

ESTIMACIÓN DE LA PREVALENCIA DE ANEMIA EN ÁREAS PEQUEÑAS: UN ENFOQUE MULTINIVEL

Alia Garrudo*, Minerva Montero**, Gisela Pita*** and María Elena Díaz***.

*Centro de Neurociencias de Cuba, Cuba.

**Instituto de Cibernética, Matemática y Física, Cuba.

***Instituto de Higiene, Epidemiología y Microbiología, Cuba.

ABSTRACT

Small area estimation seeks to remedy the problem of many surveys designed at the regional level requiring information about more local areas. Unfortunately, the sample size of such regional surveys is usually too small for direct estimation on sub-regional level. This is just the case of a health survey in the eastern region of Cuba to determine the prevalence of childhood anemia. Given this scenario, this paper proposes a strategy to generate estimates at the municipal level, using an approach based on multilevel models. In the study, the influence of individual and contextual characteristics on the prevalence of anemia is evaluated and the risk of acquiring the disease in different strata of the population under two years old is estimated. The spatial variance among different risk groups demonstrates the importance of location on individual characteristics. A graphic diagnosis of the model residuals allows to identifying vulnerable areas where implementing priority programs of food and health must be achieved by optimizing the use of limited resources.

KEYWORDS: small areas, multilevel models, logistic regression, anemia prevalence.

MSC: 62P10

RESUMEN

La estimación de áreas pequeñas tiene como objetivo remediar el problema de que, si bien muchas encuestas se diseñan a nivel regional, a menudo se requiere información sobre áreas más locales. Desafortunadamente, el tamaño de la muestra lograda por las encuestas regionales suele ser demasiado pequeña para hacer estimaciones directas en el nivel sub-regional. Este es justo el caso de una encuesta de salud realizada en la región oriental de Cuba para determinar la prevalencia de anemia infantil. Ante tal escenario, en este trabajo se propone una estrategia para generar estimaciones a nivel municipal, utilizando un enfoque basado en modelos multinivel. En el estudio se evalúa la influencia de características individuales y contextuales sobre la prevalencia de anemia y se estima el riesgo de adquirir la enfermedad en diferentes estratos de la población menor de dos años de edad. La varianza territorial entre diferentes grupos de riesgo demuestra la importancia de la localidad sobre las características individuales. Un diagnóstico gráfico de los residuos de los modelos permite identificar áreas vulnerables donde aplicar programas prioritarios de alimentación y salud, optimizando el uso de recursos limitados.

1. INTRODUCCIÓN

En la actualidad es indiscutible el papel de las encuestas como instrumento de investigación para obtener información sobre un amplio rango de tópicos de interés. Mediante las encuestas se puede obtener información no sólo de la población como un todo, sino también para varias subpoblaciones. Un problema importante es cuando en una encuesta, que se diseñó a nivel de determinada población, los tamaños de muestra de las áreas específicas son demasiados pequeños para proporcionar resultados fiables a nivel local [1],[15]. Cuando éste es el caso, no es adecuado hacer estimaciones directas, se deben considerar métodos indirectos de estimación de áreas pequeñas basados en el uso de información auxiliar a través de modelos implícitos o explícitos [2], [7] [9], [13].

El objetivo de este trabajo es proponer algunas ideas para el diseño e implementación de una estrategia de estimación de áreas pequeñas, basada en la utilización de los modelos multinivel [3],[14]. La discusión se orienta hacia el análisis de los modelos apropiados en el contexto de una encuesta de salud. La motivación parte de los resultados de una serie de encuestas realizadas por el Instituto de Higiene, Epidemiología y Microbiología para estudiar la prevalencia de anemia infantil en las cinco provincias de la región oriental de Cuba. Un objetivo derivado de esta investigación es obtener información estadística a nivel municipal para ampliar el alcance y la cobertura de los programas de prevención y control de la enfermedad.

Los estimadores de áreas pequeñas basados en los modelos multinivel se usan para conectar las diferentes áreas, frecuentemente utilizando fuentes externas de información, como censos, estadísticas corrientes de oficinas administrativas, etc. El principio subyacente de este enfoque es “pedir prestada información” desde otras áreas geográficas relacionadas para aumentar el tamaño efectivo de la muestra.

Mediante este enfoque no sólo se investigan los factores que afectan la prevalencia de la enfermedad, sino que también puede conocerse qué proporción de la variabilidad de la respuesta puede deberse al contexto del área, independientemente de las características individuales. Finalmente, podrán conocerse las diferencias en la distribución geográfica de este problema de salud, lo que resultará de gran utilidad para diseñar programas de control de la enfermedad, tanto de prevención como de intervención, además de servir para evaluar sus resultados o para establecer prioridades en el reparto de recursos.

