

Artículo original

Estimación del acuerdo entre dos métodos para medición clínica de la edad gestacional

Dr. Carlos Grandi*

RESUMEN

Introducción. En la práctica clínica es frecuente que cuando se propone un nuevo método solamente puede evaluarse por comparación con otras técnicas establecidas ("estándar de oro").

El objetivo del presente estudio fue describir un abordaje para evaluar el grado de acuerdo entre dos métodos para la medición clínica de la edad gestacional y estimar si son intercambiables.

Población, material y métodos. Diseño observacional, de cohortes. Entre 2000 y 2001 se enrolaron prospectivamente 137 embarazadas en el Hospital Materno Infantil Ramón Sardá de Buenos Aires para su control prenatal antes de la 16ª semana de gestación. Se estimó la edad gestacional por fecha de la última menstruación y por ultrasonografía. Se calcularon la media de las diferencias, su desviación estándar, el promedio entre los dos métodos para cada caso individual y el intervalo de confianza al 95% (límite del acuerdo).

Resultados. La media de la edad gestacional en la primera ultrasonografía fue de 19,6 (DE 4,9; IC 95% 18,7-20,4) semanas. La concordancia entre ambos métodos se incrementaba en función del lapso establecido y alcanzó el 71,5% (63-78) \pm 10 días y el 82,5% (75-87) \pm 2 semanas. La media de las diferencias fue de 0,35 + 1,89 (0,03-0,67) semanas y los límites del acuerdo al 95% mostraron que, independientemente de la edad gestacional, las diferencias alcanzaban 7,5 semanas.

Conclusiones. La estimación de la edad gestacional por la fecha de la última menstruación excedió la estimación por ultrasonografía en 25% y la subestimó en 20%. Los dos métodos no serían intercambiables en la práctica clínica.

Palabras clave: edad gestacional, ultrasonografía, acuerdo.

SUMMARY

Introduction. In the clinical practice, a new method frequently can only be assessed by means of its comparison against other established techniques (gold standard). The objective of the present work was to describe an approach –through graphical techniques and simple statistical calculations– to assess the agreement level between two methods for clinical measurement of gestational age and to determine if they are interchangeable.

Population, material and methods. This is an observational cohort study. Between 2000 and 2001, 137 pregnant women were prospectively enrolled at the Hospital Materno-Infantil Ramón Sardá of Buenos Aires before the 16th week of gestation. The last normal menstrual period and ultrasound scan measurements were used for estimating gestational age. Mean difference, standard deviation and 95%

confidence interval (limits of agreement) between the methods, and the mean of the two measurements were calculated and plotted for each case.

Results. Mean gestational age at the first ultrasound scan was 19.6 (SD 4.9; 95% CI 18.7-20.4) weeks. The agreement between both methods were 71.5% (95% CI 63-78) \pm 10 weeks and 82.5% (95% CI 75-87) \pm 2 weeks. Mean difference was 0.35 + 1.89 (95% CI 0.03-0.67) weeks and the limits of agreement showed that, independently of gestational age, differences reached 7.5 weeks.

Conclusions. The estimation of gestational age by the last normal menstrual period overestimated the one estimated ultrasound scan ("gold standard") by 25% and underestimated it by 20%. In the clinical practice both methods are not interchangeable.

Key words: gestational age, ultrasound, agreement.

INTRODUCCIÓN

En la práctica clínica es frecuente medir variables del cuerpo humano, como la presión arterial o la duración de un embarazo; esto puede ser extremadamente difícil o imposible de realizar directamente sin efectos adversos sobre el sujeto, por lo que el verdadero valor se desconoce.

En su lugar se han desarrollado técnicas indirectas y cuando se propone un nuevo método solamente puede evaluarse por comparación con otras técnicas establecidas, pero no podemos estar seguros que alguno de los métodos nos brinde una medida absolutamente correcta y, por consiguiente, deberíamos estimar el grado de acuerdo entre ambos.

Lo que importa conocer es en cuánto difieren los métodos, ya que si no perjudica la interpretación clínica, entonces los dos métodos podrían usarse en forma intercambiable.

