

# CONCORDANCIA INTEROBSERVADOR EN LA EVALUACIÓN DE LA ESCALA DE DESARROLLO INFANTIL BAYLEY II

## MATCHING INTEROBSERVER ASSESSMENT SCALE OF CHILD DEVELOPMENT BAYLEY II

Recibido 7 de abril 2015  
Aceptado 13 de abril 2015

### Correspondencia:

María Magdalena Álvarez Ramírez  
Manantial de San Cristóbal s/n 2 piso  
Xalapa 2000. C. P. 91700, Xalapa, Veracruz, México  
Tel. 0122 8842 1700 Ext. 19 518  
malvarez@uv.mx

### Autores:

María Magdalena Álvarez Ramírez  
*Doctorado en Neuroetología, Universidad Veracruzana*

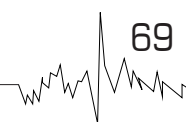
Dora Elizabeth Granados Ramos  
*Universidad Veracruzana, Facultad de Psicología-Xalapa,  
Laboratorio de Psicobiología.*

Pablo Pacheco Cabrera  
*Instituto de Neuroetología, Universidad Veracruzana*

Tania Ivonne Hernández Coloa  
*Universidad Veracruzana, Facultad de Psicología-Xalapa,  
Laboratorio de Psicobiología.*

**Palabras clave:** Concordancia, interobservador,  
acuerdo, escala Bayley II.

**Key Words:** Agreement, interobserver agreement,  
Bayley II scale.



## RESUMEN

Un elemento importante en la realización de proyectos de investigación en niños es disminuir el sesgo interobservador al realizar mediciones del desarrollo.

Cuando se asegura la validez de un instrumento de evaluación, la magnitud del error se puede determinar a través de estudios de concordancia, que tienen como propósito estimar hasta qué punto dos observadores coinciden en su evaluación.

La Escala de Desarrollo Infantil Bayley II (BSID-II) evalúa el desarrollo de los niños de 1 a 42 meses de edad, la prueba consta de tres subescalas diferenciadas: cognoscitiva, psicomotora y de comportamiento. El objetivo del presente estudio fue determinar la concordancia interobservador en la evaluación de la Escala de Desarrollo Infantil Bayley II en niños de 1 a 42 meses de edad.

Se realizó un estudio descriptivo de corte transversal con 30 niños nacidos a término sin diagnóstico de enfermedad neurológica, genética o metabólica que acudieron a consulta externa al Laboratorio de Psicobiología durante el periodo comprendido de agosto de 2013 a mayo de 2014. Se documentó la historia clínica y se evaluó con la escala BSID-II a cada participante. Se evaluó la concordancia mediante el test de Kappa media de Cohen con el software Epi dat 4.

La prevalencia de retraso global en el desarrollo de los niños fue de 67%, a partir de ello se observó concordancia tanto en las comparaciones del Índice de desarrollo mental  $k=0.9528$ , como en el Índice de desarrollo psicomotor  $k=.9023$ .

El análisis demostró concordancia entre dos profesionistas, lo cual implica que de acuerdo a las categorías no hubo sesgo en el diagnóstico. Este dato es un indicador positivo respecto al nivel de coordinación y la coincidencia en el diagnóstico, además facilita la implementación de tratamientos adecuados para cada uno de los casos evaluados.

**Palabras clave:** concordancia, interobservador, acuerdo, escala Bayley II.

## ABSTRACT

An important element when conducting research projects in children is decreasing interobserver bias when making measurements.

When the validity of an instrument of evaluation is ensured, the magnitude of error can be determined through studies of concordance, whose purpose is to estimate how much two observers agree in their assessment.

The Scale of Infant Development Bayley II evaluates the development of children 1-42 months of age. The test consists of three distinct subscales: cognitive, motor and behavioral. The aim of the present study was to determine interobserver agreement on the implementation of the Bayley Scale of Infant Development (BSID-II, 1993) in children 1-42 months of age.

A descriptive cross-sectional study was conducted with 30 children born at term without a diagnosis of neurological, genetic or metabolic diseases that attended external consultation in the lab of Psychobiology during the period august 2013 to may 2014. A clinical history was documented, and the BSDI-II scale was applied, the agreement was assessed by Cohen's Kappa test mean with Epi dat 4 software.

The prevalence of developmental delay in children was 67% concordance was observed in the mental development index  $k = 0.9528$  as well as in the psychomotor development index  $k = 0.9023$ .