2. MOTIVACIÓN

La anemia por deficiencia de hierro continúa siendo el desorden nutricional más común en Cuba entre los niños de edad preescolar, especialmente entre los residentes de las provincias orientales. A pesar de que desde 1987 se han estado implementando en el país diferentes programas de control de la anemia, la prevalencia de esta enfermedad aún es elevada en las regiones más vulnerables [10].

Estudios realizados en las provincias de Las Tunas, Holguín, Granma, Santiago de Cuba y Guantánamo en los años 2005, 2008 y 2011, mostraron los altos porcentajes de anemia infantil en la región oriental [11], [12]. Los análisis exploratorios basados en los datos de esos tres años revelaron que además de las esperadas disparidades entre diferentes grupos de riesgo, existen también disparidades en los porcentajes de anemia por provincia, donde los más afectados por esta enfermedad son los niños menores de dos años.

Los datos de la encuesta del 2011 se utilizan para motivar la construcción de modelos logísticos multinivel, capaces de predecir las estimaciones de la prevalencia de la enfermedad en áreas pequeñas (a nivel de municipio). Los objetivos específicos de este trabajo son: 1) Identificar y evaluar factores individuales que influyen en la prevalencia de anemia, 2) determinar la prevalencia de anemia en diferentes niveles de grupos de riesgo; y 3) caracterizar las disparidades geográficas de la anemia.

3. DATOS

Para el estudio se seleccionan los datos de los niños menores de dos años encuestados en el 2011. En la Tabla 1 se presentan los tamaños de muestra resultantes en cada uno de los municipios que fueron seleccionados para la encuesta.

Evidentemente, las estimaciones directas dentro de algunos de los municipios no serán confiables debido al pequeño tamaño de muestra. Los municipios aquí se consideran entonces áreas pequeñas. En casos como éstos se sugiere usar modelos estadísticos que relacionen la variable de interés con información contextual suplementaria.

Las variables individuales (y sus categorías) seleccionadas para el estudio se presentan en la Tabla 2. Un niño se considera con anemia si el valor de su hemoglobina es menor que 110g/L. Su peso al nacer se considera bajo si es menor de 2500 g. La edad se categorizó en dos grupos porque estudios anteriores ya han revelado que durante el primer año de vida hay una mayor prevalencia de anemia. La asistencia al círculo infantil se consideró como un factor de protección contra la anemia, debido a que los niños que asisten a estos centros reciben cuidados de salud y una dieta más completa.

Table 1: Tamaños de muestra resultante en cada municipio

Provincia	Municipio	Muestra
Las Tunas	Puerto Padre	44
	Tunas	99
	Jobabo	38
	Amancio	18
Holguín	Banes	33
	Holguín	71
	Calixto García	1
	Cueto	12
	Sagua de Tánamo	36
Granma	Jiguaní	25
	Bayamo	63
	Manzanillo	40
	Campechuela	21
	Niquero	14
Stgo. de Cuba	San Luis	36
	Santiago de Cuba	97
	Palma Soriano	32
	Tercer Frente	14
Guantánamo	Guantánamo	117
	Baracoa	47
	San Antonio del Sur	19
	Niceto Pérez	8

4. MODELO DE REGRESIÓN LOGÍSTICA MULTINIVEL

Los modelos de regresión convencional suponen que todas las unidades experimentales son independientes en el sentido de que las variables que pueden estar influyendo en la presencia o no de la anemia tienen el mismo efecto en todos los municipios. La modelación multinivel relaja este supuesto y permite que el efecto de las variables varíe a través de los municipios. Bajo este enfoque, los individuos y los municipios se consideran como dos niveles separados de un sistema jerárquico. Los individuos representan las unidades de nivel-1 y los municipios representan las unidades de nivel-2.

Cuando el problema comprende dos niveles, se diferencian dos tipos de covariables: las “contextuales” o de nivel-2, cuyos valores en este caso, varían de un municipio a otro (siendo el mismo para todos los individuos del mismo municipio), y las “individuales” o de nivel-1 (cuyos valores varían entre los individuos del mismo municipio). Los modelos multinivel combinan los efectos de variables de los dos niveles dentro de un único modelo, mientras se tienen en cuenta la interdependencia entre observaciones dentro de unidades de nivel-2 [4].

En el presente estudio un análisis de particular interés es estimar la probabilidad de que el i -ésimo individuo del j -ésimo municipio tenga o no anemia en función de covariables en los dos niveles. Sea Y_{ij} la variable respuesta (*Anemia*) medida en el niño i del municipio j . Cuando la variable de interés es binaria el enfoque más común para modelar la respuesta (0,1) es utilizar la función de enlace logit o transformación log-odds [5] [16]:

$$\eta_{ij} = \text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right)$$

donde $p_{ij} = P(Y_{ij} = 1)$, con $Y_{ij}|p_{ij} \sim \text{Bernoulli}(p_{ij})$.