En Perinatología es común estimar la duración de la gestación utilizando la fecha de la última menstruación (FUM) y, desde hace algunas décadas, se ha ido reemplazando por la ultrasonografía fetal (US) precoz a partir de mediciones antropométricas, considerada el estándar de oro.¹

* Consejo de Investigación, GCBA. Sector Epidemiología Perinatal y Bioestadística, Maternidad Sardá, Buenos Aires.

Correspondencia: cgrandi@intramed.net

Aclaración de intereses: Este trabajo no contó con ningún tipo de financiamiento para su realización.

Sin embargo, evaluada rigurosamente en estudios, no se ha demostrado que la mejora en la estimación de la edad gestacional (EG) con mediciones rutinarias por US precoz en poblaciones de bajo riesgo sea beneficioso para la salud materna y los resultados perinatales.^{2,3}

Se ha postulado que las mediciones ultrasonográficas no brindan mejores estimaciones de la EG que una FUM confiable. Más aún, las mediciones con ultrasonido precoces no son accesibles para un gran número de mujeres en países en desarrollo como los de nuestra región, ni para muchas de alto riesgo en países desarrollados.

El objetivo del presente estudio es describir un abordaje –mediante técnicas gráficas y cálculos estadísticos sencillos– para evaluar el grado de acuerdo entre dos métodos para la medición clínica de la edad gestacional en una población de bajo riesgo, estimar si son intercambiables.

POBLACIÓN, MATERIAL Y MÉTODOS

Se trata de un diseño observacional y descriptivo, tipo cohorte. La muestra incidental consistió en 137 embarazadas de un

estudio prospectivo observacional con el fin de evaluar el estado nutricional preconcepcional y durante la gestación admitidas en el Hospital Materno Infantil Ramón Sardá de Buenos Aires para su control prenatal antes de la 16ª semana de gestación, entre marzo de 2000 y octubre de 2001. *Criterios de exclusión:* las portadoras de embarazo múltiple, malformaciones fetales graves o enfermedad Rh. *Criterios de eliminación:* abandono del control prenatal, muerte fetal, toxemia, diabetes e hidropesía fetal.

Estimación de la edad gestacional (al primer control prenatal)

- 1). *Fecha de la última menstruación (FUM):* el intervalo entre el primer día del último período menstrual normal (confiable) y la fecha del parto.
- 2). *Ultrasonografía precoz (US):* mediante las mediciones ultrasonográficas del diámetro biparietal, la longitud cefalocaudal, longitud femoral y la circunferencia abdominal realizadas por tres ecografistas certificados independientes de la estimación de la EG por FUM antes de la 20ª semana de gestación, con un ecógrafo Voluson 530 D (Medison®) bidimensional y empleando como estándar de referencia las curvas de Hadlock.^{4,5}

Límites del acuerdo

El primer paso es calcular el coeficiente de correlación de Pearson (r), que estima el grado de relación entre los dos métodos, y el coeficiente de determinación ($R^2 = r^2$), que evalúa la proporción (%) en la que una variable estima la variabilidad de la otra.⁶

Puede anticiparse que exista una sistemática tendencia de un método de exceder al otro. Esto se denomina sesgo y se calcula mediante la media de las diferencias (\bar{d}) entre los dos métodos (FUM - US) y su desviación estándar (s_d) en cada caso individual.

Si las diferencias están normalmente distribuidas (distribución gaussiana), se espera que el 95% de estas diferencias caerán entre $\bar{d} \pm 2 s_d$ (límite de acuerdo al 95%). Estos valores definen el rango dentro del cual estarán la mayoría de las diferencias entre las mediciones por los dos métodos. La decisión sobre qué es un acuerdo aceptable es exclusivamente clínica; la estadística por sí sola no puede responder a esta pregunta.⁷

TABLA 1. Características maternas

	n	Media	DE	%	IC 95%
Nacionalidad argentina	106			78	69-83
Conurbano bonaerense	89			65	56-72
Unión estable	110			81	72-86
Estudios secundarios	82			60	51-67
Edad (años)	137	26	5,7		
Primíparas	48			35	27-43
EG al 1º control (semanas)	137	14,3	3,5		
Número de visitas prenatales	137	8,2	1,6		
Peso preconcepcional (kg)	137	60,7	12,7		
IMC (kg/m ²)	137	25,2	4,9		
Tabaquismo (> 1 cig/día)	23			17	11-24
Diabetes gestacional	4			2,8	1,1-7,2
Hipertensión arterial (todos los tipos)	10			7,7	4-13
Restricción del crecimiento fetal	5			3,5	1,5-8,2
Rotura prematura membranas	27			19,5	14-27
Amenaza del parto prematuro	3			2,5	0,7-6,2

IMC: Índice de masa corporal.

