

**LOS MODELOS MULTINIVEL EN EL ANÁLISIS DE  
FACTORES DE RIESGO DE SIBILANCIAS  
RECURRENTES EN LACTANTES. ENFOQUES  
FRECUENTISTA Y BAYESIANO.**

TESIS DOCTORAL PARA LA OBTENCIÓN DEL GRADO DE DOCTOR  
POR LA UNIVERSIDAD DE MURCIA.

Virginia Pérez Fernández, Licenciada en Ciencias y Técnicas  
Estadísticas

DIRECTORES:

**Prof. Dr. Manuel Sánchez-Solís de Querol**

Profesor titular de Pediatría de la Facultad de Medicina de la Universidad de Murcia

**Prof. Dr. Domingo Pérez Flores**

Catedrático de Bioestadística de la Facultad de Medicina de la Universidad de Murcia

**Prof. Dr. Diego Salmerón Martínez**

Profesor asociado de Bioestadística de la Facultad de Medicina de la Universidad de Murcia

*“Una vez que sabes que las jerarquías existen, tiendes a verlas por todas partes” Kreft, De Leeuw y Kim (1990) (1)*

# Agradecimientos

Quiero agradecer al Prof. García-Marcos, al Prof. Bosch y al Prof. Sánchez-Solís su paciencia y a la vez su insistencia para que esta Tesis llegara a buen puerto.

A mis compañeras Antonela y María, por haberme ayudado a superarme cuando lo que me apetecía era tirar la toalla.

A mis padres y hermanos por estar siempre ahí y creer que puedo con todo.

A mis amigas, Ana, Liliana, Mabel, Mari Carmen y Maribel por ser cada una mi “Pepito Grillo”.

Y a mis directores de Tesis Domingo, Manolo y Diego, por su inestimable ayuda para realizar este trabajo.

# **INDICE**

<b>1. Introducción.....</b>	<b>8</b>
1.1 Antecedentes.....	11
1.2 Breve desarrollo histórico.....	15
1.2.1 Componentes de la varianza.....	15
1.2.2 Análisis de la covarianza (ANCOVA).....	16
1.2.3 Modelo de pendientes como resultado: (“slopes as outcomes”).....	16
1.2.4 Modelos de coeficientes aleatorios (MCA).....	17
1.3 Análisis multinivel: Conceptos y estructuras.....	18
1. Estructura Jerárquica.....	19
2. Tipos de variables.....	20
3. Correlación intraclase (ICC).....	21
4. “Fijos vs aleatorios”.....	24
1.4 Campos de aplicación del análisis multinivel.....	25
1.5 Modelo básico de regresión multinivel: 2 niveles.....	29
1.5.1 Del Modelo clásico al modelo multinivel.....	29
1.5.2 Modelos que se derivan del modelo completamente aleatorizado.....	37
1.5.3 Asunciones en el análisis multinivel.....	41
1.5.4 Significación y ajuste de los modelos.....	42
1.5.5 Métodos de estimación de los parámetros.....	46
1.5.6 Varianza.....	48

1.5.7. Multinivel con diferentes variables respuesta y estructura.....	54
1.5.8. Muestreo.....	61
<b>2. Justificación.....</b>	<b>64</b>
<b>3. Objetivos.....</b>	<b>66</b>
<b>4. Material y Métodos.....</b>	<b>68</b>
4.1 Datos y variables.....	69
4.2.1 Estudio ESCAPE.....	69
4.2.2 Estudio EISL.....	70
4.2 Metodología.....	72
4.1.1 Modelo logístico multinivel de 2 niveles: Enfoque frecuentista.....	73
4.1.2 Modelo logístico multinivel de 2 niveles: Enfoque bayesiano.....	80
<b>5. Resultados.....</b>	<b>86</b>
<b>6. Discusión.....</b>	<b>113</b>
<b>7. Conclusiones.....</b>	<b>125</b>
<b>8. Bibliografía.....</b>	<b>127</b>
<b>9. Anexos.....</b>	<b>141</b>

# **1. INTRODUCCIÓN**



La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos (Estadística Descriptiva), siempre y cuando la variabilidad y la incertidumbre sean una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones (Inferencia estadística). Entre los desarrollos científicos más importantes del siglo 20, está el crecimiento explosivo del razonamiento estadístico y los métodos para aplicar a estudios de salud. Los ejemplos incluyen desarrollos en métodos de probabilidad para inferencia, estadística epidemiológica, pruebas clínicas, análisis de supervivencia y estadística genética. Los problemas en salud pública y la investigación biomédica han estimulado el desarrollo de estos métodos, que alternativamente han mejorado nuestra capacidad de hacer inferencias válidas de los datos. Este crecimiento ha sido posible gracias a los avances en técnicas computacionales y el desarrollo de paquetes de programas para implementar los procedimientos numéricos que han afectado a la práctica de la medicina y al uso de los métodos estadísticos que los investigadores usan.

En la actualidad existen cerca de 40 paquetes o programas específicos, estadísticos gratuitos o de libre distribución y aproximadamente 60 programas o paquetes no gratuitos.

<b>R</b>	Redatam
<b>SAS</b>	Minitab
<b>SPSS</b>	Matlab
SPAD	<b>S-PLUS</b>
<b>Stata</b>	LISREL
Statgraphics	WinQSB
	Excel

Tabla1. Paquetes estadísticos más utilizados

Los más usados en el campo de la investigación son R, SPSS, STATA, SAS, SPLUS.

Los modelos multinivel o también llamado Jerárquicos, son una respuesta a la necesidad de analizar la relación entre los individuos y el medio en el que se desenvuelven.

Existe una diversidad de nombres que se utilizan en los textos y publicaciones relacionados con este tipo de análisis; y que en algún sentido podrían confundir al lector que tiene una primera aproximación con las técnicas. Dentro de estos nombres se encuentran: Análisis Multinivel o Modelos multinivel (Multilevel Analysis or Multilevel Models), Análisis Jerárquico o Modelos Jerárquicos (Hierarchical Analysis or Hierarchical Models), Análisis Contextual (Contextual Analysis), Análisis de Regresión de Efectos Aleatorios Multinivel (Multilevel Random Effects Regression Analysis, Modelos Lineales Jerárquicos (Hierarchical Linear Models-HLM), Modelos de Efectos Mixtos (Mixed-Effects Models), Modelos de Coeficientes Aleatorios (Random Coefficient Models), etc., de todas estas denominaciones la más genérica es la de Modelos Multinivel, que refleja claramente la naturaleza jerárquica de los datos, pero no prejuzga la forma de las funciones que se utilizan (2), ni el tipo de jerarquía.

El software más completo para el análisis multinivel es el MLwiN. Existen otros programas que se pueden usar para el análisis de datos con una estructura jerárquica aunque son más limitados (3;4), como:

- HLM, <http://www.ssicentral.com/hlm/>
- BMDP 5V, <http://www.statistical-solutions-software.com/products-page/bmdp-statistical-software/>
- PROC MIXED (procedimiento de SAS), [http://support.sas.com/documentation/cdl/en/statug/63033/H/TML/default/viewer.htm#mixed\\_toc.htm](http://support.sas.com/documentation/cdl/en/statug/63033/H/TML/default/viewer.htm#mixed_toc.htm)
- STATA, <http://www.ats.ucla.edu/stat/stata/topics/MLM.htm>

Para una revisión más amplia consultar con <http://www.bristol.ac.uk/cmm/learning/mmssoftware/>

## 1.1 Antecedentes:

Los modelos Multinivel trabajan con observaciones anidadas en grupos. Los datos sociales, del comportamiento, los económicos y los de las ciencias de la salud tienen a menudo una estructura jerárquica. Un ejemplo en el campo de las ciencias de la salud son pacientes agrupados en médicos, centros de atención primaria y estos agrupados a su vez en áreas de salud, etc. Estas agrupaciones hacen que los sujetos que pertenecen al mismo grupo reciban una serie de influencias comunes; por lo tanto, las observaciones dentro de un mismo grupo no son independientes entre sí y si se aplican los métodos de análisis de regresión se viola uno de los principios básicos de la estadística frecuentista, la independencia de los datos (3), lo que se traduce en la subestimación de los errores estándar de los coeficientes de regresión. Esto implica, que los intervalos de confianza serán demasiado estrechos y los p-valores demasiado pequeños, lo que nos puede llevar a inferir que un predictor tiene un efecto “real” sobre la variable respuesta cuando podríamos atribuir dicho efecto al azar, es decir, a un aumento del riesgo  $\alpha$  o probabilidad de error tipo I. Tradicionalmente, se han usado modelos de regresión con parámetros lineales fijos para el análisis de este tipo de datos. Un estudio conocido e influyente llevado a cabo en 1970 en niños de primaria (5), concluyó que los niños expuestos a estilos llamados 'formales' de enseñar lectura, exhibieron un mayor progreso que aquellos que usaron otro método. Se analizaron los datos utilizando técnicas de regresión múltiples tradicionales, usaron como unidad de análisis a los niños de forma individual e ignoraron sus agrupaciones en profesores y en clases. Los resultados fueron estadísticamente significativos. Después Aitkin et al. (6), demostraron que cuando el análisis tuvo en cuenta correctamente la agrupación de los niños en clases, las diferencias significativas

desaparecieron y no se podía concluir que los niños que usaron un método tradicional diferirían de los otros (3).

Se pueden analizar variables de diferentes niveles, como si fueran de un único nivel común, usando la agregación o desagregación de los datos. Por ejemplo: una variable medida en un médico, tal como la edad, se puede agregar a nivel de centro de salud asignando a un centro la edad media de los médicos que allí trabajan. Una variable medible en el centro, tal como si es pequeño, mediano o grande, puede ser desagregada a nivel de médico asignando a cada médico el tamaño del centro donde trabaja. De esta forma se puede usar simplemente la correlación clásica o las técnicas de regresión, o al menos eso parece. Las consecuencias de la agregación y la desagregación no son triviales, y crean dos tipos de problemas. El primer problema es estadístico, la agregación de los datos, consiste en calcular el valor de la media en cada grupo en las variables a estudiar y a continuación realizar el análisis con las unidades de segundo orden, los grupos. Esto se traduce en una pérdida de información y de potencia estadística (aumento del riesgo  $\beta$  o probabilidad del error tipo II). Por otro lado si los datos están desagregados, es decir, si asignamos los valores de las variables de segundo orden o nivel a cada unidad individual, y a continuación se realiza el análisis en el nivel individual o primer nivel, nos encontramos con que no podemos considerar que los datos del primer nivel de un mismo grupo constituyan observaciones independientes. Si obviamos este hecho, y tratamos de estudiar diferencias entre grupos, aumentamos el riesgo de cometer un error de tipo I o riesgo  $\alpha$ , o por el contrario, ser demasiado conservador al estudiar las diferencias entre individuos (7). El segundo problema se plantea desde el punto de vista conceptual, y surge a la hora de interpretar los resultados. Si no se es muy cuidadoso en la interpretación de los resultados, se puede incurrir en “la falacia del nivel equivocado” que consiste en analizar los datos en un nivel y

extraer conclusiones a otro nivel. El caso más conocido es el de la “falacia ecológica” conocida también por el 'efecto de Robinson' (8). Robinson en su investigación, mostró datos agregados que describían la relación entre el porcentaje de negros y el nivel de analfabetismo en nueve regiones geográficas en 1930. La *correlación ecológica* (la correlación entre las variables agregadas) a nivel de región fue 0.95, pero la correlación individual entre las variables absolutas en el nivel individual era 0.20. Robinson concluyó que en la práctica una correlación ecológica es claramente distinta a su correlación individual. Por lo tanto, realizar inferencias sobre un nivel alto de la jerarquía, con resultados obtenidos del análisis realizado a un nivel más bajo es engañoso; este error es conocido como “la falacia atomista”. Alker (9) categorizó estos tipos de falacias, ver tabla2.

Nivel de estudio	Inferencia	Tipo de Falacia
<ul style="list-style-type: none"> <li>•GRUPO --&gt;</li> <li>•INDIVIDUAL --&gt;</li> <li>•INDIVIDUAL --&gt; (sin incluir variables relevantes del nivel grupo)</li> <li>•GRUPO --&gt; (sin incluir variables relevantes del nivel INDIVIDUAL)</li> </ul>	<ul style="list-style-type: none"> <li>•NIVELINDIVIDUAL --&gt;</li> <li>•NIVELGRUPO --&gt;</li> <li>•NIVELINDIVIDUAL --&gt;</li>   <li>•NIVELGRUPO --&gt;</li> </ul>	<ul style="list-style-type: none"> <li>• <b>ECOLÓGICA</b></li> <li>• <b>ATOMISTA</b></li> <li>• <b>PSICOLOGÍSTICA</b></li>   <li>• <b>SOCIOLOGÍSTICA</b></li> </ul>

Tabla2. Adaptado de Diez-Roux AV, Schwartz, S, Susser,E. Ecologic studies and ecologic variables in public health research (10)

Un error diferente, es el conocido como 'la Paradoja de Simpson' (11). La paradoja de Simpson, describe la desaparición de la significación estadística de una asociación entre variables o una comparación entre grupos, cuando los datos son desagregados por grupos. También es referida como el cambio en el sentido de una asociación entre dos

variables (cuantitativas o cualitativas), cuando se controla el efecto de una tercera variable o variable de confusión(12).

Se ha señalado también por Holt, Scott y Ewings (1980) (13), que pueden resultar en graves errores de inferencia, del análisis de datos de esquemas complejos de muestreo, si lo que se asume, es que los datos han sido obtenidos bajo un esquema de muestreo aleatorio sencillo.

## 1.2 Breve desarrollo histórico:

En esta sección se recoge un breve resumen sobre los modelos estadísticos, que más han contribuido al desarrollo del análisis multinivel.

Las herramientas desarrolladas en un área de estudio a menudo pueden ser usadas en otras áreas. Este resumen muestra que los modelos multinivel son un conglomerado de modelos conocidos, tales como, análisis de componentes de la varianza, modelos de coeficientes aleatorios,....

### 1.2.1 Componentes de la varianza:

El análisis de componentes de la varianza (y análisis de modelos mixtos) tiene una complicada y larga historia. El primer uso de esta técnica fue en astronomía (1861) por Airy (14). Pero fue el trabajo de Fisher (15) el que plantó la semilla para el crecimiento de posteriores trabajos. El procedimiento está diseñado para estimar la contribución de múltiples factores a la variabilidad de una variable dependiente  $Y$ . Se usa para analizar un experimento anidado en el cual los factores están estructurados en un estricto orden jerárquico, y en el que todos los efectos se asumen como aleatorios. En este tipo de estudios, las muestras de cada factor, son tomadas del interior de las muestras del factor inmediatamente superiores a él, y la variabilidad de  $Y$  es la suma de la variabilidad explicada por los diversos factores. Dentro de estos modelos, los componentes individuales son independientes, pero se correlacionan perfectamente dentro de los grupos. Se han desarrollado modelos de regresión aleatorios para modelar datos continuos (Bock 1983) y también medidas repetidas de datos dicotómicos (16) donde ciertas características de los datos excluyen la utilización de modelos tradicionales ANOVA.

### 1.2.2 Análisis de la covarianza (ANCOVA):

El análisis de la covarianza, es otra forma tradicional de analizar datos agrupados. Los dos niveles de análisis, están incluidos en el modelo, pero no de la misma forma. Las variables explicativas del nivel individual están relacionadas, como en los modelos de regresión, pero al mismo tiempo a las unidades del segundo nivel se les permite variar en el intercepto. Los modelos ANCOVA incorporan tanto variables cuantitativas como cualitativas y por lo tanto tiene un carácter mixto. ANCOVA evalúa el efecto del segundo nivel, corrigiendo por las diferencias pre-existentes entre las unidades de segundo nivel. Este análisis considera los efectos individuales como ruido, y se centra en el efecto de grupo. Todo lo que el ANCOVA puede hacer es estimar si existen diferencias entre los contextos, pero no nos dice por qué existen esas diferencias.

La utilización de modelos multinivel se ha visto entorpecida en el pasado por la forma cerrada de las fórmulas matemáticas para estimar los componentes de la varianza y covarianza, que han estado disponibles solamente para diseños perfectamente balanceados.

### 1.2.3 Modelo de pendientes como resultado: (“slopes as outcomes”)

Dentro del escenario de trabajo de los modelos de efectos fijos, la aproximación de “pendientes como resultados” se puede considerar como una aproximación al análisis multinivel. Fue propuesta en 1978 por Burstein et al. (17), como una alternativa a las técnicas de descomposición de la varianza de Cronbach. La idea es realizar un análisis de regresión lineal separado para cada grupo o elemento del nivel macro, seguido por la regresión de los coeficientes del primer nivel en el segundo, usando los interceptos y coeficientes como variable



respuesta junto con las variables explicativas del nivel macro o nivel superior. Pero esto no aporta nada nuevo desde el punto de vista estadístico, las regresiones intra y entre grupos no están conectadas unas con otras. Si se ignora la estructura jerárquica de los datos y se analiza cada grupo separadamente, se obtienen modelos poco parsimoniosos, y que se ocupan solo de las diferencias y no de los puntos en común.

### 1.2.4 Modelos de coeficientes aleatorios (MCA):

Fueron propuestos por vez primera en 1947 por Wald(18) y en 1950 por Rubin (19), y la revolución computacional en los 70 ayudó a que se extendiera su uso. Estos modelos pueden ser considerados una extensión de los modelos de componentes de la varianza descritos anteriormente, en los que se incluyen variables explicativas o covariables, y conceptualmente están basados en los modelos “pendientes como coeficientes”. Una diferencia entre estos dos modelos es que los MCA no estiman de forma separada los coeficientes para cada grupo o elemento de nivel macro, aunque cada grupo puede diferir de otro en el intercepto, en las pendientes o en ambos (20). En estos modelos, los coeficientes de regresión del primer nivel son tratados como variables aleatorias en el nivel macro. Esto significa que los coeficientes del primer nivel son entendidos como originadores de una distribución de probabilidad. Los parámetros más importantes de esta distribución, media y varianza, están entre el conjunto de parámetros que deben ser estimados por el modelo (2). El problema de estos modelos es que proporcionan estimaciones poco reales para grupos que tienen pocas observaciones. *Y no hay ninguna posibilidad de incluir variables de otro nivel.*

## 1.3 Análisis multinivel: Conceptos y estructuras

Para poder usar tanto coeficientes aleatorios como variables de distintos niveles, se deben utilizar modelos multinivel. El análisis Multinivel permite incluir características de grupos diferentes en modelos de comportamiento individual. Al incorporar las características de grupos al modelo multinivel, se tiene en cuenta la estructura jerárquica de los datos, se obtienen estimaciones correctas de los errores estándar y se facilita la exploración de la variación entre grupos que puede ser de interés por derecho propio. Además se pueden construir también pruebas válidas e intervalos de confianza y se pueden incorporar variables de estratificación al modelo, utilizadas en el diseño de la muestra. En datos estructurados jerárquicamente, los individuos pertenecientes al mismo grupo probablemente son más parecidos entre sí que individuos en grupos diferentes. Debido a esto, las variaciones en el resultado se pueden deber a diferencias entre grupos, y a diferencias individuales dentro de un mismo grupo.

Estos son algunos ejemplos de preguntas que pueden ser exploradas usando el análisis multinivel:

- ¿Puede una variable del campo de la salud variar a través de las distintas áreas de salud?
- ¿Son estas variaciones entre áreas explicadas por la diferencia en el acceso a los servicios de salud?
- ¿La variación total entre áreas es diferente entre la zona urbana y la rural?

Antes de adentrarnos más en el AM, es necesario hacer un resumen de los **conceptos básicos** para su entendimiento:

**1. Estructura Jerárquica:** Una jerarquía consiste en un conjunto de observaciones que conforman el nivel individual, primer nivel o micro nivel, que se encuentran anidadas dentro de un nivel superior, segundo nivel o nivel macro (esto en el caso más sencillo, en el que solo se contemplan dos niveles, aunque cabe más de un nivel macro, como por ejemplo pacientes agrupados en centros de salud, dentro de áreas de salud, y agrupados en provincias). También se suelen referir a los niveles macro como grupos o contextos (esto ocurre más en el ámbito de la educación). Ejemplos de estas estructuras son pacientes anidados en médicos o medidas repetidas anidadas dentro de personas (2;20).

*Clasificación gráfica de Estructuras Jerárquicas (21;22):*

**1. a. Estructuras jerárquicas de 2 niveles:**

*Las estructuras jerárquicas aparecen cuando tenemos una unidad del primer nivel o del nivel más bajo anidada en una y solo una unidad de segundo nivel o nivel más alto. Tipos:*

*1. a.1 Pacientes dentro de centros de salud*

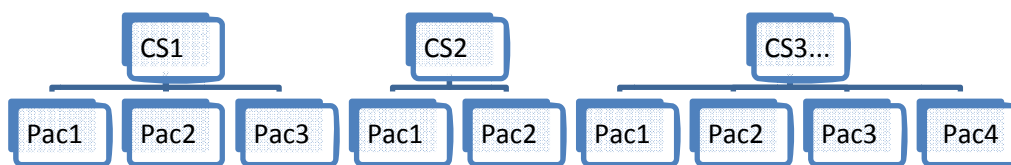


Fig. 1. Diagrama de 2 niveles anidados.

1. a.2 Medidas repetidas, datos de panel

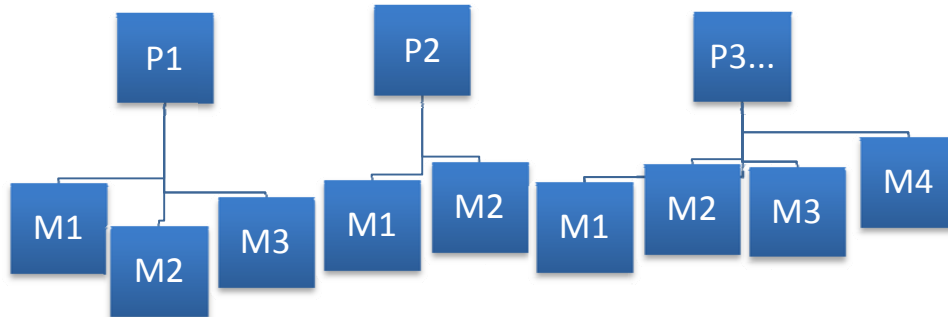


Fig. 2. Clasificación y diagrama de unidades de un diseño de medidas repetidas, mediciones anidadas en pacientes.

1. a.3 Respuestas multivariantes dentro de individuos

A veces deseamos modelar más de una respuesta. Por ejemplo, podemos querer considerar conjuntamente peso y talla de una persona, porque las dos respuestas están probablemente relacionadas.

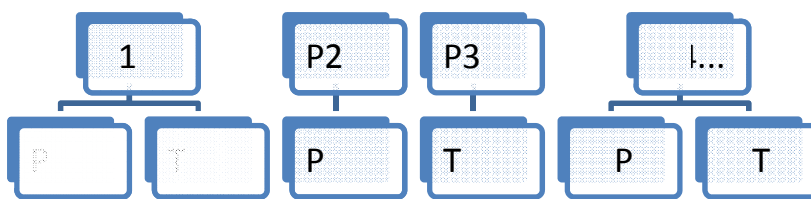


Fig. 3. Diagrama de un modelo de respuesta multivariante

**1. b Estructura de 3 niveles:**

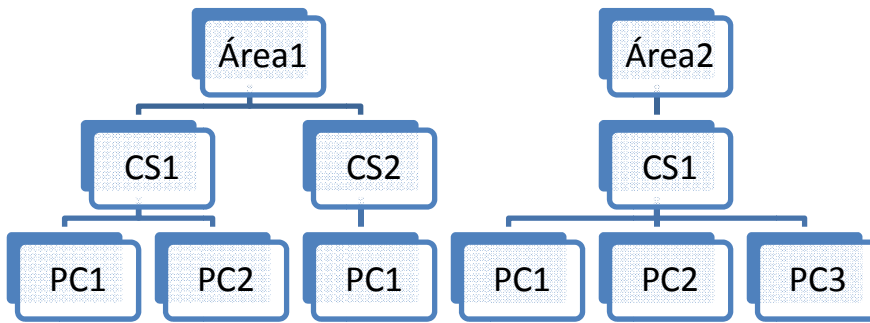


Fig. 4. Estructura jerárquica de 3-niveles, pacientes dentro de centros de salud dentro de áreas.

**1. b.1 Estudio trasversal repetido:**

Individuos dentro de cohortes y estas dentro de ciudades.

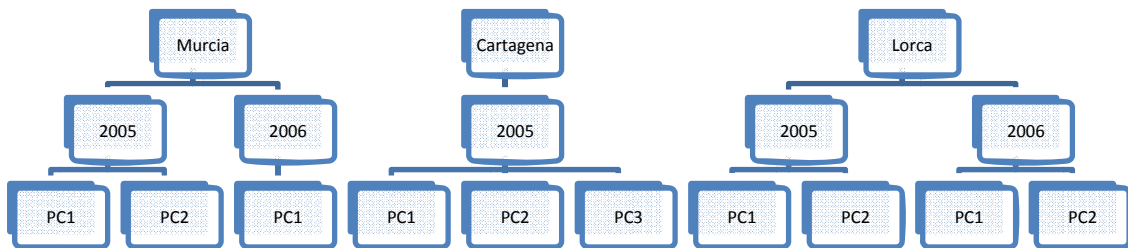


Fig. 5. Estructura jerárquica de 3-niveles

### 1. c. Estructuras no-jerárquicas:

En todos los ejemplos vistos con anterioridad, cada unidad de un nivel inferior se encontraba anidada en una y solo una unidad de nivel superior. Esto es lo que se conoce como jerarquías estrictas. Pero la realidad puede ser más complicada que esto, y podemos encontrar:

- Estructuras con clasificación cruzada

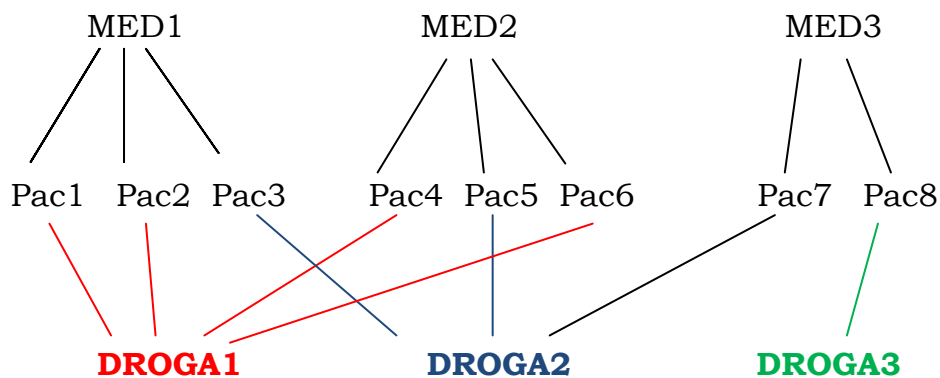


Fig6. Muestra los pacientes unidos al tipo de medicamento que toman y al médico que los atendió. El médico no está anidado dentro del tipo de droga, y esta a su vez no está anidada en el nivel médico. Clasificación cruzada.

**2. Tipos de variables:** Desde el punto de vista de la estadística, en relación a la naturaleza de los datos se habla de variables continuas, discretas, nominales y ordinales; estableciéndose una serie de modelos, procedimientos y formas de análisis según sea el tipo de datos con los que se cuenta. En multinivel, además de tomar en cuenta la naturaleza de los datos, se han establecido otro tipo de clasificaciones, que tienen que ver con el propósito del estudio y con la unidad de análisis de la que se obtiene la información. La primera clasificación agrupa la información en variables de respuesta y variables explicativas, tal como se conocen en el análisis de regresión. La segunda clasificación viene

dada por la unidad de análisis en la que se le miden las variables. Supongamos una estructura jerárquica en la población donde se identifican las siguientes unidades: pacientes dentro de hospitales, y estos a su vez agrupados dentro de áreas geográficas. Las variables que se extraen de la unidad más pequeña, en este caso los pacientes, se les denominan variables a nivel individual; las variables que se miden a las unidades que agrupan a otras unidades, por ejemplo los hospitales o áreas geográficas, se llaman variables a nivel grupo o variables contextuales. En algunos casos las variables a nivel individual y las de grupo son similares, pero medidas por diferentes indicadores, por ejemplo a nivel individual el ingreso de la persona y al nivel grupo el ingreso medio de un barrio. Por otra parte, está el caso en que la variable puede ser medida únicamente al nivel grupo, porque fue definida sólo en ese nivel, por ejemplo el número de establecimientos de salud dentro de un área geográfica. La importancia de las variables a nivel grupo es que contienen información que no ha sido recogida por las variables a nivel individual, y es por esto que se incluyen en el análisis. En multinivel, las variables de respuesta se miden al nivel individual, y las variables explicativas se miden tanto a nivel individual como de grupo (23).

**3. Correlación intraclase (ICC):** Es un indicador de la homogeneidad interna de los grupos. Es una medida de la similitud de las unidades del nivel individual y de las diferencias entre las unidades del nivel macro. Por ejemplo, se desea investigar la relación entre el nivel de colesterol y la edad, en este estudio las observaciones de los pacientes están anidadas dentro de distintos médicos. El ICC se define como la varianza entre los médicos dividida entre la varianza total, donde la varianza total es la suma de la varianza entre médicos y la varianza dentro de cada médico (24). Cuanto menor es la varianza dentro de los grupos, mayor es el índice de correlación intraclase.

**4. “Fijos vs aleatorios”:** estos términos se aplican a 3 entidades distintas:

a. Efectos fijos y efectos aleatorios. Un factor compuesto por diferentes tratamientos se dice que es un factor de efectos fijos si todos los posibles tratamientos en los que está interesado el investigador están presentes en el experimento. Un efecto aleatorio se refiere a un factor compuesto por una muestra aleatoria de entre todos los tratamientos relevantes. Por ejemplo, los efectos fijos solo permiten que se hagan inferencias acerca de los tratamientos usados en el experimento.

b. Variables fijas y aleatorias. Una variable aleatoria es una variable que toma sus valores de una distribución de probabilidad, por lo tanto, tiene una media y una varianza (que puede ser o no conocida). En general, se asume que al medir una variable aleatoria se comete un error, ya que si se realizan dos mediciones, difieren una de otra. Una variable fija es aquella que sus valores son conocidos. Un ejemplo es la variable sexo (20).

c. Coeficientes fijos y aleatorios. Son conceptos nuevos, que aparecen en relación a los modelos de coeficientes aleatorios. En los modelos de regresión lineal clásicos, se estiman los parámetros que especifican la recta de regresión, que son el intercepto y las pendientes; estos parámetros son los mismos para todos los grupos, en otras palabras, estos coeficientes son fijos. Los coeficientes aleatorios son coeficientes que se distribuyen según una función de probabilidad (20). Una regla general para los coeficientes de regresión aleatorios es que solo pueden ser considerados aleatorios en el nivel superior en el que han sido medidos.