El modelo de regresión logística de nivel-1 con un único predictor (x) adquiere la siguiente forma:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{ij} \tag{1}$$

donde β_{0j} es el intercepto para el municipio j y β_{1j} es la pendiente de η_{ij} sobre x_{ij} en el municipio j . El intercepto β_{0j} representa el valor de η_{ij} cuando el predictor es igual a cero.

Table 2: Variables de estudio

VARIABLES	CATEGORÍAS
Anemia	1 si 0 no
Edad	1 si < 1 año 0 si ≥ 1 año
Peso al Nacer	1 si < 2500 g 0 si ≥ 2500 g
Asistencia al Cículo Infantil	1 no 0 si
Anemia de la madre durante el embarazo	1 si 0 no
Lactancia exclusiva hasta los 6 meses	1 no 0 si
Zona de residencia	1 rural 0 urbana

Como cada coeficiente β_{0j} está probablemente relacionado con el municipio, este término podría explicarse a través del siguiente modelo de nivel-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j} \quad (2)$$

donde γ_{00} es el intercepto promedio de los municipios, γ_{01} es la pendiente promedio asociada a la variable predictora z_j de nivel-2 y u_{0j} es el efecto aleatorio del municipio sobre el intercepto (condicional sobre z_j). Los coeficientes representando las pendientes de nivel-2, se suponen fijos a través de los municipios:

$$\beta_{1j} = \gamma_{10} \quad (3)$$

Sustituyendo, las ecuaciones (2) y (3) de nivel-2 en los correspondientes términos del modelo (1), se tiene la siguiente ecuación combinada:

$$\eta_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + u_{0j}$$

En esta formulación, u_{0j} es el único efecto aleatorio (representa diferentes efectos entre los municipios), el resto de los coeficientes son los efectos fijos (representa el mismo efecto para todos los municipios). Las desviaciones u_{0j} se suponen errores aleatorios independientes con distribución normal y varianza σ^2 .

El log-odds puede convertirse en una probabilidad, aplicando la transformación inversa exponencial: $p_{ij} = 1/(1 + \exp(-\eta_{ij}))$. Con este escenario es razonable utilizar esta probabilidad para generar estimaciones de la prevalencia de anemia en áreas pequeñas, ya que esta probabilidad puede capturar cualquier efecto contextual que esté afectando las localidades.

Es válido aclarar que el modelo presentado hasta aquí sólo considera una única variable de nivel-1 y otra de nivel-2, pero durante el desarrollo de la modelación se pueden incluir otras variables explicativas en cualquiera de los dos niveles.

4.1. Modelo logístico multinivel con control del estrato

En ocasiones, después de realizar un análisis multinivel, las estimaciones de los parámetros asociados a las variables explicativas en el nivel contextual no dicen mucho sobre las causas de la heterogeneidad entre las áreas. El interés de la investigación no debe entonces centrarse sólo en el nivel de agregación (en este caso, los municipios), porque es posible que además de los factores que se consideraron en el diseño del estudio, existan algunos factores de confusión que también deben ser considerados.

Para el control de los posibles factores de confusión en la etapa de análisis, se propone agrupar la muestra en estratos que sean internamente homogéneos respecto al factor de confusión. Si se supone que los estratos son diferentes, entonces se debe tratar de ajustar un modelo global que describa cómo varían los parámetros de acuerdo a los cambios en los niveles del factor de confusión (o de cualquier otro factor que interactúe y sea de interés para el estudio) [8] [6].

Si se tienen L estratos, sean W_1, W_2, \dots, W_{L-1} las variables indicadoras que permiten definir la pertenencia a los L estratos, tal que:

$$W_l = \left\{ \begin{array}{ll} 1 & \text{si } \eta_{ij} \text{ es una observación obtenida en el } l - \text{ésimo estrato} \\ 0 & \text{en cualquier otro caso} \end{array} \right\}$$

Si se consideran p variables explicativas X_1, X_2, \dots, X_p , medidas en las unidades del primer nivel, se tendrá el siguiente modelo de nivel-1:

$$\eta_{ij} = \beta_{0j} + \sum_{k=1}^p \beta_{kj} x_{ki} + \sum_{l=1}^{L-1} \alpha_l w_{lij} \quad i = 1, 2, \dots, I; \quad j = 1, 2, \dots, J$$

donde w_{lij} es el valor de la variable W_l para el i -ésimo niño del j -ésimo municipio. Nótese que los coeficientes α_l , $l = 1, 2, \dots, L - 1$, son fijos; o sea, no cambian de municipio a municipio, como es el caso de los $\beta_j = (\beta_{1j}, \beta_{2j}, \dots, \beta_{pj})$, los cuales son llamados coeficientes aleatorios.

En el presente trabajo las ecuaciones en el segundo nivel se definen como en (2) y (3), pero estos modelos pueden generalizarse incluyendo múltiples predictores de nivel-2.

Al aplicar este modelo queda eliminado el “ruido” correspondiente a las variables que se utilizan para la estratificación y que representan al factor de confusión cuyo efecto en sí mismo no ha podido ser medido. Lo que interesa es eliminar su efecto en la estimación del efecto que las variables explicativas de nivel-1 tienen en la variable respuesta.

