones. Este evento también tendrá su definición y se considera relevante al momento de evaluar el riesgo de mortalidad perinatal.

A nivel de una empresa aseguradora en salud como EMSSANAR EPS-S, es importante acercarse a la probabilidad de ocurrencia de una mortalidad perinatal y/o una complicación neonatal extrema, con el fin de realizar intervenciones tempranas para prevenir un desenlace fatal, cumpliendo con ello su misión de realizar un manejo técnico del riesgo en salud.

De aquí se deriva la importancia de la aplicación de métodos estadísticos para el cálculo de este tipo de probabilidades con base en la información recolectada de eventos ya ocurridos. A esto se dedica el presente trabajo.

## **1. OBJETIVO GENERAL**

Determinar un modelo que permita cuantificar el riesgo de mortalidad perinatal y / o de complicaciones neonatales tempranas no fatales entre las gestantes afiliadas a la EPS-S EMSSANAR, residentes en la ciudad de Pasto.

## 2. MARCO TEÓRICO

### 2.1. DEFINICION DE TERMINOS

**Mortalidad perinatal (MPN):** como ya se definió en la introducción, es la que ocurre en el periodo comprendido entre las 22 semanas completas (154 días) de gestación y los siete días completos después del nacimiento (período neonatal temprano) o mayor de 500 gramos de peso.

**Nacido Vivo (NV):** se considera que un producto de la gestación es un nacido vivo cuando se cumple que hay expulsión o extracción completa del cuerpo de la madre, independientemente de la duración del embarazo de un producto de la concepción que respire o dé otra señal de vida como latidos del corazón, pulsaciones del cordón umbilical o movimientos de los músculos voluntarios, con cordón umbilical seccionado o no y placenta desprendida o no. Una muerte ocurrida sin que se cumpla a cabalidad esta definición se considera una muerte fetal.

**Complicación neonatal extrema no fatal (CNE):** este concepto se concibió originalmente para ser aplicado a las complicaciones maternas (no a las del feto o del neonato) con el nombre de Morbilidad Materna Extrema definida como “una complicación obstétrica severa que pone en riesgo la vida de la gestante y que requiere de una intervención médica urgente con el fin de prevenir la muerte” o “toda mujer admitida para cuidado intensivo durante el embarazo o los 42 días postparto”. Adaptando estos conceptos para el neonato (no sería posible aplicarlos para el feto ya que aún no es un nacido vivo), se define una CNE como una “complicación neonatal severa que pone en riesgo la vida del neonato y requiere una intervención médica urgente con el fin de prevenir la muerte o que requiere admisión en una unidad de cuidados intensivos (UCI) neonatal durante los primeros 7 días completos después del nacimiento. Para el presente estudio, se tiene como principal criterio para aceptar un caso de CNE, el requerimiento de admisión en una UCI neonatal. Esta adaptación conceptual se basa en estudios que demuestran una fuerte asociación entre los eventos de mortalidad materna y mortalidad perinatal, principalmente en los primeros 7 días de vida del neonato.

**Factor de Riesgo:** una variable que se considera relacionada con la probabilidad de que un individuo desarrolle o no una enfermedad, cuya acción se da antes de la iniciación del proceso patológico en la historia natural de la enfermedad.

## **2.2. COMENTARIOS SOBRE EL RIESGO EN SALUD**

El riesgo, en general se puede definir como la combinación de la probabilidad de un suceso y sus consecuencias (Guía ISO/CEI 73). Este concepto ha sido acuñado principalmente en el ámbito de la economía y la gerencia de negocios pero otras definiciones han adaptado este concepto más concretamente a las actividades humanas y al ámbito de la salud.

Un concepto del riesgo más antiguo en el tiempo pero teóricamente moderno con una visión hacia la evaluación y el manejo del mismo, lo brindan Andreas Klinker y Ortwin Renn, al definir el riesgo como la posibilidad de que las acciones humanas u otros eventos traigan consecuencias que afecten aspectos o cosas que los humanos valoren. Si la gran mayoría de seres humanos consideran dichas consecuencias potenciales como desagradables o no deseables, la sociedad se verá en la necesidad (u obligación) de evitarlas, reducirlas y mitigarlas, o al menos de controlar el riesgo de que ocurran. Este es el escenario donde se requiere el trabajo sobre la evaluación y el manejo (gestión) del riesgo.

Kleinbaum y Rothman, en sus libros de Epidemiología Moderna e Introducción a la metodología, definen el riesgo en salud en términos de incidencia acumulada como la probabilidad de que ocurra un evento de importancia para la salud en una población de individuos susceptibles, en un periodo de tiempo determinado, dado que esos individuos no pueden morir o desaparecer de la población fuente debido a otras causas.

El abordaje del riesgo en salud se ha venido construyendo alrededor de los posibles daños que pueda sufrir la salud de un individuo o de una población mediante diferentes modalidades de evaluación o gestión del riesgo tales como: las pruebas de tamizaje para enfermedades particulares, los programas para evaluación constante e integral de múltiples riesgos, y las técnicas de estimación cuantitativa del riesgo de enfermar o morir.

Esta última modalidad de evaluación del riesgo, basada en el conocimiento de los factores de riesgo para ciertas enfermedades o eventos que afecten la salud de un individuo o una población, permite considerar que es posible realizar una cuantificación del riesgo de presentar un evento que afecte la salud mediante el uso de técnicas estadísticas apropiadas, con el fin de orientar tempranamente la implementación de intervenciones preventivas.

### **2.3. DETERMINACION DE RIESGOS EN SALUD PERINATAL**

Se calcula que cada año se presentan 8 millones de muertes perinatales en el mundo, de las cuales 2,9 millones ocurren en los primeros 7 días del nacimiento. El 98% de ellas ocurren en países en vías de desarrollo.

En estudios llevados a cabo en países en desarrollo revelan que, si bien algunos factores de riesgo relacionados con las complicaciones trabajo de parto son los principales determinantes de la mortalidad perinatal (principalmente en poblaciones rurales), existen factores de tipo biológico, demográfico y cultural que determinan el riesgo de una mortalidad infantil temprana.

Algunos de estos factores encontrados como preedictores significativos de mortalidad infantil temprana son las pérdidas fetales o neonatales en embarazos previos, el hecho de que la embarazada cuente con un círculo social estable, el nivel educativo de los padres, y la paridad previa entre otros.

Estas variables pueden llegar a presentarse como predictoras de un resultado como una mortalidad perinatal o una complicación neonatal.

### **2.4. METODO ESTADÍSTICO**

El propósito principal de cuantificar el riesgo de un evento en salud se logra mediante la construcción de un modelo que tenga un adecuado ajuste matemático y que sea biológicamente razonable al describir la relación entre la variable dependiente (o respuesta) y la variable o conjunto de variables independientes (predictoras).

Los modelos más comúnmente usados son los de regresión lineal donde la variable dependiente y las predictoras son continuas. En otros modelos como el análisis discriminante, se desea clasificar un nuevo elemento, dadas unas variables conocidas, como parte de una población con base en el conocimiento de un amplio conjunto de elementos, con un cierto número de variables observadas, provenientes de dos o más poblaciones distintas. Sin embargo, el análisis discriminante clásico está basado en la normalidad multivariante de las variables consideradas como útiles para una clasificación adecuada. Así pues, si alguna de o varias las variables predictoras no son continuas, se requiere otros tipo de enfoques para este problema.

Ante un problema como el planteado donde la variable dependiente refleja la ocurrencia o no de un hecho (resultado dicotómico), y las variables independientes pueden ser continuas o discretas (dicotómicas o politómicas) sin

distribución conjunta normal multivariante, la regresión logística ofrece una respuesta más favorable.

### **2.4.1. La Regresión Logística**

Como su nombre lo indica, este método utiliza la función logística como función de distribución para el análisis de una variable dependiente dicotómica lo cual implica unas ventajas importantes: en primer lugar, desde el punto de vista matemático, es una función flexible y fácil de usar y, en segundo lugar, como ventaja adicional para el análisis de eventos en salud, se presta a sí misma para una interpretación clínicamente significativa.