Presentación gráfica del acuerdo

Lo primero es un dispersograma (gráfico de dispersión) de los datos con su línea de igualdad a 45° de ambos ejes sobre la cual todos los puntos deberían yacer si ambos métodos dieran siempre y exactamente el mismo valor. No se presenta la recta de regresión, ya que no estamos interesados en predecir la edad gestacional de un método por el otro. El segundo paso es graficar la diferencia (eje y) entre las mediciones por ambos métodos (FUM - US) y el promedio (eje x) de ambos métodos para cada caso que representaría, idealmente, el mejor estimador de la edad gestacional (desconocido). Finalmente, se le agrega el límite de acuerdo al 95% ($\bar{d} \pm 2 s_d$). La posible relación entre las discrepancias y el verdadero valor se calculó mediante el coeficiente de correlación de Spearman.⁶

Relación entre las diferencias y la magnitud de las mediciones

El presente abordaje asume que la media y la desviación estándar de las diferencias son iguales a lo largo del rango de las mediciones de la edad gestacional (7 a 34 semanas). En el caso de observarse aumento de la variabilidad de las diferencias con el incremento de la EG, se puede eliminar esta relación mediante la transformación logarítmica de ambas mediciones (estandarización). Los límites del acuerdo derivados de esta transformación luego se recalculan mediante los antilogaritmos, con la ventaja de que a partir de estos valores se calcula la magnitud (en %) en la que un método excede (límite superior) o subestima (límite inferior) al otro.⁸

Para el análisis de los datos se utilizaron los programas Excel® (Microsoft Office, 2003) y Statistica 5.5 (Tulsa, OK, 2000).

RESULTADOS

La *Tabla 1* muestra las características sociodemográficas y obstétricas de las madres, apreciándose que la mayoría eran jóvenes argentinas en unión estable, de tamaño normal según estándares internacionales⁹ y con instrucción primaria completa. El primer control prenatal fue precoz y el número de visitas, de 8,2 + 1,6. La edad gestacional (FUM) se registró "con dudas" en 30 casos (8,9%, IC 95% 6,3-12,4). Las patologías obstétricas fueron de baja incidencia excepto la

rotura prematura de las membranas, ya que el protocolo exigía el registro tanto antes como durante el trabajo de parto.

La *Tabla 2* muestra las características de los recién nacidos. El 96% fue de término y de peso adecuado para la edad gestacional, con un adecuado índice ponderal (2,8 + 2,2 kg/m³); fue baja la incidencia de bajo peso (< 2.500 g) y muy bajo peso (< 1.500 g) y no se registró ninguna muerte fetal.

La media de la edad gestacional en la primera ultrasonografía fue de 19,6 (DE 4,9; IC 95% 18,7-20,4) semanas, lo que correspondió a una amenorrea de 19,8 (4,9; 19,0-20,6) semanas. En 88 casos (64,2%; 56-71), la estimación por FUM superó a la US, mientras que solamente en dos (1,5%) fueron iguales.

En la *Figura 1* puede apreciarse que la concordancia entre ambos métodos se incrementaba en función del lapso establecido, alcanzó el 71,5% (63-78) ± 10 días y el 82,5% (75-87) ± 2 semanas.