5. RESULTADOS

Antes de evaluar los modelos condicionales, el primer paso del procedimiento empleado es determinar si la interdependencia de los individuos dentro de los municipios es significativa. Para esto se evalúa el modelo nulo (o sea, sin predictores):

$$\eta_{ij} = \gamma_{00} + u_{0j}$$

donde γ_{00} es el promedio poblacional de las probabilidades transformadas y u_{0j} es la desviación aleatoria de este promedio con respecto al promedio para el municipio j . La varianza estimada del efecto aleatorio del intercepto (Tabla 3) indicadora de la variación existente entre los municipios es estadísticamente significativa, lo que sugiere que la prevalencia de anemia de los niños de la región oriental varía entre municipios. Es importante señalar que el municipio Calixto García se excluyó del análisis por contar en su muestra con solo un niño.

El log-odds de tener anemia en un municipio típico ($u_{0j} = 0$) es -0.49 (Tabla 3). Esto corresponde a una probabilidad predicha de $1/(1 + \exp(-0.49)) = 0.38$. El parámetro de la varianza de nivel-2, que corresponde a la estimación de la varianza poblacional en el intercepto, puede usarse para calcular los rangos esperados de los log-odds, así, un intervalo de confianza (IC) del 95% para el log-odds de anemia de un municipio de la región oriental es de $-0.49 \pm 1.96\sqrt{0.66} = (-2.1 ; 1.1)$ unidades. Convirtiendo estos log-odds a probabilidades predichas, el IC del 95% para la probabilidad predicha de tener anemia es (0.1 ; 0.75). O sea, cuando se han considerado efectos aleatorios y no se ha incluido ningún predictor en el modelo, la tasa de anemia varía de 0.1% a 0.75% entre los municipios. Esta varianza de la probabilidad entre municipios justifica el uso de los modelos multinivel.

5.1. Estimación de los factores de riesgo

En esta sección se presenta un resumen del desarrollo de los modelos condicionales propuestos.

Modelo con predictores de nivel 1: Para iniciar el proceso de ajustar el efecto de los posibles factores de riesgo se estimaron los parámetros de un modelo con predictores de nivel-1 (modelo A). En el análisis, las relaciones encontradas entre la variable respuesta y las variables explicativas individuales no parecen cambiar de un municipio a otro, por lo que las diferencias entre las unidades de nivel-2 se expresan sólo en términos de los interceptos. En la Tabla 3 se presentan las correspondientes estimaciones de los parámetros y sus errores estándar.

El modelo A incluye todas las variables individuales consideradas en el estudio. El signo positivo de los coeficientes significativos del modelo A confirman que los niños menores de 1 año, con un bajo peso al nacer y cuyas madres tuvieron anemia durante el embarazo, contribuyen a aumentar la probabilidad de tener anemia. Entre los niños mayores de 1 año, el no asistir al círculo infantil también favorece la ocurrencia

de anemia. Note que de todas las variables, la *Edad*, es la que hace una contribución más significativa a la ecuación de regresión y que las variables *Lactancia exclusiva hasta los seis meses* y *Zona de residencia* no aparecen como variables explicativas determinantes, ya que los coeficientes estimados asociados a estas variables no son significativos. La variación del intercepto entre los municipios sigue siendo estadísticamente significativa, y podría ser explicada por variables en el segundo nivel.

Table 3: Estimación de los parámetros (y errores estándar) de los modelos: nulo, A y B.

Parámetros	Modelo nulo	Modelo A	Modelo B
Efectos fijos			
Intercepto	-0.49 (0.22)*	-1.58 (0.37)*	-1.20 (0.30)*
Edad < 1 año		1.09 (0.33)*	1.13 (0.33)*
No asistencia al Círculo x Edad ≥ 1 año		0.86 (0.30)*	0.91 (0.31)*
Peso al nacer < 2500 g		0.68 (0.28)*	0.65 (0.27)*
Zona de residencia rural		0.17 (0.17)	
No lactancia exclusiva hasta los 6 meses		-0.15 (0.16)	
Anemia de la madre durante el embarazo		0.41 (0.16)*	0.36 (0.15)*
Efecto contextual			
Hemoglobina media municipal-110g/L			-0.17 (0.02)*
Efecto aleatorio			
Varianza nivel-2	0.66 (0.29)*	0.68 (0.30)*	0.01 (0.02)

Nota: El símbolo * indica que la estimación es significativa ($p < 0.05$)

Desafortunadamente, no se cuenta con información adicional acerca de características de los municipios que posiblemente sean responsable de la variación entre ellos, no obstante, se debe explorar si a partir de los datos de la encuesta es factible construir nuevas variables de nivel-2 que puedan explicar la variación entre los municipios.