En cuanto a las variables independientes o predictoras (también llamadas *covariables*) pueden ser de cualquier naturaleza: dicotómicas, ordinales, continuas o nominales.

#### **2.4.1.1. El Modelo de Regresión Logística con respuesta Dicotómica**

Si la variable dependiente (Y) refleja la ocurrencia o no de un evento en forma dicotómica, puede admitir los siguientes valores:

$$\begin{aligned} Y=1 & \text{ "el evento ocurre"} \\ Y=0 & \text{ "el evento no ocurre"} \end{aligned}$$

Para el evento de interés, en una primera etapa de resultado dicotómico, el resultado se plantearía de la siguiente forma:

$$\begin{aligned} Y=1 & \text{ "ocurre una muerte perinatal"} \\ Y=0 & \text{ "no ocurre una muerte perinatal – hay un nacido vivo"} \end{aligned}$$

Lo que se procura mediante la regresión logística es, en principio, expresar la probabilidad de que ocurra el evento como función de ciertas variables que se presumen relevantes o influyentes  $X_1, X_2, \dots, X_k$ .

#### **2.4.1.2. El modelo univariado**

El modelo en su forma analítica más simple, cuando se incluye una sola variable dependiente se presenta así:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

donde  $\pi(x)$ , representa la probabilidad de que Y tome un valor de 1 dado el valor de la variable  $x$  al utilizar la distribución logística, y donde  $\beta_0$  es el parámetro constante y  $\beta_1$  es el parámetro asociado a la variable predictora.

Es importante anotar que si la variable dependiente Y es dicotómica (o binaria), el error en la expresión

$$y = \pi(x) + \varepsilon$$

solo puede tomar 2 valores posibles:

$$\varepsilon = 1 - \pi(x) \text{ si } Y = 1 \text{ con probabilidad } \pi(x)$$

$$\varepsilon = -\pi(x) \text{ si } Y = 0 \text{ con probabilidad } 1 - \pi(x)$$

Así,  $\varepsilon$  tiene una distribución con media 0 y varianza  $\pi(x) [1 - \pi(x)]$ . Es decir que es la distribución binomial, no la normal, la que describe la distribución de los errores y será la distribución estadística en la cual se basa el análisis.

Una transformación de  $\pi(x)$  de gran utilidad en el estudio de la regresión logística es la denominada *transformación logit*, que se define en términos de  $\pi(x)$  en la siguiente forma:

$$\begin{aligned} g(x) &= \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] \\ &= \beta_0 + \beta_1 x . \end{aligned}$$

La importancia de esta transformación radica en que  $g(x)$  conserva muchas de las propiedades deseables de un modelo de regresión lineal, por lo cual los principios que guían el análisis de la regresión lineal también guían el análisis en la regresión logística.

Para lograr el ajuste inicial de un modelo de regresión logística, el método general para obtener estimadores de los parámetros que maximizan la probabilidad de obtener los valores reales observados es el de *máxima verosimilitud*. Para aplicar este método, la *función de máxima verosimilitud* es la siguiente:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

donde  $(x_i, y_i)$  son la variable dependiente y la independiente respectivamente para el sujeto  $i$  – ésimo;  $\pi(x_i)$  es la contribución de los pares  $(x_i, y_i)$  cuando  $Y = 1$  y  $[1 - \pi(x_i)]$  es la contribución de los pares  $(x_i, y_i)$  cuando  $Y = 0$ .

Una forma matemáticamente más sencilla de trabajar esta expresión está dada en forma de logaritmo así:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\}$$

Para encontrar los valores de  $\beta$  que maximizan  $L(\beta)$ , se diferencia  $L(\beta)$  con respecto a  $\beta_0$  y  $\beta_1$  y se igualan las expresiones resultantes a cero, obteniendo así las *ecuaciones de verosimilitud* siguientes:

$$\sum [y_i - \pi(x_i)] = 0 \quad \text{y} \quad \sum x_i [y_i - \pi(x_i)] = 0$$

Estas expresiones requieren métodos iterativos para su solución los cuales se encuentran ya programados en los paquetes estadísticos.

### 2.4.1.3. El Modelo Multivariado

Dado que el modelo que se espera trabajar para la mortalidad perinatal contiene más de una variable, en adelante se presentará únicamente información relacionada con el modelo logístico multivariado.

Considerando un grupo de  $p$  variables independientes denotadas por el vector  $x' = (x_1, x_2, \dots, x_p)$ , el logit para el modelo de regresión logística múltiple está dado por la ecuación

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

en cuyo caso, el modelo de regresión logística es

$$\pi(\mathbf{x}) = \frac{e^{g(\mathbf{x})}}{1 + e^{g(\mathbf{x})}}$$

Aquí se introduce el concepto de variables Dummy. Cuando alguna de las variables independientes es de índole cualitativa o nominal, para incluirla entre las covariables requiere un tratamiento especial. Si la  $j$ -ésima variable  $x_j$  consta de  $k_j$  posibles valores, deben crearse entonces  $k_j - 1$  variables Dummy (también llamadas indicadoras o de diseño) dicotómicas que serán denotadas por  $D_{jl}$  y los coeficientes para estas variables se denotan por  $\beta_{jl}$  con  $l = 1, 2, \dots, k_j - 1$ . Así el logit para el modelo con  $p$  variables y con la  $j$ -ésima variable discreta sería

$$g(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p$$

Esta notación particular se suprimirá en adelante y cada uno de los parámetros tomará un valor  $\beta_p$ .

#### 2.4.1.4. Pruebas para la significancia de los coeficientes

Una vez se cuenta con un modelo propuesto, es necesario evaluar la significancia de las variables en el modelo.

El primer término que cumple este propósito es la **razón de verosimilitud** (también denominado *deviance*) el cual resulta útil para determinar si hay una diferencia significativa entre incluir en el modelo todas las variables y no incluir ninguna, es decir, para evaluar si las variables  $x_1, x_2, \dots, x_p$ , tomadas en conjunto, contribuyen efectivamente a explicar las modificaciones que se introducen en  $\pi(x)$ .

La razón de verosimilitud corresponde a la siguiente expresión:

$$G = -2Ln \left[ \frac{\text{verosimilitud}_{\text{del modelo ajustado}}}{\text{verosimilitud}_{\text{del modelo saturado}}} \right]$$

donde el numerador del cociente es la función de verosimilitud mencionada en la sección anterior, calculada con  $p^*$  variables y el cociente es la función de verosimilitud calculada con la totalidad de las variables  $p$  donde  $p^* < p$ . La razón  $G$  se distribuye  $\chi^2$  con  $p$  grados de libertad.

Bajo la hipótesis nula los coeficientes de las  $p$  variables dependientes, son iguales a cero. Si se rechaza la hipótesis nula se puede concluir que al menos uno de los  $p$  coeficientes es diferente de cero.

El Test de Wald se basa en un segundo estadístico  $W$  que se calcula mediante la expresión

$$W = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)}$$

donde  $\hat{\beta}_1$  es el valor del estimador del parámetro  $\beta_1$  y  $\widehat{SE}(\hat{\beta}_1)$  es un estimador de la desviación estándar del estimador del parámetro. Este mismo valor se puede calcular para cada uno de los estimadores de los parámetros.  $W$  sigue una distribución normal estándar.

Otro método para evaluar la significancia del modelo es el cálculo de **intervalos de confianza** para los parámetros. La base para la construcción de los intervalos de confianza son la obtención de los estimadores  $\hat{\beta}$  de los parámetros y su respectiva desviación estándar, datos ya obtenidos para los estadígrafos anteriores. Un intervalo de confianza para  $\hat{\beta}_1$  se calcula así:

$$\hat{\beta}_1 \pm z_{1-\alpha/2} \widehat{SE}(\hat{\beta}_1)$$

Los intervalos de confianza también pueden interpretarse por medio de diagramas de dispersión que grafiquen el logit estimado o la probabilidad  $\pi(x)$  estimada frente a una variable cuantitativa a cuyos lados se ubican los intervalos de confianza en forma de bandas.