## 1.4 Campos de aplicación del análisis multinivel:

Las ciencias sociales y de la educación han sido las pioneras en la investigación de la relación entre los individuos y el marco contextual en el que desarrollan sus vidas.

Como dijo Moellering en 1972 (25) “La Sociedad humana está organizada como modelo general en jerarquías anidadas”

El uso del análisis multinivel ha sido progresivo. Si bien en la década de los ochenta podían encontrarse algunas publicaciones en las que se refiriera el uso de tales modelos, es al final de esta década cuando se produce una gran abundancia de las mismas (2). Ver figura 7.

Como ya se ha dicho, el análisis multinivel se aplica en múltiples campos como en educación (26-32), demografía (33-40), sociología y sociología y salud (33;41-45) y, por supuesto, en el campo que nos interesa, que es el de las ciencias de la salud.

Si realizamos una búsqueda en Pubmed con las siguientes palabras clave:

*"multilevel model" OR "multilevel analysis" OR "hierarchical model" OR "multilevel modelling" OR "hierarchical linear model" OR "multi-level analysis" OR "multi-level model" OR "multi-level modeling" OR "mixed model" OR "mixed linear model" OR "mixed no linear models" OR "random effects model" OR "random coefficient models" OR "random coefficient model" OR "panel data model"*

obtenemos los siguientes resultados:

Localizados 15204 artículos, 15160 con abstract disponible en el periodo 1966-2012.

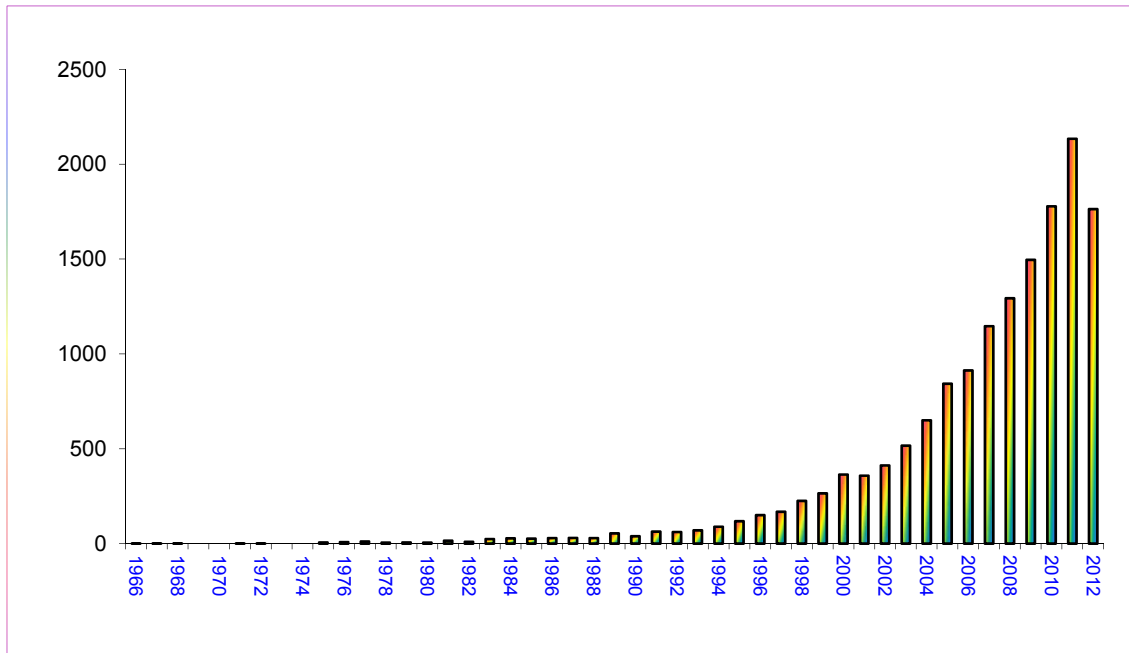


Fig7. Gráfico de publicaciones en el periodo 1966-2012

Observamos que el crecimiento de las publicaciones en los últimos 12 años ha sido exponencial, siendo 13666 de estas publicaciones correspondientes al periodo 2000-2012. Y de estos, 9747 son de investigaciones en humanos. El 26.46% de estas publicaciones son estudios con individuos menores de 18 años. Solo 31 del total de estas publicaciones son en español.

Dentro del campo de la investigación en salud pública, existe un claro interés en ver qué papel juegan los factores ambientales y su influencia en la susceptibilidad individual a una enfermedad (22). Muchos de los datos observados en las ciencias biomédicas, tienen una estructura jerárquica o anidada. Por ejemplo en los estudios de herencia en humanos y animales existe una jerarquía natural, ya que la descendencia se agrupa en familias. La descendencia de los mismos padres tiende a ser física y mentalmente más parecida que individuos elegidos al azar en una población grande. Por ejemplo los niños de la misma familia pueden tender a ser bajos, quizás porque sus padres son

bajos o puede que sea debido a un entorno común empobrecido. Muchos experimentos también crean datos estructurados en jerarquías, por ejemplo ensayos clínicos llevados a cabo en varios centros elegidos aleatoriamente (3). El análisis multinivel se ha usado en estudios sobre cáncer (46-49), enfermedades cardiovasculares (50-53), estudios de genética (54-56), enfermedades del aparato respiratorio (57-59) y en otras especialidades médicas. Esta tesis dirige su uso a la investigación de las sibilancias. Existen pocos estudios que hayan utilizado el análisis multinivel para el estudio del asma o sibilancias (43;58;60-71), no habiendo encontrado ninguno que estudie y analice los factores de riesgo de las sibilancias recurrentes en el primer año de vida.

En la tabla 3 se resume el tipo de modelo, el número de niveles y las unidades que se eligen en cada nivel de algunos de los estudios citados en el párrafo anterior.

<b>Referencia bibliográfica</b>	<b>Tipo de modelo</b>	<b>Nº de niveles</b>	<b>Unidades de análisis</b>
<b>39</b>	Poisson	2	1º nivel: municipios 2º nivel: comunidades
<b>40</b>	Logístico	3	1º nivel: pacientes 2º nivel: plan médico 3º nivel: comunidad
<b>41</b>	Logístico	4	1º nivel: pacientes 2º nivel: atención primaria 3º nivel: hospital 4º nivel: área de salud
<b>42</b>	Medidas repetidas	2	1º nivel: medidas realizadas 2º nivel: pacientes
<b>43</b>	Lineal	3	1º nivel: pacientes 2º nivel: atención primaria 3º nivel: países
<b>44</b>	Lineal	2	1º nivel: pacientes 2º nivel: médicos
<b>45</b>	Lineal	2	1º nivel: pacientes 2º nivel: distritos censales
<b>46</b>	Logístico	2	1º nivel: pacientes 2º nivel: vecindario
<b>48</b>	Logístico	2	1º nivel: pacientes 2º nivel: ciudad
<b>49</b>	Medidas repetidas	2	1º nivel: medidas realizadas 2º nivel: pacientes
<b>50</b>	Logístico	3	1º nivel: individuo 2º nivel: área 3º nivel: país
<b>51</b>	Logístico	2	1º nivel: niño 2º nivel: escuela
<b>35</b>	Multinomial	2	1º nivel: niño 2º nivel: tiempo
<b>53</b>	Logístico	2	1º nivel: enfermedad 2º nivel: paciente
<b>54</b>	Logístico	2	1º nivel: paciente 2º nivel: médico

Tabla 3. Resumen del tipo de modelo, el número de niveles y las unidades que se eligen en cada nivel.

## 1.5 Modelo básico de regresión multinivel: 2 niveles

### 1.5.1 Del Modelo clásico al modelo multinivel

El principio básico más importante que debemos considerar es que el análisis multinivel lineal puede ser visto como una extensión del análisis de regresión lineal. Vamos a partir del modelo clásico de regresión, con variable dependiente continua y aplicado, por ejemplo, a un solo contexto para familiarizarnos con la notación algebraica estándar.

Supongamos que estamos interesados en estudiar la respuesta de los individuos en relación a una variable explicativa, como puede ser la edad, con el siguiente modelo de regresión:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i / \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

Donde  $y_i$  es la respuesta para el  $i$ -ésimo individuo,  $X_{1i}$  es el valor de la variable explicativa edad para el  $i$ -ésimo individuo,  $\beta_0$  y  $\beta_1$  son los coeficientes de regresión que representan el intercepto (es el valor medio de  $y$  para individuos con  $X = 0$ ) y la pendiente (es el incremento que se produce en el valor de  $y$  cuando aumenta en una unidad el valor de  $x$ ), respectivamente; La expresión  $\beta_0 + \beta_1 x_{1i}$  es conocida como la *parte fija* del modelo y  $\varepsilon_i$  es el error aleatorio o residuo (es lo que el valor de  $y_i$  se aleja del resultado esperado de los individuos que tienen su mismo valor de  $X$ ). Se asume el supuesto de homocedasticidad, Es decir, la varianza de los residuos es siempre la misma e independiente de los valores de  $X$ . El residuo es la parte del resultado  $y_i$  que no podemos explicar mediante la parte fija de la regresión.

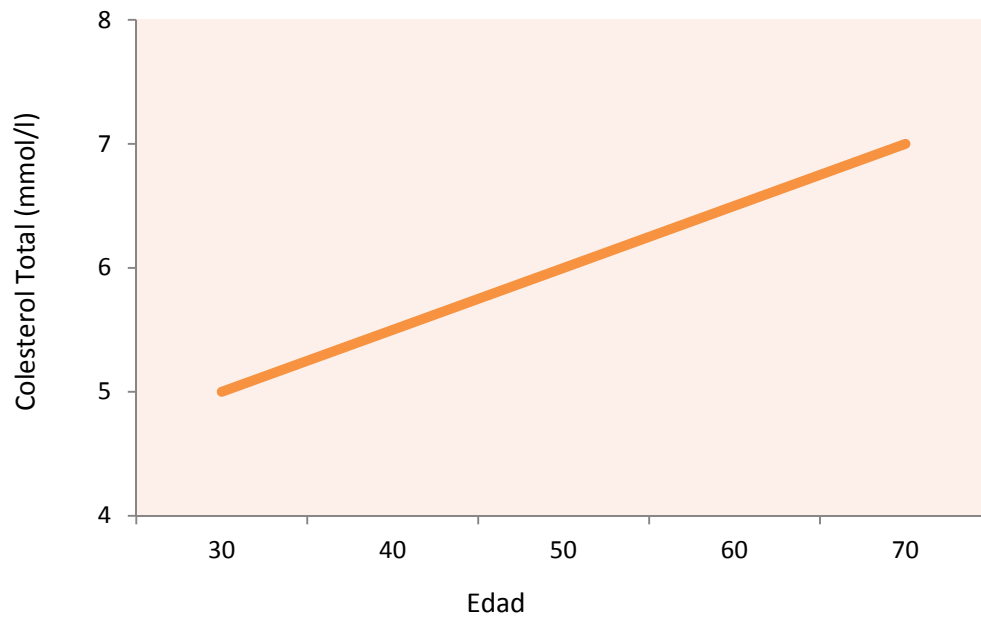


Fig. 8. Ilustración de un análisis de regresión lineal, de la relación entre colesterol total y edad.

Supongamos ahora que en nuestro estudio tenemos pacientes que “pertenecen” al mismo médico. Es lógico asumir, que las características de los pacientes que visitan a un mismo médico, difieren de aquellos pacientes a los que les corresponde otro médico. Estas diferencias pueden ser debidas al área de salud donde ejerce el médico, a ciertas características personales del médico, etc.

(2)

$$y_i = \beta_0 + \beta_1 edad_i + \beta_2 MD_{1i} + \beta_3 MD_{2i} + \beta_4 MD_{3i} + \dots + \beta_m MD_{m-1i} + \varepsilon_i$$

$i=1, \dots, n$

donde  $(\beta_2, \dots, \beta_m)$  son los coeficientes de regresión de las variables dummy que representan a los distintos médicos, siendo  $m$  el número de médicos.

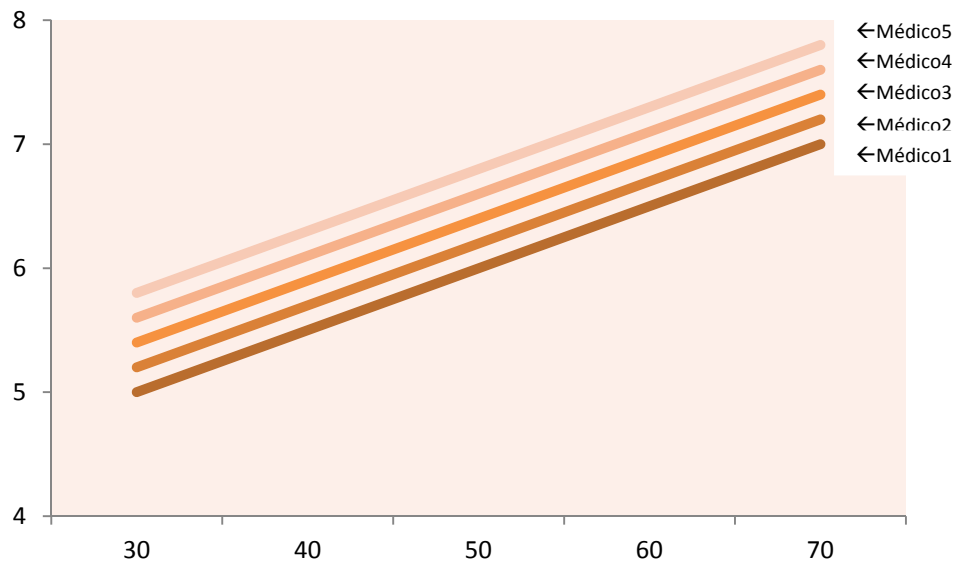


Fig. 9. Ilustración del análisis de regresión lineal, de la relación entre colesterol total y edad, corregido por la variable dicotómica médico.

Resumiendo, si tenemos un estudio con 12 médicos, es necesario estimar 11 coeficientes de regresión, en la regresión lineal. Esto supone una pérdida de poder y eficiencia del modelo, porque no existe un interés real en la diferencias de la variable dependiente respecto la variable médico y no parece muy práctico estimar una ecuación de regresión para cada médico como se ve en la figura anterior.

(3)

$$y_{i1} = \beta_{01} + \beta_{11}edad_{1i} + \varepsilon_{i1}$$

$$y_{i2} = \beta_{02} + \beta_{12}edad_{2i} + \varepsilon_{i2}$$

.

.

$$y_{i12} = \beta_{012} + \beta_{112}edad_{12i} + \varepsilon_{i12}$$

Por lo tanto una forma más eficiente de corregir por el posible efecto de la variable médico es utilizar el análisis multinivel, es decir, usar la variable médico como 2° nivel o nivel macro. Usando el vocabulario propio del análisis multinivel, decimos que las observaciones que se hacen a nivel de paciente, se encuentran anidadas dentro del nivel médico, Figura 10. Con el uso del análisis multinivel, no es necesario estimar cada intercepto por separado ( $\beta_{01}$ ,  $\beta_{02}$ , ...,  $\beta_{012}$ ), lo que se estima es el valor medio del intercepto más la varianza de los interceptos en el segundo nivel (efecto aleatorio).

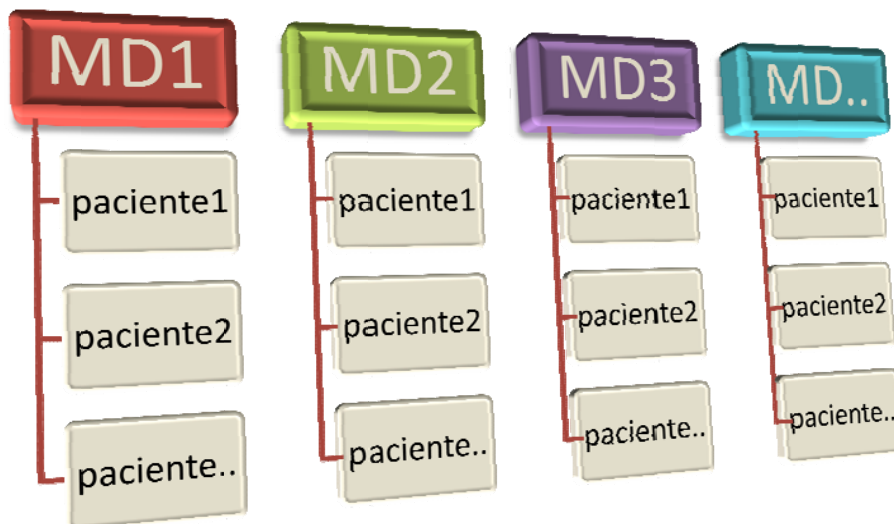


Fig. 10. Estructura multinivel de 2 niveles.

*Notación:* En general si un ítem tiene dos subíndices  $ij$ , indica que varía de individuo en individuo dentro de un mismo grupo. Si un ítem tiene únicamente un subíndice  $j$ , solamente varía entre los grupos  $j$ , pero tiene el mismo valor para todos los individuos de un mismo grupo. Y si un ítem no tiene ningún subíndice es que es constante a través de todos los individuos y grupos.



Si establecemos una jerarquía de dos niveles, donde los individuos del nivel 1 se encuentran agrupados dentro de unidades del nivel 2, se incorpora un nuevo subíndice ( $j$ ) en el modelo. La ecuación de regresión para el  $i$ -ésimo individuo dentro del  $j$ -ésimo grupo (elementos de segundo nivel), vendrá dada por:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \quad (4)$$

Esta variabilidad en el segundo nivel es la característica principal de los modelos multinivel, dando lugar a un modelo diferente al modelo tradicional (1). En la ecuación (4),  $y_{ij}$  es la respuesta para el paciente  $i$  en el médico  $j$ ;  $x_{ij}$  es el valor de la variable explicativa (por ejemplo edad) para el paciente  $i$  en el médico  $j$ ,  $\beta_{0j}$  y  $\beta_{1j}$  son el intercepto y la pendiente, respectivamente, en el médico  $j$  y  $\varepsilon_{ij}$  es la variación residual del  $i$ -ésimo paciente en el  $j$ -ésimo médico, es decir, lo que se aparta de la respuesta esperada según su edad. Se asume que tiene una distribución normal, con media 0 y varianza constante,  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ , es decir, asumimos que la varianza es homogénea para todos los médicos.

El objetivo de la modelización multinivel es considerar dentro de un mismo modelo los distintos niveles de la estructura jerárquica y conocer la variabilidad debida al segundo nivel, al ser los coeficientes  $\beta_{0j}$  y  $\beta_{1j}$  valores propios de cada médico. Estamos suponiendo la variabilidad entre contextos, con lo que estos coeficientes pasan a ser variables aleatorias en el nivel macro, y los podemos expresar como:

$$\beta_{0j} = \beta_0 + u_{0j} \quad (5)$$

$$\beta_{1j} = \beta_1 + u_{1j} \quad (6)$$

El intercepto  $\beta_{0j}$ , se descompone en  $\beta_0$  que es la respuesta media en la población de médicos y representa la parte fija de la ecuación y en  $u_{0j}$  que refleja la varianza en los interceptos de los diferentes médicos y representa la parte aleatoria del modelo en el segundo nivel.

De igual forma,  $\beta_{1j}$  puede tener variabilidad propia, y se descompone en  $\beta_1$  que es la pendiente para la población de médicos y en  $u_{1j}$  que representa la variabilidad de los coeficientes de regresión para edad para los distintos médicos, y es la parte aleatoria del modelo en el segundo nivel vinculada a la pendiente. Entonces:

$$E(\beta_{0j}) = \beta_0 \qquad E(\beta_{1j}) = \beta_1 \qquad (7)$$

$$E(u_{0j}) = E(u_{1j}) = 0$$

$$V(u_{0j}) = \sigma^2_{u0} \qquad V(u_{1j}) = \sigma^2_{u1} \qquad V(\varepsilon_{ij}) = \sigma_\varepsilon^2$$

$$\text{cov}(u_{0j}, u_{1j}) = \sigma_{u01}$$

$\sigma^2_{u0}$ ,  $\sigma^2_{u1}$  y  $\sigma_{u01}$  son la varianza entre los puntos de corte, pendientes y la covariancia entre el punto de corte y pendiente, respectivamente. La idea de variación entre contextos es esencial en los modelos multinivel, si no existiera no sería necesario el empleo de estos modelos.

La siguiente figura muestra el significado de los siguientes valores que pueden adoptar las varianzas y covarianzas del modelo:

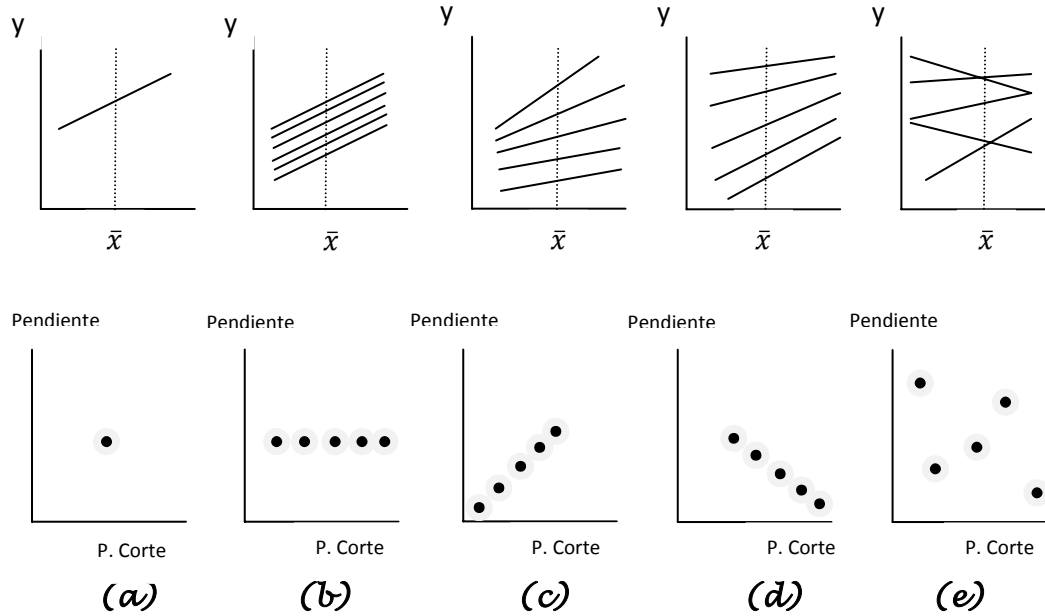


Figura 11. Variación en el nivel macro.

Si retomamos el ejemplo de los pacientes anidados en médicos que hemos estado usando hasta ahora vemos que, en la figura 11.a todos los médicos comparten la misma ecuación, es decir, para todos los médicos la relación entre la edad y la variable respuesta es la misma. Tienen la misma recta de regresión por lo tanto:

$$\sigma^2_{u0} = \sigma^2_{u1} = \sigma_{u01} = 0$$

En la figura 11.b los médicos comparten la misma pendiente, pero difieren en el intercepto, lo que significa que:

$$\sigma^2_{u1} = 0 ; \sigma^2_{u0} > 0$$

En la figura 11.c se observa que cuanto mayor es la media del médico mayor es la pendiente y también difieren en el punto de corte por lo que:

$$\sigma^2_{u1} > 0 ; \sigma^2_{u0} > 0$$

Además cuanto mayor es el punto de corte mayor es la pendiente lo que significa que:

$$\sigma_{u01} > 0$$

En la figura 11.d encontremos el caso contrario por lo tanto:

$$\sigma^2_{u1} > 0 ; \sigma^2_{u0} > 0 ; \sigma_{u01} < 0$$


La última ilustración 11.e muestra el caso en el que no hay relación alguna entre el punto de corte y la pendiente, luego:

$$\sigma^2_{u1} > 0 ; \sigma^2_{u0} > 0 ; \sigma_{u01} = 0$$


Resumiendo, el modelo de regresión representado en (4), incluyendo la estructura jerárquica de 2 niveles, se puede expresar como la suma de una *parte fija* y una *parte aleatoria*, dando lugar al modelo general (ecuación 5.8) que es un modelo llamado **modelo completamente aleatorio**:

(8)

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + ( u_{1j} x_{ij} + u_{0j} + \varepsilon_{ij} )$$



Parte fija



Parte aleatoria

Los coeficientes de la regresión  $\beta_0$  y  $\beta_1$  son usualmente referidos como parámetros fijos del modelo y el conjunto de varianzas y covarianzas como los parámetros aleatorios (72). En la ecuación 8 se contempla la posibilidad de variación tanto en puntos de corte como en pendiente, sin embargo hay modelos más sencillos y también más complejos.

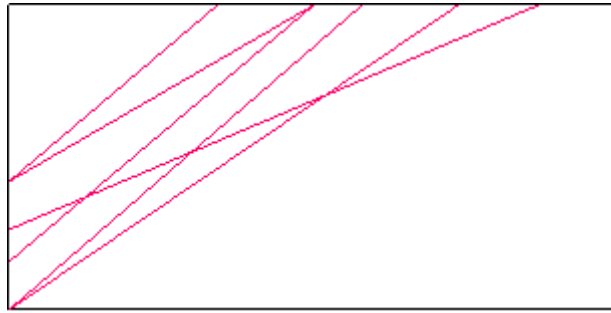


Figura 12. Modelo completamente aleatorio.

### 1.5.2 Modelos que se derivan del modelo completamente aleatorizado:

MODELO NULO: o modelo con solo intercepto, es un caso particular del 1° modelo, por lo tanto, es un modelo de interceptos aleatorios en el que no existe ninguna variable predictora.

(9)

**1° Nivel**

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

**2° Nivel**

$$\beta_{0j} = \beta_0 + u_{0j}$$



**Modelo Nulo:**

$$y_{ij} = \beta_0 + (u_{0j} + \varepsilon_{ij})$$

El modelo nulo no sirve para explicar la varianza de la variable dependiente Y, este solo descompone la varianza en dos componentes independientes:  $\sigma_\varepsilon^2$ , que es la varianza del error del 1° nivel, y  $\sigma^2_{u_0}$  que es la varianza del error del nivel macro o 2° nivel (12). Este es un modelo muy importante porque provee una partición básica de la variabilidad de los datos entre los 2 niveles (73). Usando este modelo podemos definir la correlación intraclase (CCI,  $\rho$ ) o el coeficiente de

partición de la varianza (VPC en la literatura inglesa) con la siguiente fórmula:

(10)

$$\rho = \frac{\sigma_{\mu_0}^2}{\sigma_{\mu_0}^2 + \sigma_{\varepsilon}^2}$$

Como ya se ha visto en los conceptos básicos, el coeficiente de correlación intraclase indica la proporción de la varianza explicada por la estructura de grupo en la población (74). En nuestro ejemplo de pacientes y médicos, el CCI indica la correlación de las observaciones de los pacientes pertenecientes al mismo médico, y también mide la dependencia de las observaciones de los pacientes dentro del grupo de médicos. El CCI toma valores entre 0 y 1, tal que, si es 0 no hay diferencias entre los elementos del 2º nivel ( $\sigma_{u0}^2 = 0$ ) y si el CCI es 1, no ha diferencias dentro de cada grupo ( $Y_{ij}=Y_j$  o  $\sigma_{\varepsilon}^2 = 0$ ).

Además de para el cálculo del CCI debemos usar este modelo como estrategia metodológica. Como una primera aproximación al análisis multinivel es necesario ajustar el modelo nulo para ver si la varianza es significativamente distinta de cero, ya que si no es así no tendría sentido incluir variables explicativas en el modelo, porque no habría nada que explicar (2;73).

Lo normal, es que exista varianza significativa tanto en el primer nivel como en el segundo, por ello nuestra labor será introducir variables en el modelo que reduzcan la varianza no explicada.

**MODELO DE INTERCEPTOS ALEATORIOS:** en el que se plantea que existe variación entre los puntos de corte  $\sigma_{u0}^2 > 0$  pero no hay variación entre pendientes  $\sigma_{u1}^2 = 0$  y  $\beta_1$  se asume que es el mismo para cada grupo (figura 13); El modelo multinivel más simple con una variable explicativa es:

(11)

**1° Nivel o nivel micro**    **2° Nivel o nivel macro**

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1$$



**Modelo de varianza de efectos aleatorios (variance components model):**

$$y_{ij} = \beta_0 + \beta_1x_{ij} + (u_{0j} + \varepsilon_{ij})$$

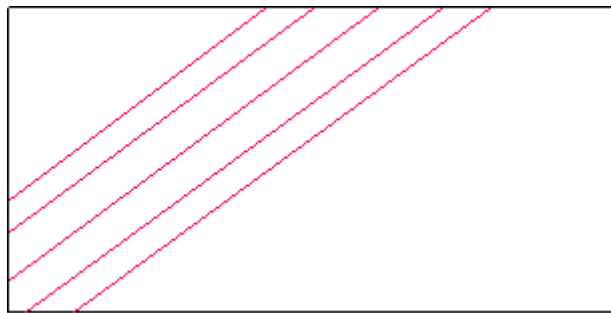


Figura 13. Representación modelo de interceptos aleatorios.