En la Figura 1 se muestran los 21 residuos de nivel-2 en orden ascendente, con sus intervalos de confianza al 95%, obtenidos a partir de las estimaciones del modelo A. Si se observan los intervalos, se aprecia que hacia los extremos del eje horizontal, hay un grupo de municipios donde los intervalos no contienen al cero. Esto significa que estos municipios difieren significativamente de los valores promedios predichos por los parámetros fijos del modelo A. En particular se destacan (hacia la derecha) los cuatro municipios de las Tunas y un municipio de Santiago de Cuba con los mayores residuos positivos del intercepto. De aquí se infiere que los municipios que ejercieron una mayor influencia en la prevalencia de anemia de la región oriental en el 2011 fueron (ordenados por la magnitud positiva de los residuos): Las Tunas, Jobabo, Puerto Padre, Amancio y Sagua de Tánamo. Los residuos por debajo del valor cero esperado (hacia la izquierda) se alternan entre los municipios de Granma y Santiago, exceptuando el municipio de San Luis, que junto a los municipios de Holgín y Guantánamo pueden considerarse como típicos en la región, ya que se encuentran en el conjunto de municipios cuyos intervalos de confianza incluyen el cero. Esto significa que los municipios que ejercieron una menor influencia en la prevalencia de la enfermedad en la región fueron (ordenados por la magnitud negativa de los residuos): Campechuela, III Frente, Jiguaní, Manzanillo, Niquero, Stgo. de Cuba, Bayamo y Palma Soriano.

Del análisis anterior se concluye que el gráfico de la Figura 1 puede ser muy útil para ayudar a identificar los municipios en la región oriental que sufren problemas de salud por su propia ubicación; estos problemas podrían deberse a condiciones socio-económicas particularmente diferentes. Por tanto, nuevos estudios son necesarios para determinar la presencia de otros factores que pudieran estar favoreciendo la prevalencia de anemia a nivel municipal, tales como deficiencia de micronutrientes en la dieta de la localidad, indicadores de obesidad por malnutrición, etc. De aquí que una importante cuestión es tratar de explicar la variabilidad entre los municipios mediante predictores de nivel-2.

El siguiente paso del proceso de modelación requiere del planteamiento de un modelo que sea razonable, tanto desde el punto de vista teórico como a través de ciertas evidencias en los datos en lo que a las suposiciones se refiere. Esto implica que debe haber un chequeo previo de los datos por medio de análisis exploratorios preliminares, los cuales proporcionarán información valiosa para un buen planteamiento del modelo.

Teniendo en cuenta que no se cuenta con otra fuente de información capaz de describir los contextos municipales se realizó un análisis exploratorio más exhaustivo de los datos originales. En la Figura 2 se muestra la

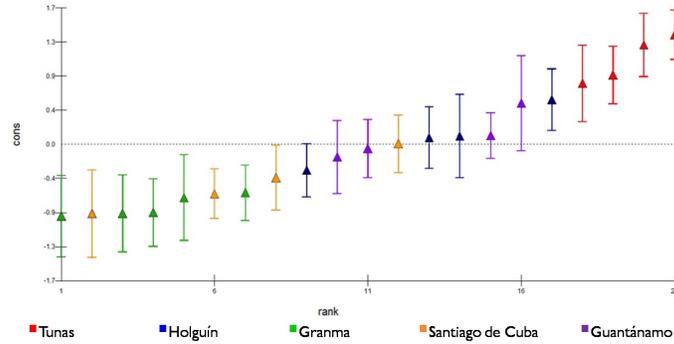


Figure 1: Residuos de los 21 municipios y sus intervalos de confianza

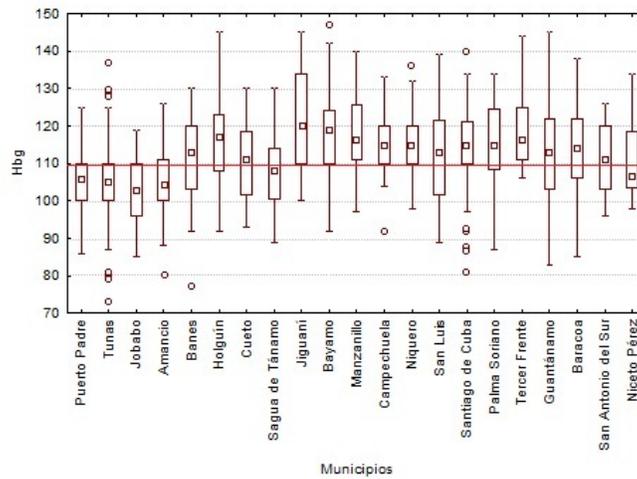


Figure 2: Distribución de la hemoglobina por municipio

distribución de la hemoglobina de los niños menores de dos años muestreados en cada municipio. La línea horizontal marca el punto de corte que se define para que un niño se considere con anemia. Es evidente que el comportamiento de la hemoglobina entre los municipios es marcadamente diferente. Se determina entonces tomar el promedio de la hemoglobina por municipio, centrada en el valor del punto de corte que se define para la anemia infantil, como un indicador del estado nutricional del municipio, capaz de explicar hasta cierto punto, las diferencias entre las unidades de nivel-2.