#### 2.4.1.5. Interpretación de los coeficientes del modelo

En el modelo de regresión logística, el valor de un estimador de un parámetro representa el cambio en el logit correspondiente al cambio de una unidad en la variable independiente.

Sin embargo, cuando la variable independiente es dicotómica se requieren otros conceptos para interpretar dicho valor, basados en los valores de las

probabilidades, en términos de la regresión logística, presentados dentro de una tabla de 2 x 2.

**Cuadro 1. Valores del modelo de regresión logística cuando la variable independiente es dicotómica**

Outcome Variable (Y)	Independent Variable (X)	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
y = 0	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Total	1.0	1.0

El concepto fundamental es el de los *odds*. Los *odds* asociados a cierto suceso se definen como la razón entre la probabilidad de que dicho suceso ocurra y la probabilidad de que no ocurra. Así, los *odds* del evento entre los individuos con valores de  $x=1$  en la variable dependiente serían  $\pi(1)/[1-\pi(1)]$ , y los *odds* asociados al mismo suceso entre los individuos con valores de  $x=0$  en la variable dependiente serían  $\pi(0)/[1-\pi(0)]$ . De aquí se desprende otro concepto, el de la razón de *odds* o el *odds ratio (OR)* que se define como la razón entre los *odds* cuando  $x=1$  y los *odds* cuando  $x=0$ , y está dado por la siguiente ecuación.

$$OR = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

Sustituyendo las expresiones de la ecuación anterior por las presentadas en el cuadro 1 se obtiene

$$\begin{aligned}
\text{OR} &= \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right) / \left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right) / \left(\frac{1}{1+e^{\beta_0}}\right)} \\
&= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} \\
&= e^{(\beta_0+\beta_1)-\beta_0} \\
&= e^{\beta_1}.
\end{aligned}$$

Así, para la regresión logística con una variable independiente dicotómica, la relación entre el *odds ratio* y el coeficiente de regresión es

$$\text{OR} = e^{\beta_1}$$

y el correspondiente estimador del OR sería  $\hat{\text{OR}} = \exp(\hat{\beta}_1)$ .

La importancia en la epidemiología del OR radica en que es una medida de asociación que indica que tanto más probable (o improbable) es que el evento en estudio esté presente entre los que tienen  $x = 1$  y  $x = 0$ . Desde el punto de vista estadístico, el manejo del OR facilita la interpretación de los parámetros estimados del modelo. También se puede calcular intervalos de confianza para los OR llevando a función exponencial el intervalo de confianza ya conocido para los estimadores de los parámetros tal como se muestra en la siguiente expresión:

$$\exp\left[\hat{\beta}_1 \pm z_{1-\alpha/2} \times \widehat{\text{SE}}(\hat{\beta}_1)\right]$$

Para el caso de variables independientes politémicas, al ser tratadas como variables *dummy*, se calculan los estimadores de los parámetros para cada una de las  $k_j - 1$  variables *dummy* obtenidas. La obtención del OR a partir de este cálculo es igual que para una variable dicotómica.

#### 2.4.1.6. Métodos para selección de variables

Si bien los métodos de *regresión paso a paso (step wise)* han perdido parte de su relevancia para la selección de las variables importantes para el modelo, es a la vez uno de los más usados y que puede complementarse bien con las pruebas de significancia para los parámetros y las pruebas de bondad de ajuste.

El método “Forward” consiste en construir sucesivos modelos de manera que cada uno difiera del modelo inmediatamente anterior en una sola variable e ir comparando los resultados de cada versión con los de la anterior.

El método “Backward” incluye todas las variables explicativas disponibles en el modelo, y paso a paso va eliminando las variables que no sean significativas en la estimación del modelo.

Estos son los métodos más utilizados. Otra forma de realizar selección de variables corriendo el modelo logístico con cada una de las variables en estudio, evaluando su significancia individual y considerar su posterior inclusión en el modelo. Finalmente, el criterio técnico del investigador también determina la inclusión o exclusión de alguna variable en el modelo.

#### 2.4.1.7. Pruebas de bondad de ajuste

Una vez se cuenta con un modelo cuyas variables se consideran significativas, cabe la pregunta: ¿qué tan bien el modelo seleccionado describe el resultado de la variable? Se considera que el modelo se “ajusta” si las medidas que resumen la distancia entre los valores observados de la variable respuesta  $Y$ , representados en un vector  $y' = (y_1, y_2, y_3, \dots, y_n)$ , y los valores predichos o “ajustados” por el modelo para la variable  $Y$ , representados por el vector  $\hat{y}' = (\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_n)$ , son pequeñas, y si la contribución hecha por cada par  $(y_i, \hat{y}_i)$ , con  $i = 1, 2, 3, \dots, n$  para dichas medidas de resumen es pequeña en relación al error del modelo. Para ello, se necesita calcular dichas medidas de resumen y las distancias entre  $y_i$  y  $\hat{y}_i$ , con lo cual se puede proceder al análisis correspondiente, incluso valiéndose de herramientas gráficas.

Es importante el concepto de *patrones de covariables* (covariate patterns) el cual se refiere a cada combinaciones valores de las covariables en el modelo.

- **Chi-cuadrado de Pearson**

Es una medida de resumen de la bondad de ajuste basada en el cálculo de los residuales de Pearson, dados por la siguiente expresión:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}$$

donde el valor  $m_j$  está relacionado con el número de patrones de covariables. La sumatoria de estos residuales se convierte en el estadístico chi-cuadrado de Pearson:

$$X^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2$$

- **El deviance**

Es otro estadístico basado en el cálculo de sus propios residuales que están dados por la siguiente expresión:

$$d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[ y_j \ln \left( \frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left( \frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}$$

El estadístico basado en los residuales *deviance*, es:

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2$$

Para los estadísticos mencionados hasta el momento, bajo el supuesto de encontrarnos con un modelo ajustado, la distribución es chi-cuadrado con  $J - (p + 1)$  grados de libertad.

- **El Test de Hosmer – Lemeshow**

Se realiza mediante una agrupación de los valores predichos en 10 grupos cuyos puntos de corte corresponden, generalmente, a los deciles de las probabilidades estimadas. Luego se construye una tabla donde se ubica frente a cada grupo, el número de valores observados y de valores esperados para cada una de las categorías de la variable respuesta. La simple inspección de los valores de esta tabla, cuando los valores observados y esperados son similares, ya puede dar cuenta de una adecuada bondad de ajuste del modelo, pero se ofrece una ecuación para calcular un estadístico (de Hosmer – Lemeshow):

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}$$

donde  $n'_k$  es el número total de sujetos en el  $k$  – ésimo grupo,  $c_k$  es el número de patrones de covariables en el  $k$  – ésimo decil,

$$o_k = \sum_{j=1}^{c_k} y_j$$

es el número de respuestas entre los patrones de covariables, y

$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

es la probabilidad estimada promedio.

La distribución del estadístico es aproximada a un chi-cuadrado con  $g - 2$  grados de libertad, donde  $g$  es el número de agrupaciones del test, y la interpretación del p-valor obtenido es que a medida que éste incrementa, indica que el modelo tiene un mejor ajuste.

- **Las tablas de clasificación**

Son una forma intuitiva de resumir los resultados de un modelo ajustado de regresión logística. En esta tabla se cruza la variable respuesta dicotómica contra otra variable dicotómica construida a partir de los valores de las probabilidades estimadas por el modelo logístico, cuyo punto de corte generalmente es 0,5. El cuadro 2 presenta un modelo genérico de clasificación basado en un modelo de regresión logística.

**Cuadro 2. Tabla de Clasificación para un modelo de Regresión Logística**

Clasificados	Casos Observados		Total
	MPN	Nacido Vivo	
MPN	a	b	a+b
Nacido Vivo	c	d	c+d
Total	a+c	b+d	a+b+c+d

A simple vista se podría decir que si los valores en las casillas a y d son elevados, el modelo tiene un buen ajuste ya que tiende a dar como resultado probabilidades cercanas a las observadas.