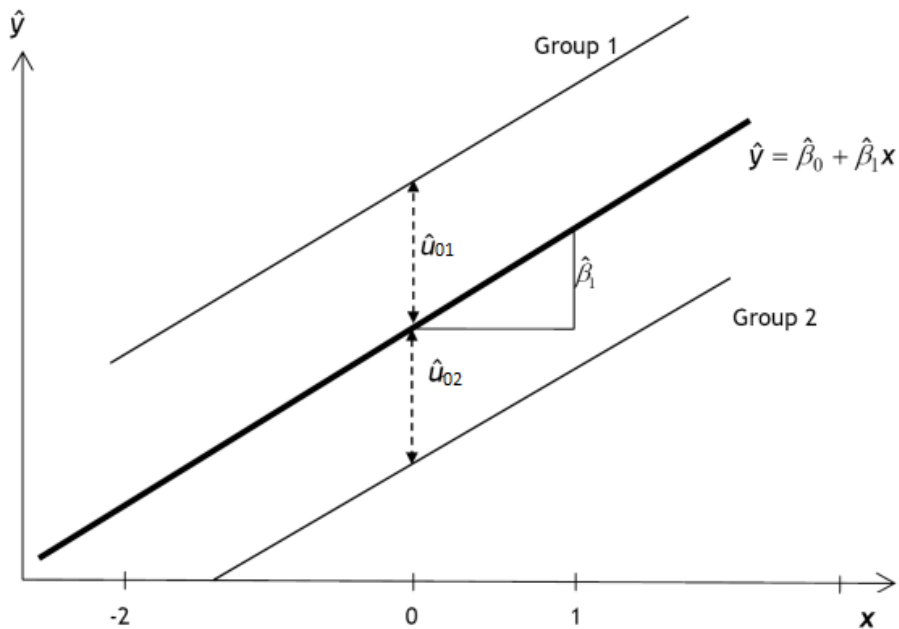


Figura 14. Líneas de predicción para un modelo de intercepto aleatorio para dos unidades de nivel 2

MODELO DE PENDIENTES ALEATORIAS (solo): en el que se plantea que existe variación entre las pendientes  $\sigma^2_{u1} > 0$  pero no hay variación entre los puntos de corte  $\sigma^2_{u0} = 0$ , por lo tanto;

(12)

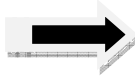
**1° Nivel**

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$$

**2° Nivel**

$$\beta_{0j} = \beta_0$$

$$\beta_{1j} = \beta_1 + u_{1j}$$



**Modelo de pendientes aleatorias:**

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{1j} x_{ij} + \varepsilon_{ij})$$

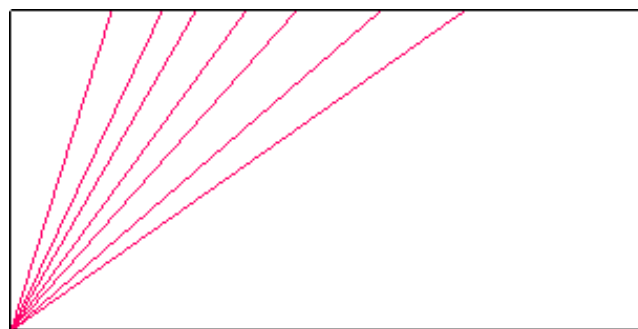


Figura 15. Representación modelo de pendientes aleatorias.

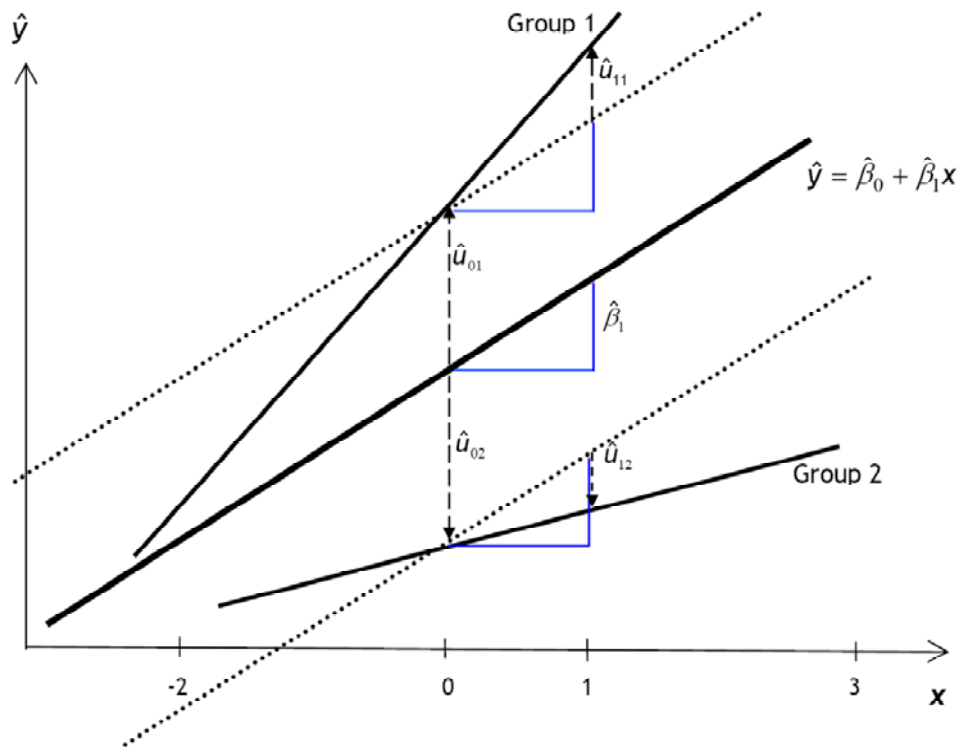


Fig16. Líneas de predicción de un modelo de pendientes aleatorias, para dos grupos.



### 1.5.3 Asunciones en el análisis multinivel

Como ya se ha dicho los modelos multinivel son una extensión de los modelos de regresión tradicionales, por ello, comparten básicamente los supuestos de aplicación (2;20), es decir, todas las asunciones para análisis de regresión lineal 'estándar' también se aplican o son válidas para análisis multinivel. Por lo tanto una variable dependiente se debería distribuir Normalmente. Al igual que en el Análisis de regresión lineal, se puede comprobar la normalidad por medio de gráficos de normalidad (vea para detalles: Goldstein y Healy 1995; Goldstein y Spiegelhalter, 1996.) La asunción de independencia de las observaciones no se aplica en estos modelos, porque la razón para realizar un análisis multinivel en primer lugar, es que las observaciones de los datos que van a ser analizados están correlacionadas.

Una asunción adicional para análisis de multinivel, y que se ha mencionado anteriormente, es que los interceptos y las pendientes aleatorias deben estar normalmente distribuidos. Esta asunción puede relajarse con distribuciones distintas de la distribución normal, aunque puede complicar el procedimiento de estimación.

Además de verificar las asunciones del análisis de multinivel, puede ser también importante comprobar si existen outliers en nuestros datos. Debido a la estructura multinivel de los datos, podemos encontrar outliers en los diferentes niveles. Por ejemplo, las observaciones de pacientes pueden influir en la relación total que se analiza o pueden ser outliers en el nivel paciente. Por otra parte una única observación de un paciente, puede ser también un outlier para un médico en particular a quien ese paciente 'pertenece', en otras palabras, la observación paciente puede ser un outlier al nivel de médico (75-77).

Snijders y Brosker (73) plantean las siguientes preguntas, que nos ayudaran a comprobar los supuestos:

- i. La parte fija del modelo ¿contiene las variables adecuadas?
- ii. La parte aleatoria del modelo ¿contiene las variables adecuadas?
- iii. Los residuos del primer nivel, ¿están normalmente distribuidos?
- iv. Los coeficientes aleatorios del segundo nivel, ¿están normalmente distribuidos?
- v. Los coeficientes aleatorios del segundo nivel, ¿tienen una matriz de varianzas-covarianzas constante?

### 5.4 Significación y ajuste de los modelos

Como en cualquier otro modelo de regresión, la interpretación de los modelos multinivel depende de (2;74):

- ✓ La significación de los coeficientes de regresión.
- ✓ Como de bien el modelo ajusta los datos.

La teoría estadística que hay detrás del modelo de regresión multinivel es complicada. A partir de los datos observados, se quieren estimar los parámetros del modelo multinivel:

1. Los coeficientes de regresión
2. Los componentes de la varianza.

Los estimadores más utilizados en el análisis de regresión multinivel son los estimadores de máxima verosimilitud (Maximum likelihood estimators, (MLE) (74;78). El objetivo de la estimación de máxima verosimilitud es encontrar un estimador del parámetro, dependiente de los datos conocidos, más cercano al verdadero valor del parámetro. Es

decir, dado un conjunto de datos y el modelo probabilístico subyacente, la estimación de máxima verosimilitud, toma el valor del parámetro que da lugar a la distribución con la que los datos son más probables.

Los estimadores de ML, nos van a ayudar a contestar a las siguientes preguntas:

### 1. *¿Este predictor es estadísticamente significativo?*

El procedimiento de máxima verosimilitud, produce, errores estándar, para la mayoría de las estimaciones. La significación de un predictor viene dada por la ratio entre el estimador del parámetro y su error típico. Este test es conocido como el Test de Wald (79). Esta distribución del estadístico de Wald sirve para aceptar o rechazar la hipótesis nula establecida sobre el estimador del parámetro  $\beta$  ( $H_0: \beta = 0$ ) Se verifica que:

$$\frac{\hat{\beta}}{\hat{s}_{\beta}} \sim N(0,1) \text{ o lo que es equivalente } \left(\frac{\hat{\beta}}{\hat{s}_{\beta}}\right)^2 \sim \chi^2_{1g.l.}$$

En la práctica, una regla general para determinar la significación de un predictor es que si,

(13)

$$Z = \frac{\text{parámetro}}{\text{error estandar}} > 2 \rightarrow p < 0.05$$

**2. ¿Aporta este modelo (con  $x$  predictores) información significativa comparado con el modelo nulo o con otro modelo alternativo?**

Los procedimientos de máxima verosimilitud también producen un estadístico llamado *Deviance*. Este estadístico indica como de bien ajusta el modelo a los datos. Si llamamos  $L_1$ , al valor del máximo de la función de verosimilitud (likelihood) en la estimación de los parámetros del modelo 1, entonces se define la deviance:

$$Dev = -2 \ln(L_1) \equiv -2 \log(\text{likelihood}) \quad (14)$$

Como ya se ha visto anteriormente, si en el modelo nulo no hay varianza estadísticamente distinta de 0 en los niveles contemplados, ningún modelo que se derive de este añadiendo variable explicativas, mejorará el ajuste, ya que, la varianza del intercepto no es significativamente distinta de 0. Si dos modelos están anidados, es decir, un modelo se obtiene a partir de otro más general eliminando parámetros de este último, entonces podemos compararlos. Para llevar a cabo la comparación hacemos uso de la deviance de cada modelo, ya que, la diferencia de las deviances se distribuye como una chi-cuadrado con los grados de libertad iguales a la diferencia del número de parámetros estimados en los modelos que estamos comparando, bajo la hipótesis nula de que ambos modelos son iguales.

**Ejemplo:** Supongamos que tenemos dos modelos  $M_1$  con  $m_1$  parámetros y  $M_2$  con  $m_2$  parámetros.

(15)

$$D_1 = -2 \ln(L_1) \text{ y } D_2 = -2 \ln(L_2)$$

$$D = -2 \ln \frac{L_2}{L_1} \sim \chi_{m_2 - m_1}^2$$

Si la diferencia es significativa, nos quedamos con el  $M_2$ , y sino con  $M_1$ , es decir, los parámetros que aparecen en el modelo 2 y no en el modelo 1 son significativamente distintos de 0 y, por tanto, las correspondientes variables de ajuste estarán asociadas de forma significativa con la variable respuesta.

### 3. ¿Cómo comparamos modelos no anidados?(74)

Si los modelos que queremos comparar no están anidados, el principio de parsimonia nos indica que deberíamos escoger el modelo más simple. Pero también podemos usar el *Criterio de información de Akaike* (80), conocido como AIC. Para un modelo de regresión multinivel el AIC se calcula a partir del valor de la deviance (Dev), y el número de parámetros estimados (q):

(16)

$$AIC = Dev + 2q$$

El AIC, es un índice de ajuste general, que asume que los modelos que se están comparando ajustan el mismo conjunto de datos, y usan un mismo método de ajuste. Un criterio de ajuste similar es el *Criterio de información bayesiana de Schwarz* (80;81), conocido como BIC y que viene dado por:

(17)

$$BIC = Dev + q \cdot LN(N)$$

Al igual que la deviance, cuanto menor es el valor de AIC y del BIC mejor es el ajuste. Tanto el AIC como el BIC, penalizan a los modelos con un elevado número de parámetros, pero el BIC impone una mayor penalización para la mayoría de tamaños muestrales, por ello, para los modelos multinivel con diferentes tamaños muestrales y varios niveles, y por lo tanto, el criterio AIC es más recomendable que el BIC.

### **1.5.5 Métodos de estimación de los parámetros**

Como ya se ha explicado anteriormente, en un modelo multinivel hay dos tipos de parámetros: fijos y aleatorios. Los parámetros fijos corresponden a los efectos medios en la población, y son las pendientes y el intercepto. Los aleatorios corresponden a las varianzas y covarianzas de todos los niveles (2).

A la hora de estimar dichos parámetros, debemos distinguir entre métodos y algoritmos de estimación. Un método de estimación consiste en un conjunto de reglas y principios cuya aplicación da lugar a una ecuación o ecuaciones que ponen en relación los datos con el parámetro buscado. Los algoritmos son métodos matemáticos, que por medio de iteraciones sucesivas, permiten obtener soluciones para dichas ecuaciones.

MÉTODOS DE ESTIMACIÓN		ALGORITMOS DE ESTIMACIÓN
Máxima verosimilitud(ML)*	Tanto los coeficientes de regresión como los componentes de la varianza se incluyen en la función de verosimilitud(12).	<i>Fisher Scoring</i> Se basa en la mejor aproximación de la función de verosimilitud y se puede usar con ML y REML.
Máxima verosimilitud restringida (REML)	Solo los componentes de la varianza se incluyen en la función de verosimilitud.	<i>Expectation-Maximization (EM)</i> Calcula estimaciones máximo-verosímiles en casos en los que existen datos perdidos. Pero es muy lento. Se usa para REML.
		<i>Iterative Generalized Least Squares (IGLS)</i> Refinamiento secuencial del procedimiento basado en Mínimos Cuadrados Generalizados. Produce estimadores sesgados de los parámetros aleatorios debido a que no tiene en cuenta la varianza muestral de la parte fija del modelo.
		<i>Restricted IGLS (RIGLS)</i> Para estimar los parámetros usa el método REML, y se pueden conseguir estimadores insesgados. Aconsejable para muestras pequeñas.
Máxima verosimilitud(ML)*	Para modelos no lineales con variables Dicotómicas:	<i>Cuasi-verosimilitud Marginal (MQL(82) es)</i> Linealiza la partes fija del valor predecido de la variable dependiente.
	utilizan expansiones de la serie de Taylor para linealizar el modelo multinivel,	<i>Cuasi-verosimilitud Penalizada (PQL )</i> Linealiza ambas partes (fija y aleatoria) del valor predecido de la variable dependiente.
Estimación Bayesiana	Combina la distribución a priori (integra conocimientos previos de los parámetros) con la función de verosimilitud, y produce una distribución a posteriori, que describe la incertidumbre de los parámetros después de observar los datos.	<i>Método Monte Carlo por Cadena de Markov</i> MCMC, Métodos de Monte Carlo por Cadenas de Markov. Simulación de cadenas de Markov convergentes hacia la distribución a posteriori de los parámetros.

Tabla 4. Métodos y algoritmos de estimación

\*En la práctica la diferencia entre ML y REML no es muy grande, pero cuando el número de unidades de nivel macro es inferior a 30, es preferible el método REML (83).

### 1.5.6 Varianza

En la regresión múltiple, la varianza explicada mide la proporción de la varianza total de la variable respuesta (Y) que es explicada por la relación lineal que existe entre Y y las variables explicativas del modelo ( $X_i$ ). Para medir dicha proporción usamos el llamado coeficiente de determinación, más conocido por  $R^2$ , que toma valores entre 0 y 1. Un valor próximo a 1 se interpreta como un buen ajuste del modelo.

Pero ¿cómo se mide en una regresión multinivel, la varianza explicada por el modelo?

Los modelos multinivel permiten dividir la varianza total en diferentes componentes de variación según los distintos niveles de agrupación de los datos (84). Por ejemplo en la investigación del asma en la infancia podemos considerar el estudio de los factores de riesgo de la proporción de asmáticos en la infancia, y en este caso podríamos dividir la varianza en variación entre y variación dentro de la unidades del nivel superior (ciudad de residencia, por ejemplo).

Supongamos que partimos de un modelo de dos niveles de intercepto aleatorio y con un predictor  $x_{ij}$ , tratado como un efecto fijo:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij} \quad \mathbf{(A)}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\text{var}(u_{0j}) = \sigma^2_{u0} \quad \text{var}(e_{ij}) = \sigma^2_{e0}$$

$$\text{var}(y_{ij} | x_{1ij}) = \sigma^2_{u0} + \sigma^2_{e0}$$



En este apartado vamos a ver una aproximación a los métodos más usados, que son los de Golstein H. et al. (85) y Snijders T.A.B. et al. (73).

#### 1.5.6.1 MÉTODO PROPUESTO POR GOLSTEIN.

En el modelo **A** podemos resumir la ‘importancia’ del nivel centro, como la proporción de la varianza total explicada, que se conoce como “coeficiente de partición o división de la varianza” (variance partition coefficient) VPC(86) y viene dado por la fórmula:

(18)

$$VCP = \frac{\text{varianza residual del 2º nivel}}{\text{varianza residual del 1º nivel} + \text{varianza residual del 2º nivel}}$$

$$\Rightarrow VPC = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2}$$

En el caso de un **modelo de intercepto aleatorio** el VPC, también mide la correlación residual entre las respuestas de dos niños en la misma ciudad de residencia, y en este caso es conocido también como “Coeficiente de Correlación Intraclase,  $\rho$ ”, y se usa comúnmente como una medida de la importancia de considerar que los datos están agrupados o estructurados jerárquicamente.

El VPC es útil en el caso que tengamos un modelo con una única fuente de variación en cada nivel, pero lo es menos, en un **modelo de coeficientes aleatorios**.

Supongamos que tenemos el siguiente modelo:

(19)

<b>1° Nivel</b>	<b>2° Nivel</b>
$y_{ij} = \beta_{0j}x_0 + \beta_{1j}x_{1ij}$	$\beta_{0j} = \beta_0 + u_{0j} + e_{ij}$
$x_0 = 1$	$\beta_{1j} = \beta_1 + u_{1j}$

$$\text{Var}(u_{0j}) = \sigma_{u_0}^2, \quad \text{Var}(u_{1j}) = \sigma_{u_1}^2,$$

$$\text{Cov}(u_{0j}, u_{1j}) = \sigma_{u_01}, \quad \text{Var}(e_{0ij}) = \sigma_e^2$$

$$VPC = \frac{\sigma_{u_0}^2 + 2\sigma_{u_01}x_{1ij} + \sigma_{u_01}^2x_{1ij}^2}{\sigma_{u_0}^2 + 2\sigma_{u_01}x_{1ij} + \sigma_e^2}$$

En este caso el VPC no es igual a la correlación entre unidades. Dados dos niños  $i_{1j}$  e  $i_{2j}$ , la correlación entre  $x_{i_{1j}}$  y  $x_{i_{2j}}$  viene dada por:

(20)

$$\frac{(\sigma_{u_0}^2 + \sigma_{u_01}(x_{1i_{1j}} + x_{1i_{2j}}) + \sigma_{u_1}^2x_{1i_{1j}}x_{1i_{2j}})}{\sqrt{(\sigma_{u_0}^2 + 2\sigma_{u_01}x_{1i_{1j}} + \sigma_{u_1}^2x_{1i_{1j}}^2 + \sigma_e^2)(\sigma_{u_0}^2 + 2\sigma_{u_01}x_{1i_{2j}} + \sigma_{u_1}^2x_{1i_{2j}}^2 + \sigma_e^2)}}$$

1.5.6.2 MÉTODO PROPUESTO POR SNIJDERS.

Snijders y Bosker (73) proponen otra definición de la proporción de varianza explicada, y la llaman *Reducción de la proporción del error de la predicción*.

En el marco de los modelos multinivel de 2 niveles, se puede elegir entre predecir el valor de Y para un individuo en un grupo, o predecir el valor medio de Y para un grupo. Lo que da lugar a dos formas de medir la proporción de varianza explicada:

- i. *Reducción de la proporción del error de predicción de un valor individual.*

Cuando desconocemos el valor de x para un individuo, la mejor predicción es  $E(Y_i)$  y la varianza del error de predicción es  $\text{var}(Y_i)$ . En cambio cuando conocemos el valor de x la mejor predicción es:

$$\sum_{h=0}^q \beta_h x_{hij} \quad \text{y el error de predicción es} \quad Y_{ij} - \sum_{h=0}^q \beta_h x_{hij} = u_{0j} + e_{ij}$$

Por lo tanto la varianza del error de predicción es:

(21)

$$\text{var} \left( Y_{ij} - \sum_{h=0}^q \beta_h x_{hij} \right) = \text{var}(u_{0j} + e_{ij}) = \sigma_{\mu_0}^2 + \sigma_e^2$$

Y La *reducción de la proporción de la varianza del error de predicción* para el nivel 1 será:

(22)

$$R_1^2 = 1 - \frac{\text{var}(Y_{ij} - \sum_{h=0}^q \beta_h x_{hij})}{\text{var}(Y_{ij})} = 1 - \frac{\sigma_{\mu_0}^2 + \sigma_e^2}{\text{var}(Y_{ij})}$$

Lo normal es desconocer el valor de estos parámetros, por ello la mejor forma para estimar  $R_1^2$  es usar las estimaciones de de dichos parámetros.

(23)

$$\hat{R}_1^2 = 1 - \frac{(\hat{\sigma}_{\mu_0}^2 + \hat{\sigma}_e^2)_N}{(\hat{\sigma}_{\mu_0}^2 + \hat{\sigma}_e^2)_A}$$

N = varianza del modelo nulo

A = varianza del modelo alternativo (con al menos un predictor)

- ii. *Reducción de la proporción del error de predicción de la medida de grupo*

La proporción de varianza explicada en el nivel 2 se puede definir como, la reducción en la proporción de la varianza del error de predicción de la media  $\bar{Y}_j$ , de una unidad de nivel 2 elegida al azar.

Supongamos que partimos del modelo de 2 niveles del apartado anterior A.

Si conocemos los valores de todos los predictores  $x_{hij}$  para todos los  $i$  del grupo  $j$ , entonces la mejor predicción de  $\bar{Y}_j$  es el valor de la regresión  $\sum_{h=0}^q \beta_h x_{h,j}$  y la varianza del error de la predicción:

(24)

$$var\left(\bar{Y}_j - \sum_{h=0}^q \beta_h x_{h,j}\right) = \sigma_{\mu_0}^2 + \frac{\sigma_e^2}{n_j}$$

/  $n_j$  el número de unidades de nivel 1 en el grupo  $j$ .

Entonces a partir de estos datos definimos la reducción en la proporción de la varianza del error de predicción de  $\bar{Y}_j$  como:

(25)

$$R_2^2 = 1 - \frac{\text{var}(\bar{Y}_j - \sum_{h=0}^q \beta_h x_{h,j})}{\text{var}(\bar{Y}_j)} =$$

Y su estimación:

(26)

$$\hat{R}_2^2 = 1 - \frac{(\hat{\sigma}_{\mu_0}^2 + \frac{\hat{\sigma}_e^2}{n})_N}{(\hat{\sigma}_{\mu_0}^2 + \frac{\hat{\sigma}_e^2}{n})_A}$$

N = varianza del modelo nulo

A = varianza del modelo alternativo (con al menos un predictor)

La cantidad de varianza explicada en un segundo nivel es un único valor. Pero es posible que cada grupo j tenga un  $n_j$  distinto, entonces ¿Qué valor se debe usar? Se puede usar cualquier valor que sea considerado *a priori*, representativo de las unidades de nivel 2. Si los valores de  $n_j$  varían mucho en la población, se puede usar la media armónica

(27)

$$\frac{N}{\sum_j \frac{1}{n_j}}$$

Respecto a  $R_1^2$  y  $R_2^2$ , sus valores poblacionales no pueden ser menores de cero. En cambio sus estimaciones pueden aumentar su valor al eliminar un predictor o disminuir al incluir un nuevo predictor, esto puede ser debido al azar o por una mala especificación de la parte fija del modelo. Estos cambios en los valores de  $\hat{R}_1^2$  y  $\hat{R}_2^2$  en una dirección equivocada sirven de diagnóstico para el investigador, para detectar posibles errores de especificación de la parte fija.

El cálculo de la varianza explicada en modelos de pendientes aleatorias, implica un aumento en la complejidad de las fórmulas de  $R_1^2$  y  $R_2^2$  y se pueden consultar en Snijders et al (1994) (87).

### **1.5.7. Multinivel con diferentes variables respuesta y estructura**

Hasta ahora se ha explicado el análisis multinivel desde el punto de vista de modelos con variable dependiente continua y una estructura simple donde los individuos del primer nivel están anidados dentro de las unidades de segundo nivel, ver figura 1. Una de las mayores ventajas de este tipo de análisis es que puede ser usado también con otro tipo de variables respuesta y otras estructuras jerárquicas.

Una posible estructuración de los modelos multinivel, es dividirlos según el tipo de variable respuesta:

- 1) Variable dependiente continua (modelo lineal):
  - a) Medidas repetidas → Cuando se repiten las mediciones en los mismos individuos, por ejemplo hacer un seguimiento de la tensión arterial a un grupo de individuos, se establece una jerarquía de dos niveles, el segundo nivel son los individuos y las repeticiones de la medición son las unidades del primer nivel. Hay distintos modelos que utilizan mediciones repetidas.
    - i) Estudios de crecimiento (caso particular de los estudios longitudinales): las curvas de crecimiento representan un proceso de desarrollo que se produce en función del tiempo, una forma adecuada de modelarlas radica en describir los

valores esperados de las observaciones como funciones polinómicas del tiempo (3). En un modelo de dos niveles para medidas repetidas, las unidades del Nivel 1 especifican Curvas de crecimiento únicas (polinómicas) o trayectorias de crecimiento, para cada individuo (Unidad del nivel 2) (88)

(28)

$$y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + \beta_{2j}t_{ij}^2 + \beta_{3j}t_{ij}^3 + \dots + \beta_{pj}t_{ij}^p + e_{ij} / \beta_{pj} = \beta_p + u_{pj}$$

para aquellos coeficientes que se asumen aleatorios y  $\beta_{pj} = \beta_p$  para aquellos coeficientes no aleatorios

$$u_p \sim N(0, \Omega_u) \quad / \Omega_u = \begin{pmatrix} \sigma_{u0}^2 & \dots & \sigma_{u0p} \\ \vdots & \ddots & \vdots \\ \sigma_{u0p} & \dots & \sigma_{up}^2 \end{pmatrix} \forall p: \beta_{pj} \text{ coeficiente aleatorio}$$

donde  $y_{ij}$  representa la medida de la variable dependiente para el sujeto j en la ocasión i.

- ii) Datos de respuesta multivariante: Las respuestas multivariantes surgen, cuando hay más de una variable dependiente, una variable respuesta medida en el nivel 1 (medidas repetidas) y una o más variables respuesta medidas en el nivel 2 (nivel individuo) (89). Podemos extender el modelo 6.1, añadiendo una var. Dependiente

(29)

$$y_{(p+1)j}^{(2)} = \alpha_{(p+1)} + u_{(p+1)j}$$

/el superíndice 2 indica que es una variable respuesta definida en el nivel 2, y  $\alpha_{(p+1)}$  es el intercepto de la parte fija.

- b) Diseños cruzados: La asunción que las estructuras de datos son puramente jerárquicas es a menudo una simplificación. Los individuos pueden pertenecer simultáneamente a más de una agrupación de un nivel jerárquico dado. Esto lleva a una estructura clasificada cruzada (89-91). Por ejemplo, podemos clasificar a un enfermo, por el centro de salud al que acude o por el vecindario donde reside, tal que, ambas clasificaciones se estiman como aleatorias. Y se puede modelar:

(30)

$$y_{i(j_1, j_2)} = (X\beta)_{i(j_1, j_2)} + u_{j_1} + u_{j_2} + e_{i(j_1, j_2)}$$

$$j_1=1, \dots, J_1 \quad j_2=1, \dots, J_2$$

donde un paciente  $i$ , pertenece a la combinación de c.salud  $j_1$  y vecindario  $j_2$ . La parte aleatoria del modelo, tiene 2 términos residuales del nivel 2  $u_{j_1}$  y  $u_{j_2}$ .

- c) Meta-análisis: nos encontramos con un caso de análisis de datos agrupados; en el que los pacientes (nivel 1) se agrupan en bloques constituidos por cada estudio particular (nivel 2) (89;92). Para un sujeto  $i$ , en el estudio  $j$  que recibió el tratamiento  $h$ , podemos escribir un modelo básico para la respuesta  $y_{hij}$  como:

(31)

$$y_{hij} = (X\beta)_{ij} + \alpha t_{ij} + u_j + e_{ij}$$

$$j=1, \dots, J, \quad i=1, \dots, n_j$$

$$u_j \sim N(0, \sigma_u^2) \quad e_{ij} \sim N(0, \sigma_e^2)$$

$(X\beta)_{ij}$  son la función lineal de las covariables para el sujeto  $i$  en el estudio  $j$ ,  $u_j$  es el efecto aleatorio del estudio  $j$  y  $e_{ij}$  es el residuo para el sujeto  $i$  en el estudio  $j$ . El término  $t_{ij}$  es una variable dicotómica que hace referencia al tratamiento.



## 2) Variable dependiente discreta(modelo no lineal):

Aunque los modelos multinivel fueron desarrollados originalmente para variables de respuesta con distribución normal y bajo los supuestos de una distribución normal de los errores en cada individuo, estos métodos han sido generalizados para situaciones en las que la variable de respuesta es binomial, nominal u ordinal y para procesos donde la probabilidad del evento es pequeña y se puede modelar con una distribución de Poisson. Casos particulares son los siguientes.

a) Modelo logístico: Se usa cuando nuestra variable dependiente toma dos valores. Es una extensión de los modelos multinomiales estándar. Siendo el modelo multinivel más simple, para una estructura jerárquica de 2 niveles, con una variable independiente, el intercepto aleatorio y link logit, es el siguiente:

(32)

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + u_{0j}$$

$$/ u_{0j} \sim N(0, \sigma_{u_0}^2) \quad \text{y} \quad \pi_{ij} = P(y_{ij} = 1)$$

Estos modelos se ampliarán, en la parte de metodología.

- b) Modelo Multinomial: Usamos el análisis Multinomial logístico multinivel, cuando la variable dependiente es una variable categórica con más de dos categorías. Este modelo es una generalización de del modelo logit multinivel, y para una estructura de dos niveles, donde la categoría de referencia de la variable respuesta es t, se define:

(33)

$$\pi_{ij(s)} = P(y_{ij} = 1)$$

$$\log\left(\frac{\pi_{ij(s)}}{\pi_{ij(t)}}\right) = \beta_{0(s)} + \beta_{1(s)}x_{ij} + u_{0(s)j} \quad , S = 1, \dots, t-1$$

$$/ u_{0(s)j} \sim N(0, \sigma_{u_{0(s)}}^2) \quad \text{y} \quad \sum_{h=1}^t \pi_{ij(h)} = 1.$$

La ecuación 33, especifica un modelo para las restantes t-1 categorías, si t es 2 entonces se reduce a un modelo logit multinivel.