The analysis demonstrated concordance between the two professionals, which means that according to the categories no bias in occurred. This data is a positive indicator with respect to the level of coordination and the agreement in the diagnosis and also facilitates the implementation of an appropriate treatment for each of the cases evaluated.

**Keywords:** agreement, interobserver agreement, Bayley II scale.



## Planteamiento del problema

Diversos organismos internacionales involucrados en el desarrollo infantil han resaltado la importancia que tiene el diagnóstico y la atención temprana para promover la expresión de las capacidades físicas, mentales y sociales del niño.<sup>1</sup>

Las investigaciones realizadas en el campo de la psicología han puesto de manifiesto la importancia del desarrollo de la inteligencia, la personalidad, y el comportamiento social en la primera infancia, por lo que, si los niños no reciben la atención y el cuidado adecuado a sus necesidades, se puede obstaculizar su desarrollo.

La evaluación del desarrollo en las etapas tempranas es un elemento relevante en las estrategias de salud que se realiza a través de escalas que miden las áreas psicomotora y cognoscitiva. Es imprescindible asegurar la confiabilidad de los resultados de las evaluaciones a través de la concordancia interobservadores. Si existe concordancia, se potencializarán las estrategias diagnósticas, terapéuticas y de evaluación de los profesionistas, así como la oportuna atención en casos de rezago en el desarrollo.

## INTRODUCCIÓN

El desarrollo adecuado en la infancia es indispensable para la construcción de capital humano y social. Dicho proceso inicia con la fecundación y envuelve aspectos que van desde el crecimiento físico hasta la maduración neurológica, cognoscitiva, social y afectiva del niño.<sup>2</sup> Factores adversos tanto biológicos como socioculturales pueden alterar el desarrollo de los niños y producir un daño; a la probabilidad de que esto ocurra se la denomina riesgo para el desarrollo.<sup>3</sup> Dentro de los factores biológicos se pueden considerar la prematurez, bajo peso al nacimiento, malnutrición, edad de la madre y enfermedades mentales en los cuidadores o padres. Como factores socioculturales se incluyen los problemas socioeconómicos, bajo nivel educativo y la ausencia de alguno de los padres.

El seguimiento del desarrollo de los niños durante los tres primeros años de vida es fundamental debido al proceso de maduración cerebral que tiene lugar en esta etapa ya que en este periodo sensible pueden presentarse dificultades en el desarrollo.<sup>5</sup> Durante éste proceso tiene gran importancia la

plasticidad cerebral, que consiste en la capacidad cerebral de discriminar, asociar, aprender y modular sensaciones a partir de un estímulo para las conductas adaptativas.<sup>4, 6</sup>

La detección precoz de problemas del desarrollo requiere de un proceso continuo de vigilancia durante los primeros años de vida. Las alteraciones en el desarrollo en las etapas tempranas pueden asociarse a problemas en etapas posteriores como bajo desempeño y deserción escolar, problemas neurológicos, psiquiátricos, emocionales y conductuales, disminución en las habilidades sociales y problemas de aprendizaje. Es por ésta razón que la evaluación y seguimiento del desarrollo en los niños cobra importancia como estrategia de prevención de secuelas del desarrollo.<sup>7</sup>

Para ésta investigación se utilizó la Escala de Desarrollo Infantil Bayley II (BSID-II), instrumento de evaluación ampliamente utilizado para determinar el índice de desarrollo a niños de 1 a 42 meses de edad, que consta de 3 escalas: mental, psicomotora y del comportamiento.<sup>8</sup>



La escala mental consta de 178 elementos; evalúa las capacidades de memoria, generalización, clasificación, vocalizaciones, lenguaje y habilidades sociales. Se expresa en puntuaciones típicas o índices de desarrollo mental (IDM) que presentan una media de 100 y una desviación de 15.<sup>8,9</sup>

La escala psicomotora consta de 111 elementos; valora el control de los grupos musculares gruesos y finos, los resultados se expresan en puntuaciones típicas o índices de desarrollo psicomotor (IDP) con puntuaciones medias de 100 y desviaciones típicas de 15.