También es necesario tener en cuenta que la clasificación es sensible al tamaño relativo de los grupos de componentes, favoreciendo en la clasificación al grupo

- **El área bajo la curva ROC**

La curva ROC describe la probabilidad de detectar una señal verdadera (sensibilidad) y una falsa señal ( $1 -$  especificidad) para un rango entero de posibles puntos de corte; El área bajo la curva ROC, que tiene un rango entre cero y uno, provee una medida de la capacidad del modelo para discriminar entre

los sujetos que presentan la respuesta de interés frente a aquellos que no la presentan.

Como regla general para evaluar la capacidad discriminadora del modelo a través del área bajo la curva se puede asumir que una discriminación aceptable se da si ésta toma valores mayores a 0,7.

- **Coefficiente de determinación**

Cox y Snell en 1989 introdujeron el concepto de coeficiente de determinación para modelos lineales generalizados. Con él, se pretende cuantificar mediante un valor comprendido entre 0 y 1 la bondad del ajuste; siendo mayor la variación explicada por el modelo mientras más alto es el valor de esta medida:

$$R^2 = 1 - \left( \frac{L(\hat{\beta}_1)}{L(\hat{\beta})} \right)^{2/n}$$

- **Coefficiente de correlación de Pearson**

En el mismo sentido que la medida anterior, también se presenta el coeficiente de correlación de Pearson el cual se calcula mediante esta expresión

$$r^2 = \frac{\left[ \sum_{i=1}^n (y_i - \bar{y})(\hat{\pi}_i - \bar{\pi}) \right]^2}{\left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right] \times \left[ \sum_{i=1}^n (\hat{\pi}_i - \bar{\pi})^2 \right]}$$

Donde  $\bar{y} = \bar{\pi} = n_1 / n$ .

Otra medida denominada la **suma de cuadrados similar a la regresión lineal** es

$$R_{ss}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\pi}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

En general, los coeficientes  $R^2$  para regresión logística pueden ayudar en la etapa de construcción del modelo para comparar 2 modelos similares. Sin embargo, dados sus valores bajos en la mayoría de los casos, no se recomienda presentarlos abiertamente para evitar confusiones con la concepción que se tiene de este estadístico en el ámbito de la regresión lineal.

## 2.5. ESTUDIOS DE CASOS Y CONTROLES Y APLICACIÓN DE REGRESION LOGÍSTICA

Los estudios de Casos y Controles son una de las principales formas de investigación epidemiológica debido a su agilidad y bajo costo. Son también una de las áreas donde más se utiliza la regresión logística como herramienta de análisis. En estos estudios un grupo de individuos, denominados “casos”, tienen una condición dada (usualmente una enfermedad) y se identifican con  $Y = 1$ , son elegidos para compararlos con un grupo de sujetos, denominado “controles” que no tienen dicha condición, identificados con  $Y = 0$ . Para unos y otros se registran datos presentes o pasados (denominados “factores de exposición” asociados con las  $p$  variables explicativas) que se suponen relevantes a los efectos de desembocar en la condición que se estudia.

Si se ajusta un modelo logístico multivariado para las  $p$  variables explicativas, se obtienen las estimaciones  $\beta_0, \beta_1, \dots, \beta_p$ . Sin embargo, la función logística como tal *no permite* estimar  $P(Y=1)$  para un perfil dado. Ello es debido a que la muestra integrada por los casos y los controles seleccionados para el estudio, salvo raras excepciones, no es representativa de la población. Para poder usar las estimaciones de dichos parámetros a que da lugar el estudio de casos y controles, con el fin de estimar la probabilidad del suceso en estudio, es necesario hacer una corrección a la estimación de  $\beta_0$ , el cual se propone en la siguiente fórmula:

$$\pi(x) = \frac{1}{1 + e^{-\beta_0^* - \beta_1 x_1 - \dots - \beta_p x_p}}$$

donde

$$\beta_0^* = \beta_0 + \ln \frac{P_1}{P_2}$$

y donde, a su vez,  $P_1$  es la fracción de muestreo con que se eligieron los casos y  $P_2$  es la fracción de muestreo para los controles. De aquí resulta que si  $n_1$  y  $n_2$  son el número de casos y de controles, respectivamente, y si  $I$  es la incidencia de la población estudiada, entonces:

$$\frac{P_1}{P_2} = \frac{n_1}{n_2} \cdot \frac{1 - I}{I}$$

Ocurre con frecuencia que sólo un pequeño número de casos está disponible para el estudio. En esta circunstancia la razón de selección de los controles por cada caso no debería ser de 1:1 sino de 2, 3 o hasta 4 controles por caso. El incremento del poder estadístico con razones mayores a 4:1 es muy bajo.

### **3. METODOLOGÍA**

#### **3.1. TIPO DE ESTUDIO**

Se desarrolla un estudio analítico con base en el diseño epidemiológico de casos y controles teniendo para cada modelo estudiado la definición de caso y realizando una selección aleatoria, sin pareamiento, de 4 controles por cada caso. El análisis estadístico y la determinación del modelo logístico más adecuado se realiza mediante el uso de los paquetes estadísticos SPSS 15.0 y Statgraphics Centurión.

#### **3.2. RECOLECCION DE LA INFORMACION**

Toda la información recolectada corresponde a los eventos ocurridos durante el año 2006 ya que con ello se garantiza que la información está completa y disponible para su análisis. Además, todos los registros del año 2006 se filtraron teniendo en cuenta los criterios de afiliación la EPS-S Emssanar y que el municipio de residencia de la madre sea Pasto.

La información sobre los nacidos vivos se obtuvo del Sistema de Estadísticas Vitales del Instituto Departamental de Salud de Nariño. La misma institución brindó la información de los casos de mortalidad perinatal obtenidos de la base de datos del Sistema de Vigilancia Epidemiológica (SIVIGILA) Departamental; el documento base para el registro de información en el SIVIGILA es la ficha única de notificación para mortalidad perinatal. Los datos de las complicaciones neonatales tempranas, se obtuvieron de las historias clínicas recopiladas en el Área de Atención al Usuario de la EPS-S EMSSANAR para todos los casos de los recién nacidos que requirieron manejo en una UCI neonatal en el período de tiempo mencionado. Para estos casos, también se obtuvo la información básica de la ficha de notificación obligatoria para mortalidad perinatal. Se obtuvo un total de 12 casos de mortalidad perinatal y 12 casos de complicación neonatal temprana, cuya totalidad se tomarán como “casos”. De la población de potenciales “controles”, es decir los 1206 nacidos vivos, se obtiene aleatoriamente 48 registros para lograr la relación de 4 controles por cada caso. No se realizó pareamiento entre casos y controles mediante otras variables confusoras.

### 3.3. VARIABLES

Teniendo en cuenta que el modelo pretende diseñarse con base en variables que puedan obtenerse desde las primeras consultas prenatales, que sean clínicamente significativas para explicar alguno de los eventos en estudio sin incluir variables de resultados de exámenes paraclínicos (de laboratorio), se seleccionaron y homologaron las variables que cumplieran con estos criterios y que se encuentren en los certificados de nacidos vivos, en las fichas de Mortalidad Perinatal y en los registros clínicos de los pacientes de la UCI neonatal. En la Cuadro 3 aparecen las variables que se tuvieron en cuenta para generar los modelos estadísticos con su correspondiente codificación.

**Cuadro 3. Descripción de las variables**

Descripción	Nombre	Escala	Codificación/Valores
Mortalidad Perinatal	MPN	Nominal/Dicotómica	0 = Nacido Vivo 1 = Mortalidad Perinatal
Ocurre una Complicación Neonatal (requiere UCI)	UCI	Nominal/Dicotómica	0 = Nacido Vivo sin complicación 1 = Complicación Neonatal en UCI
Edad de la Madre (al momento del parto)	ED_MADRE	Razón	15 – 54
Convivencia de la madre	CONV	Nominal/Dicotómica	0 = Unión (Pareja) Estable 1 = Pareja no estable / sin pareja
Nivel de Escolaridad	ESCOL	Ordinal	1 = Primaria o inferior 2 = Secundaria 3 = Superior
Nivel de Escolaridad (Dicotómica) *	ESCOL_DIC	Nominal/Dicotómica	0 = Secundaria o superior 1 = Primaria o inferior
Zona de Residencia	ZONA_RES	Nominal/Dicotómica	0 = Urbana 1 = Rural
Número de embarazos previos	EMB_PREV	Razón	0 – 11 (Rango de la variable)
Número de abortos previos	ABORTO	Razón	0 – 10 (Rango de la variable)
Presencia de abortos previos (Abortos Dicotómica) *	ABOR_PREV	Nominal/Dicotómica	0 = Sin abortos previos 1 = 1 o más abortos previos

\* Modificaciones de las variables Nivel de Escolaridad y Número de abortos previos respectivamente

## 4. RESULTADOS

Se utilizaron los paquetes de software para análisis estadístico SPSS Versión 15.0, Statgraphics Centurion y las herramientas estadísticas de Excell 2003.