- c) Modelo Poisson: La variable dependiente es una variable de conteo, como por ejemplo, número de sibilancias en el último año o número de ataques de asma en el último mes. Son una extensión de los modelos de regresión Poisson. Los datos de conteo son no negativos, por lo tanto si intentamos ajustar un modelo normal a estos datos, podemos obtener predicciones negativas, por ello es preferible usar el logaritmo de los datos, usando una función link log. Un modelo Poisson de efectos mixtos modeliza el número esperado de sibilancias o ataques de asma en un intervalo tiempo (conocido como variable de exposición u “offset”,  $t_{ij}$ ). Esta variable se debe incluir en el modelo como el  $\log(t_{ij})$ . Si la variable dependiente de conteo fue

recogida en el mismo periodo para toda unidad  $i$  de primer nivel en cada unidad  $j$  de segundo nivel entonces el modelo log de intercepto aleatorio es el siguiente:

(34)

$$E(y_{ij}) = \lambda_{ij},$$

$$\log(\lambda_{ij}) = \beta_{0j} + \beta_1 x_{ij}, \quad \beta_{0j} = \beta_0 + u_{0j} \quad u_{0j} \sim N(0, \sigma_{u_0}^2)$$

Si el periodo de recogida no es el mismo entonces el modelo log de intercepto aleatorio es el siguiente:

(35)

$$\log(\lambda_{ij}) = \log(t_{ij}) + \beta_{0j} + \beta_1 x_{ij}, \quad \beta_{0j} = \beta_0 + u_{0j} \quad u_{0j} \sim N(0, \sigma_{u_0}^2)$$

- d) Modelos de supervivencia: Cuando hablamos de análisis de supervivencia nos referimos al análisis del tiempo de seguimiento (T) de cada unidad de observación hasta que ocurre un fenómeno predefinido (muerte, por ejemplo). Puede suceder que en alguna o algunas unidades de observación no haya ocurrido el evento de estudio en el momento de cerrar el estudio, entonces se habla de “censuras” o “casos censurados”. Existen distintas formulaciones para modelizar datos de supervivencia (93;94). Vamos a mostrar la aproximación que se deriva del modelo ordinal y que usa una función de respuesta log-log (94). Supongamos que, para cada unidad de nivel 1, la observación continúa hasta un tiempo  $\underline{y}_{ij}$ , el tiempo de supervivencia es un vector de 0 excepto el último elemento que tomo el valor de  $\underline{d}_{ij}$ . Si el evento ocurre  $\underline{d}_{ij} = 1$  y si la observación es censurada  $\underline{d}_{ij} = 0$ . Definimos la probabilidad de fallar, en el tiempo  $t$  como  $\pi_{ij(t)} = P(\underline{y}_{ij} = t \mid \underline{y}_{ij} \geq t)$  de forma

análoga se puede definir  $1 - \pi_{ij(t)}$  como la probabilidad de sobrevivir más allá del tiempo  $t$ . Entonces podemos formular el modelo log log donde las covariables pueden variar a través del tiempo como:

(36)

$$\log[-\log(1 - \pi_{ij(t)})] = x'_{ij(t)}\mathbf{B} + z'_{ij}\mu'_j$$

La parte fija del modelo la compone  $x'_{ij(t)}$  es el vector de coeficientes que varían a través del tiempo y  $\mathbf{B}$  el vector de coeficientes. La parte aleatoria del modelo la forma el vector de coeficientes de los efectos aleatorios  $z'_{ij}$  y el vector de efectos aleatorios  $\mu'_j$ .

### 1.5.8. Muestreo:

El muestreo es una herramienta que se usa en la investigación científica. Todo estudio ya sea observacional o experimental, lleva implícito en la fase de diseño la determinación del tamaño muestral necesario para la ejecución del mismo (95;96).

Existe bastante literatura acerca del cálculo del tamaño muestral en estudios multinivel (2;24;73;97-101). En los estudios más sencillos, de dos niveles, se deben estimar dos tamaños muestrales distintos:

1. El tamaño de la muestra de las unidades del 1° nivel ( $n_j$ )
2. El tamaño de la muestra de las unidades del 2° nivel ( $J$ )

El tamaño total de la muestra viene dado por:

(37)

$$\sum_{j=1}^J n_j$$

En los estudios multinivel, generalmente el principal problema es determinar el tamaño muestral de las unidades del nivel grupo o segundo nivel, ya que este suele ser más pequeño que el tamaño de muestra del nivel individual (102).

En general, para calcular el número de unidades, individuos o pacientes necesarios en un estudio multinivel, lo primero es calcular un tamaño muestral para un muestreo aleatorio simple. Si queremos comparar las medias del grupo intervención y del grupo control podemos usar la siguiente fórmula:

(38)

$$N_1 = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 * \sigma^2 * (r + 1)}{d^2 * r}$$

Donde:

- $N_1$  = sujetos necesarios en el grupo intervención
- $Z_t$  = Valor Z de la distribución normal que deja a su izquierda un área de  $t / t= 1 - \alpha/2, 1 - \beta$ , siendo  $\alpha$  la probabilidad de cometer el error tipo I y  $\beta$  la probabilidad de cometer el error tipo II
- $\sigma^2$  = Varianza de la variable respuesta
- $d$  = Valor mínimo de la diferencia en media que, si existiera, se desea detectar con una probabilidad  $1-\beta$ .
- $r = \frac{N_0}{N_1}$  razón del número de sujetos entre los grupos comparados. ( $N_0$  tamaño de la muestra en el grupo control)

Si la variable respuesta es dicotómica, se puede usar la siguiente ecuación:

(39)

$$N_1 = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 * \bar{p}(1 - \bar{p}) * (r + 1)}{(p_1 - p_0)^2 * r}$$

$$\bar{p} = \frac{p_1 + (r * p_0)}{1 + r}$$

Donde:

- $N_1$  = sujetos necesarios en el grupo intervención
- $Z_t$  = Valor Z de la distribución normal que deja a su izquierda un área de  $t / t= 1 - \alpha/2, 1 - \beta$
- $\bar{p}$  = media ponderada de  $p_0$  y  $p_1$
- $r = \frac{N_0}{N_1}$  razón del número de sujetos entre los grupos comparados. ( $N_0$  tamaño de la muestra en el grupo control)
- $p_1$  =proporción de casos en el grupo intervención
- $p_0$  =proporción de casos en el grupo control

Después de calcular el tamaño muestral aleatorio simple es necesario multiplicar por un factor de corrección el cual tiene en cuenta el muestreo en dos etapas (103). En un muestreo aleatorio simple el error estándar de la media viene dado por la siguiente fórmula:

$$\text{error est.} = \frac{\text{desviación estandar}}{\sqrt{\text{tamaño muestral}}}$$

Supongamos que tenemos  $N$  macro-unidades cuyo tamaño es  $n$ , entonces el tamaño total de la muestra es  $Nn$ . El *efecto del diseño* es el factor de corrección y es un número que indica cuando debemos ajustar el denominador de la fórmula anterior para tener en cuenta el cambio en el diseño (pasar de una etapa a dos); se define como el cociente entre la varianza obtenida con el nuevo diseño muestral y la varianza obtenida con el muestreo aleatorio simple para la misma población (73;103).

El efecto del diseño para un muestreo de dos etapas con igualdad de tamaño en las macro-unidades o unidades de segundo nivel es:

$$\text{Efecto del diseño} = 1 + (n - 1)\rho$$

Donde  $\rho$  es el coeficiente de correlación intraclase (CCI).

Hay autores que proponen distintas reglas de oro, estas reglas son a menudo opiniones personales basadas en la experiencia :

1. Si se está interesado en los efectos fijos del modelo, 10 grupos en el segundo nivel serán suficientes. Si el interés está en los efectos contextuales como mínimo serán necesarios 30 grupos (102).
2. La regla del 30/30, los investigadores deben esforzarse para obtener una muestra de al menos 30 grupos con 30 individuos por grupo como mínimo (104).

## **2. JUSTIFICACIÓN**



Los modelos estadísticos son herramientas que nos permiten adentrarnos en la complejidad de los fenómenos que investigamos. La naturaleza de las observaciones estudiadas, será de gran importancia a la hora de elegir el método estadístico más apropiado para abordar su análisis (105). No es necesario recurrir a técnicas complejas de análisis por capricho sino tratar de usar los modelos más sencillos posibles que reproduzcan la complejidad del mundo que tratamos de explicar. Aunque la mayoría de las veces un problema complejo va de la mano de una técnica compleja también.

Ante el caso de datos con una estructura definida, como es el caso de datos estructurados jerárquicamente, el enfoque estadístico se plantea desde el punto de vista multinivel, donde se han desarrollado procedimientos estadísticos basados principalmente en la construcción de modelos, que permiten analizar el comportamiento de las respuestas individuales, considerando la información de grupo.

El uso de esta técnica estadística se inició en el marco de las Ciencias sociales y la Educación. La aplicación en el campo de las Ciencias de la Salud ha sido posterior y ha sufrido un aumento exponencial en la última década aunque aún no es una técnica de uso habitual.

Las sibilancias recurrentes en la infancia tienen una gran importancia sobre la morbilidad y mortalidad en ese grupo de edad, tanto en países en vías de desarrollo como en países desarrollados (106). Son una de las enfermedades respiratorias por las que más se acude al hospital en el primer año de vida, lo que tiene un elevado impacto económico (107).

En el estudio multicentrico ISAAC sobre el asma y la alergia en la infancia, se ha visto que la prevalencia de esta enfermedad varía según la ciudad o país donde se estudie (108). Por ello es de interés de esta tesis estudiar los factores de riesgo de las sibilancias recurrentes teniendo en cuenta la estructura jerárquica que conforman los niños dentro de sus ciudades de residencia.

### **3. OBJETIVOS**

Los datos del “Estudio Internacional de Sibilancias en Lactantes” (EISL) (109) presentan una estructura anidada, niños anidados en sus ciudades de residencia, por lo que creemos interesante analizar los factores de riesgo de las sibilancias recurrentes el primer año de vida con los modelos multinivel y nos planteamos en este estudio los siguientes objetivos.

### **Objetivo principal:**

EL objetivo principal de esta tesis es presentar la aportación que suponen los modelos multinivel para el adecuado tratamiento estadístico de datos estructurados jerárquicamente en el estudio de los factores de riesgo de las sibilancias recurrentes en lactantes.

### **Objetivos secundarios:**

2. Comprobar como el análisis multinivel agiliza el proceso de cálculo y mejora la estimación de los factores de riesgo de las sibilancias recurrentes en el primer año de vida frente a estudios anteriores.
3. Estudiar si el enfoque bayesiano para el análisis multinivel mejora la estimación obtenida con el enfoque frecuentista.

## **4. Material y Métodos:**

### 4.1 Datos y variables:

En este apartado se van a usar dos bases de datos, la del estudio ESCAPE (110) que usaremos como ejemplo de que no siempre es posible realizar un análisis multinivel y la base del estudio internacional de sibilancias en el lactante (EISL) (109;111).

#### ESTUDIO ESCAPE:

El objetivo de este estudio era conocer cómo afectan las distintas estaciones del año a la calidad de vida (CV) de los niños asmáticos.

Para ello fueron reclutados por sus pediatras, cuatro grupos distintos de niños asmáticos de 7-14 años, durante cada estación del año, Octubre 2003, Enero 2004, Abril 2004 y Julio 2004. Su CV se midió por medio de la versión española validada (112;113) del Paediatric Asthma Quality of Life Questionnaire (PAQLQ). El cuestionario está compuesto por 23 ítems los cuales miden síntomas, limitaciones de las actividades y la función emocional del niño. Otros factores que se recogieron fueron: gravedad del asma, atopia, tratamiento médico, inmunoterapia, obesidad, hábito de fumar en los padres y medidas ambientales de control alérgico. El Análisis estadístico que se usó fue un modelo de regresión logística, siendo el score de PAQLQ (score global) la variable dependiente (primer tercil =0 vs tercer tercil =1) y las siguientes variables independientes: sexo, tipo de residencia (urbana, rural, intermedia), área geográfica (meseta, costa Mediterránea, costa Norte, tipo de alergia, estación del año, inmunoterapia, severidad del asma, obesidad, mascotas y cubiertas antialérgicas.

Para el análisis multinivel la variable dependiente que se va a usar es el score global (primer tercil vs tercer tercil). Se va a estudiar si la siguiente estructura multinivel tiene sentido:

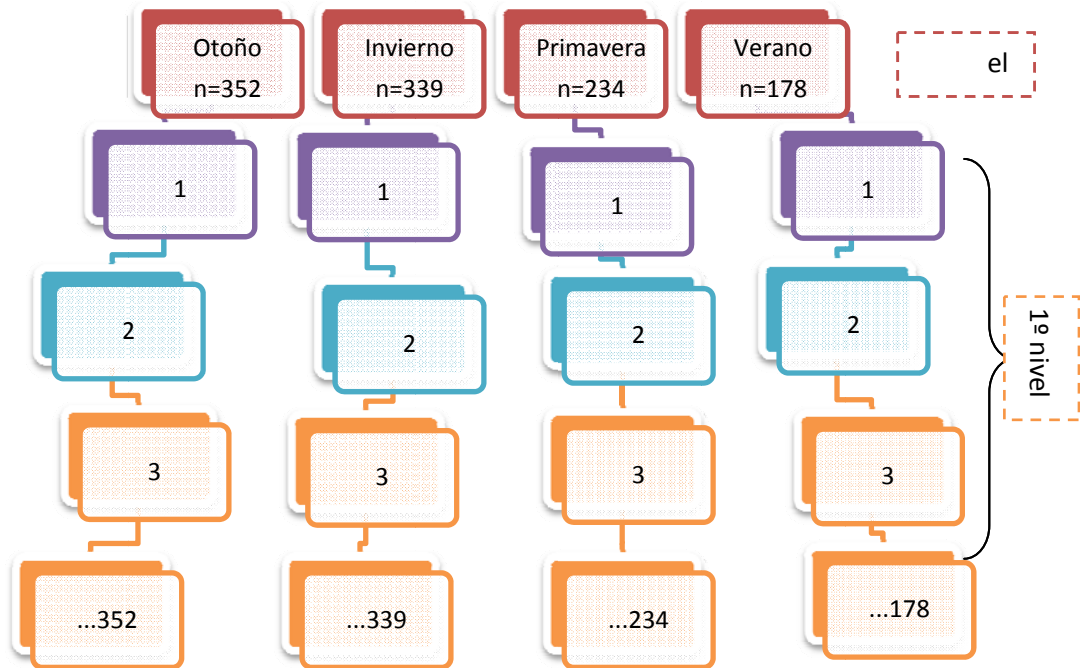


Fig17. Estructura jerárquica estudio ESCAPE

**1º nivel:** niños

**2º nivel:** estaciones del año

**Variable dependiente:** PUNTUACIÓN GLOBAL

### ESTUDIO EISL:

El estudio EISL es un estudio multicentrico, transversal, poblacional que se ha llevado a cabo usando métodos estandarizados y un cuestionario (ANEXO 1) previamente validado (114-116) basado en el cuestionario del Estudio Internacional de Asma y Alergia en la Infancia (ISAAC) (<http://isaac.auckland.ac.nz>). El reclutamiento se realizó en los centros de salud de atención primaria, localizados en el área geográfica

## MATERIAL Y MÉTODOS

del centro EISL correspondiente (ver tabla 5), en el momento en el que los niños acudían para la revisión del niño sano y/o para la administración de la vacuna correspondiente según el programa de vacunas de cada país. El periodo de inclusión de niños en el estudio se mantuvo abierto entre enero de 2005 y enero de 2007. Se invitó a participar en el estudio a los padres o cuidadores en el momento que acudían al centro de salud para una revisión o la administración de la vacuna correspondiente alrededor de la edad de 1 año, haciendo especial énfasis en que todas las preguntas del cuestionario hacían referencia a eventos que habían ocurrido en los últimos 12 meses de vida de su hijo (107). El objetivo de este estudio era el análisis de los factores de riesgo/protección de las sibilancias recurrentes

Variables: Las sibilancias se definieron como la respuesta positiva a la pregunta: “¿Ha tenido su hijo/a sibilancias o pitos en el pecho en los primeros 12 meses de vida?”

Las sibilancias recurrentes, nuestra variable dependiente, fueron definidas como tener 3 o más episodios de sibilancias durante el primer año de vida. Los antecedentes familiares de asma y rinitis se definieron como la respuesta positiva de los padres a las preguntas ¿Tiene el padre o la madre asma?, ¿Tiene el padre o la madre rinitis? Se definió eccema infantil como la respuesta positiva a la pregunta: ¿ha tenido su hijo/a manchas rojas que pican, y aparecen y desaparecen en cualquier parte de su cuerpo excepto alrededor de los ojos, nariz y en la zona del pañal, durante sus primeros 12 meses de vida? Los resfriados comunes se definieron como: haber tenido episodios cortos de resfriado con goteo de nariz, estornudos, obstrucción nasal, tos suave, con o sin fiebre. Otras variables que se han usado en el análisis son el hábito tabáquico de los padres, fumar durante el embarazo, si el niño/a asistió a la guardería en el primer años de vida, lactancia materna exclusiva (más de tres meses vs. 3 o menos), n° de hermanos, n° de personas que viven

en la casa del niño, si hay moho en la casa, si tienen mascotas y nivel de estudios de la madre.

### 4.2 Metodología:

A través de los datos de EISL, vamos a corroborar uno de los objetivos primordiales de esta tesis, es decir, ver como los modelos multinivel mejoran el análisis de los factores de riesgo/protección de la sibilancias recurrentes en el primer año de vida respecto a otro tipo de análisis.

Para el análisis desde el enfoque frecuentista se ha usado el programa estadístico MLwiN v2.25 ©Centre for multilevel modelling, University of Bristol <http://www.bristol.ac.uk/cmm/> . MLwiN es un programa que permite la especificación y el análisis de un amplio rango de modelos multinivel (117). Para el enfoque bayesiano se ha usado el programa Winbugs, WINBUGS, es un software dentro del proyecto BUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling) para el análisis bayesiano de modelos estadísticos complejos usando métodos de Monte Carlo por Cadenas de Markov (MCMC), es de distribución gratuita y ha sido desarrollado por un equipo de estadísticos y programadores del *Medical research Council Biostatistics Research Unit* en Cambridge <http://www.mrc-bsu.cam.ac.uk/bugs/> .

Para el análisis de los datos se va a usar el modelo de regresión multinivel logístico de 2 niveles tanto para el método de estimación frecuentista de la cuasi-verosimilitud penalizada (PQL) (118), como con el método bayesiano usando la estimación de Monte Carlo por cadenas de Markov (MCMC), que como está descrito en la literatura en algunos casos produce aproximadamente las mismas estimaciones que el método PQL y en la mayoría mejora las estimaciones de los efectos aleatorios (119). Ambos modelos se van a comparar entre si y a su vez



con el análisis realizado anteriormente en que se usó un meta-análisis con efectos aleatorios, en el que se consideran dos posibles fuentes de variabilidad, la variabilidad intra-estudio y la variabilidad entre-estudios, que se incorporan al estimador combinado a través de los pesos correspondientes (120;121). Los datos que se usaron en el meta-análisis fueron los resultados de aplicar para cada centro un modelo lineal generalizado con distribución binomial y link logit, para calcular los OR ajustados de los factores de riesgo/protección, siendo las sibilancias recurrentes la variable dependiente.

Además se va analizar un modelo de pendientes aleatorias y un modelo que contiene variables de segundo nivel usando el enfoque frecuentista y el enfoque bayesiano.

A continuación se presentan los dos enfoques multinivel que hemos usado para nuestro análisis.

### **4.2.1 Modelo logístico multinivel de 2 niveles: Enfoque frecuentista.**

El objetivo principal de un regresión logística es predecir la probabilidad  $\pi_i$  de que ocurra un evento  $Y$ , para un individuo  $i$ , en función de un determinado número de variables.

Un modelo general para una variable respuesta dicotómica ( $Y_i$ ) y una variable explicativa  $x_i$  es:

$$f(\pi_i) = \beta_0 + \beta_1 x_i / P(Y = 1) = \pi_i \quad (40)$$

siendo  $f(\pi_i)$  una transformación de  $\pi_i$  llamada link.

Las funciones link (122;123) más conocidas son:

1. Link logit, donde  $f(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$
2. Link probit, donde  $f(\pi_i) = \Phi(\pi_i)$  es la función de densidad acumulada de la distribución normal.
3. Link log-log, donde  $f(\pi_i) = \log(-\log(1 - \pi_i))$

Si queremos extender nuestro modelo para tener en cuenta la estructura jerárquica de 2 niveles en nuestros datos, comenzaremos ajustando el modelo nulo con intercepto aleatorio (123). Nuestra variable respuesta es  $y_{ij}$  toma el valor 1 individuo  $i$  tiene la respuesta  $Y$  el centro  $j$  y 0 si no. Si usamos el link logit, nuestro modelo nulo de dos niveles para una variable respuesta dicotómica queda de la siguiente manera:

(41)

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \log \text{odds} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} \\ \beta_{0j} &= \beta_0 + u_{0j} / u_{0j} \sim N(0, \sigma_{u_0}^2) \end{aligned}$$

Modelo de intercepto aleatorio:

Si queremos incluir en el modelo potenciales variables explicativas de la variable respuesta  $X_1, \dots, X_p$ , obtenemos el siguiente modelo:

(42)

$$\begin{aligned} \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \beta_0 + \sum_{p=1}^p \beta_p x_{(p)ij} + u_{0j} \\ / u_{0j} &\sim N(0, \sigma_{u_0}^2) \end{aligned}$$

El intercepto  $\beta_{0j}$  está formado por dos componentes: un efecto fijo  $\beta_0$ , igual para todos los centros, y un efecto aleatorio  $u_{0j}$  específico para cada centro (unidad de segundo nivel)  $j$ .

En el caso más simple en que tan solo hay una variable explicativa la formulación del modelo es:

(43)

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{(1)ij} + u_{0j}$$

$$/ u_{0j} \sim N(0, \sigma_{u_0}^2)$$

$$\hat{\pi}_{ij} = \frac{\exp(\beta_0 + \beta_1 x_{(1)ij} + u_{0j})}{1 + \exp(\beta_0 + \beta_1 x_{(1)ij} + u_{0j})}$$

$\beta_0$  se interpreta como el intercepto de conjunto en la relación entre el log-odds y x. El intercepto para una determinada unidad de nivel 2 j es  $\beta_0 + u_{0j}$  que será mayor o menor que el intercepto de conjunto dependiendo de si  $u_{0j}$  es mayor o menor de cero. Como en el caso del modelo de respuesta continua,  $u_{0j}$  es conocido como el efecto aleatorio de grupo, el residuo de grupo o el residuo de 2º nivel. La varianza del intercepto entre grupos es  $var(u_{0j}) = \sigma_{u_0}^2$ , se conoce como la varianza residual entre grupos o varianza residual de segundo nivel. Y en el modelo con variables explicativas siempre la varianza no explicada de nivel 2. Las varianzas por definición son no negativas, por ello cuando testamos la hipótesis nula  $H_0: \sigma_{u_0}^2 = 0$  la hipótesis alternativa debe ser unilateral  $H_1: \sigma_{u_0}^2 > 0$ , por lo tanto probabilidad de que el estadístico Z sea mayor o igual que una variable chi-cuadrado con tantos grados de libertad como parámetros haya en el modelo, hay que dividirla por 2 (73).

El modelo multinivel para respuestas binarias se puede derivar también a través de una variable latente de contextualización (73;123). Asumimos que existe una variable continua  $y^*_{ij}$  subyacente a  $y_{ij}$  y así podemos formular el llamado modelo *umbral*, que permite la representación:

(44)

$$y_{ij} = \begin{cases} 1 & \text{si } y^*_{ij} \geq 0 \\ 0 & \text{si } y^*_{ij} < 0 \end{cases}$$

Teniendo en cuenta esta representación podemos escribir el siguiente modelo de 2 niveles de intercepto aleatorio para la variable inobservada  $y^*_{ij}$ , como para cualquier variable continua:

$$y^*_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + \varepsilon^*_{ij}) \quad (45)$$

Para que represente un modelo de regresión logística, los residuos de primer nivel de la variable subyacente  $y^*$ , deben tener una distribución logística. Lo que significa que:

- $P(\varepsilon^*_{ij} < x) = \text{logistic}(x) \forall x$
- La media de los residuos del primer nivel es 0
- La varianza es  $\frac{\pi^2}{3} = 3.29$

Cuando se asume que  $\varepsilon^*_{ij}$  tiene esta distribución, el modelo logístico visto en el apartado anterior, es equivalente al modelo *umbral* definido aquí.

Coeficiente de partición de la varianza en términos de  $y^*$ :

La fórmula de la VPC es la siguiente:

$$VCP = \frac{\text{varianza residual del 2º nivel}}{\text{varianza residual del 1º nivel} + \text{varianza residual del 2º nivel}}$$

Y mide la proporción del total de la varianza residual que es debida a la variación entre grupos.

Como hemos visto en la sección anterior, podemos generalizar un modelo logístico de interceptos aleatorios como un modelo *umbral* para una variable latente, este nuevo modelo tiene la misma forma que un modelo de intercepto aleatorio para una variable respuesta continua, por lo que:

$$VPC = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{\varepsilon^*}^2} \quad (46)$$

donde  $\sigma_{\varepsilon^*}^2 = 3.29$  para un modelo logit.

Este VCP, se interpreta como la proporción de la varianza residual total en la propensión de estar en la categoría 1, que es debida a las diferencias entre grupos.

Para modelos con variable dependiente discreta, se recomienda el uso de métodos estimación de los parámetros de cuasi-verosimilitud (123). Estos procedimientos que ya nombramos en el cuadro de algoritmos, tabla 4, usan métodos de lianearización basados en la expansión de series de Taylor, que transforma un modelo con variable dependiente discreta en un modelo con la variable respuesta continua. Se pueden usar dos tipos de algoritmos, la cuasi-verosimilitud marginal (MQL) y la cuasi-verosimilitud predictiva (PQL). Ambos algoritmos pueden incluir tanto términos de 1° orden como de 2° orden de las series de expansión de Taylor. El procedimiento MQL de 1° orden, proporciona una aproximación cruda y se pueden obtener estimaciones sesgadas, especialmente si el tamaño de la muestra dentro de las unidades del segundo nivel es pequeño. El procedimiento PLQ de 2° orden, mejora la aproximación (124), aunque es un método menos estable y puede dar problemas de convergencia. Por ello se aconseja empezar el análisis con el algoritmo MQL de 1° orden, para obtener los valores de inicio del algoritmo PQL de 2° orden (3;125).

Modelo de pendientes aleatorias:

Un modelo de intercepto aleatorio asume que la relación entre Y y X es la misma para cada grupo, es decir la pendiente  $\beta_1$  es fija entre los grupos. Podemos relajar esta relación permitiendo que la pendiente de una o más covariables varíe aleatoriamente entre grupos, es decir permitiendo un *Modelo de pendientes aleatorias de 2 niveles*.

Modelo de pendiente aleatoria: la siguiente formulación es para una única pendiente aleatoria; la extensión al caso de dos o más pendientes aleatorias es sencilla:

(47)

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \beta_{1j}X_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} \quad / \quad u_{0j} \sim N(0, \sigma_{u_0}^2)$$

$$\beta_{1j} = \beta_1 + u_{1j} \quad / \quad u_{1j} \sim N(0, \sigma_{u_1}^2) \quad \text{y} \quad \Omega_u = \begin{pmatrix} \sigma_{u_0}^2 & \sigma_{u_{01}} \\ \sigma_{u_{01}} & \sigma_{u_1}^2 \end{pmatrix}$$

$\sigma_{u_{01}}$  es la covarianza de los efectos aleatorios.

$\Omega_u$  es la matriz de varianzas covarianzas de los efectos aleatorios.

La pendiente de la relación lineal entre X y el log-odds de Y=1 es  $\beta_1 + u_{1j}$  para el grupo j. En comparación con el modelo de interceptos aleatorios el modelo de pendientes aleatorias introduce dos nuevos parámetros respecto del  $\sigma_{u_1}^2$  y  $\sigma_{u_{01}}$ . Se pueden comparar los dos modelos llevando a cabo el test estadístico donde la hipótesis nula es que  $\sigma_{u_1}^2 = 0$  y  $\sigma_{u_{01}} = 0$ .

Para un modelo de pendiente aleatoria y una variable explicativa la varianza residual es:

(48)

$$\begin{aligned} \text{var}(u_{0j} + u_{1j}x_{ij}) &= \text{var}(u_{0j}) + 2x_{ij}\text{cov}(u_{0j}, u_{1j}) + x_{ij}^2v \\ &= \sigma_{u_0}^2 + 2\sigma_{u_0u_1}x_{ij} + \sigma_{u_1}^2x_{ij}^2 \end{aligned}$$

Existen varios métodos para valorar la significación respecto al valor 0 de la varianza entre unidades de segundo nivel, y dependen del algoritmo usado para ajustar el modelo. Para el procedimiento de la cuasi-verosimilitud (PQL2), usaremos el test de Wald. Cuando usamos este test para estimar una varianza residual, obtenemos un test crudo, porque depende de la asunción de que la varianza estimada se distribuya normalmente. Sin embargo, nos da una indicación acerca de la fuerza de la evidencia para el efecto de grupo.

Una alternativa a estos procedimientos es usar el análisis bayesiano con métodos de Monte Carlo por Cadenas de Markov (MCMC). Este procedimiento posee la ventaja de que en muestras pequeñas tiene en cuenta la incertidumbre asociada a la estimación de los parámetros aleatorios y proporciona medidas precisas de incertidumbre, lo que a su vez es importante debido a que los métodos de máxima verosimilitud tienden a sobreestimar la precisión al ignorar la incertidumbre (126).

### Modelo de intercepto aleatorio con una variable contextual:

Una ventaja de los modelos multinivel es la posibilidad de explorar los efectos de los predictores del nivel 2 mientras simultáneamente se incluyen efectos aleatorios, para permitir efectos de variables no observadas del 2º nivel. A los coeficientes del segundo nivel usualmente se les denomina efectos contextuales. El modelo más sencillo sería un modelo de intercepto aleatorio con una variable explicativa de primer nivel  $X_{1ij}$  y una variable de segundo nivel:

(49)

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2j}$$

$$\beta_{0j} = \beta_0 + u_{0j} \quad / \quad u_{0j} \sim N(0, \sigma_{u_0}^2)$$

Las covariables de segundo nivel ayudan a explicar solo la varianza residual de grupo o residuo de segundo nivel.

### 4.2.2 Modelo logístico multinivel de 2 niveles: Enfoque bayesiano.

El punto de partida en inferencia paramétrica es un modelo de probabilidad para los datos  $x$ , es decir, supondremos que los datos son la realización de una variable aleatoria con una distribución de probabilidad que depende de parámetros. Generalmente dicho modelo vendrá definido mediante una función de densidad  $f(x|\theta)$  parametrizada por el parámetro  $\theta$  multidimensional

$$x \sim f(x|\theta)$$

El objetivo es la estimación de los parámetros  $\theta$ , es decir, los datos han sido generados según el modelo para algún valor desconocido de  $\theta$  y el objetivo es estimar dicho valor a partir de la observación de los datos.