FIGURA 1. Estimación de la edad gestacional. Concordancia entre FUM y US precoz (18^a - 20^a sem.). Sardá 2000-2002, n= 137

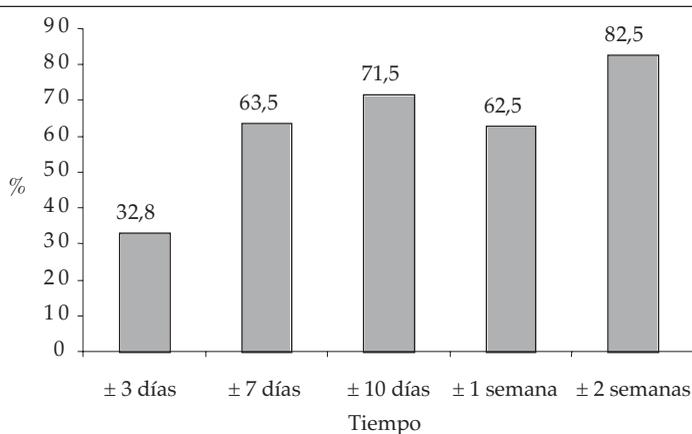


Tabla 2: Características neonatales

	n	Media	DE	%	IC 95%
EG (semanas)	137	39,2	1,8		
Peso al nacer (g)	137	3.283	500		
Índice ponderal (kg/m ³)	137	2,8	2,2		
Pretérmino (< 37 semanas)	3			2,1	0,9-4,8
Bajo peso (< 2.500 g)	5			3,6	1,5-8,2
Muy bajo peso (< 1.500 g)	1			0,7	0,01-0,4
PEG (< percentilo 10)	6			4,3	2,0-9,2

PEG: pequeños para la edad gestacional

El dispersograma de ambos métodos (*Figura 2*) muestra que, aunque los puntos se agrupan alrededor de la línea de igualdad con algunos valores extremos, no fue posible calcular la diferencia entre ellos. Usualmente el siguiente paso es calcular el coeficiente de correlación de Pearson; para los datos de la *Figura 2*, $r = 0,927$ (IC 95% 0,898-0,947; $p = 0,0001$), con rechazo de la hipótesis nula ($r = 0$) y concluyendo con seguridad que las estimaciones de la EG por FUM y US están relacionadas ($R^2 = 86,3\%$), pero esta elevada correlación no significa que ambos métodos concuerden.

Más informativo para el clínico es estimar las diferencias entre ambos métodos (FUM-US) que deben tener, aproximadamente, una distribución normal. La distribución de las diferencias presentó un ligero sesgo negativo (-0,43, EE= 0,20), atribuible a que las diferencias y el promedio de ambos métodos podrían estar relacionados (véase más adelante).

La media de las diferencias ($\bar{d} \pm s_d$) fue $0,35 \pm 1,89$ (IC 95% 0,03-0,67) semanas y los límites del acuerdo al 95% fueron: $0,35 \pm (2 \times 1,89)$, o sea 4,11 y -3,41 semanas para el límite superior e inferior, respectivamente. Las diferencias entre las mediciones por ambos métodos (FUM-US) para cada caso y el promedio de ambos métodos con los límites del acuerdo (*Figura 3*), demostraron que, independientemente de la edad gestacional, las diferencias alcanzaban 7,5 semanas. Este intervalo es muy amplio y refleja la gran variación de las diferencias.

El coeficiente de correlación por rangos de Spearman, fue de $r = -0,058$ ($p = 0,497$) lo cual señala prácticamente la ausencia de relación. La *Figura 3* también permite observar la presencia de observaciones extremas (outliers); solamente 8 casos (5,8%) estuvieron fuera de los límites.

Se asume que la media y la desviación estándar de las diferencias entre FUM y US son constantes a lo largo del rango de la EG. La distribución de las diferencias presentó un ligero sesgo negativo atribuible a un aumento de su variabilidad a medida que se incrementa la edad gestacional; esto puede apreciarse en la *Figura 2*.

Para eliminar esta relación se procedió a la transformación logarítmica (\log_n) de ambas mediciones antes del análisis. La *Figura 4* muestra la transformación logarítmica de los

datos con la media de las diferencias contra el promedio de la EG con su límite del acuerdo al 95%. De manera similar a la *Figura 3*, no se observó que las diferencias variaran sistemáticamente en ninguna dirección sobre el rango de las mediciones de la EG. Los valores extremos tampoco difirieron y fueron solamente de 6 casos (4,3%).