### 4.1. MODELO 1:

Se realiza un análisis de regresión logística con respuesta dicotómica de acuerdo con la siguiente codificación para la variable dependiente:

$$Y=1 \text{ "ocurre una muerte perinatal"}$$

$$Y=0 \text{ "no ocurre una muerte perinatal – nacido vivo"}$$

Con el fin de seleccionar las variables que se ajusten mejor al modelo se lleva a cabo una regresión logística univariada con cada una de las variables independientes revisando su significancia en el modelo (Ver Cuadro 4).

**Cuadro 4. Modelos Logísticos Univariados para la MPN**

Variable	B	E.T.	Wald	Gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED_MADRE	-0,038353	0,05202	0,5436	1	0,46095	0,96237	0,8690889	1,065669
CONV(1)	-0,356675	0,660763	0,29138	1	0,58934	0,7	0,1917149	2,555878
ESCOL			9,92554	2	0,00699			
ESCOL(1)	21,10758	17974,84	1,4E-06	1	0,99906	1,5E+09	0	.
ESCOL(2)	18,43031	17974,84	1,1E-06	1	0,99918	1E+08	0	.
ESCOL_DIC(1)	-2,822461	0,847311	11,0961	1	0,00087	0,05946	0,0112977	0,312935
ZONA_RES(1)	-0,194156	0,689812	0,07922	1	0,77836	0,82353	0,2130641	3,183083
EMB_PREV	-0,631809	0,382221	2,73238	1	0,09833	0,53163	0,2513399	1,124492
ABORTO	1,81529	0,752405	5,82088	1	0,01584	6,14286	1,4057958	26,84222
ABOR_PREV(1)	-1,81529	0,752405	5,82088	1	0,01584	0,16279	0,0372547	0,711341

Del cuadro 4 se deduce que existe evidencia de que las variables que las variables ESCOL\_DIC, ABOR\_PREV y EMB\_PREV tienen alguna asociación ( $p < 0,10$ ) con la Mortalidad Perinatal. Se decide manejar las variables ESCOL\_DIC y ABOR\_PREV correspondientes a la forma dicotómica de ESCOL y ABORTO respectivamente ya que, para la primera, se presenta una mejor significancia que en su forma de variable indicadora (dummy) y para la segunda, el análisis en términos de una regresión logística resulta más práctico al manejar

una variable dicotómica, sin que se aumente el p-valor con respecto a la forma cuantitativa.

Al realizar el análisis logístico multivariado con la totalidad de las variables (Ver Cuadro 5) se aprecia nuevamente que la significancia para el test de Wald en las variables ESCOL\_DIC y ABOR\_PREV permiten inferir alguna asociación con la variable dependiente.

**Cuadro 5. Modelo Logístico Multivariado la MPN**

Variable	B	E.T.	Wald	Gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED_MADRE	0,044194	0,094422	0,21907	1	0,63975	1,04519	0,8686023	1,257667
E_CIVIL(1)	1,399539	0,998541	1,96444	1	0,16104	0,24671	0,0348521	1,746415
ESCOL_DIC(1)	3,682869	1,226263	9,01998	1	0,00267	0,02515	0,0022738	0,278193
ZONA_RES(1)	0,359161	1,047912	0,11747	1	0,73179	1,43213	0,1836528	11,16775
EMB_PREV	0,750027	0,653516	1,31717	1	0,2511	0,47235	0,131218	1,700361
ABOR_PREV(1)	3,667418	1,418104	6,68813	1	0,00971	0,02554	0,0015855	0,411484
Constante	3,948311	2,470355	2,55449	1	0,10998	51,8477		

Finalmente, al usar el método *Step – wise* hacia delante para la selección de variables se obtiene que, en tres pasos, selecciona tres variables (ESCOL\_DIC, ABOR\_PREV y CONV).

**Cuadro 6. Modelo Logístico Multivariado seleccionado por el método Step – Wise hacia delante para la MPN**

	Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1	ESCOL_DIC(1)	-2,82246	0,84731	11,0961	1	0,00087	0,05946	0,011298	0,312935
	Constante	-0,09531	0,43693	0,04758	1	0,82732	0,90909		
Paso 2	ESCOL_DIC(1)	-3,45142	1,12726	9,37448	1	0,0022	0,0317	0,00348	0,288797
	ABOR_PREV(1)	-2,75465	1,19141	5,3458	1	0,02077	0,06363	0,00616	0,657352
	Constante	2,245792	1,19686	3,5209	1	0,0606	9,44789		
Paso 3	CONV(1)	-1,88513	0,96848	3,7888	1	0,0516	0,15181	0,022747	1,013141
	ESCOL_DIC(1)	-3,90261	1,19906	10,5932	1	0,00114	0,02019	0,001925	0,211172
	ABOR_PREV(1)	-3,51422	1,30728	7,22641	1	0,00718	0,02977	0,002296	0,385967
	Constante	4,150715	1,60096	6,72179	1	0,00952	63,4794		

Finalmente, dados los resultados de los cuadros anteriores, es claro que las variables ESCOL\_DIC y ABOR\_PREV se incluirán en un modelo preliminar final. Dado el resultado de la primera evaluación y su relevancia clínica, se incluirá también EMB\_PREV. Aún cuando el nivel de significancia por el test de Wald para el coeficiente de ED\_MADRE es  $p > 0,10$  en todas las pruebas antes realizadas, esta variable también será incluida en el modelo por su relevancia clínica. No se incluye la variable CONV debido a que solo apareció como significativa en la selección mediante el modelo Step – wise y a que resulta menos relevante que las otras desde el punto de vista clínico.

También se utiliza el modelo con las 6 posibles interacciones que se generarían con las 4 variables seleccionadas, observando que ninguna de ellas resulta significativa para ( $p > 0,05$ ). Además, no hay elementos de juicio clínico para pensar que la presencia de alguna de dichas variables modifique el valor de otra.

El Cuadro 7 presenta los parámetros del modelo con estas 4 variables y con sus estadísticos de prueba:

**Cuadro 7. Modelo Logístico Multivariado preliminar final para la MPN**

Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED MADRE	0,040151	0,08975	0,20015	1	0,6546	1,04097	0,873061	1,241166
ESCOL_DIC	3,3059	1,13016	8,55662	1	0,00344	27,2731	2,976755	249,8762
EMB_PREV	-0,88378	0,61169	2,08751	1	0,14851	0,41322	0,124598	1,370404
ABOR_PREV	3,320148	1,38582	5,73986	1	0,01658	27,6644	1,829403	418,3449
Constante	-3,33817	2,0446	2,66564	1	0,10254	0,0355		

Los valores de Exp(B) en las variables ESCOL\_DIC y ABOR\_PREV que el tener un bajo grado de escolaridad incrementa en 27 veces la probabilidad de que una gestante en el grupo pueda llegar a una Mortalidad Perinatal y el haber tenido al menos un aborto previo incrementa en 27 veces la esa misma probabilidad, con intervalos de confianza, aunque son amplios, no pasan por 1. Los valores de Exp(B) para ED\_MADRE y EMB\_PREV son cercanos a 1 y los intervalos de confianza pasan por 1 por lo cual no se consideran estadísticamente significativos.