Mientras que la inferencia *frecuentista* se basa fundamentalmente en la función de verosimilitud  $f(x|\theta)$  para realizar estimaciones (inferencias) sobre el parámetro, el punto de vista bayesiano trata al parámetro como una variable aleatoria especificando sobre ella una distribución de probabilidad  $f(\theta)$  llamada distribución a priori (127-129).

La distribución de probabilidad de una variable aleatoria refleja información sobre los valores que puede tomar dicha variable. Si una variable tiene una distribución normal  $N(0,1)$ , los valores de dicha variable estarán alrededor de 0, será muy raro (aunque no imposible) observar valores superiores a 4 o inferiores a -4 e igualmente probable observar valores superiores a 2 que inferiores a -2.



Por tanto, especificar una distribución a priori sobre  $\theta$  es una forma de establecer información sobre  $\theta$  a priori, es decir, antes de observar los datos  $x$ .

Tenemos entonces dos variables aleatorias, los datos  $x$  y el parámetro  $\theta$ . La distribución de probabilidad de los datos, si el valor del parámetro es  $\theta$  (distribución condicional), viene dada por  $f(x|\theta)$ , mientras que la distribución del parámetro viene dada por  $f(\theta)$ , la distribución a priori. Entonces la distribución de  $\theta$  condicionada a haber observado los datos, que contiene la información sobre  $\theta$  cuando disponemos de los datos  $x$ , viene dada por el teorema de Bayes

(50)

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}$$

y se llama distribución a posteriori  $\theta$ . Hemos usado la notación genérica  $f(\cdot)$  para denotar la distribución de una variable.

Por tanto, toda la información que los datos aportan sobre el parámetro  $\theta$  viene recogida en la distribución a posteriori en términos de probabilidades. Con la distribución a posteriori tenemos mucho. Si pensamos en las distribuciones de probabilidad como en una herramienta para cuantificar la información sobre la variable en cuestión, la distribución a posteriori es la herramienta que necesitamos para hacer inferencia (obtener información) sobre los parámetros. Para esto necesitamos describir la distribución a posteriori. Como es una distribución de probabilidad, su descripción puede hacerse mediante cuantiles, media y varianza. Por ejemplo, para parámetros unidimensionales, entre el cuantil 0.025 y el cuantil 0.975 se encuentra el parámetro con probabilidad a posteriori  $0.95 = 0.975 - 0.025$ , formando dichos cuantiles los extremos de lo que llamaremos intervalo de credibilidad al 95% para dicho parámetro. La probabilidad de que el parámetro esté en dicho intervalo es 0.95, a diferencia del intervalo de

confianza al 95% frecuentista, que contiene al parámetro con probabilidad 0.95, es decir, si obtenemos 100 veces datos en las mismas condiciones y calculamos los 100 intervalos de confianza al 95% correspondientes, entonces aproximadamente 95 de los 100 intervalos contienen al parámetro y 5 de los intervalos no lo contienen. En la práctica solo tenemos una vez los datos y por tanto decimos que confiamos al 95% en que el intervalo calculado con los datos contenga al verdadero valor del parámetro.

### Métodos MCMC para calcular las características de la distribución a posteriori:

En este punto aparece el problema de calcular características de la distribución a posteriori para hacer inferencia. Por ejemplo, la media a posteriori de un parámetro utilizada como estimador de dicho parámetro suele requerir el cálculo de integrales en dimensiones elevadas. Si el parámetro de interés es un  $OR = e^\beta$  siendo  $\beta$  una de las componentes del parámetro multidimensional  $\theta$ , entonces la media a posteriori del parámetro  $OR$  viene dada por la integral

$$\int e^\beta f(\theta|x)d\theta \tag{51}$$

y en general dicha integral no puede calcularse de manera exacta. El cálculo de dichas características de la distribución a posteriori requiere métodos numéricos. Los métodos más utilizados en estadística para este propósito se conocen como métodos de Monte Carlo por Cadenas de Markov (MCMC) (130;131). Dichos métodos se basan en la simulación de una cadena de Markov apropiada que tiene a la distribución a posteriori como distribución estacionaria (límite): si  $\{\theta(1), \theta(2), \dots\}$  es una cadena de Markov con distribución límite la distribución a posteriori, entonces

(52)

$$\int e^{\beta} f(\theta|x) d\theta = \lim_k \frac{1}{k} \sum_{i=1}^k e^{\beta^{(i)}}.$$

En la práctica se simula una cadena de Markov apropiada de gran longitud  $k$  y la media a posteriori de  $OR = e^{\beta}$  se aproxima por  $\sum_{i=1}^k e^{\beta^{(i)}}/k$ , y análogamente cuantiles y varianzas.

Los algoritmos más conocidos para obtener cadenas de Markov con una determinada distribución límite son el algoritmo de Metropolis-Hasting y Gibbs sampling (131). Existen programas informáticos que implementan de estos algoritmos; el más popular es WinBugs v1.4 <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.

Convergencia de los métodos MCMC

La función bugs del programa estadística R v2.13 <http://www.r-project.org/> actúa como interface entre R y el programa WinBUGS que simula cadenas de Markov en paralelo con distribución estacionaria la distribución a posteriori. Esta interface permite manejar los resultados obtenidos cómodamente.

Es una buena práctica que las cadenas empiecen cada una en puntos distintos. Cuando las cadenas que empiezan en puntos distintos no se *mezclan* es señal de que no se ha alcanzado la convergencia y se necesitan cadenas más largas o adoptar otras estrategias.

El estadístico  $\hat{R}$  monitoriza esta convergencia para cada parámetro. En la práctica cuando  $\hat{R} < 1.1$  para todos los parámetros, la convergencia es aceptable. Otro indicador de convergencia es el número efectivo de simulaciones: **n.eff**. Al menos **n.eff** debe ser 100 para todos los parámetros.

Enfoque bayesiano de los modelos multinivel logísticos:

Para aplicar la metodología bayesiana en un modelo multinivel de dos niveles hay que especificar la distribución a priori sobre los parámetros del modelo. Consideremos por ejemplo el siguiente modelo “umbral” que permite la representación de la variable dicotómica mediante una variable latente:

(44)

$$y_{ij} = \begin{cases} 1 & \text{si } y^*_{ij} \geq 0 \\ 0 & \text{si } y^*_{ij} < 0 \end{cases}$$

Teniendo en cuenta esta representación podemos escribir el siguiente modelo de 2 niveles de intercepto aleatorio para la variable inobservada  $y^*_{ij}$ , como para cualquier variable continua:

(53)

$$y^*_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + \varepsilon_{ij}$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Los parámetros del modelo son  $\beta_0, \beta_1, \sigma_{u_0}$  y  $\sigma_\varepsilon$ , es decir,  $\theta = (\beta_0, \beta_1, \sigma_{u_0}, \sigma_\varepsilon)$ . Los parámetros  $\beta_0$  y  $\beta_1$  son los parámetros de la parte fija del modelo y por similitud con el modelo de regresión simple, para estos parámetros vamos a suponer una distribución con muy poca información

$$\beta_0 \sim N(0, 10^4) \quad \beta_1 \sim N(0, 10^4)$$

## MATERIAL Y MÉTODOS

y para los parámetros  $\sigma_{u0}$  y  $\sigma_\varepsilon$  podemos usar una distribución uniforme en el intervalo  $(0,100)$ . La distribución a priori  $\theta$  es el producto de las distribuciones a priori para cada uno de los parámetros  $\beta_0, \beta_1, \sigma_{u0}$  y  $\sigma_\varepsilon$ . Con estas distribuciones a priori lo que reflejamos es que esperamos que los parámetros  $\beta_0$  y  $\beta_1$  estén en el rango  $(-100,100)$ , y que los parámetros  $\sigma_{u0}$  y  $\sigma_\varepsilon$  estén en el rango  $(0,100)$ . Si las estimaciones a posteriori no están en estos rangos significa que las distribuciones a priori elegidas no son apropiadas y habría que ampliar los rangos (132).

El modelo multinivel considerado es uno de los más simples, sin embargo se requiere de métodos MCMC para realizar la estimación a posteriori de  $\theta$ .

## **5. RESULTADOS**

## RESULTADOS

### ESCAPE:

El número de pediatras que reclutaron niños ha variado a lo largo del periodo de estudio: 132 en otoño, 128 en invierno, 89 en primavera y 74 en verano. Esto representa una pérdida aproximada de 44% en el reclutamiento de los pediatras. El número de niños con datos válidos en cada temporada fue: 345 en otoño, 334 en invierno, en la primavera de 232 y 176 en verano. Las características demográficas de la población en su conjunto y de cada grupo de niños en cada estación del año se muestran en la Tabla 4.

	Otoño n=352	Invierno n=339	Primavera n=234	Verano n=178	Todas las estaciones n=1103
Edad(años)media (SD)	10.4 (2.0)	10.2 (2.0)	10.3 (2.1)	10.6 (2.1)	10.3 (2.0)
Genero masc.	216 (62.6)	208 (62.3)	153 (65.9)	99 (56.3)	676 (62.2)
Peso, kg	41.5 (12.8)	40.9 (13.0)	41.2 (13.3)	43.0 (13.6)	41.5 (13.1)
Talla, cm	144.5 (13.7)	143.2 (13.0)	143.4 (13.3)	145.4 (14.0)	144.0 (13.5)
Residencia, n (%)					
<i>Rural</i>	63 (18.7)	70 (21.1)	38 (16.7)	22 (12.9)	193 (18.1)
<i>Intermedia</i>	87 (25.8)	77 (23.3)	44 (19.4)	45 (26.5)	253 (23.8)
<i>Urbana</i>	187 (55.5)	184 (55.6)	145 (63.9)	103 (60.6)	619 (58.1)
Zona Climática					
<i>Meseta</i>	182 (51.7)	165 (48.7)	109 (46.6)	89 (50.0)	545 (49.4)
<i>Costa Mediterránea</i>	74 (21.0)	70 (20.6)	40 (17.1)	23 (12.9)	207 (18.8)
<i>Costa Norte</i>	96 (27.3)	104 (30.7)	85 (36.3)	66 (37.1)	351 (31.8)
Tipo de Alergeno					
<i>Polvo</i>	86 (24.4)	72 (21.2)	49 (20.9)	54 (30.3)	261 (23.7)
<i>Polen</i>	31 (8.8)	29 (8.6)	21 (9.0)	30 (16.9)	111 (10.1)
<i>Moho</i>	13 (3.7)	12 (3.5)	8 (3.4)	5 (2.8)	38 (3.4)
<i>Múltiples</i>	156 (44.3)	150 (44.2)	107 (45.7)	53 (29.8)	466 (42.2)
Asistencia médica					
<i>Consulta extra-hospitalaria</i>	93 (26.6)	88 (26.0)	67 (28.9)	52 (29.4)	300 (27.3)
<i>Atención primaria</i>	257 (73.4)	250 (74.0)	165 (71.1)	125 (70.6)	797 (72.7)
Inmunoterapia	261 (23.7)	106 (30.1)	74 (21.8)	41 (17.5)	40 (22.5)
Tratamiento con ICS	394 (35.7)	135 (38.4)	122 (36.0)	74 (31.6)	63 (35.4)
Severidad del Asma					
<i>Episódica infrecuente</i>	121 (37.1)	143 (46.7)	105 (47.9)	68 (42.0)	437 (43.1)
<i>Episódica frecuente</i>	123 (37.7)	97 (31.7)	71 (32.4)	71 (43.8)	362 (35.7)
<i>Moderada persistente</i>	78 (23.9)	60 (19.6)	41 (18.7)	23 (14.2)	202 (19.9)
<i>Severa persistente</i>	4 (1.2)	6 (2.0)	2 (0.9)	0 (0.0)	12 (1.2)
Obesidad	32 (9.7)	26 (8.3)	24 (10.8)	15 (8.8)	97 (9.3)
Mascotas					
<i>Perro</i>	56 (16.2)	61 (18.2)	36 (15.9)	33 (18.5)	186 (17.1)
<i>Gato</i>	22 (6.4%)	26 (7.8%)	11 (4.8%)	11 (6.2%)	70 (6.5%)
Cubiertas antialérgicas					
<i>Almohada</i>	69 (19.6%)	79 (23.3%)	59 (25.2%)	34 (19.1%)	241 (21.8%)
<i>Edredón</i>	87 (24.7%)	87 (25.7%)	63 (26.9%)	44 (24.7%)	281 (25.5%)
Fumar en casa	163 (46.8%)	156 (46.2%)	113 (48.3%)	78 (44.1%)	510 (46.5%)

Tabla 4. Características demográficas de los participantes

En primer lugar se ha ajustado el modelo nulo de intercepto aleatorio  
 En la figura 18 y en tabla 5 se recogen el modelo nulo y las estimaciones de los parámetros fijos y aleatorios del modelo.

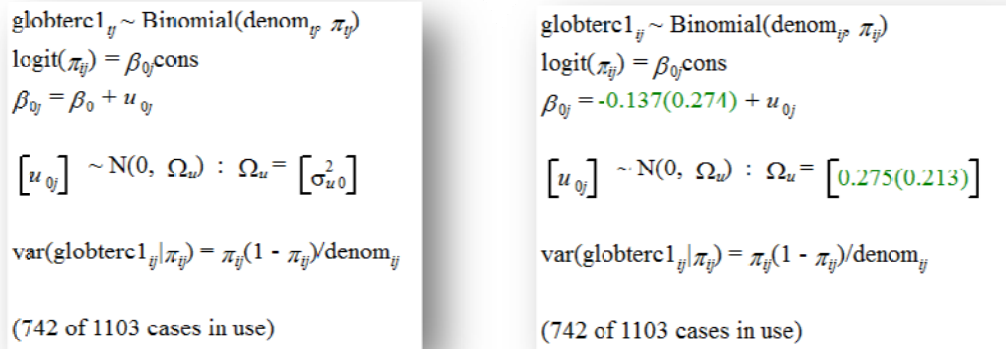


Fig18. Modelo nulo, estudio ESCAPE.

	<b>Estimación del parámetro</b>	<b>Error estándar</b>
$\beta_0$ (intercepto)	-0,137	0,274
$\sigma^2_u$ (varianza entre estaciones)	0,275	0,213

Tabla 5. Estimación PQL-2 de los parámetros fijos y aleatorios.

La varianza entre estaciones del año en el log-odds de estar en el tercil superior de la puntuación global se estima como 0,275 con un error estándar de 0,213.

El estadístico de Wald =  $(0,275/0,213)^2 = 1,670$  el cual se compara con el valor de una variable chi-cuadrado con un grado de libertad ( $n^\circ$  de parámetros en el modelo) y se obtiene un p-valor unilateral = 0.110.



### EISL

En este estudio, participaron 16 centros EISL de 6 países distintos. Se eliminaron 771 cuestionarios de entre aquellos que no habían contestado a la variable principal (tener sibilancias el primer año de vida), o no se habían recogido el número de episodios de sibilancias, o aquellos que pertenecían a niños mayores de 24 meses o menores de 12, el total de niños que se han incluido en el análisis es 28687. La tasa de respuesta fue mayor del 70% en todos los centros excepto en Valencia que fue del 61%. La participación fue del 52.3% de niños y un 47.7% de niñas y la edad media de toda la muestra fue de  $13.4 \pm 1.7$  meses.

En la figura 19 se muestra la estructura jerárquica de 2 niveles:

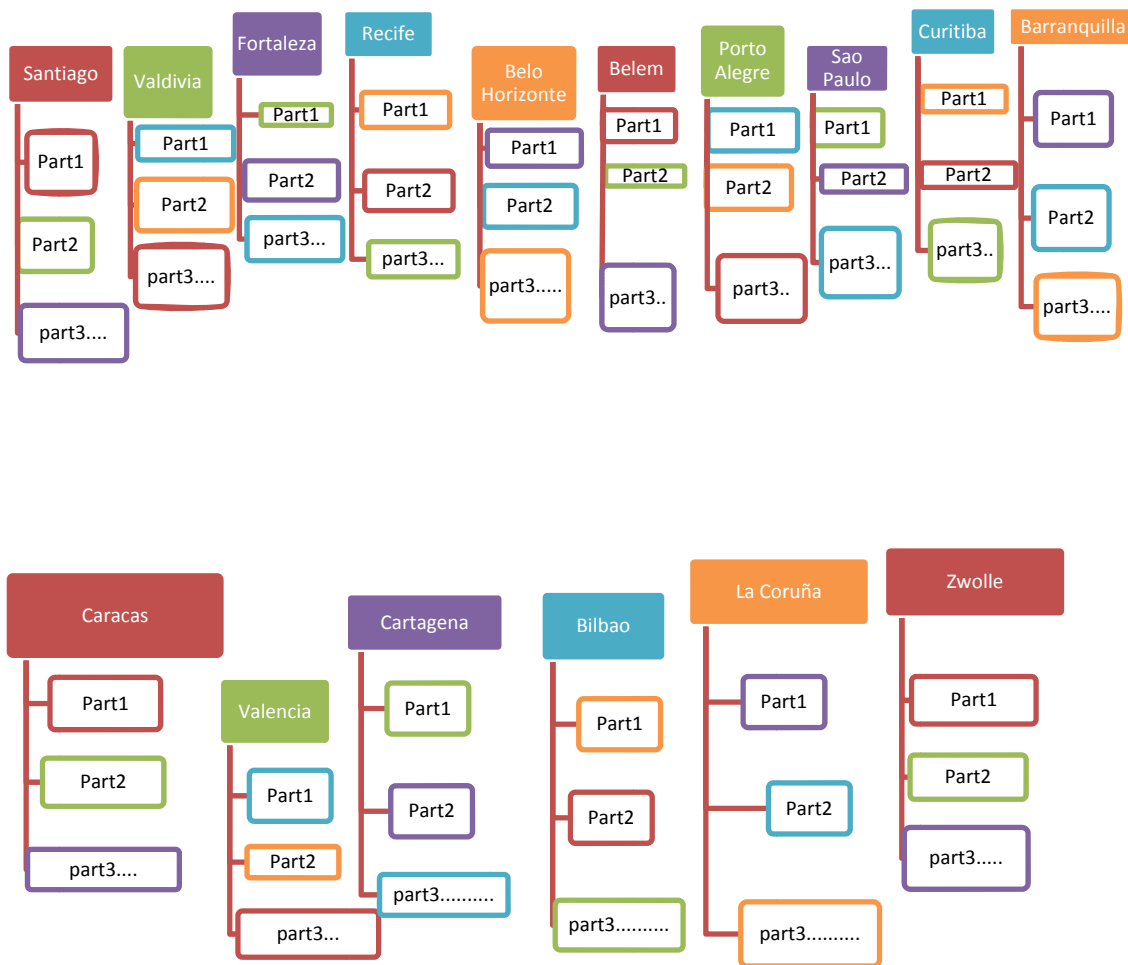


Fig19. Estructura jerárquica estudio EISL.

**1º nivel:** niños

**2º nivel:** centros

**Variable dependiente:** SIBILANCIAS RECURRENTE

En la tabla 6 se recogen los centros participantes en Latinoamérica y Europa, el tamaño de la muestra en cada centro y la prevalencia de sibilancias recurrentes por centro.

La tabla 7 contiene las prevalencias de los factores de riesgo/protección para las sibilancias recurrentes en los distintos centros. Y la tabla 8

muestra el proceso secuencial de introducción de variables en el modelo hasta llegar al modelo nº13, que es el modelo final.

País			Sibilancias recurrentes(%)
Centro	N		
Chile			
Santiago	2988		22.0
Valdivia	3075		21.1
Brasil			
Fortaleza	1209		22.1
Recife	1063		25.0
Belo Horizonte	2532		27.4
Belem	3029		21.9
Porto Alegre	1016		36.3
Sao Paulo	1012		26.7
Curitiba	3003		22.6
Colombia			
Barranquilla	1688		15.9
Venezuela			
Caracas	3009		17.5
	<b>Total America Latina</b>	<b>23624</b>	<b>23.7</b>
España			
Valencia	886		12.1
Cartagena	1172		16.2
Bilbao	996		18.6
La Coruña	930		13.8
Holanda			
Zwolle	1079		13.80
	<b>Total Europa</b>	<b>5063</b>	<b>15.0</b>
	<b>Total</b>	<b>28687</b>	<b>22.2</b>

Tabla 6. Tabla descriptiva de los centros participantes.

## RESULTADOS

	<u>Family history</u>				Fumar en el embarazo	Resfriados 3 primeros meses	Guardería	Lactancia materna >3m	Hermanos (RIQ)	Personas En casa (RIQ)	Moho	Estudios Universitarios de la madre	Mascot.
	Niños	Asma	Rinitis	eczema									
<i>Chile</i>													
<i>Santiago</i>	49.7	15.1	19.9	44.2	9.0	44.7	7.4	79.7	0-2	4-6	22.9	58.6	55.4
<i>Valdivia</i>	59.2	20.3	25.1	45.1	8.7	40.0	15.2	74.3	0-2	4-5	31.4	61.2	44.0
<i>Brazil</i>													
<i>Fortaleza</i>	53.7	35.4	48.5	52.7	9.8	35.7	3.8	62.0	0-2	3-6	16.2	25.5	35.6
<i>Recife</i>	53.1	40.0	35.9	58.9	12.3	39.8	26.3	59.9	0-2	3-5	30.8	32.5	30.7
<i>Belo Horizonte</i>	50.8	41.4	52.8	59.0	13.1	36.7	11.2	63.8	0-2	3-5	35.5	33.6	48.4
<i>Belem</i>	49.7	18.6	29.9	58.8	4.7	39.8	1.4	78.8	0-2	4-7	26.1	26.9	42.7
<i>Porto Alegre</i>	51.7	37.7	66.4	61.1	19.6	45.1	21.5	7.1	0-2	3-5	44.0	3.9	51.6
<i>Sao Paulo</i>	50.6	13.0	48.0	44.2	15.6	32.9	19.6	62.7	0-2	3-5	32.7	41.0	30.7
<i>Curitiba</i>	50.7	21.0	55.8	54.1	14.0	38.8	25.1	69.5	0-2	3-5	30.8	39.5	50.7
<i>Colombia</i>													
<i>Barranquilla</i>	53.6	29.9	48.7	53.4	3.5	39.5	5.4	68.8	0-2	6-8	41.7	38.3	40.7
<i>Venezuela</i>													
<i>Caracas</i>	53.4	44.0	46.0	39.6	8.2	43.3	18.0	54.4	0-2	4-6	28.7	49.6	37.3
<b>Latin America Total</b>	<b>52.4</b>	<b>27.7</b>	<b>40.9</b>	<b>50.9</b>	<b>9.9</b>	<b>40.1</b>	<b>13.3</b>	<b>66.2</b>	<b>0-2</b>	<b>3-6</b>	<b>30.2</b>	<b>41.1</b>	<b>44.2</b>
<i>Spain</i>													
<i>Valencia</i>	52.8	10.4	26.4	10.6	14.7	27.5	16.2	47.6	0-1	3-4	3.2	73.3	23.2
<i>Cartagena</i>	53.1	13.2	23.7	14.5	21.0	31.0	12.0	40.3	0-1	3-4	12.8	73.4	28.6
<i>Bilbao</i>	52.1	12.7	18.3	12.1	16.2	21.4	28.8	55.6	0-1	3-4	8.4	89.6	12.8
<i>La Coruña</i>	49.9	14.1	26.6	12.7	15.7	24.4	22.0	42.1	0-1	3-4	16.4	84.2	28.3
<i>The Netherlands</i>													
<i>Zwolle</i>	51.4	13.2	44.8	22.7	9.2	43.6	40.1	45.4	0-1	3-4	5.6	97.9	56.2
<b>Europe Total</b>	<b>51.9</b>	<b>12.8</b>	<b>28.2</b>	<b>14.7</b>	<b>15.4</b>	<b>30.0</b>	<b>23.9</b>	<b>46.0</b>	<b>0-1</b>	<b>3-4</b>	<b>9.4</b>	<b>83.8</b>	<b>30.4</b>
<b>Total</b>	<b>52.3</b>	<b>25.1</b>	<b>38.6</b>	<b>44.7</b>	<b>10.9</b>	<b>38.4</b>	<b>15.2</b>	<b>62.6</b>	<b>0-2</b>	<b>3-6</b>	<b>26.6</b>	<b>48.6</b>	<b>41.8</b>

Tabla 7. Prevalencia (%) de los factores de riesgo/protección usados en el análisis

<b>Covariables</b>	Modelo nulo	M. 1	M. 2	M. 3	M. 4	M. 5	M. 6	M. 7	M. 8	M. 9	M. 10	M. 11	M. 12	M. 13
Género		*	*	*	*	*	*	*	*	*	*	*	*	*
Antecedentes familiares de asma			*	*	*	*	*	*	*	*	*	*	*	*
Antecedentes familiares de rinitis				*	*	*	*	*	*	*	*	*	*	*
Eccema infantil					*	*	*	*	*	*	*	*	*	*
Fumar durante el embarazo						*	*	*	*	*	*	*	*	*
Resfriados 3 primeros meses							*	*	*	*	*	*	*	*
Asistencia a guardería								*	*	*	*	*	*	*
Lactancia materna >3m									*	*	*	*	*	*
Hermanos											*	*	*	*
Nº de personas en casa												*	*	*
Moho													*	*
Estudios Universitarios de la madre														*
Mascotas en casa														*

Tabla 8. Proceso de estimación secuencial, que consta de 13 modelos más el modelo nulo

Se inicia el análisis ajustando el modelo nulo o modelo vacío de 2 niveles, este es el modelo que solo contiene los efectos del intercepto y el centro. Para ver paso a paso como se ajusta un modelo en MLwiN consultar el manual (117).

```
sibrecuij ~ Binomial(denomij, πij)
logit(πij) = β0jcons
β0j = β0 + u0j

[u0j] ~ N(0, Ωu) : Ωu = [σu02]

var(sibrecuij|πij) = πij(1 - πij)/denomij

(28572 of 28687 cases in use)
```

```
sibrecuij ~ Binomial(denomij, πij)
logit(πij) = β0jcons
β0j = -1.374(0.090) + u0j

[u0j] ~ N(0, Ωu) : Ωu = [0.125(0.046)]

var(sibrecuij|πij) = πij(1 - πij)/denomij

(28572 of 28687 cases in use)
```

Fig. 20. Modelo Nulo EISL.

Podemos calcular el Test de wald para contrastar la hipótesis nula que dice  $\hat{\sigma}_{u0}^2 = 0$  :

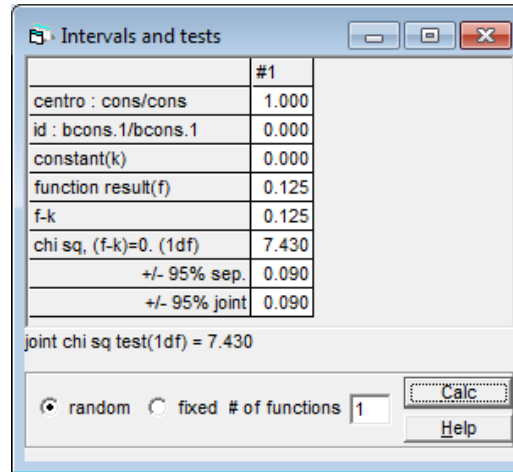


Fig. 21. Salida del MlwiN, que muestra el valor del estadístico del test de Wald para el modelo 0.

$$Z = (0,125/0,046)^2 = 7.38 \sim X^2 \text{ con 1g.l.} \rightarrow p=0,006 \text{ y un p-valor unilateral } p=0.003$$

$$VPC = 0.125 / (0.125 + 3.286) = 0.037$$

Ahora vamos a incorporar en el modelo una variable explicativa, incluimos la variable sexo que toma el valor 0 para niñas y el valor 1 para niños.

**Modelo 1**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + 0.357(0.030)\text{sexo}_1_{ij}$$

$$\beta_{0j} = -1.568(0.092) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.125(0.046)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

(28515 of 28687 cases in use)

Fig. 22. Salida del MlwiN, para el modelo 1.

El coeficiente de partición de la varianza es  $0.125/(0.125+3.286)= 0.037$

A continuación vamos a ir añadiendo una a una las variables explicativas en el modelo.



**Antecedentes de asma familiar:**

**Modelo2**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.355(0.030) \text{sexo\_1}_{ij} + 0.683(0.033) \text{asmfam\_1}_{ij}$$

$$\beta_{0j} = -1.755(0.085) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.104(0.039) \end{bmatrix}$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(28124 of 28687 cases in use)

Fig. 23. Salida del MlwiN, para el modelo 2.

**Antecedentes familiares de rinitis: Modelo 3**

$$\text{sibre}cu_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + 0.353(0.030)\text{sexo}_1_{ij} + 0.612(0.034)\text{asmfam}_1_{ij} + 0.280(0.032)\text{rinfam}_1_{ij}$$

$$\beta_{0j} = -1.850(0.083) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.097(0.036) \end{bmatrix}$$

$$\text{var}(\text{sibre}cu_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(27916 of 28687 cases in use)

Fig. 24. Salida del MlwiN, para el modelo 3.

**Eccema infantil: Modelo 4**

$$\text{sibre}cu_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + 0.346(0.030)\text{sexo}_1_{ij} + 0.591(0.034)\text{asmfam}_1_{ij} + 0.243(0.033)\text{rinfam}_1_{ij} + 0.461(0.032)\text{dermaler}_1_{ij}$$

$$\beta_{0j} = -2.019(0.073) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.069(0.026) \end{bmatrix}$$

$$\text{var}(\text{sibre}cu_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(27719 of 28687 cases in use)

Fig. 25. Salida del MlwiN, para el modelo 4.

**Fumar en el embarazo: Modelo 5**

$$\begin{aligned} \text{sibrecu}_{ij} &\sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_{0j} \text{cons} + 0.346(0.031)\text{sexo}_{1ij} + 0.582(0.035)\text{asmfam}_{1ij} + 0.251(0.033)\text{rinfam}_{1ij} + 0.459(0.032)\text{dermaler}_{1ij} + \\ &\quad 0.433(0.045)\text{tabembar}_{1ij} \\ \beta_{0j} &= -2.078(0.072) + u_{0j} \\ [u_{0j}] &\sim N(0, \Omega_u) : \Omega_u = [0.066(0.025)] \\ \text{var}(\text{sibrecu}_{ij}|\pi_{ij}) &= \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij} \\ &(27633 \text{ of } 28687 \text{ cases in use}) \end{aligned}$$

Fig. 26. Salida del MlwiN, para el modelo 5.