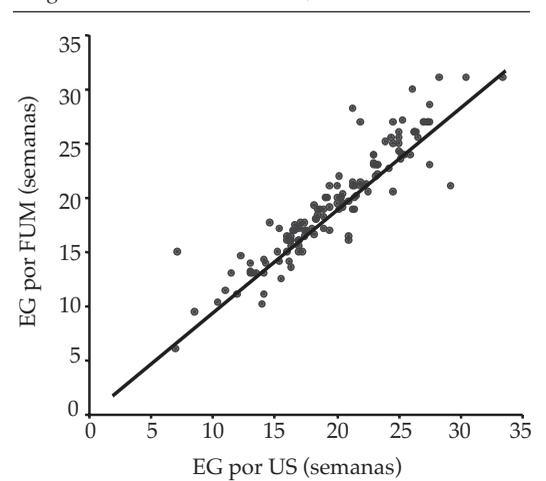
La media de las diferencias (\bar{d}) fue 0,018 ($s_e = 0,11$) y los límites del acuerdo, -0,20 y 0,23. Es necesario tomar los antilogaritmos de los límites del acuerdo para relacionarlos con la escala original de las mediciones, estos valores fueron 0,81 y 1,25 (adimensionales). Esto significa que en el 95% de los casos, la estimación de la edad gestacional por FUM excedió la estimación por US en 25% (1,25-1) y la subestimó en 20% (1-0,80).

DISCUSIÓN

Este estudio muestra un nuevo abordaje para evaluar el grado de acuerdo entre dos métodos para la medición clínica de la edad gestacional en una población hospitalaria y estimar si son intercambiables, lo cual redundaría en una importante reducción de los costos asistenciales.

Este abordaje se basa en la filosofía que plantea que, para comparar dos métodos, es necesario cuantificar las diferencias entre las mediciones individuales y, por consiguiente, no existiría lugar para el análisis basado en pruebas de hipótesis. Para el clínico, el acuerdo no es algo que esté presente o ausente, sino la magnitud de las diferencias.

FIGURA 2. Estimación de la edad gestacional. Concordancia entre FUM y US precoz con la línea de igualdad. Sardá 2000-2002, $n = 137$



La primera conclusión es que ni el uso del coeficiente de correlación por sí sólo (como se muestra aquí) ni otros análisis estadísticos ampliamente utilizados (como regresión, comparación de medias, ecuaciones estructurales o correlación intraclase)¹⁰ indican que dos métodos concuerden; solamente se puede hablar de “perfecto acuerdo” si todos los puntos de la *Figura 2* se distribuyeran sobre la línea de igualdad. Además, cuanto mayor sea el rango de las

observaciones mayor será el coeficiente r , lo que agregado al elevado nivel de significación, muestra que los dos métodos están relacionados, pero son irrelevantes a la pregunta sobre el acuerdo.

Las ventajas de este estudio son que el diseño fue prospectivo bajo protocolo, lo que disminuye los sesgos, se basó en una población de bajo riesgo (como lo demuestran los datos de las *Tablas 1 y 2*) y con un control prenatal usual, lo que aseguraría la represen-

FIGURA 3. Relación entre las diferencias y el promedio de la estimación de la edad gestacional por FUM y ecografía con su límite del acuerdo al 95%. Sardá 2000-2002, n= 137