Se procede ahora a evaluar las pruebas de bondad de ajuste para el modelo con las 4 variables mencionadas, encontrando que la prueba Chi cuadrado arroja un *p-valor* < (0,001) con lo cual se puede afirmar que el modelo en su conjunto con las 4 variables predictoras resulta significativo para explicar el resultado de P(MPN=1).

**Cuadro 8. Prueba de Chi cuadrado del modelo preliminar final para MPN**

	Chi-cuadrado	Gl	Sig.
Paso 1	26,737	4	,000

Los coeficientes de determinación, si bien no tienen la interpretación que tendrían en una regresión lineal, permiten hacer una aproximación a la explicación que las 4 variables presentes en el modelo aportan para los resultados de P(MPN=1). Para evaluar su valor como pruebas de bondad de ajuste, es posible afirmar que los valores obtenidos por este modelo en el R cuadrado de Cox y Snell y en el de Nagelkerke, se encuentran entre los más altos de todos los modelos probados hasta obtener el resultado final, superados escasamente por los resultados del modelo de tres variables sugerido por el método *Step – wise* hacia delante.

**Cuadro 9. Coeficientes de determinación del modelo preliminar final para MPN**

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	33,312(a)	,360	,569

A primera vista, el resultado de la prueba de Hosmer-Lemeshow no resulta significativo e indicaría que el modelo seleccionado no describe adecuadamente los resultados de P(MPN=1). Sin embargo en la tabla de contingencia para esta prueba (Ver Cuadro 10), en varios de los grupos de valores esperados se encuentra el valor nulo (cero). Esto es una limitante para la adecuada interpretación y valor de esta prueba. Sin embargo, se sigue la sugerencia de los autores Khan y Tempos quienes indican que se coteje los valores esperados y observados mediante simple inspección y se evalúe el grado de concordancia entre unos y otros en la tabla de contingencia para la prueba. Con esto se logra verificar que sí existe una concordancia en general entre los valores esperados y los observados.

**Cuadro 10. Estadístico de Prueba y Tabla de Contingencias para el Test de Hosmer y Lemeshow para MPN**

Paso	Chi-cuadrado	Gl	Sig.
1	29,508	8	,000

	MPN = 0		MPN = 1		Total
	Observado	Esperado	Observado	Esperado	
Paso 1 1	5	5,964	1	,036	6
2	7	6,899	0	,101	7
3	6	5,871	0	,129	6
4	6	5,818	0	,182	6
5	6	5,803	0	,197	6
6	6	5,690	0	,310	6
7	5	4,846	1	1,154	6
8	5	3,634	1	2,366	6
9	2	2,929	4	3,071	6
10	0	,544	5	4,456	5

La tabla de clasificación muestra un porcentaje global correcto del 93,3%, el cual se eleva a expensas de una adecuada clasificación de los controles nacidos vivos. Mientras que solo un tercio de los casos son clasificados adecuadamente con un punto de corte de la probabilidad predicha de 0,5. Estos porcentajes fueron los más altos en los distintos modelos que se generaron para la selección de variables.

**Cuadro 11. Tabla de clasificación del modelo preliminar final para MPN**

Observado		Pronosticado		
		MPN		Porcentaje correcto
		0	1	
Paso 1	MPN	0	1	
		48	0	100,0
		4	8	66,7
	Porcentaje global			93,3

a El valor de corte es ,500

Se realiza el ajuste para el parámetro constante de la fórmula de regresión logística para un estudio de casos y controles teniendo en cuenta que la tasa de prevalencia de la mortalidad perinatal para el período en estudio fue de 9,9 x 1000 nacidos vivos (equivalente a 0,0099):

$$\beta_0^* = \beta_0 - \ln \frac{P_1}{P_2} = -3,3381 - \ln \left( \frac{12}{48} \cdot \frac{1 - 0,0099}{0,0099} \right) = -6,5571$$

Este valor se reemplaza en la fórmula del modelo que finalmente se acepta como válido para este caso y con base en esto se obtiene el siguiente modelo final:

$$P(MPN=1) = \frac{1}{1 + \exp^{-g(x)}}$$

con

$$g(x) = 6,5571 - 0,04 * ED\_MADRE - 3,3059 * ESCOL\_DIC + 0,8837 * EMB\_PREV - 3,32 * ABOR\_PREV$$

Con base en este modelo, se calculan las probabilidades  $P(MPN=1)$  para cada uno de los individuos de la muestra encontrando que estas disminuyen con respecto a los valores del modelo sin ajustar. Con el fin de verificar la capacidad de este modelo ajustado para predecir los casos que presenten una MPN, se realiza el análisis de la curva ROC, encontrando un área bajo la curva de 0,894 con un  $p$ -valor  $< 0,05$ , lo cual lleva a considerar que el modelo tiene un aceptable poder de discriminación.

**Cuadro 12. Área bajo la curva ROC para el modelo ajustado de MPN**

Área	Error típ	Sig. asintótica(b)	Intervalo de confianza asintótico al 95%	
			Límite superior	Límite inferior
,894	,076	,000	,746	1,043

b Hipótesis nula: área verdadera = 0,5

Utilizando los valores de sensibilidad y especificidad arrojados por el análisis de la curva ROC, se evalúa cuál puede ser el punto de corte óptimo para efectos de clasificación de las probabilidades predichas por el modelo ajustado donde la sensibilidad y la especificidad sean máximas encontrando que dicho punto es  $P(MPN=1) = 0,2644$ . Los valores predichos por el modelo ajustado por encima de este punto, podrían clasificarse dentro del grupo de casos y los valores por debajo de dicho punto, dentro del grupo de controles.

#### 4.2. MODELO 2:

Se realiza un análisis de regresión logística con respuesta dicotómica de acuerdo con la siguiente codificación para la variable dependiente:

Y=1 "ocurre una complicación neonatal"  
 Y=0 "nacido vivo sin ocurrencia de una complicación neonatal"

Al igual que en el modelo 1, se utiliza una regresión logística univariada con cada una de las variables independientes revisando su significancia en el modelo (Ver cuadro 13). Se encuentra que ninguna de las variables tiene un valor de  $p < 0,05$  y únicamente la variable ESCOL\_DIC tiene una significancia que podría resultar aceptable en primera instancia para incluir a ESCOL\_DIC en el modelo.

**Cuadro 13. Modelos Logísticos Univariados para UCI**

Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED_MADRE	-0,08583	0,05859	2,14608	1	0,14293	0,91775	0,818193	1,029426
CONV(1)	-0,69315	0,65352	1,12496	1	0,28885	0,5	0,138898	1,799882
ESCOL			3,28708	2	0,19329			
ESCOL(1)	1,003302	1,2073	0,69061	1	0,40596	2,72727	0,255901	29,06599
ESCOL(2)	-0,24686	1,19635	0,04258	1	0,83652	0,78125	0,074896	8,149367
ESCOL_DIC(1)	-1,21302	0,67177	3,26063	1	0,07096	0,2973	0,079686	1,109173
ZONA_RES(1)	-0,21131	0,73843	0,08189	1	0,77476	0,80952	0,190403	3,441801
EMB_PREV	-0,20554	0,29284	0,49268	1	0,48273	0,8142	0,458639	1,445424
ABORTO	-19,9266	17974,8	1,2E-06	1	0,99912	2,2E-09	0	.
ABOR_PREV(1)	19,9266	17974,8	1,2E-06	1	0,99912	4,5E+08	0	.

Al utilizar un modelo multivariado con todas las variables (ver cuadro 14), se observa que no hay variables que tengan un  $p$ -valor  $< 0,05$ .