**Tener resfriados en los primeros 3 meses de vida: Modelo 6**

$$\begin{aligned} \text{sibrecu}_{ij} &\sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_{0j} \text{cons} + 0.352(0.032)\text{sexo}_{1ij} + 0.521(0.036)\text{asmfam}_{1ij} + 0.261(0.034)\text{rinfam}_{1ij} + 0.424(0.033)\text{dermaler}_{1ij} + \\ &\quad 0.369(0.047)\text{tabembar}_{1ij} + 0.868(0.031)\text{cold3mos}_{1ij} \\ \beta_{0j} &= -2.395(0.075) + u_{0j} \\ [u_{0j}] &\sim N(0, \Omega_u) : \Omega_u = [0.067(0.026)] \\ \text{var}(\text{sibrecu}_{ij}|\pi_{ij}) &= \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij} \\ &(26597 \text{ of } 28687 \text{ cases in use}) \end{aligned}$$

Fig. 27. Salida del MlwiN, para el modelo 6.

**Asistir a la guardería: Modelo 7**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.354(0.032)\text{sexo}_{1ij} + 0.523(0.036)\text{asmfam}_{1ij} + 0.242(0.034)\text{rinfam}_{1ij} + 0.415(0.033)\text{dermaler}_{1ij} + 0.366(0.047)\text{tabembar}_{1ij} + 0.896(0.032)\text{cold3mos}_{1ij} + 0.762(0.042)\text{scuna}_{1ij}$$

$$\beta_{0j} = -2.549(0.080) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.078(0.030)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(26523 of 28687 cases in use)

Fig. 28. Salida del MlwiN, para el modelo 7.

**Lactancia materna al menos 3 meses: Modelo 8**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.357(0.032)\text{sexo}_{1ij} + 0.529(0.036)\text{asmfam}_{1ij} + 0.239(0.034)\text{rinfam}_{1ij} + 0.406(0.033)\text{dermaler}_{1ij} + 0.339(0.048)\text{tabembar}_{1ij} + 0.891(0.032)\text{cold3mos}_{1ij} + 0.775(0.042)\text{scuna}_{1ij} + -0.223(0.034)\text{lact3mes}_{1ij}$$

$$\beta_{0j} = -2.412(0.089) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.094(0.035)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(26122 of 28687 cases in use)

Fig. 29. Salida del MlwiN, para el modelo 8.

**Nº de personas que viven en la casa: Modelo 9**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.360(0.032)\text{sexo}_{1ij} + 0.526(0.036)\text{asmfam}_{1ij} + 0.244(0.034)\text{rinfam}_{1ij} + 0.400(0.033)\text{dermaler}_{1ij} + 0.324(0.048)\text{tabembar}_{1ij} + 0.880(0.032)\text{cold3mos}_{1ij} + 0.790(0.042)\text{scuna}_{1ij} + -0.220(0.034)\text{lact3mes}_{1ij} + 0.051(0.007)\text{npers}_{ij}$$

$$\beta_{0j} = -2.643(0.095) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.096(0.036)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(26065 of 28687 cases in use)

Fig. 30. Salida del MlwiN, para el modelo 9.

**Nº de hermanos del niño/a: Modelo 10**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.356(0.033)\text{sexo}_{1ij} + 0.519(0.037)\text{asmfam}_{1ij} + 0.230(0.035)\text{rinfam}_{1ij} + 0.410(0.034)\text{dermaler}_{1ij} + 0.299(0.049)\text{tabembar}_{1ij} + 0.876(0.033)\text{cold3mos}_{1ij} + 0.782(0.043)\text{scuna}_{1ij} + -0.221(0.035)\text{lact3mes}_{1ij} + 0.044(0.007)\text{npers}_{ij} + 0.051(0.012)\text{hnos}_{ij}$$

$$\beta_{0j} = -2.658(0.096) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.095(0.036)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(25498 of 28687 cases in use)

Fig. 31. Salida del MlwiN, para el modelo 10.

**Si hay moho en la casa del niño/a: Modelo 11**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.361(0.033)\text{sexo}_{1ij} + 0.518(0.037)\text{asmfam}_{1ij} + 0.219(0.035)\text{rinfam}_{1ij} + 0.391(0.034)\text{dermaler}_{1ij} +$$

$$0.280(0.049)\text{tabembar}_{1ij} + 0.864(0.033)\text{cold3mos}_{1ij} + 0.782(0.043)\text{scuna}_{1ij} + -0.221(0.035)\text{lact3mes}_{1ij} +$$

$$0.041(0.007)\text{npers}_{ij} + 0.048(0.012)\text{hnos}_{ij} + 0.276(0.036)\text{moho}_{1ij}$$

$$\beta_{0j} = -2.694(0.094) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.089(0.034)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(25375 of 28687 cases in use)

Fig. 32. Salida del MlwiN, para el modelo 11.

**Nivel educativo de la madre: Modelo 12**

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.360(0.033)\text{sexo}_{1ij} + 0.513(0.037)\text{asmfam}_{1ij} + 0.231(0.035)\text{rinfam}_{1ij} + 0.387(0.034)\text{dermaler}_{1ij} +$$

$$0.271(0.050)\text{tabembar}_{1ij} + 0.859(0.033)\text{cold3mos}_{1ij} + 0.793(0.043)\text{scuna}_{1ij} + -0.220(0.035)\text{lact3mes}_{1ij} +$$

$$0.040(0.007)\text{npers}_{ij} + 0.041(0.013)\text{hnos}_{ij} + 0.270(0.036)\text{moho}_{1ij} + -0.150(0.037)\text{rec_niv}_{1ij}$$

$$\beta_{0j} = -2.606(0.094) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.082(0.031)]$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(25204 of 28687 cases in use)

Fig. 33. Salida del MlwiN, para el modelo 12.

**MODELO FINAL, MODELO 13, contiene las mismas variables explicativas que el modelo del artículo de factores de riesgo:**

$$\begin{aligned} \text{sibrecu}_{ij} &\sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_{0j} \text{cons} + 0.367(0.033)\text{sexo}_{1ij} + 0.518(0.037)\text{asmfam}_{1ij} + 0.236(0.035)\text{rinfam}_{1ij} + 0.398(0.034)\text{dermaler}_{1ij} + \\ &\quad 0.268(0.050)\text{tabembar}_{1ij} + 0.859(0.033)\text{cold3mos}_{1ij} + 0.789(0.043)\text{scuna}_{1ij} + -0.223(0.035)\text{lact3mes}_{1ij} + \\ &\quad 0.040(0.008)\text{npers}_{ij} + 0.043(0.013)\text{hnos}_{ij} + 0.271(0.036)\text{moho}_{1ij} + -0.143(0.037)\text{rec_niv}_{1ij} + -0.063(0.034)\text{mascot}_{1ij} \\ \beta_{0j} &= -2.597(0.094) + u_{0j} \end{aligned}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.079(0.030) \end{bmatrix}$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(24873 of 28687 cases in use)

Fig. 34. Salida del MlwiN, para el modelo 13.

$$VPC = \frac{0.079}{0.079 + 3.286} = 0.023$$

Podemos calcular el Test de wald para contrastar la hipótesis nula  $\hat{\sigma}_{u_0}^2 = 0$  :

	#1
centro : cons/cons	1.000
id : bcons.1/bcons.1	0.000
constant(k)	0.000
function result(f)	0.079
f-k	0.079
chi sq, (f-k)=0. (1df)	6.953
+/- 95% sep.	0.059
+/- 95% joint	0.059

joint chi sq test(1df) = 6.953

random  
  fixed # of functions

El valor del estadístico es 6.953, el cual se compara con el valor de una variable chi-cuadrado con 1 grado de libertad y se obtiene un p-valor igual a 0.008, si lo dividimos entre 2, obtenemos el p-valor unilateral 0.004 por lo tanto podemos decir que la varianza entre centros es estadísticamente distinta de 0.

La tabla 9 muestra los resultados del análisis de los factores de riesgo de las sibilancias recurrentes en el primer año de vida, utilizando la regresión logística meta-análisis, y la regresión logística multinivel desde el punto de frecuentista y desde el punto de vista bayesiano.



		Regresión logística Meta-análisis (A)		Mod. multinivel logit con intercepto aleatorio			
				(B) PQL-2 (Frecuentista)		(C) MCMC(Bayesiano)	
		ORa	IC95%	ORa	IC95%	ORa	IC95%†
<b>Efectos fijos</b>							
Género	<i>Masc</i>	1.55	1.38-1.75	1.44	1.35-1.54	1.45	1.35-1.54
Antecedentes familiares de asma		1.93	1.69-2.20	1.68	1.56-1.80	1.68	1.56-1.81
Antecedentes familiares de rinitis		1.42	1.25-1.62	1.27	1.18-1.36	1.27	1.18-1.35
Eccema infantil		1.73	1.56-1.92	1.49	1.39-1.59	1.49	1.39-1.59
Fumar durante el embarazo		1.48	1.28-1.72	1.31	1.19-1.44	1.31	1.19-1.44
Resfriados 3 primeros meses		3.14	2.69-3.65	2.36	2.21-2.52	2.36	2.22-2.53
Asistencia a guardería		2.66	2.21-3.21	2.20	2.02-2.39	2.21	2.03-2.41
Lactancia materna >3m		0.79	0.72-0.87	0.80	0.75-0.86	0.80	0.74-0.86
Hermanos		1.07	1.02-1.13	1.04	1.02-1.07	1.04	1.03-1.06
Nº de personas en casa		1.04	1.02-1.07	1.04	1.02-1.06	1.04	1.02-1.07
Moho		1.41	1.25-1.60	1.31	1.22-1.41	1.31	1.21-1.40
Estudios Universitarios de la madre		0.83	0.70-0.94	0.87	0.81-0.93	0.87	0.81-0.94
Mascotas en casa		0.96	0.87-1.05	0.94	0.88-1.00	0.94	0.88-1.00
<i>Intercepto</i>				-2.597	0.094	-2.593	0.099
<b>Efectos aleatorios</b>							
<i>varianza entre centros*</i> $\sigma_{\mu_0}^2$				0.079	0.030	0.108	0.051

Tabla9. Comparación de los resultados de los 3 modelos de análisis. Todas las variables son significativas p<0.001, excepto mascotas.

\*Para el intercepto y los efectos aleatorios se facilita la estimación del parámetro y su error estándar.† Intervalo de credibilidad

**Modelo de pendientes aleatorias: MODELO 14**

El siguiente modelo incluye la pendiente aleatoria para la variable cold3mos

$$\begin{aligned} \text{sibrecu}_{ij} &\sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij}) \\ \text{logit}(\pi_{ij}) &= \beta_0 \text{cons} + \beta_1 \text{sexo}_{1ij} + \beta_2 \text{asmfam}_{1ij} + \beta_3 \text{rinfam}_{1ij} + \beta_4 \text{dermaler}_{1ij} + \beta_5 \text{tabembar}_{1ij} + \beta_6 \text{cold3mos}_{1ij} + \beta_7 \text{scuna}_{1ij} + \\ &\quad \beta_8 \text{lact3mes}_{1ij} + \beta_9 \text{npers}_{ij} + \beta_{10} \text{hmos}_{ij} + \beta_{11} \text{moho}_{1ij} + \beta_{12} \text{rec\_niv}_{1ij} + \beta_{13} \text{mascot}_{1ij} \\ \beta_{0j} &= \beta_0 + u_{0j} \\ \beta_{6j} &= \beta_6 + u_{6j} \end{aligned}$$

$$\begin{bmatrix} u_{0j} \\ u_{6j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u06} & \sigma_{u6}^2 \end{bmatrix}$$

$$\text{var}(\text{sibrecu}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(24873 of 28687 cases in use)

Fig. 36. Salida del MlwiN, para el modelo 13 con una pendiente aleatoria para cold3mos.

Vemos como aparece un subíndice j en el coeficiente  $\beta_6$  y se añade un nuevo término al modelo  $u_{6j}$ .

El resultado es el siguiente:

$$\text{sibrecu}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + 0.367(0.033)\text{sexo}_{1ij} + 0.504(0.038)\text{asmfam}_{1ij} + 0.242(0.036)\text{rinfam}_{1ij} + 0.394(0.035)\text{dermaler}_{1ij} +$$

$$0.260(0.050)\text{tabembar}_{1ij} + \beta_{6j}\text{cold3mos}_{1ij} + 0.800(0.044)\text{scuna}_{1ij} + -0.223(0.035)\text{lact3mes}_{1ij} + 0.041(0.008)\text{npers}_{ij} +$$

$$0.041(0.013)\text{hmos}_{ij} + 0.264(0.036)\text{moho}_{1ij} + -0.129(0.037)\text{rec_niv}_{1ij} + -0.060(0.034)\text{mascot}_{1ij}$$

$$\beta_{0j} = -2.597(0.125) + u_{0j}$$

$$\beta_{6j} = 0.811(0.141) + u_{6j}$$

$$\begin{bmatrix} u_{0j} \\ u_{6j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.183(0.069) & \\ -0.187(0.080) & 0.294(0.112) \end{bmatrix}$$

$$\text{var}(\text{sibrecu}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

(24873 of 28687 cases in use)

Fig. 37. Salida del MlwiN, para el modelo multinivel completamente aleatorizado.

En la tabla 8 se pueden ver los resultados del análisis, en el cual se ha añadido al modelo de intercepto aleatorio una pendiente aleatoria para la variable “cold3mos”.

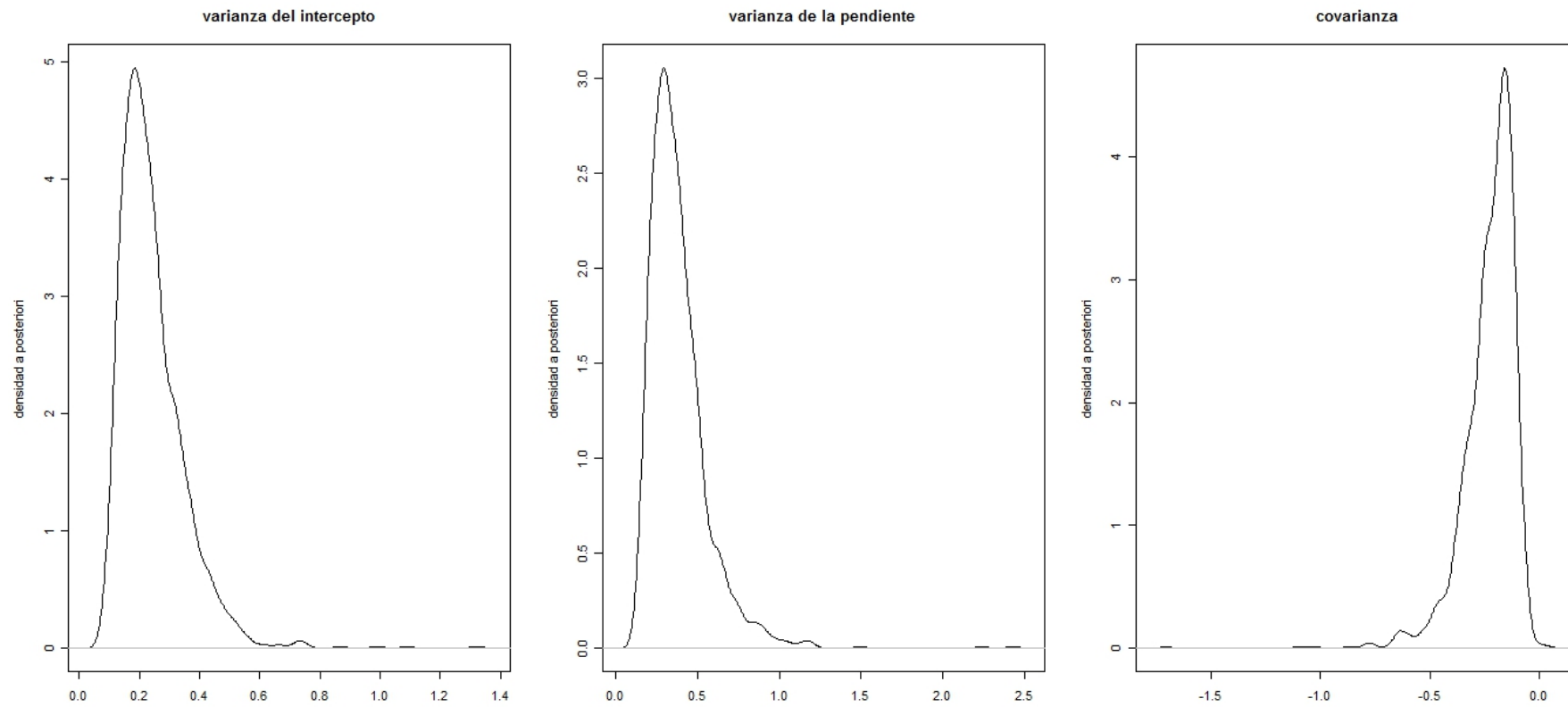


Fig. 38. Representación gráfica de las densidades a posteriori de los 3 parámetros para el modelo con pendiente aleatoria para el modelo multinivel completamente aleatorizado.

		<b>Pendiente aleatoria</b>			
		<b>(1) PQL</b>		<b>(2) MCMC</b>	
		ORa	IC95%	ORa	IC95%†
<b>Efectos fijos</b>					
Género	<i>Masc</i>	<b>1.44</b>	<b>1.35-1.54</b>	<b>1,44</b>	<b>1,35-1,54</b>
Ant. Fam. de asma		<b>1.66</b>	<b>1.54-1.78</b>	<b>1,65</b>	<b>1,53-1,78</b>
Ant. fam. de rinitis		<b>1.27</b>	<b>1.19-1.37</b>	<b>1,27</b>	<b>1,19-1,36</b>
Eccema infantil		<b>1.48</b>	<b>1.38-1.59</b>	<b>1,48</b>	<b>1,38-1,58</b>
Fumar en el embarazo		<b>1.30</b>	<b>1.18-1.43</b>	<b>1,30</b>	<b>1,18-1,43</b>
Asistencia a guardería		<b>2.23</b>	<b>2.04-2.43</b>	<b>2,23</b>	<b>2,05-2,42</b>
Lactancia materna >3m		<b>0.80</b>	<b>0.75-0.86</b>	<b>0,80</b>	<b>0,75-0,86</b>
Hermanos		<b>1.04</b>	<b>1.02-1.07</b>	<b>1,04</b>	<b>1,03-1,06</b>
Nº de personas en casa		<b>1.04</b>	<b>1.03-1.06</b>	<b>1,04</b>	<b>1,02-1,07</b>
Moho		<b>1.30</b>	<b>1.21-1.40</b>	<b>1,30</b>	<b>1,21-1,40</b>
Estudios Univ. de la madre		<b>0.88</b>	<b>0.82-0.95</b>	<b>0,88</b>	<b>0,82-0,95</b>
Mascotas en casa		<b>0.94</b>	<b>0.88-1.01</b>	<b>0,94</b>	<b>0,88-1,01</b>
		$\hat{\beta}_6$	IC95%	$\hat{\beta}_6$	IC95%†
Resf. 3 primeros meses		<b>0.811</b>	<b>0.545-1.087</b>	<b>0.870</b>	<b>0.582-1.139</b>
Intercepto*		-2.597	0.125	-2.642	0.124
<b>Efectos aleatorios*</b>					
varianza del intercepto $\sigma_{\mu_0}^2$		0.183	0.047-0.318	0.242	0.111-0.485
varianza de pendiente $\sigma_{\mu_6}^2$		0.294	0.074-0.513	0.381	0.162-0.833
Cov. del intercepto-pendiente $\sigma_{u06}$		-0.187	-0.343;-0.030	-0.228	-0.519;-0.076

Tabla10. Modelo de Pendiente aleatoria (se deriva del modelo 13) \*En cada modelo para el intercepto, pendiente aleatoria y los efectos aleatorios se facilita la estimación del parámetro y su error estándar. † Intervalo de credibilidad

Podemos comparar los dos modelos llevando a cabo el siguiente test estadístico donde la hipótesis nula es que  $\sigma_{\mu_6}^2$  y  $\sigma_{\mu_{06}}$  son iguales a 0. Para ello usamos la aproximación del test de Wald.

	#1	#2
centro : cons/cons	0.000	0.000
centro : cold3mos_1/cons	1.000	0.000
centro : cold3mos_1/cold3mos_1	0.000	1.000
id : bcons.1/bcons.1	0.000	0.000
constant(k)	0.000	0.000
function result(f)	-0.187	0.294
f-k	-0.187	0.294
chi sq, (f-k)=0. (1df)	5.539	6.839
+/- 95% sep.	0.156	0.220
+/- 95% joint	0.195	0.275

joint chi sq test(2df) = 6.849

random
  fixed # of functions

Fig. 39. Salida del MlwiN, contraste de hipótesis  $H_0: \sigma_{\mu_6}^2 = 0$  y  $\sigma_{\mu_{06}} = 0$

El valor del estadístico es 6.849, el cual se compara con el valor de una variable chi-cuadrado con 2 grados de libertad y se obtiene un p-valor (bilateral) igual a 0.032 y un p-valor unilateral 0.016.

**Añadir una o varias variables explicativas de segundo nivel:**

Cuando aún queda varianza entre centros sin explicar, otra opción es probar a introducir en el modelo final una o más variables contextuales del nivel centro, como es, si el centro es sudamericano o europeo y la latitud del centro y ver si esta explica parte de la variabilidad entre centros.

$sibreca_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$   
 $\text{logit}(\pi_{ij}) = \beta_{0j} \text{cons} + 0.367(0.033)\text{sexo}_{1_{ij}} + 0.517(0.037)\text{asmfam}_{1_{ij}} + 0.235(0.035)\text{rinfam}_{1_{ij}} + 0.395(0.034)\text{dermaler}_{1_{ij}} +$   
 $0.270(0.050)\text{tabembar}_{1_{ij}} + 0.859(0.033)\text{cold3mos}_{1_{ij}} + 0.791(0.043)\text{scuna}_{1_{ij}} + -0.224(0.035)\text{lact3mes}_{1_{ij}} +$   
 $0.040(0.008)\text{npers}_{ij} + 0.043(0.013)\text{hmos}_{ij} + 0.270(0.036)\text{moho}_{1_{ij}} + -0.139(0.037)\text{rec}_{niv}_{1_{ij}} + -0.063(0.034)\text{mascot}_{1_{ij}} +$   
 $0.157(0.158)\text{lat}_{eur}_{1_j}$   
 $\beta_{0j} = -2.706(0.142) + u_{0j}$   
 $[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.075(0.028)]$   
 $\text{var}(sibreca_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$   
 (24873 of 28687 cases in use)

Fig. 40. Salida del MlwiN, para el modelo multinivel de intercepto aleatoria con variable de segundo nivel lat\_eur.

**Modelo de intercepto aleatorio con dos variables de 2° nivel aleatorias: MODELO 15**

$$\text{sibreca}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\begin{aligned} \text{logit}(\pi_{ij}) = & \beta_{0j} \text{cons} + 0.367(0.033)\text{sexo}_{1ij} + 0.519(0.037)\text{asmfam}_{1ij} + 0.236(0.035)\text{rinfam}_{1ij} + 0.396(0.034)\text{dermaler}_{1ij} + \\ & 0.268(0.050)\text{tabembar}_{1ij} + 0.858(0.033)\text{cold3mos}_{1ij} + 0.788(0.043)\text{scuna}_{1ij} + -0.226(0.035)\text{lact3mes}_{1ij} + \\ & 0.040(0.008)\text{npers}_{ij} + 0.043(0.013)\text{hmos}_{ij} + 0.269(0.036)\text{moho}_{1ij} + -0.144(0.037)\text{rec}_{niv}_{1ij} + -0.066(0.034)\text{mascot}_{1ij} + \\ & -0.609(0.251)\text{lat}_{eur}_{1j} + -0.013(0.004)\text{latitud}_j \end{aligned}$$

$$\beta_{0j} = -2.135(0.198) + u_{0j}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [0.040(0.016)]$$

$$\text{var}(\text{sibreca}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij}) / \text{denom}_{ij}$$

(24873 of 28687 cases in use)

Fig. 41. Salida del MlwiN, para el modelo multinivel de intercepto aleatorio con variables de segundo nivel lat\_eur y latitud.

$$\text{VPC} = 0.040 / (0.040 + 3.286) = 0.012$$



<b>Mod. logit con intercepto aleatorio y 2 variables contextuales</b>					
		<b>(B) PQL-2 (Frecuentista)</b>		<b>(C) MCMC(Bayesiano)</b>	
		ORa	IC95%	ORa	IC95%†
<b>Variables de 1° nivel</b>					
Género	<i>Masc</i>	<b>1.443</b>	<b>1.353-1.540</b>	<b>1.443</b>	<b>1.350-1.542</b>
Antecedentes familiares de asma		<b>1.680</b>	<b>1.563-1.807</b>	<b>1.680</b>	<b>1.560-1.809</b>
Antecedentes familiares de rinitis		<b>1.266</b>	<b>1.182-1.356</b>	<b>1.265</b>	<b>1.183-1.354</b>
Eccema infantil		<b>1.486</b>	<b>1.390-1.588</b>	<b>1.486</b>	<b>1.389-1.590</b>
Fumar durante el embarazo		<b>1.307</b>	<b>1.185-1.442</b>	<b>1.310</b>	<b>1.191-1.441</b>
Resfr. 3 primeros meses		<b>2.358</b>	<b>2.211-2.516</b>	<b>2.366</b>	<b>2.214-2.527</b>
Asistencia a guardería		<b>2.199</b>	<b>2.021-2.392</b>	<b>2.201</b>	<b>2.021-2.398</b>
Lactancia materna >3m		<b>0.798</b>	<b>0.745-0.854</b>	<b>0.797</b>	<b>0.744-0.854</b>
Hermanos		<b>1.041</b>	<b>1.025-1.057</b>	<b>1.041</b>	<b>1.025-1.056</b>
N° de personas en casa		<b>1.044</b>	<b>1.018-1.071</b>	<b>1.044</b>	<b>1.018-1.070</b>
Moho		<b>1.309</b>	<b>1.219-1.404</b>	<b>1.307</b>	<b>1.218-1.402</b>
Estudios Univ. De la madre		<b>0.866</b>	<b>0.805-0.931</b>	<b>0.870</b>	<b>0.809-0.935</b>
Mascotas en casa		<b>0.936</b>	<b>0.876-1.001</b>	<b>0.937</b>	<b>0.880-0.998</b>
<i>Intercepto*</i>		-2.135	0.098	0.069	0.036
<b>Variables de 2° nivel</b>					
Latinoamerica vs Europa		<b>0.544</b>	<b>0.333-0.890</b>	<b>0.563</b>	<b>0.278-1.139</b>
Latitud		<b>0.987</b>	<b>0.979-0.995</b>	<b>0.987</b>	<b>0.977-0.998</b>
<i>varianza entre centros*</i>	$\sigma_{\mu_0}^2$	0.040	0.016	0.069	0.036

Tabla 11. Modelos de interceptos aleatorios con las variables continente donde está el centro y latitud del centro. \*En cada modelo para el intercepto y los efectos aleatorios se facilita la estimación del parámetro y su error estándar. † Intervalo de credibilidad

## **6. Discusión**

### ESCAPE:

Los resultados del análisis multinivel para los datos del estudio ESCAPE proporcionan una estimación de la varianza residual entre estaciones del año de 0.0275 con un error estándar de 0.213. Para testar la significación de la varianza del segundo nivel se ha usado el test de Wald, ya que el método de estimación que se ha usado ha sido la cuasi-verosimilitud de segundo orden. El valor del estadístico del test es 1,670 que se corresponde con un p-valor unilateral = 0.110. Por lo tanto, no podemos concluir que exista una variación significativa entre estaciones en la proporción de individuos que se encuentran en el tercil superior de la puntuación global. El objetivo fundamental de la modelización multinivel es analizar los datos teniendo en cuenta la estructura anidada de estos, pero en este caso podemos afirmar que no tiene sentido el análisis multinivel con esta base de datos. Este es un ejemplo de que, aunque mentalmente se pueda intuir una estructura jerárquica de los datos, no siempre es procedente un análisis multinivel.

### EISL:

En el análisis de los datos del estudio EISL “Estudio Internacional de Sibilancias en Lactantes” se ha evaluado inicialmente si tiene sentido tener en cuenta su estructura jerárquica, dado que los datos de los individuos se encuentran anidados en centros.

Por una parte, la variación que existe en la prevalencia de las sibilancias recurrentes en los distintos centros en ambos continentes, por ejemplo Porto Alegre en Brasil tiene una prevalencia del 36.3% frente a un 16.2% de prevalencia en Cartagena (España), sugiere la influencia de factores medioambientales y/o ecológicos regionales o locales, lo que apoya la idea de que es necesario explorar la estructura anidada de los datos, ya que individuos que pertenecen al mismo centro y que tienen una zona de residencia similar, están expuestos a la misma influencia de factores medioambientales y/o ecológicos.

Por otra parte, las diferencias encontradas en la prevalencia de sibilancias son parecidas a las descritas en los estudios ISAAC (“International study of asthma and allergies in childhood”), donde la prevalencia de síntomas de asma actual en niños (6-7 años) y adolescentes (13-14 años) fue significativamente menor en España y Portugal, que en los países de habla hispana y portuguesa de América Latina (133).

Para ajustar el modelo nulo del análisis multinivel frecuentista, hemos usado primero la estimación MQL1, y se han usado las estimaciones MQL como valores de inicio de estimación PQL2, evitando así problemas de convergencia (118).

La OR de tener sibilancias recurrentes en un centro (con  $u_{0j} = 0$ ) se estima como  $\exp(\hat{\beta}_{0j}) = \exp(-1.374) = 0.253$ . El valor del intercepto para un centro  $j$  es  $-1.374 + u_{0j}$ , donde la estimación de la varianza de  $u_{0j}$  es  $\hat{\sigma}_{u_0}^2 = 0.125$  con un error estándar de 0.046. La pregunta subyacente consiste en si es o no necesario permitir que el intercepto sea diferente entre centros. Para ello usamos el test de Wald. Desde el punto de vista teórico no es correcto realizar un test de Wald sobre parámetros de la varianza, pero nos da una indicación acerca de si la varianza del intercepto es ‘importante’ (24). El test nos indica que  $\hat{\sigma}_{u_0}^2$  es estadísticamente distinto de 0, por lo tanto, podemos concluir que existe variación significativa entre centros que tienen sibilancias recurrentes. Y en este caso sí que tiene sentido realizar un análisis jerárquico de dos niveles.

Como se ha visto en la parte teórica, el coeficiente de correlación intraclase se estima como el cociente entre la varianza entre grupos y la varianza total. Como en un modelo logístico la varianza total no está directamente disponible, una alternativa es usar la ecuación propuesta por Tom Snijders et al (73). Si aplicamos esta ecuación a los datos del modelo nulo, obtenemos un valor de 0.037 para el coeficiente de

partición de la varianza. Por lo que, el 3.7% de la varianza residual de la sibilancias recurrentes es atribuible a las diferencias existentes entre centros.