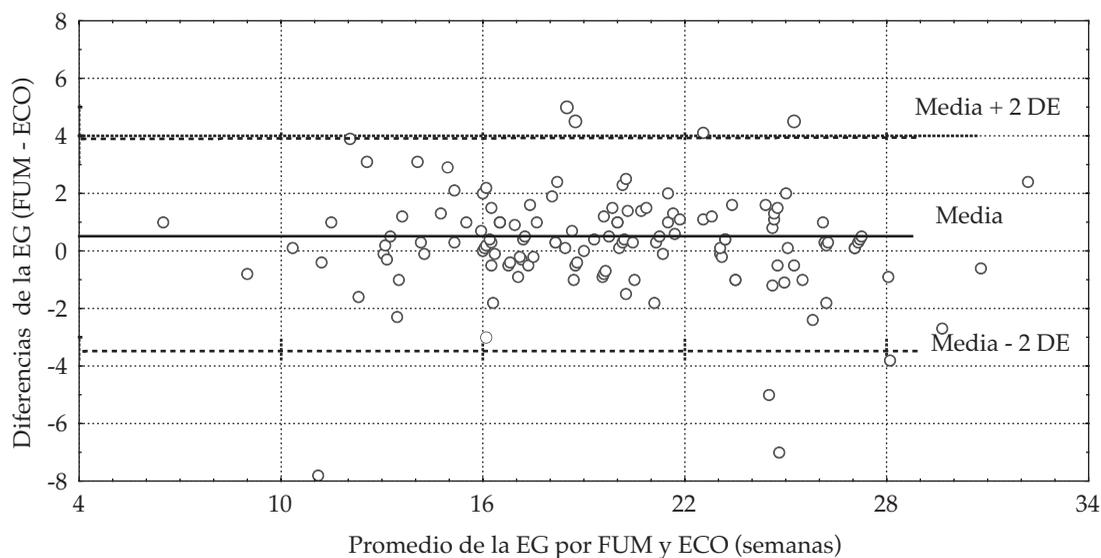
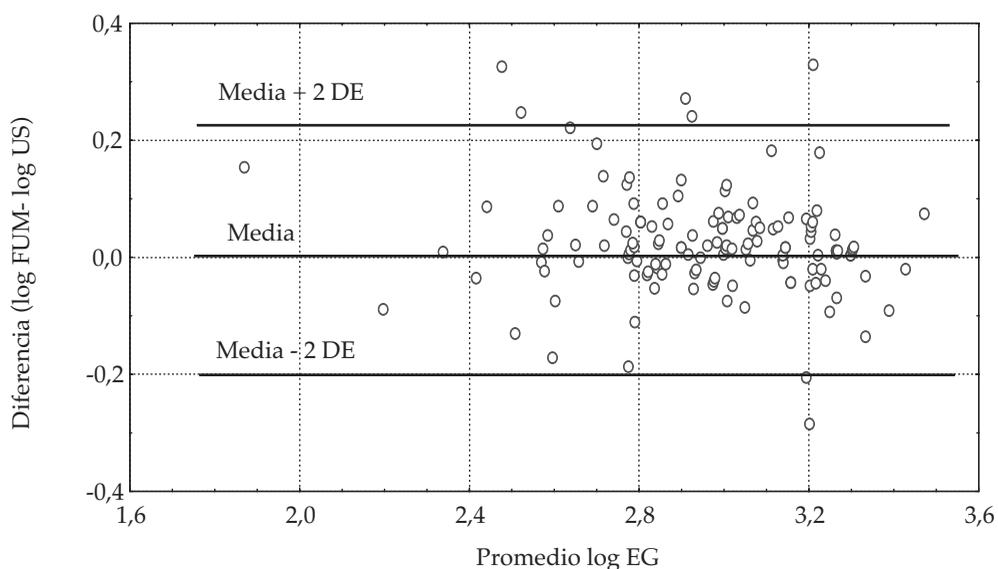


FIGURA 4. Relación entre las diferencias y el promedio de la estimación de la edad gestacional luego de la transformación logarítmica, con su límite del acuerdo al 95%. Sardá 2000-2002, n= 137



tatividad de la población que se asiste en la mayoría de los centros de nuestra región.

Las limitaciones fueron la gran variabilidad de las diferencias (ilustrada por la desviación estándar de las diferencias $-s_d$) y la fecha de la última menstruación "dudosa" en 30 casos; al reanalizar los datos y excluir estos casos, los límites del acuerdo no se modificaron.

Los intervalos del acuerdo son muy amplios, tanto utilizando una escala común como luego de la transformación logarítmica de los datos. Así, la mayoría de las estimaciones de la edad gestacional por FUM excedieron la estimación por US en 25% y la subestimaron en 20%. Esto señala que, bajo la interpretación más optimista, existen considerables discrepancias entre los dos métodos y que el grado del acuerdo no sería aceptable para la práctica clínica, por lo que se descarta que sean intercambiables.