La escala de comportamiento evalúa las características de la conducta que el niño tiene en las actividades, con el evaluador, en el espacio de evaluación y con el padre o principal cuidador durante la prueba. Califica las áreas de atención/alertamiento en los primeros seis meses, orientación/interacción, regulación emocional y cualidad motora de 1 a 42 meses de edad.<sup>9</sup>

Los índices de confiabilidad de la prueba Bayley II, son de 87 para la escala mental y 78 para la escala psicomotora. La sensibilidad de 0.75-0.90 y especificidad de 0.76-0.91. La escala de comportamiento tiene rangos de calificación basada en una escala percentilar: 1-10 no óptimo, 11-25 cuestionable y 26-99 dentro de los límites normales.<sup>10</sup>

En el proceso de validación de instrumentos, es importante considerar el grado de acuerdo o desacuerdo entre las mediciones que uno o más profesionistas realizan a otra persona. A esto se le denomina, concordancia, que ayuda a comprobar la confiabilidad de la evaluación, identificar índices de desarrollo y sus posibles alteraciones.<sup>11</sup>

La concordancia interobservador, es cuando dos o más profesionistas realizan una evaluación a un participante de manera independiente y simultánea. Proceso que consiste en revisar la técnica de

evaluación, realizar evaluaciones simultáneas, calificar la escala individualmente, comparar resultados obtenidos por cada profesionista y contrastar los resultados con un experto.

La precisión de la evaluación puede verse afectada por dos factores fundamentales: la variación propia del instrumento o procedimiento, debido a la calidad y calibrado del mismo y la variación condicionada por el entrenamiento, formación y capacidad del profesionista. A esto se le llama error o sesgo del observador. A medida que el error es menor, mayor es su precisión y más válido será el resultado.<sup>12</sup>

La precisión de las observaciones puede evaluarse de tres formas, al comparar:

- a) Un observador consigo mismo, concordancia intraobservador: cuando un profesional realiza dos o más evaluaciones a una misma persona con el propósito de conocer o confirmar un diagnóstico con el uso de una técnica.
- b) Un observador con un estándar de oro: estudia el grado de concordancia de un reactivo utilizado en pruebas de tamizaje frente al resultado de una prueba de confirmación realizada por un experto.
- c) Varios observadores, concordancia interobservadores: Si dos o más observadores realizan una entrevista de manera independiente y es casi simultánea a una persona, se puede estimar el grado de acuerdo o concordancia.

La medida que se utiliza para evaluar la concordancia interobservadores, es el índice de acuerdo observado o Kappa de Cohen que se califica con los valores mínimos y máximo de Kappa.<sup>13</sup>

El índice de concordancia, es la primera aproximación al acuerdo interobservador, expresa el



porcentaje de acuerdo entre ellos, es decir, en qué medida hubo coincidencia en la evaluación entre los observadores con relación al total de elementos examinados.

Para estimar la concordancia, se usa la prueba Kappa de Cohen donde se consideran adecuados los valores superiores a 0,50. Para el cálculo de la concordancia es necesario conocer la prevalencia esperada del fenómeno estudiado, el número de profesionistas o evaluaciones y el número de categorías o casillas diagnósticas posibles.<sup>11</sup>

El índice de Kappa relaciona el acuerdo potencial que alcanzan los observadores independientemente de lo que se atribuye al azar. Se calcula la diferencia

entre la proporción del acuerdo observado y lo esperado por azar, si esto es igual a cero el grado del acuerdo observado se puede atribuir completamente al azar. Sin embargo, si la diferencia es positiva, es decir, mayor a cero, indica que el grado del acuerdo observado es mayor de lo que se puede atribuir al azar.<sup>14</sup>

El valor de Kappa que debe considerarse como indicador de buena concordancia depende del tipo, complejidad y prevalencia del problema que se esté estudiando. Landis y Koch propusieron en 1977 una escala de interpretación del valor de Kappa que considera aceptable un valor mayor o igual a 40 y excelentes los valores superiores a 0,75 (Ver tabla 1).<sup>15</sup>

**Tabla 1.** Interpretación cualitativa de los valores de Kappa.

Interpretación	Valores de Kappa
0 y 0,20	Deficiente
0,21 y 0,40	Pobre
0,41 y 0,60	Aceptable
0,61 y 0,80	Buena
0,81 y 1,00	Excelente

**Fuente:** Campo-Arias A, Herazo E. Concordancia intra- e inter-observadores. Rev. Colomb. Psiquiat. 2010; 39 (2): 423-430.

El valor máximo de Kappa es 1 y el valor mínimo es 0, una concordancia total entre los observadores se produce cuando el acuerdo observado es del 100%.