Al añadir al modelo la variable explicativa sexo, que toma el valor 0 para niñas y el valor 1 para niños, no ha habido ningún cambio en la estimación de la varianza entre centros, lo que sugiere que la distribución de niños '1' y niñas '0' es similar entre centros. El coeficiente de partición de la varianza es 0.037, el 3.7% de la variación residual en la propensión a tener sibilancias recurrentes es atribuible a características de la variable centro desapercibidas. El coeficiente de la variable explicativa es 0.0357 y es significativamente distinto de 0 con un p-valor < 0.001.

Al introducir predictores en el primer nivel, es muy frecuente que disminuya la varianza residual en el segundo nivel, ya que es posible que la variable explicativa no tenga la misma distribución en todas las unidades de segundo nivel (2;73). Por ello, al introducir el resto de variables explicativas se observan variaciones en la varianza entre centros. En el caso de la variable "Antecedentes de asma familiar" la varianza entre centros ha bajado de 0.125 a 0.104.

El Modelo final es el que contiene las mismas variables explicativas que el artículo de García-Marcos et al. (107) y se corresponde con el modelo 13. El valor del coeficiente de partición de la varianza ha pasado de 0.037 a 0.023, este resultado muestra que aproximadamente el 2.3% de la varianza residual de las sibilancias recurrentes es atribuible las diferencias entre centros, después de controlar por las variables explicativas.

En todos los modelos obtenemos que los dos factores de riesgo más elevados han sido la asistencia a la guardería y tener resfriados en los 3 primeros meses de vida. Otros factores de riesgo han sido el ser varón, haber fumado durante el embarazo, tener moho en el domicilio del niño,

número de hermanos del niño y número de personas que habitan en el domicilio del niño. La lactancia materna durante más de 3 meses y que la madre tenga estudios universitarios han sido factores de protección. El único efecto fijo cuyo coeficiente no es significativo ha sido tener mascotas en casa.

En los 3 modelos, el sexo masculino es un factor de riesgo de tener sibilancias recurrentes en el primer año de vida. Como se ha visto también en otros estudios (106;107;134) los niños son más susceptibles a padecer asma en edades tempranas.

Es interesante la asociación de los antecedentes familiares de asma y rinitis con las sibilancias recurrentes, similares en los 3 modelos. En cuanto a la influencia de la historia familiar de asma o rinitis no existe una evidencia concluyente, pero esta asociación refuerza los resultados obtenidos en estudios previos (133;135).

Tener eccema infantil también se asocia positivamente con tener sibilancias recurrentes, este resultado es congruente con lo encontrado por Martínez et al. (136).

En todos los modelos se observa que fumar en el embarazo es un riesgo para desarrollar sibilancias recurrentes en el primer año de vida. El hábito tabáquico durante el embarazo se relaciona con el bajo peso al nacimiento (137). Lo que se relaciona con una peor función pulmonar e incrementa el riesgo de sibilancias durante los primeros meses de vida (138).

Está descrito en la literatura que las enfermedades virales respiratorias severas y la atopia son factores de riesgo para las sibilancias en la infancia (139;140). Con los tres métodos de análisis se obtiene que 'tener resfriados en los 3 primeros meses de vida' es el mayor factor de riesgo para las sibilancias, resultado que está en consonancia con lo publicado hasta el momento. Los factores ambientales que están relacionados con la exposición a virus, como son la asistencia a la guardería y el tener hermanos, influyen en la prevalencia de tener

sibilancias. Herr M. et al. (141) Y Stern D.A. et al. (142) también llegaron a estas conclusiones en sus respectivos estudios.

La lactancia materna más allá de los 3 primeros meses, es un factor de protección. El papel protector que juega la lactancia materna es ya conocido y hay una amplia bibliografía al respecto (143;144).

El tener moho en la casa donde vive el niño es otro factor de riesgo de tener sibilancias recurrentes en el primer año de vida. Los niños pasan la mayor parte del tiempo en el interior, tanto en sus casas como en las escuelas. Los hongos de interior son importantes contaminantes biológicos que afectan a la salud de los niños (145-147).

En la tabla 9 se han mostrado los resultados de los distintos análisis estadísticos llevados a cabo. Desde el punto de vista del enfoque frecuentista hemos analizado los datos usando un meta-análisis de efectos aleatorios (modelo A) y un análisis multinivel de 2 niveles mediante el algoritmo de la cuasi-verosimilitud penalizada de 2º orden (modelo B). Desde el punto de vista del enfoque bayesiano hemos usado un análisis multinivel de 2 niveles mediante el Método Monte Carlo por Cadenas de Markov (C), para los efectos fijos se ha usado una distribución a priori normal de media 0 y varianza 10000. Para la desviación típica del efecto aleatorio se ha usado una uniforme en el intervalo (0,100).

Las estimaciones del modelo A difieren ligeramente del modelo B y C, el primer modelo obtiene estimaciones más pequeñas para los  $\beta$  de casi todas las variables excepto para las variables lactancia materna, nº de personas en casa y estudios universitarios de la madre que permanecen prácticamente invariables. Si nos fijamos en los intervalos de confianza al 95% para el enfoque frecuentista y en los intervalos de credibilidad al 95% para el enfoque bayesiano vemos que los dos resultados de los modelos multinivel proporcionan intervalos más estrechos, lo que significa que el análisis multinivel mejora la precisión de las

estimaciones. Además de que estos modelos proporcionan el porcentaje de la varianza residual de las sibilancias recurrentes que es atribuible a las diferencias entre centros, después de controlar por las variables explicativas.

Rodriguez G. y Goldman N. (148) realizaron un estudio en el que se planteaban el coste de ignorar la estructura jerárquica de los datos en modelos con variable respuesta binaria. En este se comparaba una regresión logística ordinaria con una regresión multinivel con intercepto aleatorio (enfoque frecuentista). En este estudio mostraban que las estimaciones de los efectos fijos  $\beta_p$  eran bastante similares entre los dos modelos, en cambio, sí que había diferencia en la estimación de los errores estándar. Breslow y Clayton (118) obtuvieron resultados similares con anterioridad.

Si comparamos el modelo B, representante del enfoque frecuentista frente al modelo C, representante del enfoque bayesiano, observamos que las estimaciones de las odds ratio de los efectos fijos son prácticamente iguales, tanto en las estimaciones de los coeficientes como las estimaciones de los errores estándar, lo que se traduce también en unos intervalos de confianza en muchos de los casos iguales. Esta coincidencia de resultados es de esperar debido al gran tamaño muestral de las unidades de primer nivel.

En cambio la estimación de la varianza entre centros es ligeramente mayor en el enfoque bayesiano 0.108 frente a 0.079 del enfoque frecuentista. En este caso es difícil decir qué estimación es mejor, si la del enfoque frecuentista que usa el algoritmo PQL-2 o el enfoque bayesiano que usa el método MCMC. Goldstein y Rasbash (149) en 1996 usaron los datos de Rodriguez G. y Goldman N para simular 200 conjuntos de datos, partieron del conocimiento del verdadero valor de los parámetros

$$\beta_0 = 0.665, \beta_1 = 1.0, \beta_2 = 1.0, \beta_3 = 1.0, \sigma_{u_2} = 1.0, \sigma_{u_3} = 1.0$$



y obtuvieron que las estimaciones PQL de segundo orden se aproximaban considerablemente al verdadero valor de las estimaciones, excepto en la estimación de la varianza de segundo nivel donde había una sub-estimación del 20%:

$$\beta_0 = 0.660, \beta_1 = 0.965, \beta_2 = 0.968, \beta_3 = 1.002, \sigma_{u_2} = 0.802, \quad \sigma_{u_3} = 0.968$$

Engel B. en 1998 (150) sugirió que antes problemas reales, el método de Montecarlo por Cadenas de Markov puede producir estimaciones más aceptables aunque hay que tener cuidado en la elección de las distribuciones a priori de los componentes de la varianza (151).

Browne W. y Draper D. (119) en 2006 realizaron un estudio de comparación entre el método bayesiano y el método basado en la máxima verosimilitud para ajustar modelos multinivel logísticos de efectos aleatorios. Para ello se basaron, como Goldstein et al., en los datos utilizados por Rodríguez G. y Goldman N. (148) y partiendo de los verdaderos valores de los parámetros generaron 500 muestras replicando las estructura de los datos de Rodríguez y Goldman. En este caso los resultados para un modelo de efectos aleatorios con 3 niveles y una pendiente fija y dos pendientes aleatorias fueron:

Método	$\beta_0 = 0.66$	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$	$\sigma_{u_2} = 1$	$\sigma_{u_3} = 1$
<b>PQL-2</b>	0.612 (0.009)	0.945 (0.009)	0.958 (0.005)	0.942 (0.011)	0.888 (0.009)	0.568 (0.010)
<b>Bayesiano</b>	0.655	1.015	1.031	0.982	1.023	1.130
<b>U(0,∞)</b>	(0.010)	(0.010)	(0.005)	(0.013)	(0.011)	(0.016)

Browne W. y Draper D. (119) concluyen que para modelos de efectos aleatorios de 2 niveles ambos métodos, el basado en la verosimilitud (PQL-2) y el método bayesiano, proporcionan estimaciones aproximadamente insesgadas y ambos experimentan dificultades a la hora de obtener los intervalos de confianza/credibilidad en muestras pequeñas, aunque el algoritmo PQL-2 es mejor en esos casos. Y para modelos de 3 niveles las estimaciones obtenidas con el método PQL-2 son sesgadas, sobre todo la estimación de la varianza de segundo nivel y los métodos bayesianos proporcionan estimaciones calibradas para todos los parámetros del modelo. Esto nos hace pensar que sería interesante para nuestro estudio de las sibilancias recurrentes aumentar el número de niveles.

Una de las principales ventajas del procedimiento de estimación de la quasi-verosimilitud penalizada, es que incluso con bases de datos muy grandes y un gran número de parámetros la carga computacional es ligera frente al método MCMC que tiene un coste computacional muy elevado, y el tiempo de cálculo aumenta con la complejidad del modelo. Pero gracias a mejoras en la velocidad del hardware y en la eficiencia de los algoritmos de Monte Carlo, el enfoque bayesiano para los modelos multinivel se ha convertido en una herramienta muy interesante.

Después de utilizar un modelo de interceptos aleatorios se ha planteado el modelo 14 como respuesta a la pregunta de si es o no necesario permitir que la relación entre los resfriados y las sibilancias recurrentes sea diferente para los centros. O en otras palabras ¿es necesario permitir una pendiente aleatoria para “tener resfriados los primeros 3 meses de vida”. Al permitir que la pendiente de esta variable sea aleatoria obtenemos un modelo completamente aleatorio. Con este modelo se parte del supuesto que los centros variarán tanto en la probabilidad de tener sibilancias recurrentes (variación aleatoria de los puntos de corte) como en la relación que se establece en cada centro

entre tener resfriados los 3 primeros meses y la probabilidad de tener sibilancias recurrentes (variación aleatoria de las pendientes) (2).

La extensión de un modelo de intercepto aleatorio a uno de pendientes aleatorias introduce dos nuevos términos en la parte aleatoria del modelo:

$\sigma_{\mu_6}^2$ , *variación entre pendientes*

$\sigma_{\mu_06}$ , *covarianza entre el intercepto y la pendiente.*

Podemos ver en la tabla 10 los resultados de los factores de riesgo/protección de las sibilancias recurrentes obtenidos tanto con el enfoque frecuentista como con el enfoque bayesiano. Como en el caso del modelo de interceptos aleatorios las estimaciones de los efectos fijos son similares en los dos enfoques.

En el enfoque frecuentista para comprobar si tiene sentido permitir la pendiente aleatoria para la variable “*resfriados en los 3 primeros meses de vida*” podemos llevar a cabo el siguiente test estadístico donde la hipótesis nula es que  $\sigma_{\mu_6}^2$  y  $\sigma_{\mu_06}$  son iguales a 0. Usamos la aproximación del test de Wald y teniendo en cuenta que  $\sigma_{\mu_6}^2$  no puede ser negativo, se obtiene un p-valor unilateral de 0.016 y por lo tanto podemos concluir que el efecto de cold3mes varía de verdad entre centros.

En el enfoque bayesiano para comprobar si tiene sentido permitir la pendiente aleatoria para la variable “*resfriados en los 3 primeros meses de vida*” debemos observar la figura 38 donde se ve la representación gráfica de las densidades a posteriori de los 3 parámetros. Como los intervalos de credibilidad no contienen al 0 entonces podemos afirmar que tiene sentido permitir la pendiente aleatoria para esta variable.

Para un centro cualquiera  $j$ , el efecto de haber tenido resfriados en los tres primeros meses de vida sobre el log-odds de tener sibilancias recurrentes es  $0.811 + \mu_{6j}$  para el modelo 1 y  $0.870 + \mu_{6j}$  para el modelo 2. Dependiendo del valor de  $\mu_{6j}$ , las diferencias en un centro dado serán mayores o menores que 0.811.

Para la interpretación de la covarianza, lo más importante es el signo. En el modelo 1 y 2 el signo es negativo, lo que indica una relación inversa entre el intercepto aleatorio y la pendiente aleatoria y se traduce en que para un centro con un intercepto relativamente alto tendrá una pendiente relativamente baja.

El enfoque bayesiano proporciona estimaciones de los efectos aleatorios distintas al del enfoque frecuentista como hemos ya citado Goldstein y Rasbash (149) afirman en su estudio que el algoritmo PQL-2 subestima la estimación de la varianza, por ello nos podríamos quedar con los resultados del enfoque bayesiano.

Uno de los mayores beneficios de usar el análisis multinivel es la posibilidad de incluir conjuntamente variables del primer nivel y del segundo nivel. En algunos casos podemos estar interesados en estudiar si las variables de segundo nivel pueden explicar la variación entre centros, es decir, es posible que otra parte de la diferencia entre las probabilidades de tener sibilancias recurrentes de los centros se deba a que cada uno de ellos tiene distintas características.

La inclusión de las variables contextuales, latitud del centro y si el centro está en Latino-América o en Europa, conlleva una sustancial reducción de la varianza entre centros. En el modelo B (enfoque frecuentista) de la tabla 9 y de la tabla 11, vemos que ha pasado de 0.079 a 0.040, y en el modelo C de la tabla 9 y de la tabla 11, vemos que ha pasado de 0.108 a 0.069, lo que significa que la variación entre centros en la probabilidad de tener sibilancias recurrentes el primer año de vida se asocia con la latitud del centro y si está en Europa o en Latino-América en ambos enfoques.

## **7. Conclusiones**

## CONCLUSIONES

1° Los modelos multinivel mejoran la precisión de las estimaciones de los parámetros respecto a los modelos que no tienen en cuenta la estructura jerárquica de los datos.

2° Los modelos multinivel aportan información adicional acerca del porcentaje de varianza residual que explican los centros.

3° El análisis multinivel mantiene la significación estadística de los factores de riesgo/protección encontrados con el modelo que no ha tenido en cuenta la estructura jerárquica de los datos.

4° La inclusión de las variables contextuales o variables de segundo nivel (latitud, y Latino-América vs. Europa) mejora el porcentaje de la varianza residual de tener sibilancias recurrentes en el primer año de vida que es debido a las diferencias entre centros.

5° El enfoque bayesiano proporciona resultados numéricos similares, en cuanto a las estimaciones de los parámetros y la precisión de las mismas.

6° Para futuros estudios, se sugiere ampliar el número de unidades de segundo nivel y el número de niveles, para corroborar que el enfoque bayesiano mejora el enfoque frecuentista.

## **8. Bibliografía**

- (1) Kreft I, De Leeuw J, Kim KS. Comparing Four Statistical Packages for Hierarchical Linear Regression. CA UoCP, editor. 311. 1990. Los Ángeles.

Ref Type: Report

- (2) Gaviria-Soto JL, Castro-Morera MC. Modelos Jerárquicos Lineales. Madrid: 2005.

- (3) Golstein H. Multilevel Statistical Models. Hodder Arnold, editor. 3. 2003.

Ref Type: Serial (Book, Monograph)

- (4) De Leeuw J, Kreft I. Software for Multilevel Analysis. <http://escholarship.org/uc/item/2xn1n7df> . 2011. UC Los Angeles: Departments of statistics, UCLA.

Ref Type: Electronic Citation

- (5) Bennett N. Teaching styles and pupil progress. London: 1976.

- (6) Aitkin M, Longford N. Statistical Modelling Issues in School Effectiveness Studies. Journal of the Royal Statistical Society Series A (General) 1986; 149:1-43.

- (7) Moerbeek M. The consequence of ignoring a level of nesting in multilevel analysis. Multivariate Behavioral Research 2004; 39(1):129-149.

- (8) Robinson WS. Ecological Correlations and the Behavior of Individuals. American Sociological Review 1950;(15):351-357.

- (9) Alker H. A typology of ecological fallacies. In: Dogan M, Rokkam S e, editors. Social ecology. Boston: The MIT Press, 1969: 69-86.

- (10) Diez Roux AV, Schwartz S, Susser E. Ecologic studies and ecologic variables in public health research. In: Oxford University Press, editor. *The Oxford textbook of public health*. London: 2002: 493-509.

- (11) Lindley DV, Novick MR. The Role of Exchangeability in Inference. Annals of Statistics 1981; 9(1):45-58.

- (12) Hox JJ. Applied multilevel analysis. Amsterdam: 1995.

- (13) Holt D, Scott AJ, Ewings PD. Chi-Squared Tests with Survey Data. Journal of the Royal Statistical Society Series A (General) 1980; 143(3):303-320.



- (14) Airy GB. On the Algebraic and Numerical Theory of Errors of Observations and the Combination of Observations. London: 1861.
- (15) Fisher RA. Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 1925; 22:700-725.
- (16) Gibbons R, Bock R. Trend in correlated proportions. Psychometrika 1987; 52(1):113-124.
- (17) Burstein L, Linn RL, Capell FJ. Analyzing Multilevel Data in the Presence of Heterogeneous Within-Class Regressions. Journal of Educational and Behavioral Statistics 1978; 3(4):347-383.
- (18) Wald A. A Note on Regression Analysis. The Annals of Mathematical Statistics 1947; 18(4):586-589.
- (19) Rubin H. Statistical Inference in Dynamic Economic Models. New York: Wiley, 1950.
- (20) Kreft I, De Leeuw J. Introducing multilevel modeling. London: 1998.
- (21) Browne W, Golstein H, Rabash L. Multiple membership multiple classification (MMMC) models. Statistical Modelling 2001; 1:103-124.
- (22) Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Soc Sci Med 1998; 46(1):97-117.
- (23) Diez Roux AV. A glossary for multilevel analysis. J Epidemiol Community Health 2002; 56(8):588-594.
- (24) Twisk JWR. Applied multilevel Analysis. Cambridge: 2006.
- (25) Moellering H, Tobler W. Geographical Variances. Geographical Analysis 1972; 4(1):34-50.
- (26) Griew P, Page A, Thomas S, Hillsdon M, Cooper AR. The school effect on children's school time physical activity: the PEACH Project. Prev Med 2010; 51(3-4):282-286.
- (27) Wium N, Burgess S, Moore L. Brief report: multilevel analysis of school smoking policy and pupil smoking behaviour in Wales. J Adolesc 2011; 34(2):385-389.
- (28) Saab H, Klinger D. School differences in adolescent health and wellbeing: findings from the Canadian Health Behaviour in School-aged Children Study. Soc Sci Med 2010; 70(6):850-858.

- (29) Wold B, Torsheim T, Currie C, Roberts C. National and school policies on restrictions of teacher smoking: a multilevel analysis of student exposure to teacher smoking in seven European countries. *Health Educ Res* 2004; 19(3):217-226.
- (30) Boerboom TB, Mainhard T, Dolmans DH, Scherpbier AJ, Van Beukelen P, Jaarsma AD. Evaluating clinical teachers with the Maastricht clinical teaching questionnaire: how much 'teacher' is in student ratings? *Med Teach* 2012; 34(4):320-326.
- (31) Opdenakker MC, Minnaert A. Relationship between learning environment characteristics and academic engagement. *Psychol Rep* 2011; 109(1):259-284.
- (32) Ireson J, Hallam S, Plewis I. Ability grouping in secondary schools: effects on pupils' self-concepts. *Br J Educ Psychol* 2001; 71(Pt 2):315-326.
- (33) Prince SA, Kristjansson EA, Russell K, Billette JM, Sawada M, Ali A et al. A multilevel analysis of neighbourhood built and social environments and adult self-reported physical activity and body mass index in Ottawa, Canada. *Int J Environ Res Public Health* 2011; 8(10):3953-3978.
- (34) Merlo J, Viciano-Fernandez FJ, Ramiro-Farinas D. Bringing the individual back to small-area variation studies: A multilevel analysis of all-cause mortality in Andalusia, Spain. *Soc Sci Med* 2012.
- (35) Jackson AL, Davies CA, Leyland AH. Do differences in the administrative structure of populations confound comparisons of geographic health inequalities? *BMC Med Res Methodol* 2010; 10:74.
- (36) Gallego B, Westbrook MT, Dunn AG, Braithwaite J. Investigating patient safety culture across a health system: multilevel modelling of differences associated with service types and staff demographics. *Int J Qual Health Care* 2012; 24(4):311-320.
- (37) Bentley R, Jolley D, Kavanagh AM. Local environments as determinants of walking in Melbourne, Australia. *Soc Sci Med* 2010; 70(11):1806-1815.
- (38) Antai D, Wedren S, Bellocco R, Moradi T. Migration and child health inequities in Nigeria: a multilevel analysis of contextual- and individual-level factors. *Trop Med Int Health* 2010; 15(12):1464-1474.

- (39) Haines VA, Beggs JJ, Hurlbert JS. Neighborhood disadvantage, network social capital, and depressive symptoms. *J Health Soc Behav* 2011; 52(1):58-73.
- (40) McMurdo ME, Argo I, Crombie IK, Feng Z, Sniehotta FF, Vadiveloo T et al. Social, environmental and psychological factors associated with objective physical activity levels in the over 65s. *PLoS One* 2012; 7(2):e31878.
- (41) Jaime Castillo JM. Actitudes cívicas y dimensiones de la ciudadanía democrática en Europa. *Revista Española de Investigaciones Sociológicas* 2009; 125:47-80.
- (42) Diehl K, Schneider S. How relevant are district characteristics in explaining subjective health in Germany? A multilevel analysis. *Soc Sci Med* 2011; 72(7):1205-1210.
- (43) Baydar N, Kieckhefer G, Joesch JM, Greek A, Kim H. Changes in the health burden of a national sample of children with asthma. *Soc Sci Med* 2010; 70(2):321-328.
- (44) Dragano N, Bobak M, Wege N, Peasey A, Verde PE, Kubinova R et al. Neighbourhood socioeconomic status and cardiovascular risk factors: a multilevel analysis of nine cities in the Czech Republic and Germany. *BMC Public Health* 2007; 7:255.
- (45) Bentley R, Kavanagh AM, Subramanian SV, Turrell G. Area disadvantage, individual socio-economic position, and premature cancer mortality in Australia 1998 to 2000: a multilevel analysis. *Cancer Causes Control* 2008; 19(2):183-193.
- (46) Fukuda Y, Nakaya T, Nakao H, Yahata Y, Imai H. Multilevel analysis of solar radiation and cancer mortality using ecological data in Japan. *Biosci Trends* 2008; 2(6):235-240.
- (47) Meersman SC, Breen N, Pickle LW, Meissner HI, Simon P. Access to mammography screening in a large urban population: a multi-level analysis. *Cancer Causes Control* 2009; 20(8):1469-1482.
- (48) Robertson R, Campbell NC, Smith S, Donnan PT, Sullivan F, Duffy R et al. Factors influencing time from presentation to treatment of colorectal and breast cancer in urban and rural areas. *Br J Cancer* 2004; 90(8):1479-1485.
- (49) King MT, Kenny P, Shiell A, Hall J, Boyages J. Quality of life three months and one year after first treatment for early stage breast cancer: influence of treatment and patient characteristics. *Qual Life Res* 2000; 9(7):789-800.

- (50) Ludt S, Wensing M, Szecsenyi J, van Lieshout J, Rochon J, Freund T et al. Predictors of health-related quality of life in patients at risk for cardiovascular disease in European primary care. *PLoS One* 2011; 6(12):e29334.
- (51) Verpooten GA, Aerts A, Coen N, Vancayzeele S, Hermans C, Bowles J et al. Antihypertensive effectiveness of aliskiren for the 'real-world' management of hypertension: multilevel modelling of 180-day blood pressure outcomes (the Belgian DRIVER Study). *Int J Clin Pract* 2011; 65(1):54-63.
- (52) Kavanagh A, Bentley RJ, Turrell G, Shaw J, Dunstan D, Subramanian SV. Socioeconomic position, gender, health behaviours and biomarkers of cardiovascular disease and diabetes. *Soc Sci Med* 2010; 71(6):1150-1160.
- (53) Sundquist J, Malmstrom M, Johansson SE. Cardiovascular risk factors and the neighbourhood environment: a multilevel analysis. *Int J Epidemiol* 1999; 28(5):841-845.
- (54) Hallman DM, Friedel VC, Eissa MA, Boerwinkle E, Huber JC, Jr., Harrist RB et al. The association of variants in the FTO gene with longitudinal body mass index profiles in non-Hispanic white children and adolescents. *Int J Obes (Lond)* 2012; 36(1):61-68.
- (55) Lee MH, Yang HI, Jen CL, Lu SN, Yeh SH, Liu CJ et al. Community and personal risk factors for hepatitis C virus infection: a survey of 23,820 residents in Taiwan in 1991-2. *Gut* 2011; 60(5):688-694.
- (56) Finan PH, Zautra AJ, Davis MC, Lemery-Chalfant K, Covault J, Tennen H. Genetic influences on the dynamics of pain and affect in fibromyalgia. *Health Psychol* 2010; 29(2):134-142.
- (57) Pattenden S, Armstrong BG, Houthuijs D, Leonardi GS, Dusseldorp A, Boeva B et al. Methodological approaches to the analysis of hierarchical studies of air pollution and respiratory health--examples from the CESAR study. Central European Study on Air pollution and Respiratory Health. *J Expo Anal Environ Epidemiol* 2000; 10(5):420-426.
- (58) Pikhart H, Prikazsky V, Bobak M, Kriz B, Celko M, Danova J et al. Association between ambient air concentrations of nitrogen dioxide and respiratory symptoms in children in Prague, Czech Republic. Preliminary results from the Czech part of the SAVIAH Study. Small Area Variation in Air Pollution and Health. *Cent Eur J Public Health* 1997; 5(2):82-85.

- (59) Dahal GP, Johnson FA, Padmadas SS. Maternal smoking and acute respiratory infection symptoms among young children in Nepal: multilevel analysis. *J Biosoc Sci* 2009; 41(6):747-761.
- (60) Campbell SM, Hann M, Hacker J, Burns C, Oliver D, Thapar A et al. Identifying predictors of high quality care in English general practice: observational study. *BMJ* 2001; 323(7316):784-787.
- (61) Francis NA, Butler CC, Hood K, Simpson S, Wood F, Nuttall J. Effect of using an interactive booklet about childhood respiratory tract infections in primary care consultations on reconsulting and antibiotic prescribing: a cluster randomised controlled trial. *BMJ* 2009; 339:b2885.
- (62) Patel MM, Chillrud SN, Correa JC, Hazi Y, Feinberg M, Kc D et al. Traffic-related particulate matter and acute respiratory symptoms among New York City area adolescents. *Environ Health Perspect* 2010; 118(9):1338-1343.
- (63) Pefoyo AJ, Laurier C, Rivard M. [Determinants of geographic variations in asthma medication in Quebec]. *Rev Mal Respir* 2010; 27(1):49-62.
- (64) Pikhart H, Bobak M, Kriz B, Danova J, Celko MA, Prikazsky V et al. Outdoor air concentrations of nitrogen dioxide and sulfur dioxide and prevalence of wheezing in school children. *Epidemiology* 2000; 11(2):153-160.
- (65) Sternthal MJ, Jun HJ, Earls F, Wright RJ. Community violence and urban childhood asthma: a multilevel analysis. *Eur Respir J* 2010; 36(6):1400-1409.
- (66) Suh M, Kim HH, Choi DP, Kim KW, Sohn MH, Ha KH et al. Association between body mass index and asthma symptoms among Korean children: a nation-wide study. *J Korean Med Sci* 2011; 26(12):1541-1547.
- (67) Sunyer J, Atkinson R, Ballester F, Le Tertre A, Ayres JG, Forastiere F et al. Respiratory effects of sulphur dioxide: a hierarchical multicity analysis in the APHEA 2 study. *Occup Environ Med* 2003; 60(8):e2.
- (68) Tatalovich Z, Wilson JP, Milam JE, Jerrett ML, McConnell R. Competing definitions of contextual environments. *Int J Health Geogr* 2006; 5:55.
- (69) Ungvari I, Hullam G, Antal P, Kiszal PS, Gezsi A, Hadadi E et al. Evaluation of a partial genome screening of two asthma

- susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. PLoS One 2012; 7(3):e33573.
- (70) Villamizar LA, Lopez AB, Ortiz HC, Velazquez JN, Cala LM. [Incidence of respiratory symptoms and the association with air pollution in preschoolers: a multilevel analysis]. Cad Saude Publica 2010; 26(7):1411-1418.
- (71) Zuidgeest MG, van Dijk L, Spreeuwenberg P, Smit HA, Brunekreef B, Arets HG et al. What drives prescribing of asthma medication to children? A multilevel population-based study. Ann Fam Med 2009; 7(1):32-40.
- (72) De la Cruz F. Modelos Multinivel. Rev Per Epidemiol 2008; 12(3):1-8.
- (73) Snijdders TAB, Bosker RJ. Multilevel Analysis. An introduction to basic and advanced multilevel modeling. London: 1999.
- (74) Hox JJ. Multilevel analysis. Techniques and applications. New Jersey: 2002.
- (75) ATKINSON AC. Masking unmasked. Biometrika 1986; 73(3):533-541.
- (76) Barnett V, Lewis T. Outliers in Statistical Data. 3 ed. 1994.
- (77) Lawrence AJ. Deletion Influence and Masking in Regression. Journal of the Royal Statistical Society Series B (Methodological) 1995; 57(1):181-189.
- (78) Edwards AWF. Likelihood. 1972.
- (79) Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 1943; 54(3):426-482.
- (80) Akaike H. [Data analysis by statistical models]. No To Hattatsu 1992; 24(2):127-133.
- (81) Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics 1978; 6(2):461-464.
- (82) Golstein H. Nonlinear multilevel models, with an application to discrete response data. Biometrika 1991; 78(1):45-51.