Además nuestros resultados son consistentes con otros estudios que postularon que la estimación de la EG por FUM es superior, en promedio, a la basada en la ultrasonografía precoz.¹¹⁻¹³ Las explicaciones más citadas para esta sistemática diferencia incluyen el recordatorio dudoso de la FUM y la mayor frecuencia de ovulación tardía (> 14 días) en relación a la ovulación precoz (< 14 días).

Creemos que el método propuesto es fácilmente entendible por la mayoría de los usuarios, requiere los programas usuales de computación y cubriría las situaciones clínicas habituales. Además, permite evaluar la reproducibilidad (precisión) de los métodos cuando existen dos o más mediciones en el mismo sujeto.

CONCLUSIONES

La estimación de la edad gestacional por FUM excedió la estimación por US en 25% y la subestimó en 20%. Esto señala que los dos métodos no son intercambiables en la práctica clínica.

Agradecimientos

A los siguientes profesionales que participaron en el equipo de investigación: Dr. Hugo A. Sola (Nutrición Infantil), Dr. Pablo Durán (Epidemiología Nutricional), Dr.

Guillermo Luchtenberg (Neonatología), Dr. Eduardo Negri (Obstetricia), Lic. Susana Checa (Socióloga), Lic. Elsa Schwartzman (Socióloga), Lic. Nutr. María T. Rinaudo (Nutrición), Dra. María Perego (Bioquímica), Dra. Graciela Briozzo (Bioquímica) y Obs. Alicia Pécora (Obstetricia), a la Dra. Mónica Rittler por la revisión crítica del manuscrito. ■

BIBLIOGRAFÍA

1. Taipale P, Hiilesmaa V. Predicting delivery date by ultrasound and last menstrual period in early gestation. *Obstet Gynecol* 2001; 97:189-94.
2. Crowther C, Kornman L, O'Callaghan S, George K, Furness M, Wilson K. Is an ultrasound assessment of gestational age at the first antenatal visit of value? A randomized clinical trial. *Br J Obstet Gynaecol* 1999; 106:1273-9.
3. Villar J, Ezcurra E, Gurtner de la Fuente V, Campodónico L. Preterm delivery syndrome: the unmet need. En: Keirse M, editors, *New Perspectives for the effective treatment of preterm labor. An International consensus. Research & Clinical Forums.* Wells Medical, England 1994; 16:9-33.
4. Grandi C, López F. Estimación de la edad gestacional: Revisión de la literatura. *Rev Hosp Mat Inf Ramón Sardá* 2004; 23(3):138-43.
5. Hadlock F, Harrist R, Martínez-Poyer. In utero analysis of fetal growth: a sonographic weight standard. *Radiol* 1991; 181:129-133.
6. Hirsch R, Riegelman R. *Statistical operations. Analysis of health research data.* 1st ed. Cambridge (USA): Blackwell Science, 1996.
7. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 8:307-10.
8. Bland J, Altman D. Measuring agreement in method comparison studies. *Stat Met Med Res* 1999; 8:135-160.
9. Institute of Medicine (United States). Committee on Nutritional Status during Pregnancy and Lactation, Food and Nutrition Board. Subcommittee on Nutrition during pregnancy. Part 1, Weight Gain; part II, Nutrient Supplements. Washington, DC: National Academy Press, 1990.
10. Altman D. *Practical statistics for medical research.* 1st ed. London: Chapman & Hall, 1991:396-403.
11. Grandi C. Epidemiología de la prematuridad. Factores de riesgo y métodos de detección precoz. En: Programa de Actualización en Neonatología. Buenos Aires: Ed. Panamericana, 2004:69-116.
12. Yang H, Kramer M, Platt R, Blondel B, Bréart G, Morin I, et al. How does early ultrasound scan estimation of gestational age lead to higher rates of preterm birth. *Am J Obstet Gynecol* 2002; 186:433-437.
13. Barros F, Victora C, Barros A, Santos I, Albernaz E, Matijasevich A, et al. The challenge of reducing neonatal mortality in middle-income countries: findings from three Brazilian birth cohorts in 1982, 1993, and 2004. *Lancet* 2005; 365:847-854.