Modelo con predictores de nivel 1 y 2: Los parámetros estimados del modelo incluyendo la variable contextual (Hemoglobina media municipal-110g/L) se muestran en la última columna de la Tabla 3 (modelo B). El signo negativo del coeficiente asociado a esta variable revela que la probabilidad de tener anemia es mayor si el niño procede de un municipio con una hemoglobina promedio por debajo de 110g/L. Es notable cómo la varianza estimada pierde significación con respecto a la estimada en el modelo A, lo que corrobora que la variación más importante entre los municipios se debe principalmente a unos cuantos municipios atípicos con los peores y mejores resultados promedios de hemoglobina en la región.

Note que las estimaciones del modelo B se obtuvieron después de excluir las variables que no resultaron significativas en el modelo A. La variable *Edad* sigue teniendo un papel determinante al explicar las diferencias individuales en la anemia. El resto de las variables mantienen las mismas relaciones con la respuesta.

5.2. Estimación de la prevalencia de anemia en grupos de riesgo

En general, los resultados obtenidos hasta aquí, brindan un escenario de confianza para predecir, a partir de las variables explicativas consideradas, la probabilidad de tener anemia entre los niños representativos de la distribución de la población en estudio. Para calcular los log-odds de tener anemia en diferentes grupos de riesgo de un municipio típico ($u_{0j} = 0$) de la región oriental se ignora el término aleatorio del modelo y la predicción se obtiene evaluando valores específicos de las variables explicativas de nivel-1 en el modelo marginal.

El modelo se formuló considerando las covariables que resultaron significativas en el modelo B, pero no se incluyó la variable *Asistencia al Círculo Infantil* debido a que su efecto protector sobre la anemia se ve restringido por la ausencia de estos centros en muchos de los municipios de la región y a que sólo los niños mayores de un año son los que pueden asistir al Círculo. El resto de los factores (*Edad, Peso al nacer y Anemia de la madre durante el embarazo*) son variables que pueden ser identificadas por el sistema de salud para redirigir las estrategias de prevención y control en función de disminuir el riesgo de la enfermedad.

Cada subpoblación de niños definidos por las variables *Edad, Peso y Anemia de la madre durante el embarazo* en cada municipio, tiene su propio riesgo de anemia. En la Tabla 4 se muestran las correspondientes probabilidades de tener anemia para cada grupo en un municipio típico, calculadas a través de la transformación inversa de los log-odds. En la última columna aparecen los correspondientes intervalos de confianza predictivos para los municipios de la población. La predicción del número de niños en cada subpoblación puede estimarse multiplicando el riesgo de anemia predicho por el número de niños en la subpoblación.

Table 4: Estimaciones de la prevalencia de anemia en grupos de riesgo de un municipio típico

Edad	Peso	Anemia madre	Probabilidad	IC 95%
< 1 año	< 2500 g	si	0.65	0.29 - 0.86
		no	0.55	0.24 - 0.82
	> 2500 g	si	0.46	0.18 - 0.77
		no	0.39	0.14 - 0.71
≥ 1 año	< 2500 g	si	0.54	0.23 - 0.82
		no	0.47	0.19 - 0.78
	> 2500 g	si	0.39	0.14 - 0.71
		no	0.32	0.11 - 0.65

5.3. Disparidades geográficas de la anemia

Para estimar las probabilidades de tener anemia en cada municipio se considera el término de área y se ignora el término aleatorio del modelo. La predicción se obtiene evaluando valores específicos de las variables explicativas en el modelo marginal (sin el efecto aleatorio). En la Figura (3) se representa mediante una escala de colores la distribución espacial de la anemia en el grupo de mayor riesgo (niños menores de un año, con bajo peso al nacer y cuyas madres tuvieron anemia durante el embarazo) para los municipios considerados en el estudio. Los municipios más oscuros representan las áreas con mayor prevalencia de anemia para el grupo de mayor riesgo. Por otro lado, los municipios con colores más claros se asocian a áreas con menor prevalencia de anemia. Los coloreados en gris son los municipios de los que no se tienen datos.

Si se contara con variables suplementarias de otra fuentes, que ofrezcan información de todos los municipios, se podría hacer estimaciones para cada uno de los municipios, ya que se está suponiendo que los municipios seleccionados representan una muestra aleatoria de la población de municipios. O sea, las tasas locales de anemia pueden ser predichas sobre la base del conocimiento de la composición socio-demográfica de la población a través de los modelos que mejor describan cómo una variable dependiente (en este caso la anemia) responde a variables predictoras a nivel de individuo y a nivel de área.