- (83) Bryk AS, Raudenbush SW. Hierarchical linear models: Applications and data analysis methods. Newbury park, CA: Sage Publications, 1992.
- (84) Browne WJ, Subramanian SV, Jones K, GOLDSTEIN H. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2005; 168(3):599-613.
- (85) GOLDSTEIN H, Browne W, Rasbash J. Partitioning Variation in Multilevel Models. *Understanding Statistics* 2002; 1(4):223-231.
- (86) GOLDSTEIN H, Browne W, Rasbash J. Partitioning Variation in Multilevel Models. *Understanding Statistics* 2002; 1(4):223-231.
- (87) SNIJDERS TAB, BOSKER RJ. Modeled Variance in Two-Level Models. *Sociological Methods & Research* 1994; 22(3):342-363.
- (88) Van der leeden R. Multilevel analysis of repeated measures data. *Quality and Quantity* 1998; 32:15-29.
- (89) Golstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Statist Med* 2002; 21(21):3291-3315.
- (90) Rasbash J, Golstein H. Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model. *Journal of Educational and Behavioral Statistics* 1994; 19(4):337-350.
- (91) Raudenbush SW. A Crossed Random Effects Model for Unbalanced Data with Applications in Cross-Sectional and Longitudinal Research. *Journal of Educational Statistics* 1993; 18(4):321-349.
- (92) Goldstein H, Yang M, Omar R, Turner R, Thompson S. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2000; 49(3):399-412.
- (93) Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Statist Med* 2002; 21(21):3291-3315.
- (94) Golstein H. *Handbook of Multilevel analysis*. New York: Springer, 2007.
- (95) Hulley SB. *Diseño de la investigación clínica. Un enfoque epidemiológico*. Barcelona: 1993.

- (96) Perez-Fernandez V. Sampling and questionnaires. *Allergol Immunopathol (Madr)* 2008; 36(4):228-233.
- (97) SNIJDERS TAB, BOSKER RJ. Standard Errors and Sample Sizes for Two-Level Research. *Journal of Educational and Behavioral Statistics* 1993; 18(3):237-259.
- (98) Lee EW, Dubin N. Estimation and sample size considerations for clustered binary responses. *Stat Med* 1994; 13(12):1241-1252.
- (99) Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics* 1997; 53(3):937-947.
- (100) Cohen MP. Determining Sample Sizes for Surveys with Data Analyzed by Hierarchical Linear Models. *Journal of Official Statistics* 1998; 14(3):267-275.
- (101) Plewis I, Hurry J. A Multilevel Perspective on the Design and Analysis of Intervention Studies. *Educational Research and Evaluation* 1998; 4(1):13-26.
- (102) Maas CJM, Hox JJ. Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 2005; 1(3):86-92.
- (103) Cochran WG. *Sampling Techniques*. 3 ed. New York: 1977.
- (104) Kreft I. *Are Multilevel Techniques Necessary?: An overview, including Simulation Studies*. Los Angeles: 1996.
- (105) Pita-Fernandez S. *Uso de la estadística y la epidemiología en atención primaria*. Jarpyo Editores SA, editor. 115-161. 1997. Madrid, Universidad de Alicante. *Manual de metodología de trabajo en atención primaria*.
- Ref Type: Serial (Book, Monograph)
- (106) Mallol J, Andrade R, Auger F, Rodriguez J, Alvarado R, Figueroa L. Wheezing during the first year of life in infants from low-income population: a descriptive study. *Allergol Immunopathol (Madr)* 2005; 33(5):257-263.
- (107) Garcia-Marcos L, Mallol J, Sole D, Brand PL. International study of wheezing in infants: risk factors in affluent and non-affluent countries during the first year of life. *Pediatr Allergy Immunol* 2010; 21(5):878-888.
- (108) Pearce N, Ait-Khaled N, Beasley R, Mallol J, Keil U, Mitchell E et al. Worldwide trends in the prevalence of asthma symptoms:



- phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax* 2007; 62(9):758-766.
- (109) Mallol J, Garcia-Marcos L, Sole D, Brand P. International prevalence of recurrent wheezing during the first year of life: variability, treatment patterns and use of health resources. *Thorax* 2010; 65(11):1004-1009.
- (110) Garcia-Marcos L, Carvajal U, I, Escribano MA, Fernandez BM, Garcia dLR, Tauler TE et al. Seasons and other factors affecting the quality of life of asthmatic children. *J Investig Allergol Clin Immunol* 2007; 17(4):249-256.
- (111) Mallol J, Garcia-Marcos L, Aguirre V, Martinez-Torres A, Perez-Fernandez V, Gallardo A et al. The International Study of Wheezing in Infants: questionnaire validation. *Int Arch Allergy Immunol* 2007; 144(1):44-50.
- (112) Tauler E, Ferrer M, Vall O, Alonso J. [Validation of the Spanish version of the Pediatric Asthma Quality of Life Questionnaire in children with asthma]. *Med Clin (Barc )* 2002; 118(11):439.
- (113) Tauler E, Vilagut G, Grau G, Gonzalez A, Sanchez E, Figueras G et al. The spanish version of the paediatric asthma quality of life questionnaire (PAQLQ): metric characteristics and equivalence with the original version. *Qual Life Res* 2001; 10(1):81-91.
- (114) Bianca AC, Wandalsen GF, Miyagi K, Camargo L, Cezarin D, Mallol J et al. International Study of Wheezing in Infants (EISL): validation of written questionnaire for children aged below 3 years. *J Investig Allergol Clin Immunol* 2009; 19(1):35-42.
- (115) Mallol J, Garcia-Marcos L, Aguirre V, Martinez-Torres A, Perez-Fernandez V, Gallardo A et al. The International Study of Wheezing in Infants: questionnaire validation. *Int Arch Allergy Immunol* 2007; 144(1):44-50.
- (116) Chong Neto HJ, Rosario N, Dela Bianca AC, Sole D, Mallol J. Validation of a questionnaire for epidemiologic studies of wheezing in infants. *Pediatr Allergy Immunol* 2007; 18(1):86-87.
- (117) Rabash J, Steele F, Browne WJ, Golstein H. *A User's Guide to MLwiN v2.10*. Bristol: 2009.
- (118) Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *J Am Statist Assoc* 1993;(88):9-25.

- (119) Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 2006; 1(3):473-514.
- (120) Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999; 18(3):321-359.
- (121) Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychological Methods* 1998; 3(4):486-504.
- (122) Olsson U. Generalized linear models, an applied approach. Sweden: 2002.
- (123) Guo G, Zhao H. Multilevel Modeling for Binary Data. *Annual Review of Sociology* 2000; 26:441-462.
- (124) GOLDSTEIN H, Rasbash J. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society Series A-Statistics in Society* 1996; 159:505-513.
- (125) BRESLOW NC. APPROXIMATE INFERENCE IN GENERALIZED LINEAR MIXED MODELS.
- (126) Golstein H. Hierarchical Linear Models: Problems and Prospects. *Journal of Educational and Behavioral Statistics* 1995; 20(2):201-204.
- (127) Berger JO. Statistical decision theory and Bayesian Analysis. 2nd revised ed. New York: 1993.
- (128) Box GEP, Tiao GC. Bayesian inference in statistical Anlysis. 1973.
- (129) Robert C. The Bayesian Choice: from Decision-Theoric Foundations to Computational Implementation. 2nd ed. New york: 2001.
- (130) Casella G, Robert C. Monte Carlo statistical methods. New York: 1999.
- (131) Spiegelhalter WR. Markov Chain Monte Carlo in practice. 1995.
- (132) Gelman A, Hill J. Data anlysis using regression and multilevel/hierarchical models. 2006.
- (133) Mallol J, Clayton T, Asher I, Williams H, Beasley R. ISAAC Findings In Children Aged 13-14 Years - An Overview. *Allergy and Clinical Inmunology International* 1999; 11(5):176-182.

- (134) Sherriff A, Peters TJ, Henderson J, Strachan D. Risk factor associations with wheezing patterns in children followed longitudinally from birth to 3(1/2) years. *Int J Epidemiol* 2001; 30(6):1473-1484.
- (135) Belanger K, Beckett W, Triche E, Bracken MB, Holford T, Ren P et al. Symptoms of wheeze and persistent cough in the first year of life: associations with indoor allergens, air contaminants, and maternal history of asthma. *Am J Epidemiol* 2003; 158(3):195-202.
- (136) Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med* 1995; 332(3):133-138.
- (137) Haddow JE, Knight GJ, Palomaki GE, Kloza EM, Wald NJ. Cigarette consumption and serum cotinine in relation to birthweight. *Br J Obstet Gynaecol* 1987; 94(7):678-681.
- (138) Brooke AM, Lambert PC, Burton PR, Clarke C, Luyt DK, Simpson H. The natural history of respiratory symptoms in preschool children. *Am J Respir Crit Care Med* 1995; 152(6 Pt 1):1872-1878.
- (139) Kusel MM, Keadze T, Johnston SL, Holt PG, Sly PD. Febrile respiratory illnesses in infancy and atopy are risk factors for persistent asthma and wheeze. *Eur Respir J* 2012; 39(4):876-882.
- (140) Wu P, Dupont WD, Griffin MR, Carroll KN, Mitchel EF, Gebretsadik T et al. Evidence of a causal role of winter virus infection during infancy in early childhood asthma. *Am J Respir Crit Care Med* 2008; 178(11):1123-1129.
- (141) Herr M, Just J, Nikasinovic L, Foucault C, Le Marec AM, Giordanella JP et al. Influence of host and environmental factors on wheezing severity in infants: findings from the PARIS birth cohort. *Clin Exp Allergy* 2012; 42(2):275-283.
- (142) Stern DA, Morgan WJ, Halonen M, Wright AL, Martinez FD. Wheezing and bronchial hyper-responsiveness in early childhood as predictors of newly diagnosed asthma in early adulthood: a longitudinal birth-cohort study. *Lancet* 2008; 372(9643):1058-1064.
- (143) Just J, Belfar S, Wanin S, Pribil C, Grimfeld A, Duru G. Impact of innate and environmental factors on wheezing persistence during childhood. *J Asthma* 2010; 47(4):412-416.

- (144) Chulada PC, Arbes SJ, Jr., Dunson D, Zeldin DC. Breast-feeding and the prevalence of asthma and wheeze in children: analyses from the Third National Health and Nutrition Examination Survey, 1988-1994. *J Allergy Clin Immunol* 2003; 111(2):328-336.
- (145) Bueso A, Figueroa M, Cousin L, Hoyos W, Martinez-Torres AE, Mallol J et al. Poverty-associated risk factors for wheezing in the first year of life in Honduras and El Salvador. *Allergol Immunopathol (Madr)* 2010; 38(4):203-212.
- (146) Iossifova YY, Reponen T, Ryan PH, Levin L, Bernstein DI, Lockey JE et al. Mold exposure during infancy as a predictor of potential asthma development. *Ann Allergy Asthma Immunol* 2009; 102(2):131-137.
- (147) Cho SH, Reponen T, LeMasters G, Levin L, Huang J, Meklin T et al. Mold damage in homes and wheezing in infants. *Ann Allergy Asthma Immunol* 2006; 97(4):539-545.
- (148) Rodriguez G, Goldman N. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 1995; 158(1):73-89.
- (149) Goldstein H, Rasbash J. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society Series A-Statistics in Society* 1996; 159:505-513.
- (150) Engel B. A Simple Illustration of the Failure of PQL, IRREML and APHL as Approximate ML Methods for Mixed Models for Binary Data. *Biom J* 1998; 40(2):141-154.
- (151) Hobert JP, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* 1996; 91(436).

## **9. Anexos**

Cuestionario EISL



U.D. Pediatría  
Facultad de Medicina  
Universidad de Murcia



Región de Murcia  
Consejería de Sanidad  
Dirección General de Salud Pública

**ESTUDIO SOBRE SALUD RESPIRATORIA DURANTE EL PRIMER AÑO DE VIDA**

Nombre y apellidos del niño/a \_\_\_\_\_

Dirección: \_\_\_\_\_ Ciudad: \_\_\_\_\_

Teléfono: \_\_\_\_\_ Nombre del Pediatra del niño: \_\_\_\_\_

Centro de Salud al que pertenece el niño: \_\_\_\_\_

<b>1. Fecha de hoy</b>	<b>DIA</b>	<b>MES</b>	<b>AÑO</b>		
<input type="text" value="__/__/__"/>	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7
			2005	2006	2007
<b>2. Fecha de nacimiento del niño/a</b>	<b>DIA</b>	<b>MES</b>	<b>AÑO</b>		
<input type="text" value="__/__/__"/>	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9	<input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6
			2004	2005	2006
<b>3. Persona que rellena la encuesta</b> (madre=1, padre=2, otra=3, madre y padre=4) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4		
<b>4. Sexo de su hijo/a</b> (niño=1, niña=2) .....			<input type="radio"/> 1 <input type="radio"/> 2		
<b>5. Edad de su hijo/a</b> (en meses, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....			<input type="text" value="__ m"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>6. ¿Cuánto pesó su hijo/a al nacer?</b> (menos de 1500 gramos=1, de 1500 a 1999 gramos=2, de 2000 a 2499 gramos=3, de 2500 a 3499 gramos=4, más de 3500 gramos=5, no lo recuerdo=6) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6		
<b>7. ¿Cuánto pesa su hijo/a actualmente?</b> (en kilogramos, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....			<input type="text" value="__ kg"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>8. ¿Cuánto midió su hijo/a al nacer?</b> (en centímetros, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....			<input type="text" value="__ cm"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>9. ¿Cuánto mide su hijo/a actualmente?</b> (en centímetros, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....			<input type="text" value="__ cm"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>10. Raza o etnia de su hijo/a</b> (blanca=1, gitana=2, norteafricana=3, subsahariana=4, india hispanoamericana=5, asiática=6, otra=7) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7		
<b>11. ¿Nació su hijo/a en España?</b> (sí=S, no=N) .....			<input type="radio"/> S <input type="radio"/> N		
<b>11.1. Si ha respondido "NO", ¿en qué país nació su hijo/a?</b> (por favor escríbalo en el recuadro) .....			<input type="text" value="País _____"/>		
<b>12. Lugar de nacimiento del padre</b> (España=1, Europa, excepto España=2, África=3, América del Norte=4, América del Sur y Central=5, Asia=6, Australia/Oceania=7) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7		
<b>13. Lugar de nacimiento de la madre</b> (España=1, Europa, excepto España=2, África=3, América del Norte=4, América del Sur y Central=5, Asia=6, Australia/Oceania=7) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7		
<b>14. ¿Ha tenido su hijo/a silbidos o pitos en el pecho en sus primeros 12 meses de vida?</b> (sí=S, no=N) ...			<input type="radio"/> S <input type="radio"/> N		
<b>15. ¿Cuántos episodios de silbidos o pitos en el pecho ha tenido su hijo/a en sus primeros 12 meses de vida?</b> (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si nunca ha tenido pitos marque "00") .....			<input type="text" value="__ e"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>16. ¿A qué edad tuvo su hijo/a el primer episodio de silbidos o pitos en el pecho?</b> (en meses, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si nunca ha tenido pitos marque "00") .....			<input type="text" value="__ m"/> <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9 <input type="radio"/> 0 <input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 <input type="radio"/> 4 <input type="radio"/> 5 <input type="radio"/> 6 <input type="radio"/> 7 <input type="radio"/> 8 <input type="radio"/> 9		
<b>17. Cuando su hijo/a tiene pitos en el pecho ¿Suele comenzar por un resfriado normal?</b> (sí=1, no=2, nunca ha tenido pitos=3) .....			<input type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3		

18. ¿Se le han desencadenado o empeorado a su hijo/a los pitos en el pecho al moverse más (gatear, correr...) o al enfadarse o reír? (sí=1, no=2, nunca ha tenido pitos=3) .....	1 2 3
19. Los pitos en el pecho de su hijo/a ¿afectan a la alimentación del niño/a? (sí=1, no=2, nunca ha tenido pitos=3) .....	1 2 3
20. Los pitos en el pecho de su hijo/a ¿limitan las actividades diarias de los padres del niño/a? (sí=1, no=2, nunca ha tenido pitos=3) .....	1 2 3
21. Los pitos en el pecho de su hijo/a ¿han modificado la vida familiar? (sí=1, no=2, nunca ha tenido pitos=3) .....	1 2 3
22. ¿Ha recibido su hijo/a tratamiento con medicamentos inhalados para dilatarle los bronquios (broncodilatadores) en nebulizaciones o inhalaciones? Ejemplo: Ventolín, Terbasmin, Berodual, Berotec, etc. (sí=1, no=2, ns/nc=3) .....	1 2 3
23. ¿Ha recibido su hijo/a tratamiento con corticoides inhalados? Ejemplo: Flixotide, Becloasma, Pulmicort, Flunilol, Inhalacor, Budesonida, Becotide, Pulmictan, Ribujet, etc. (sí=1, no=2, ns/nc=3) ..	1 2 3
24. ¿Ha recibido su hijo/a tratamiento con otros medicamentos como antileucotrienos (Singulair) o ketotifeno (Zasten, Ketasma)? (sí, tomó antileucotrienos=1, sí, tomó ketotifeno=2, sí, tomó ambos medicamentos=3, no=4, no sabe=5) ..	1 2 3 4 5
25. En los primeros 12 meses de vida ¿cuántas veces se ha despertado usted por la noche debido a la tos con ahogos o silbidos/pitos en el pecho de su hijo/a? (nunca=1, raras veces (menos de 1 vez al mes)=2, algunas veces (algunas semanas en algunos meses)=3, frecuentemente (dos o más noches por semana, casi todos los meses)=4) .....	1 2 3 4
26. En los primeros 12 meses de vida, ¿han sido las sibilancias o silbidos/pitos en el pecho tan importantes como para tener que llevarlo a un Servicio de Urgencias? (sí=S, no=N) .....	S N
27. En los primeros 12 meses de vida, ¿han sido las sibilancias o silbidos/pitos en el pecho tan importantes que usted notara que su hijo/a tenía mucha dificultad para respirar y se ahogaba? (sí=S, no=N) .....	S N
28. ¿Ha estado su hijo/a hospitalizado por pitos, silbidos o ahogos en el pecho en sus primeros 12 meses de vida? (no=0, 1 vez=1, 2 veces=2, 3 veces=3, 4 veces=4, 5 veces=5, 6 veces=6, 7 veces=7, 8 veces=8, 9 o más veces=9) .....	0 1 2 3 4 5 6 7 8 9
29. ¿Le ha dicho un médico alguna vez que su hijo/a tiene asma? (sí=S, no=N) .....	S N
30. ¿Ha tenido su hijo/a neumonía o bronconeumonía en sus primeros 12 meses de vida? (sí=S, no=N) .....	S N
31. ¿Ha estado su hijo/a hospitalizado por neumonía o bronconeumonía en sus primeros 12 meses de vida? (sí=S, no=N) .....	S N
32. ¿Ha tenido su hijo/a, en sus primeros 12 meses de vida, manchas rojas en la piel que pican y que aparecen y desaparecen, en cualquier parte del cuerpo, excepto alrededor de la boca y la nariz y excepto en el área del pañal? (sí=S, no=N) .....	S N
33. ¿Fuma la madre (o tutora) del niño/a actualmente? (sí=S, no=N) .....	S N
33.1. ¿Cuántos cigarrillos al día fuma la madre (o tutora) del niño/a? (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no fuma marque "00") .....	0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
34. ¿Fuma el padre (o tutor) del niño/a actualmente? (sí=S, no=N) .....	S N
34.1. ¿Cuántos cigarrillos al día fuma el padre (o tutor) del niño/a? (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no fuma marque "00") .....	0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
35. ¿Fumó la madre del niño durante el embarazo del niño/a? (sí=S, no=N) .....	S N
35.1. Si responde "SÍ" marque en qué trimestre, puede marcar varias respuestas (1º trimestre=1, 2º trimestre=2, 3º trimestre=3) .....	1 2 3
36. ¿Cuántos de los que viven en la casa fuman cigarrillos, incluyendo a los padres? .....	0 1 2 3 4 5 6 7 8 9
37. ¿Tiene asma el padre del niño/a? (sí=S, no=N) .....	S N
38. ¿Tiene asma la madre del niño/a? (sí=S, no=N) .....	S N
39. ¿Tiene asma algún hermano/a del niño/a? (sí=S, no=N) .....	S N
40. ¿Tiene alergia nasal (rinitis alérgica, fiebre del heno) el padre del niño/a? (sí=S, no=N) .....	S N
41. ¿Tiene alergia nasal (rinitis alérgica, fiebre del heno) la madre del niño/a? (sí=S, no=N) .....	S N
42. ¿Tiene alergia nasal (rinitis alérgica, fiebre del heno) algún hermano/a del niño/a? (sí=S, no=N) .....	S N
43. ¿Tiene alergia en la piel (dermatitis alérgica) el padre del niño/a, excluyendo dermatitis alérgica de contacto? (sí=S, no=N) .....	S N
44. ¿Tiene alergia en la piel (dermatitis alérgica) la madre del niño/a, excluyendo dermatitis alérgica de contacto? (sí=S, no=N) .....	S N
45. ¿Tiene alergia en la piel (dermatitis alérgica) algún hermano/a del niño/a, excluyendo dermatitis alérgica de contacto? (sí=S, no=N) .....	S N
46. ¿Se le han realizado al padre del niño/a, alguna vez, pruebas de alergia en piel o en sangre? (no=1; sí, y mostró alergia=2; sí, pero no mostró alergia=3) .....	1 2 3

47. ¿Se le han realizado a la madre del niño/a, alguna vez, pruebas de alergia en piel o en sangre? (no=1; sí, y mostró alergia=2; sí, pero no mostró alergia=3) ..... 1 2 3

48. ¿Se le han realizado a algún hermano/a del niño/a, alguna vez, pruebas de alergia en piel o en sangre? (no=1; sí, y mostró alergia=2; sí, pero no mostró alergia=3) ..... 1 2 3

49. ¿Ha ido su hijo/a a la guardería en sus primeros 12 meses de vida? (sí=S, no=N) ..... S N

50. ¿A qué edad comenzó su hijo/a a ir a la guardería? (en meses, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no ha ido todavía a una guardería marque "00") .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

51. ¿Con qué frecuencia ha ingerido su hijo/a cualquiera de los siguientes productos (NO elaborados en casa) durante sus primeros 12 meses de vida?: yogur, flan, natillas, petit suisse, patatas fritas envasadas, mermeladas, chocolate, refrescos, zumos de sobre, en brick o botella, néctar, etc.. (nunca=1, una vez al mes=2, una vez por semana=3, todos los días de la semana=4) ..... 1 2 3 4

52. En la casa que vive el niño/a, ¿qué tipo de combustible se usa, predominantemente, para la calefacción?(electricidad=1, gas central=2, estufa de gas=3, queroseno=4, carbón=5, madera=6, otra=7) ..... 1 2 3 4 5 6 7

53. En la casa que vive el niño/a, ¿qué tipo de combustible se usa, predominantemente, para cocinar? (electricidad=1, gas=2, queroseno=3, carbón=4, madera=5, otra=6) ..... 1 2 3 4 5 6

54. ¿Tiene aire acondicionado en la casa que vive el niño/a? (sí=S, no=N) ..... S N

55. De las siguientes mascotas, marque las que tenía en su casa cuando nació su hijo/a (no tenía mascotas=1, perro=2, gato=3, aves=4, conejo/hamster=5, otra=6) ..... 1 2 3 4 5 6

56. De las siguientes mascotas, marque las que tiene actualmente en la casa que vive el niño/a (no tengo mascotas=1, perro=2, gato=3, aves=4, conejo/hamster=5, otra=6) ..... 1 2 3 4 5 6

57. La casa en la que vive el niño/a, ¿está enmoquetada? (sí=S, no=N) ..... S N

58. La casa en la que vive el niño/a, ¿tiene baño completo (lavabo, inodoro, ducha/bañera) en el interior? (sí=S, no=N) ..... S N

59. La cocina de la casa en la que vive el niño/a (el lugar dónde se prepara la comida) ¿está dentro de la casa? (sí=S, no=N) ..... S N

60. ¿Tiene teléfono (fijo o móvil) en la casa en la que vive el niño/a? (sí=S, no=N) ..... S N

61. Marque, por favor, el nivel de educación alcanzado por la madre del niño/a (educación básica, primaria o ninguna (8 años o menos)=1, educación media o secundaria incompleta (9-11 años)=2, educación media o secundaria completa y superior (12 y más años)=3, educación universitaria=4) ..... 1 2 3 4

62. ¿Cuántos meses alimentó a su hijo/a exclusivamente con leche materna (sin leches adaptadas, cereales, zumos de frutas u otros alimentos como papillas, etc.. ? (en meses, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no le dió leche materna marque "00") .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

63. ¿Cuántos resfriados (estornudos, tos, moquillo nasal como agua, con o sin fiebre) ha tenido su hijo/a en sus primeros 12 meses de vida? (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no ha tenido resfriados marque "00") .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

64. ¿Cuántos meses tenía su hijo/a cuando se resfrió por primera vez? (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no se ha resfriado nunca marque "00") .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

65. ¿Considera usted que su hijo/a vive en una zona con contaminación atmosférica (humos de fábricas, tráfico intenso de vehículos, etc..) (sí=S, no=N) ..... S N

65.1. Si ha respondido "SI", marque lo que considere oportuno (mucho=1, moderado=2, poco=3) ..... 1 2 3

66. ¿Hay moho (hongos) o manchas de humedad en la casa que vive el niño/a? (sí=S, no=N) ..... S N

67. ¿Tiene su hijo/a las vacunas correspondientes a su edad completas? (sí=S, no=N) ..... S N

68. Número de hermanos/as que tiene su hijo/a (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro. Si no tiene hermanos marque "00") .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

69. ¿Cuántas personas, adultos y niños, viven en total en la casa en la que vive el niño/a actualmente ? (por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9

70. Ocupación del padre (directivos, administradores, licenciados=1, otros directivos téc. medios, diplomados=2, cuadros intermedios, administrativos=3, trabajadores manuales cualificados=4, trabajadores manuales semicualificados=5, trabajadores no cualificados=6, otros casos, mal especificados=7, actualmente no trabaja=8) ..... 1 2 3 4 5 6 7 8

71. Ocupación de la madre (directivos, administradores, licenciados=1, otros directivos téc. medios, diplomados=2, cuadros intermedios, administrativos=3, trabajadores manuales cualificados=4, trabajadores manuales semicualificados=5, trabajadores no cualificados=6, otros casos, mal especificados=7, actualmente no trabaja=8) ..... 1 2 3 4 5 6 7 8

72. Edad de la madre del niño (en años, por favor, marque las bolas correspondientes a la derecha y escríbalo también en el recuadro) .....  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9



73. ¿Qué se utiliza, *predominantemente*, en la cocina de la casa en la que vive el niño/a para freír? (aceite de oliva=1, mantequilla=2, margarina=3, otro tipo de aceite=4) ..... ① ② ③ ④

74. Durante el embarazo de su hijo/a, ¿con qué frecuencia comió o bebió lo siguiente? (nunca u ocasionalmente=1, una o dos veces por semana=2, tres o más veces por semana=3, no lo comió por intolerancia o alergia=4, no sabe=5)

	Nunca Ocasion.	1-2 veces sem.	3 o más sem.	Alergia	NS
74.1. Carne (ternera, pollo, cordero, conejo, cerdo, ...) .....	①	②	③	④	⑤
74.2. Hamburguesas cocinadas en casa .....	①	②	③	④	⑤
74.3. Comida rápida: 74.3.1. Pizzas precocinadas, platos precocinados .....	①	②	③	④	⑤
74.3.2. Hamburguesas en burgers, perritos, etc. ....	①	②	③	④	⑤
74.3.3. Frituras: croquetas, palitos merluza, etc. ....	①	②	③	④	⑤
74.4. Pescado blanco .....	①	②	③	④	⑤
74.5. Pescado azul .....	①	②	③	④	⑤
74.6. Fruta fresca/zumo natural .....	①	②	③	④	⑤
74.7. Verdura fresca .....	①	②	③	④	⑤
74.8. Ensaladas .....	①	②	③	④	⑤
74.9. Verdura cocinada .....	①	②	③	④	⑤
74.10. Legumbres .....	①	②	③	④	⑤
74.11. Cereales, incluido pan .....	①	②	③	④	⑤
74.12. Pasta .....	①	②	③	④	⑤
74.13. Arroz .....	①	②	③	④	⑤
74.14. Mantequilla .....	①	②	③	④	⑤
74.15. Margarina .....	①	②	③	④	⑤
74.16. Frutos secos, o mantequilla de cacahuete/avellana .....	①	②	③	④	⑤
74.17. Patatas cocinadas en casa (incluidas patatas fritas) .....	①	②	③	④	⑤
74.18. Patatas fritas de bolsa y snacks (gusanitos, fritos, etc...) .....	①	②	③	④	⑤
74.19. Leche .....	①	②	③	④	⑤
74.20. Yogur .....	①	②	③	④	⑤
74.21. Huevos .....	①	②	③	④	⑤
74.22. Bollería industrial, galletas .....	①	②	③	④	⑤
74.23. Bebidas con alcohol .....	①	②	③	④	⑤
74.24. Bebidas gaseosas .....	①	②	③	④	⑤

75. ¿Durante cuánto tiempo en total tomó la madre del niño/a anticonceptivos orales antes de quedarse embarazada del niño/a al que se refiere este cuestionario? (nunca=0, menos de 1 año=1, de 1 a 3 años=2, de 4 a 6 años=3, más de 6 años=4) ..... ① ② ③ ④

76. Durante el embarazo del niño/a al que se refiere esta encuesta, ¿con qué frecuencia tomó la madre del niño paracetamol (termalgin, gelocatil, etc.)? (nunca o menos de 1 vez al mes=1, de 1 a 4 veces al mes=2, más de 1 vez a la semana=3) ..... ① ② ③

76.1. Si tomó alguna vez paracetamol durante el embarazo ¿porqué causa lo hizo? (cefalea/migraña=1, fiebre=2, dolor muscular=3, otra causa=4) ..... ① ② ③ ④

77. Durante el embarazo y parto del niño/a al que se refiere esta encuesta, ¿tuvo alguna de las siguientes complicaciones? Por favor, responda "SI" o "NO" en cada caso (sí=S, no=N)

77.1. Hipertensión .....	③ ④	77.8. Sufrimiento fetal, hipoxia..	③ ④
77.2. Amenaza de aborto .....	③ ④	77.9. Prematuridad .....	③ ④
77.3. Infecciones .....	③ ④	77.10. Vueltas de cordón .....	③ ④
77.4. Diabetes gestacional .....	③ ④	77.11. Trauma obstétrico en el niño .....	③ ④
77.5. Mala presentación .....	③ ④	77.12. Forceps, ventosa .....	③ ④
77.6. Rotura prematura de aguas.	③ ④	77.13. Cesárea .....	③ ④
77.7. Problemas de la placenta ....	③ ④		