**Cuadro 14. Modelo Logístico Multivariado para la UCI**

Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED_MADRE	-0,09091	0,07502	1,46858	1	0,22557	0,9131	0,788246	1,057727
CONV(1)	-0,49857	0,72091	0,47829	1	0,4892	0,6074	0,147855	2,495227
ZONA_RES(1)	0,636591	0,83966	0,5748	1	0,44836	1,89003	0,364545	9,799068
EMB_PREV	0,112467	0,38907	0,08356	1	0,77253	1,11904	0,521995	2,398951
ESCOL_DIC(1)	-1,35424	0,78809	2,95285	1	0,08573	0,25814	0,055085	1,209715
ABOR_PREV(1)	19,35919	17846,1	1,2E-06	1	0,99913	2,6E+08	0	.
Constante	-17,9665	17846,1	1E-06	1	0,9992	1,6E-08		

Al usar el método Step – wise hacia delante para la selección de variables, no se selecciona ninguna variable para el modelo. Si se aplica el método Step – wise hacia atrás, se conservan dos variables en el modelo que son ED\_MADRE y ESCOL\_DIC después de 5 pasos. Sin embargo, los estimadores de los parámetros no resultan significativos ( $p$ -valor de 0,119 y 0,059 respectivamente).

**Cuadro 15. Modelo Logístico Multivariado seleccionado por el método Step – Wise hacia atrás para UCI**

	Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 5	ED_MADRE	-0,09419	0,06041	2,43107	1	0,11895	0,91011	0,808487	1,024506
	ESCOL_DIC(1)	-1,31896	0,69909	3,55961	1	0,0592	0,26741	0,067939	1,052557
	Constante	1,756547	1,56547	1,25902	1	0,26184	5,7924		

Tal como se hizo en el modelo 1, se incluye la variable EMB\_PREV debido a su importancia clínica. No se incluye la variable ABOR\_PREV debido a su valor en el test de Wald es muy bajo lo cual genera un *p-valor* cercano a 1, lo cual se explica al revisar la base de datos en la cual en ninguno de los “casos” de complicación neonatal temprana se presentaron abortos previos. Se comprueba la posible significancia estadística de las 3 interacciones posibles entre las 3 variables que se incluyen en el modelo observando que ninguna de ellas resulta significativa para ( $p > 0,05$ ). Tampoco hay elementos de juicio clínico para pensar que la presencia de alguna de dichas variables modifique el valor de otra.

Con estas consideraciones el modelo previo se construiría con las tres variables de las cuales se aprecia la estimación de sus parámetros y sus respectivos estadísticos de prueba.

**Cuadro 16. Modelo Logístico Multivariado preliminar final para UCI**

Variable	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
							Inferior	Superior
ED_MADRE	-0,0852	0,07161	1,4154	1	0,23416	0,91833	0,798075	1,056709
ESCOL_DIC(1)	-1,34671	0,71106	3,58703	1	0,05823	0,26009	0,064547	1,048062
EMB_PREV	-0,07565	0,34163	0,04904	1	0,82475	0,92714	0,474626	1,811089
Constante	1,707895	1,56522	1,19062	1	0,2752	5,51734		

El valor de Exp(B), evaluado en términos de OR, muestra que los riesgos relativos para cada una de estas tres variables resultan cercanos al valor de 1, a excepción del valor en ESCOL\_DIC que podría insinuar que la baja escolaridad es un factor protector para presentar un evento UCI. Sin embargo, los 3 intervalos de confianza contienen al número 1, lo cual muestra que los OR para este modelo no son estadísticamente significativos.

Se procede ahora a evaluar las pruebas de bondad de ajuste para el modelo con las cuatro variables mencionadas, encontrando que la prueba Chi cuadrado arroja un *p-valor*  $> (0,05)$  con lo cual se puede afirmar que el modelo en su conjunto con las 3 variables predictoras no resulta significativo para explicar el resultado de  $P(MPN=1)$ .

**Cuadro 17. Prueba de Chi cuadrado del modelo preliminar final para UCI**

	Chi-cuadrado	gl	Sig.
Paso 1	6,121	3	,106

Los valores de R cuadrado de Cox y Snell y en el de Nagelkerke obtenidos para este modelo son bajos en general, aún cuando son los más altos al compararlos con los modelos probados incluyendo a todas las variables e incluso al modelo que incluía las 2 variables por el método de Step-wise hacia atrás. Esto indica un bajo ajuste del modelo que debe corroborarse con otras medidas como las revisadas a continuación.

**Cuadro 18. Coeficientes de determinación del modelo preliminar final para UCI**

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	53,927(a)	,097	,153

La prueba de Hosmer-Lemeshow proporciona evidencia para aceptar la hipótesis nula de que el modelo propuesto describe adecuadamente los resultados de P(MPN=1), dado que el *p-valor* es mayor que 0,05. Sin embargo, en la tabla de contingencia para esta prueba (Ver cuadro 20), se aprecia que no existe una buena concordancia entre los valores observados los esperados, principalmente entre los valores del grupo UCI=1, lo cual hace pensar que no hay un adecuado ajuste del modelo a los datos reales. También cabe la consideración de que el valor del estadístico de Hosmer – Lemeshow pierda algo de su validez al presentarse varias celdas de valores observados con valor cero.

**Cuadro 19. Estadístico de Prueba y Tabla de Contingencias para el Test de Hosmer y Lemeshow para UCI**

Paso	Chi-cuadrado	gl	Sig.
1	10,098	8	,258

	UCI = 0		UCI = 1		Total
	Observado	Esperado	Observado	Esperado	

Paso 1	1	6	5,726	0	,274	6
	2	5	5,488	1	,512	6
	3	6	5,347	0	,653	6
	4	6	6,071	1	,929	7
	5	7	5,886	0	1,114	7
	6	4	4,855	2	1,145	6
	7	6	4,732	0	1,268	6
	8	3	4,386	3	1,614	6
	9	2	3,650	4	2,350	6
	10	3	1,860	1	2,140	4

La tabla de clasificación muestra un porcentaje global correcto del 76,7%, a expensas principalmente de una adecuada clasificación de los controles nacidos vivos. Solo uno de los 12 casos de Complicación Neonatal Temprana que llega a UCI se clasificó correctamente (el 8,3%) con un punto de corte de la probabilidad predicha de 0,5. El modelo que incluía las dos variables seleccionadas por el método Step-wise hacia atrás apenas mejora el porcentaje global de clasificación correcta hasta un 78,3% pero a expensas de clasificar adecuadamente un “control” de nacidos vivos sin complicaciones.

**Cuadro 20. Tabla de clasificación del modelo preliminar final para UCI**

Observado		Pronosticado		Porcentaje correcto
		UCI		
		0	1	
Paso 1	UCI	0		
		45	3	93,8
		11	1	8,3
	Porcentaje global			76,7

a El valor de corte es ,500

Se realiza el ajuste para el parámetro constante de la fórmula de regresión logística para un estudio de casos y controles teniendo en cuenta que la tasa de prevalencia de las complicaciones neonatales tempranas es la misma que para la mortalidad perinatal para el período en estudio, y fue de 9,9 x 1000 nacidos vivos (equivalente a 0,0099).

$$\beta_0^* = \beta_0 - \ln \frac{P_1}{P_2} = 1,7078 - \ln \left( \frac{12}{48} \cdot \frac{1 - 0,0099}{0,0099} \right) = -1,51108$$

Este valor se reemplaza en la fórmula del modelo que finalmente se acepta como válido para este caso y con base en esto se obtiene el siguiente modelo final:

$$P(UCI = 1) = \frac{1}{1 + \exp^{g(x)}}$$

con

$$g(x) = 1,5108 + 0,0852 * ED\_MADRE + 1,3467 * ESCOL\_DIC + 0,0756 * EMPB\_PREV$$

Finalmente se realiza el análisis de la curva ROC con los datos correspondientes a este modelo, encontrando un área bajo la curva de 0,748 lo cual lleva a considerar que el modelo tiene un adecuado poder de discriminación.

**Cuadro 21. Área bajo la curva ROC para el modelo ajustado de UCI**

Área	Error típ.	Sig. asintótica(b)	Intervalo de confianza asintótico al 95%	
			Límite superior	Límite inferior
,748	,081	,008	,589	,907

b Hipótesis nula: área verdadera = 0,5

Utilizando los valores de sensibilidad y especificidad arrojados por el análisis de la curva ROC, se evalúa cuál puede ser el punto de corte óptimo para efectos de clasificación de las probabilidades predichas por el modelo ajustado donde la sensibilidad y la especificidad sean máximas encontrando que dicho punto es  $P(UCI=1) = 0,1828$ . Los valores predichos por el modelo ajustado por encima de este punto, podrían clasificarse dentro del grupo de casos y los valores por debajo de dicho punto, dentro del grupo de controles.