El índice Kappa mide el grado de acuerdo entre los profesionistas, no la calidad de la observación, por lo que no procede considerar a uno de los observadores como estándar, sino al experto en la evaluación o procedimiento. El acuerdo interobservador, es útil para medir datos categóricos.<sup>13</sup> El diagnóstico en desarrollo es por lo general cualitativo o categórico, se establece si quien se evaluó, reúne

los puntajes para un desempeño esperado para su edad o etapa de desarrollo.

La concordancia intraobservador e interobservador puede ser real o producto del azar. Por ello, además de la concordancia observada, es necesario estimar hasta qué grado el acuerdo se debe al azar o a la probabilidad, por eso es preciso considerar la sumatoria de las concordancias observadas menos la sumatoria de las concordancias atribuibles al azar entre el total de observaciones, menos la sumatoria de concordancias atribuibles al azar<sup>15</sup>, de acuerdo a la siguiente fórmula:

$$k = \frac{[(\sum \text{concordancias observadas}) - (\sum \text{concordancias atribuibles al azar})]}{[(\text{total de observaciones}) - (\sum \text{concordancias atribuibles al azar})]}$$

**Fuente:** Cerda K, Villarroel L. evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. Bioestadística Rev Chil Pediatr 2008; 79(1): 54-58.

Con esto se logra un mejor grado de precisión, por lo tanto un mejor intervalo de confianza.<sup>13, 14</sup>

El objetivo del presente estudio fue determinar la concordancia interobservador en la evaluación de la Escala de Desarrollo Infantil Bayley II en niños de 1 a 42 meses de edad.

## MATERIAL Y MÉTODOS

Se realizó un estudio transversal descriptivo de la concordancia interobservador en la evaluación de la Escala de Desarrollo Infantil Bayley II (BSID-II)<sup>9</sup> en el Laboratorio de Psicobiología de la Facultad de Psicología de la Universidad Veracruzana. La muestra estuvo conformada por 30 niños nacidos a término que fueron evaluados después de obtener el consentimiento informado de sus padres o cuidador principal.

Para ésta investigación se consideró la concordancia interobservador lograda en las escalas mental y psicomotora.

Con los padres o cuidadores principales de los niños se elaboró una historia clínica en la que se anotaron datos de identificación, información sociodemográfica y antecedentes patológicos pre, peri y pos natales.



La evaluación de la BSID-II se realizó por un profesional capacitado, en un espacio libre de estímulos distractores.

El análisis de la concordancia interobservador se integró en diez pasos:

1. Revisión sistemática de la Escala de Desarrollo Infantil Bayley II, a través de exploración y análisis de cada uno de los reactivos.
2. Evaluación con la escala por las dos observadoras con supervisión de un experto.
3. Realización de la escala a niños que acudieron a consulta durante el período del estudio.
4. Calificación individual de la evaluación de la escala por cada una de las evaluadoras.
5. Comparación de resultados entre observadoras para cuantificar acuerdos y discrepancias en la calificación de la escala.
6. Observación de la videograbación de la evaluación.
7. Revisión de la calificación por cada evaluadora.
8. Comparación de calificaciones finales entre las observadoras.
9. Revisión y retroalimentación del proceso por el experto.
10. Emisión de resultados.

## RESULTADOS

La muestra se conformó por 30 niños, 15 del sexo femenino y 15 del sexo masculino. La edad promedio de los niños fue de 17 meses con una DE  $\pm$  10 meses. En el 67% de la muestra evaluada el desempeño fue ligeramente retrasado tanto en la escala psicomotora como en la escala mental.

En relación a la concordancia interobservador, en la evaluación de la escala mental se presentó solo un desacuerdo entre las profesionistas (Ver tabla 2).

**Tabla 2.** Acuerdos y desacuerdos en la calificación de la escala mental.

Observador 1				
Calificación de Desempeño	Acelerado	Normal	Ligeramente retrasado	Significativamente retrasado
Acelerado	2 (+)	1 (-)	0	0
Normal	0	7 (+)	0	0
Ligeramente Retrasado	0	0	11 (+)	0
Significativamente Retrasado	0	0	0	9 (+)

Se señalan como acuerdos (+) y desacuerdos (-) interobservador.

Al realizar la Escala de desarrollo psicomotora las evaluadoras presentaron dos desacuerdos (Ver tabla 3).

**Tabla 3.** Acuerdos y desacuerdos en la calificación de la escala psicomotora.

Observador 1				
Calificación de Desempeño	Acelerado	Normal	Ligeramente retrasado	Significativamente retrasado
Acelerado	1 (+)	0	0	0
Normal	1 (-)	7 (+)	0	0
Ligeramente Retrasado	0	1 (-)	12 (+)	0
Significativamente Retrasado	0	0	0	8 (+)

Se señalan como acuerdos (+) y desacuerdos (-) interobservador.