Teniendo en cuenta la dificultad para el planteamiento de nuevos modelos y retomando la Figura (2), pero ahora haciendo distinción entre las provincias, es evidente que hay un cierto comportamiento homogéneo de la respuesta dentro de las provincias. Este comportamiento se manifiesta también en el mapa de la Figura (3). Para estudiar el efecto de la provincia, éstas se consideran como estratos diferentes y se incluyen en el modelo como predictores de nivel-2 con efectos fijos. Esto permite estimar la probabilidad de tener anemia en

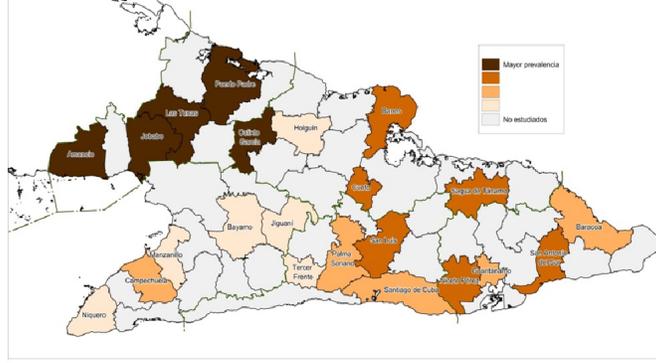


Figure 3: Distribución geográfica de la anemia por municipio

cada región eliminando el efecto contextual que tienen las provincias sobre los factores de riesgo. Tomando la provincia Granma como referencia, se crean cuatro variables indicadoras y se formula un modelo con control de los estratos obteniéndose las siguientes estimaciones:

$$\eta_{ij} = -1.59 + 0.36(\text{edad})_{ij} + 0.73(\text{peso})_{ij} + 0.44(\text{Santiago})_j + 0.94(\text{Guantánamo})_j + 0.99(\text{Holguín})_j + 2.21(\text{Tunas})_j$$

A partir de este modelo se obtienen las estimaciones de la prevalencia provincial de anemia en niños menores de un año con bajo peso al nacer, que es el grupo más vulnerable. La Tabla (5) muestra estas estimaciones y los correspondientes intervalos del confianza al 95%.

Table 5: Estimaciones de la prevalencia de anemia provincial en niños menores de un año con bajo peso al nacer

Provincia	Probabilidad	IC 95%
Las Tunas	0.84	0.74 - 0.91
Guantánamo	0.61	0.46 - 0.75
Holguín	0.61	0.44 - 0.75
Santiago de Cuba	0.48	0.32 - 0.65
Granma	0.37	0.23 - 0.53

Los efectos fijos de las provincias fueron todos significativos. Este resultado implica que cada provincia tiene un efecto diferente sobre la prevalencia de anemia. O sea, el análisis revela evidencias de efectos, no solo a nivel municipal, sino también a nivel provincial. Como era de esperar, el mayor valor estimado del efecto de la provincia sobre la prevalencia de anemia en la región oriental corresponde a la provincia Las Tunas. En la Figura (4) se muestra un mapa en el cual se aprecia claramente las disparidades geográficas de la prevalencia la anemia por provincia.

6. DISCUSIÓN Y CONCLUSIONES

Este estudio propone una estrategia para la estimación de áreas pequeñas en una investigación de salud. Los resultados de la investigación ponen de manifiesto la importancia de los modelos multinivel para estudiar la prevalencia de anemia en áreas pequeñas (en este caso, municipios), utilizando datos de una encuesta regional (nivel provincial).

Debido a la insuficiencia de datos en áreas específicas es difícil obtener estimaciones directas fiables de la prevalencia de la enfermedad a nivel municipal. En este trabajo se propone generar estimaciones de áreas pequeñas utilizando un enfoque basado en la regresión logística multinivel. La filosofía de la propuesta es incluir efectos aleatorios a nivel de área en los modelos de regresión para tener en cuenta la variación entre municipios, más allá de lo explicado por la información de covarianza.

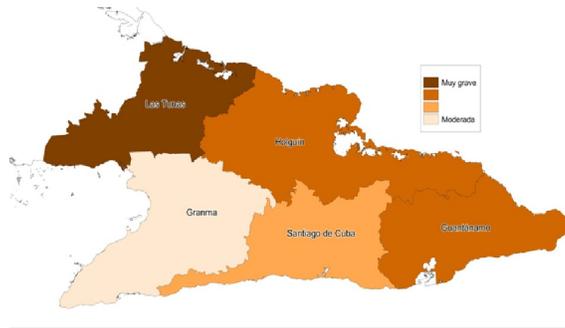


Figure 4: Distribución geográfica de la anemia por provincia

Los modelos formulados permiten explicar claramente el efecto de las variables individuales sobre la prevalencia de anemia en la región oriental, además de describir las diferencias entre los grupos de riesgo de los diferentes municipios. La información contextual adicional se utiliza para explicar parte de la variabilidad entre los municipios.