## 5. DISCUSION

En el modelo 1, en el proceso de selección de variables se parte del hecho de que dos de ellas aparecen como significativas para modelo tanto en las pruebas univariadas como en las multivariadas; dos variables que tienen alguna relevancia clínica se agregan a las dos primeras que son estadísticamente relevantes. Por otra parte en el modelo 2 ninguna de las variables seleccionadas alcanzó un  $p$ -valor  $< 0,05$  en el Test de Wald, lo cual ya supone una desventaja para este modelo.

Las pruebas de bondad de ajuste evaluadas para el modelo 1 no descartan que el hecho de que éste tenga un buen ajuste entre los valores predichos y los observados en la realidad. Además se aprecia consistencia entre los valores de las distintas pruebas de bondad de ajuste, con la única excepción del Test de Hosmer – Lemeshow cuyas consideraciones ya se hicieron en los resultados. Estas mismas pruebas para el modelo 2, no aportan evidencia suficiente para sustentar el hecho de que hubiera un buen ajuste entre los valores predichos por éste y los observados.

Los valores de Exp (B) (equivalentes a los OR) en las variables ESCOL\_DIC y ABOR\_PREV en el modelo 1 resultan estadísticamente significativos y concuerdan con la evidencia epidemiológica de que una baja escolaridad y los antecedentes de abortos son factores de riesgo para la mortalidad perinatal. Los valores de Exp (B) para las otras dos variables del modelo 1 no resultaron significativos así como los valores para las variables del modelo 2.

Las diferencias entre los resultados obtenidos para los modelos 1 y 2 se atribuyen al hecho de que los valores de algunas variables para los “casos” de Complicaciones Fetales Tempranas cambiaron sustancialmente frente a los mismos entre los “casos” de Mortalidad Perinatal. Tal es el caso de dos de las variables ESCOL\_DIC y ABOR\_PREV que resultaban significativas para el modelo 1 y que ya no lo fueron en el modelo 2, y cuyos valores se pueden apreciar en el cuadro 23. Las demás variables presentaron unos valores similares, aún las dos variables numéricas, pero cabe recordar que todas ellas no resultaban estadísticamente significativas para el modelo.

**Cuadro 22. Valores para las variables ESCOL\_DIC y ABOR\_PREV en los Modelos 1 y 2**

Variable	MODELO 1		MODELO 2	
	Casos con valor = 0	Casos con valor = 1	Casos con valor = 0	Casos con valor = 1
ESCOL_DIC(1)	2	10	6	6
ABOR_PREV(1)	7	5	12	0

El uso del ajuste del modelo propuesto para el diseño de casos y controles se propuso con el fin de generar un mecanismo válido para utilizar muestras de controles y no la población total de potenciales controles, ya que en un estudio posterior donde se requiera levantar la información mediante encuestas o búsqueda de registros clínicos, se necesitaría una estrategia que disminuya costos operativos y además permita una buena predicción de la probabilidad de llegar a un evento adverso. El uso del ajuste del modelo logístico resulta bastante práctico aunque presenta la dificultad de que los valores  $P(MPN=1)$  disminuyen notablemente.

El mecanismo para hallar el nuevo punto de corte para el diseño ajustado mediante los resultados de la sensibilidad y la especificidad de la prueba para distintos puntos de corte aportados por el análisis de la curva ROC puede, aliviar el efecto de la disminución en el valor de  $P(Y=1)$  que se da con el modelo ajustado.

## 6. CONCLUSIONES

Frente al objetivo del trabajo, entre los dos modelos obtenidos para cuantificar el riesgo de mortalidad perinatal y / o de complicaciones neonatales tempranas no fatales el que mejor podría cumplir con este propósito es el modelo 1 el cual permitiría evaluar únicamente el riesgo (probabilidad) de que una gestante afiliada a la EPS-S Emssanar, residente en la ciudad de Pasto, pueda presentar, en algún momento, una muerte perinatal. Dentro del modelo 1, las variables ESCOL\_DIC y ABOR\_PREV que son una clasificación dicotómica del nivel de escolaridad y los antecedentes de abortos de las gestantes evaluadas, son las variables estadísticamente más significativas para el modelo y que, a su vez, determinan un riesgo relativo muy alto frente al evento de una muerte perinatal.

Una posibilidad que permita incluir el evento de las complicaciones neonatales tempranas no fatales es generar un modelo utilizando la regresión logística multinomial. Se hicieron algunos ensayos de este método con la población total de nacidos vivos frente a los casos de MPN y UCI pero aún no se halló la evidencia técnica para sustentar el uso de un diseño muestral específico y con validez epidemiológica, tal como el de casos y controles con relación a la regresión logística con respuesta dicotómica, para utilizar una muestra de la población de más de dos categorías de “controles” y poder ajustar un modelo estadístico.

Para estudios similares planeados a futuro, se sugiere la posibilidad de buscar otras variables de origen clínico y paraclínico que han demostrado ser factores de riesgo para eventos relacionados con morbilidad perinatal tales como la ganancia del peso materno y el nivel de hemoglobina materno durante el primer trimestre de gestación, con el propósito de generar modelos logísticos alternos y con mejor ajuste. Estas variables podrían resultar significativas dentro de un nuevo modelo.

## BIBLIOGRAFÍA

Agresti, A.: An introduction to Categorical Data Analysis, New York, NY: John Wiley and sons, 1996.

Alcaldía de Medellín, Secretaría de Salud. Modelos para el Análisis de la Mortalidad Materna y Perinatal. Medellín, 2005.

Breslow, N.E., and Day, N.E. Statistical Methods in Cancer Research. Vol. 1: The analysis of case – control studies. International Agency for research on Cancer. Lyon, France. 1980.

Cabrales J.A., Saenz L., Grau M.A., de Rojas L., González Y., Pina N.y Lizano M. Factores de riesgo de bajo peso al nacer en un hospital cubano, 1997–2000. *Rev Panam Salud Publica/Pan Am J Public Health* 12(3), 2002:180-184

Federation of European Risk Management Associations. Estándares de Gerencia de Riesgos. Brussels, 2003.

Hosmer, D.W. y Lemeshow, S: Applied logistic regression, Second Edition. John Wiley and sons. New York, 2000.

Instituto Departamental de Salud de Nariño. Indicadores en salud 2006. Pasto, 2007.

Instituto Nacional de Salud. Protocolo de Vigilancia de la Mortalidad Perinatal. Bogotá, D.C. 2007.

Katz J., West Jr K., Khatry S., Christian P., LeClerq S., Kimbrough E., Ram S.

Klinke, A., Renn, O.: A new approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based strategies. *Risk Analysis*, Vol.22, Nº 6, 2002.

Kusiako T., Ronsmans C., Van der Paal L. Perinatal mortality attributable to complications of childbirth on Matlab, Bangladesh. *Bulletin of the World Health Organization*, 2000, 78 (5):621-627.

Montgomery, Douglas C. P. Introduction to linear regression analysis. John Wiley & Sons. 2000

Organización Panamericana de la Salud. Clasificación estadística internacional de enfermedades y problemas relacionados con la salud. Décima revisión (CIE-10). Washington, D.C. 1995.

Peña, D. Análisis Multivariado. Editorial McGraw - Hill. Madrid 1999

Prado G., Collazos M., Villaba J.C.: Investigación en Salud, Factores Sociales. Editorial McGraw - Hill. Bogotá D.C., 1997.

Risk factors for early infant mortality in Sarlahi district, Nepal. *Bulletin of the World Health Organization*, 2003, 81 (10):717-725.

Silva, L.C: Excursión a la Regresión Logística en Ciencias de la Salud. Editorial Diaz de Santos. Madrid, 1995.

Vogt, T.M.: Risk Assessment and Health Hazard Appraisal. *Ann. Rev. Public Health*. 1981. 2:31-47.

World Health Organization. Millenium Development Goals, the health indicators: scope, definitions and measurement methods. Geneve, 2003.