Posterior a la evaluación de la Escala de Desarrollo Infantil Bayley II, se calculó el índice de concordancia para ambas evaluadoras, utilizando el índice de Kappa con 10, 20 y 30 sujetos. Se observó que el valor de índice incrementó a medida que aumentó el número de sujetos evaluados (Ver tablas 4 y 5).

**Tabla 4.** Índice de concordancia en la escala mental.

Casos	Kappa	EE	I C (95%)	P
10	0.8571	0.1291	.6041-10000	0.0000
20	0.9240	0.0714	.7840-10000	0.0000
30	0.9528	0.0458	.8631-10000	0.0000

EE corresponde al error estándar, IC intervalo de confiabilidad y p el valor de significancia.

**Tabla 5.** Índice de concordancia en la escala psicomotora.

Casos	Kappa	EE	I C (95%)	P
10	0,6364	0.2038	.2370-10000	0.0043
20	0,8438	0.1020	.6439-10000	0.0000
30	0,9023	0.0659	.7732-10000	0.0000

EE corresponde al error estándar, IC intervalo de confiabilidad y p el valor de significancia.

En la escala mental se encontró una concordancia interobservador de .95 ( $p=0.0000$ ) y .90 ( $p=0.0000$ ) en la escala psicomotora que corresponden a la clasificación de excelente en los criterios mencionados anteriormente.<sup>12</sup>





## DISCUSIÓN

Estudios clínicos resaltan la importancia de la concordancia en el diagnóstico como el de Landa y Goñi<sup>14</sup> quienes destacan la utilidad de realizar dentro de la capacitación del personal encargado de las evaluaciones, procesos de confiabilidad, para disminuir los sesgos en la realización de las escalas clínicas. Por lo que es recomendable centrarse en el logro de la concordancia interobservadores en las evaluaciones de desarrollo dirigidas a poblaciones en etapas tempranas.

La confiabilidad interobservador evidenciada por los valores del índice de Kappa caracterizan una concordancia excelente entre las profesionistas en cuanto a los criterios de evaluación de la Escala de Desarrollo Infantil Bayley II.

Se debe mencionar que con la confiabilidad interobservador no se está evaluando la validez de los resultados, sino el grado de acuerdo en la medida obtenida, por lo que debe considerarse que en todo proceso de medición los puntajes obtenidos nunca estarán libres de errores.<sup>17</sup>

Los errores que afectan las medidas obtenidas en una evaluación pueden ser aleatorios y sistemáticos. Los errores aleatorios derivan de factores que afectan la medición de la variable de forma accidental en toda la muestra, adicionando variabilidad a los datos, sin afectar el desempeño promedio del grupo y los errores sistemáticos son aquellos errores que pueden ser causados por cualquier factor que afecte la medición de la variable en toda la muestra y tiende a sobreestimar o subestimar el valor verdadero anulando la validez, por eso se considera como sesgo de medición.<sup>18</sup>

Los resultados encontrados mostraron la confiabilidad entre evaluadores y la escala de medición, indicando su precisión y la posibilidad de utilización para la evaluación del desarrollo de los niños de 1 a 42 meses de edad.

Se logró el objetivo planteado, de determinar la concordancia interobservador en la evaluación de la Escala de Desarrollo Infantil Bayley II en niños de 1 a 42 meses de edad. El valor del índice de Kappa obtenido de la evaluación de la escala,

permitió mostrar que a medida que se adquirió experiencia en el proceso de evaluación, la concordancia interobservador fue más eficiente incrementándose el valor de los acuerdos.

Los resultados en la presente investigación coinciden con lo reportado por Cerda y Villarroel, en cuanto a que, a mayor índice de concordancia interobservadores, menor es el error debido a la variabilidad en las mediciones entre los profesionistas.<sup>15</sup>

Cabe resaltar que la confiabilidad y la concordancia no son propiedades fijas de los instrumentos de medida, pero si, el producto de la interacción entre la escala, los participantes y el contexto de la evaluación. Para que esta propiedad no se vea afectada en contextos diversos es importante establecer el control sobre las variables que interfieren en el proceso de medición. Para esto, los observadores deben ser capacitados sobre la técnica a utilizar, criterios de evaluación y calificación.