Las estimaciones de áreas pequeñas obtenidas bajo el enfoque multinivel puede ser una muy útil información a la hora de establecer prioridades y objetivos de intervención en programas locales de salud. Cuando se van a establecer objetivos de intervención en programas locales de salud, es importante detectar grupos de riesgo en los que los gobiernos provinciales puedan establecer prioridades. Por ejemplo, en el caso de Cuba, el gobierno prioriza, para los grupos de riesgo identificados, la venta subvencionada de alimentos ricos en hierro y otros micronutrientes.

En este trabajo se comprobó que los factores de riesgo más importantes entre los niños menores de un año, considerado el grupo más vulnerable, son: tener un bajo peso al nacer y que la madre haya tenido anemia durante el embarazo. Las disparidades entre provincias muestran, además, que “la localidad” contribuye significativamente en la prevalencia de anemia. Se recomienda entonces ampliar las descripciones de los contextos para identificar exhaustivamente el factor discriminante entre los municipios y/o provincias. La idea es utilizar modelos estadísticos que relacionen la variable de interés con información contextual adicional, por ejemplo, datos de censo y/o registros administrativos de puntos centinelas para áreas pequeñas.

En resumen, la estrategia propuesta no sólo permitirá el desarrollo de mejores modelos explicativos y predictivos, sino que puede crear las bases para una mejor política de evaluación y vigilancia.

RECEIVED: JULY, 2015
REVISED: OCTOBER

REFERENCES

- [1] COCHRAN G. (1977): **Sampling Techniques**. (3^{ra} ed). New York: John Wiley and Sons.
- [2] GHOSH M. and RAO J. N. K. (1994): Small Area Estimation: An Appraisal. **Statist. Sci**, 9(1): 55-76.
- [3] GOLDSTEIN H. (2002) **Multilevel Statistical Models**. (3^{ra} ed). London: Hodder Arnold.
- [4] HOX, J. (2010) **Multilevel analysis. Techniques and applications**. 2nd Edition, New York: Routledge.
- [5] KHAN H. R. and SHAW E. (2011): Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data. **Journal of Data Science**, 9: 93-110.
- [6] MONTERO, M. CASTELL y E. DÍAZ, M. (2000): Un modelo multinivel multivariado con control de un factor de confusión. **Revista Multiciencia** Vol.4 - No. 1 - ASSER-São Carlos, 137-144.
- [7] NGUYEN P., HAUGHTON D., HUDSON I., BOLAND J. (2010) Multilevel models and small area estimation in the context of Vietnam living standards surveys. **42^{èmes Journées de Statistique, Marseille, France}**. < *inria* – 00494741 >

- [8] OJEDA, M. M. y TORRES, M. T. (1997): Un modelo de regresión jerárquica para análisis de datos de una muestra por conglomerados estratificada. **Agrociencia**, 30(4): 587-596.
- [9] PFEFFERMAN, D. (2002) Small area estimation - New developments and directions. **International Statistical Review.**; 70(1): 125-143.
- [10] PITA G., JIMÉNEZ S. (2011): La Anemia por deficiencia de hierro en la población infantil de Cuba. Brechas por cerrar. **Revista Cubana de Hematología, Inmunología y Hemoterapia**, 27(2): 179-195.
- [11] PITA G., JIMÉNEZ S., BASABE B., MACÍAS C., SELVA L., HERNÁNDEZ C., CRUZ M., HERRERA R., O'FARRILL R., CALDERIUS I., PAUL K., LEYVA M. (2013). El bajo consumo de alimentos ricos en hierro y potenciadores de su absorción se asocia con anemia en preescolares cubanos de las provincias orientales 2005-2011. **Rev Chil Nutr**, 40(3).
- [12] PITA G. , JIMÉNEZ S., BASABE B., GARCÍA R., MACÍAS C., SELVA L., HERNÁNDEZ C., CRUZ M., HERRERA R, O'FARRILL R., CALDERIUS I., PAUL K., LEYVA M., AROCHA C., HERRERA D. (2014): Anemia in Children under Five Years Old in Eastern Cuba, 2005-2011. **MEDICC Review**, January 2014; 16(1): 16-23.
- [13] RAO, J.N.K. (2003): **Small area estimation**. Wiley series in survey methodology.
- [14] RAUDENBUSH, S.W. and BRYK A.S. (2002): **Hierarchical linear models: Applications and data analysis methods**. (2nd Edition). Thousand Oaks, CA: Sage.
- [15] SARNDAL, C., SWENSSON B., and WRETMAN J. (1992): **Model assisted survey sampling**. Springer-Verlag New York.
- [16] ZHANG X, ONUFRAK S, HOLT JB, CROFT JB. A Multilevel Approach to Estimating Small Area Childhood Obesity Prevalence at the Census Block-Group Level. *Prev Chronic Dis* 2013;10:120252. DOI: <http://dx.doi.org/10.5888/pcd10.120252>