A partir de la estadística, es importante aclarar que el índice de Kappa no muestra la cualidad de la medición sino el nivel de concordancia o acuerdo entre ambos profesionistas. Puede ser que el acuerdo interobservadores sea excelente ( $k=80-100$ ) sin embargo pueden estar equivocados respecto al diagnóstico real, por lo que los valores deben ser supervisados por un experto en la evaluación, manejo estadístico e interpretación de la prueba utilizada.

El uso de la concordancia interobservador en el ámbito clínico permite asegurar que los procedimientos sean eficaces, accesibles, sencillos de análisis y de uso amplio.

## CONCLUSIONES

El nivel de concordancia interobservador en la Escala de Desarrollo Infantil Bayley II del presente estudio fue excelente.

El nivel de concordancia interobservador alcanzado facilitó la coordinación en el trabajo, permitió la precisión en la obtención de índices de desarrollo en las escalas mental y psicomotora, así como la identificación temprana de alteraciones en el desarrollo.



## REFERENCIAS BIBLIOGRÁFICAS

1. UNICEF. Convención sobre los derechos del niño. 2014. <http://www.unicef.org/spanish/crc/> Consultado el 14 de enero de 2014.
2. Figueiras A, Neves de Souza I, Ríos V, Benguigui Y. Manual para la vigilancia del desarrollo infantil (0-6 años) en el contexto de AIEPI. Organización Panamericana de la Salud Washington, D.C.: OPS, 2011.
3. Andraca I, Pino P, La Parra A, Rivera F, Castillo M. Factores de riesgo para el desarrollo psicomotor en lactantes nacidos en óptimas condiciones biológicas. *Rev. Saúde Pública*. 1998; 32 (2): 479-487.
4. Johnson Mark. El cerebro en Desarrollo. En: La primera infancia en perspectiva. Child and Youth Studies Group. The open University, Reino Unido 2012.
5. Hernández-Muela S, Mulas F, Mattos L. Plasticidad neuronal funcional. *Revista de Neurología*. Instituto Valenciano de Neurología. 2004; Supl (1): 558-568.
6. Zuluaga J. Desarrollo Comportamental y comunicativo. Neurodesarrollo y Estimulación. Editorial Médica. 2001.
7. Pinto L. Lo maravilloso y mágico del neurodesarrollo humano. *Rev Chil Pediatr* 2008; 79 Supl (1): 18-20.
8. Quezada-Villalobos L, Soto-García I, Escobar-Cabello M, & López-Suárez A. Confiabilidad inter-observador de la Escala Psicomotora Infantil de Alberta en niños de término y pretérmino de la provincia de Talca - Chile. *Rev. Cienc. Salud*. 2010; 8 (2): 21-32
9. Bayley N. Escalas Bayley de Desarrollo Infantil. BSID. 1969/1977. Madrid: TEA.
10. Bayley N. Bayley Scales of Infant Development. Second Edition (BSID-II). San Antonio: Hartcourt Brace & Company; 1993.
11. Pérez-López J, et al. Las escalas Bayley BSID-I frente a BSID-II como instrumento de evaluación en Atención Temprana. *Anales de psicología*. 2012; 28 (2): 484-489.
12. Rizzoli-Córdoba A, et al. Validación de un instrumento para la detección oportuna de problemas de desarrollo en menores de 5 años en México. *Bol. Med. Hosp. Infant. Mex*. 2013; 70(3): 195-208.
13. Campo-Arias A, Herazo E. Concordancia intra- e inter-observadores. *Rev. Colomb. Psiquiat*. 2010; 39 (2): 423-430.
14. Cortés-Reyes E, Rubio-Romero J, Gaitán-Duarte H. Métodos estadísticos de evaluación de la concordancia y la reproducibilidad de pruebas diagnósticas. *Revista Colombiana de Obstetricia y Ginecología*. 2010; 61: 247-255.
15. Landa N, Goñi A, García de Jalón E, López-Goñi J. Concordancia en el diagnóstico entre pediatría y salud mental. *An. Sist. Sanit. Navar*. 2009; 32 (2): 161-168.
16. Cerda J, Villarroel L. Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. *Rev Chil Pediatr*. 2008; 79 (1): 54-58.
17. Morales Á. Epidemiología Clínica: Investigación Clínica aplicada. Bogotá: Editorial Médica Internacional 2004; 107-109.
18. Trochim WMK. Research methods: knowledge bases. 2006. Cincinnati (OH): Atomic Dog Publishing. <http://www.socialresearch-methods.net/kb/reliable.php>. Consultado: 01 de marzo 2015

