

Variabilidad interobservador en la evaluación de cardiotocogramas intraparto y su efecto en las decisiones clínicas: un tema de seguridad en la atención obstétrica

Inter-examination variation in the evaluation of intrapartum cardiotocograms and its effect in clinical decisions: an issue of security in obstetrician care

Luis Alberto Villanueva-Egan,¹ Alejandra Grajeda-Campa²

RESUMEN

Introducción. La cardiotocografía intraparto ha demostrado desde su incorporación altas variabilidades intra e interobservador que afectan el desempeño de la prueba. El objetivo del estudio es evaluar la reproducibilidad, el rendimiento y los efectos de la prueba sobre las decisiones obstétricas.

Material y métodos. Dos especialistas con experiencia y dos médicos residentes del último año de la especialidad, evaluaron en forma independiente 70 registros cardiotocográficos intraparto, provenientes en partes iguales de recién nacidos con Apgar ≥ 7 y Apgar < 7 .

Resultados. El análisis de la concordancia entre observadores reveló entre los especialistas un valor de Kappa de 0.43 (IC95% 0.25-0.60) y entre residentes de 0.55 (IC95% 0.36-0.73). Del análisis global de los 4 observadores se obtuvo un valor de Kappa de 0.48 (IC95% 0.36-0.73). El desempeño de la prueba demostró una alta heterogeneidad entre los 4 observadores, independientemente de su nivel de experiencia.

Conclusiones. En las condiciones reales de uso por este grupo, la cardiotocografía intraparto no tiene ninguna utilidad clínica por sus altas tasas de falsos positivos y falsos negativos, independientemente de la experiencia del observador.

Palabras clave: cardiotocogramas, trabajo de parto, acuerdo interobservador, frecuencia cardíaca fetal.

ABSTRACT

Introduction. The Intrapartum Cardiotocography has demonstrated, from its incorporation, its high intra and inter-examination variabilities that affect the performance of the test. The goal of the study is to evaluate the reproducibility, the performance and the effects of the test over the obstetric decisions.

Material and methods. Two specialists with experience and two medical residents of the last year of the specialty, evaluated independently 70 intrapartum cardiotocographic records, coming in equal parts from new borns with Apgar ≥ 7 and Apgar < 7 .

Results. The concordance analysis between observers revealed among the specialists a Kappa value of 0.43 (IC95% 0.25-0.60) and among the residents 0.55 (IC95% 0.36-0.73). Of the global analysis of the 4 observers a Kappa value of 0.48 (IC95% 0.36-0.73) was obtained. The performance of the test demonstrated high heterogeneity among the 4 observers, regardless of their level of experience.

Conclusions. In real conditions of use by this group, the intrapartum cardiotocography does not have any clinical usefulness due to its high rate of false positives and negatives, regardless of the observers experience.

Key words: cardiotocograms, labour, inter-observer agreements, fetal cardiac frequency.

¹ Dirección de Investigación de la Comisión Nacional de Arbitraje Médico (CONAMED).

² Médica especialista en Ginecología y Obstetricia en práctica privada.

Folio: 196/12 Artículo recibido: 20-06-2012 Artículo reenviado: 30-07-2012 Artículo aceptado: 07-09-2012

Correspondencia: Dr. Luis Alberto Villanueva Egan. Director de Investigación. CONAMED. Mitla 250 piso 8 esq. Eje 5 Sur (Eugenia) Col. Vértiz Narvarte, Del. Benito Juárez. México D.F. 03020 correo electrónico: lvillanueva@conamed.gob.mx.

INTRODUCCIÓN

La asfixia perinatal, que en algunos casos es el resultado de la mala calidad de la atención obstétrica, es una causa importante de morbi-mortalidad fetal y neonatal, incluyendo secuelas neurológicas de distinta gravedad y grado de recuperación, que incluyen a la parálisis cerebral. Para ello, desde su introducción en la década de los sesenta, se ha generalizado la idea de que mediante la monitorización electrónica continua de la frecuencia cardíaca fetal (FCF) durante el trabajo de parto, es posible identificar precozmente a los fetos afectados, o potencialmente afectados, por la falta de oxígeno (hipoxia fetal).¹ No obstante, la tranquilidad que le ofrece al obstetra, la también llamada cardiotocografía (CTG) ha sido objeto de múltiples críticas dirigidas a la prueba en sí y a su aplicación generalizada. Entre ellas destacan: (i) la alta tasa de falsos positivos para asfixia, lo que conduce a la sobreindicación de la cesárea, con el riesgo latente de prematuridad iatrogénica más los inherentes a la operación cuando se realiza en condiciones de urgencia; (ii) la falta de reducción, en forma significativa, de la parálisis cerebral, la mortalidad neonatal u otras medidas estándar de bienestar neonatal; (iii) la falta de criterios estandarizados de interpretación; (iv) el desacuerdo en relación a los algoritmos para la intervención frente a patrones específicos; y, finalmente, (v) los señalamientos sobre la alta variabilidad inter e intrabobservador en la interpretación de los registros cardiotocográficos (RCTG) ante e intraparto.² Este último aspecto se acentúa cuando quien la realiza, no posee los conocimientos ni la experiencia necesaria, tal como ocurre en otras pruebas diagnósticas que dependen en mayor medida de la apreciación del observador. En ocasiones los resultados de las pruebas varían no sólo por cuándo, cómo o dónde se realizaron, sino por quién las realizó.

No obstante su utilidad para la mejora en la toma de decisiones durante la atención obstétrica, en nuestro medio son escasos o ausentes los estudios de reproducibilidad de las pruebas médicas más frecuentemente utilizadas en el ejercicio de la especialidad.

El objetivo de este estudio es evaluar la reproducibilidad, el rendimiento y los efectos sobre las decisiones clínicas de la monitorización electrónica intraparto de la FCF cuando es realizada por médicos del servicio de obstetricia de un hospital público del Distrito Federal, al que por solicitud de los participantes, y en respeto a su confidencialidad, se omitirá su nombre y en adelante se denominará solamente como "Hospital".

MATERIAL Y MÉTODOS

Se incluyeron los expedientes de las mujeres con em-

barazo a término que fueron atendidas en la Unidad Toco-Quirúrgica del Hospital en el período comprendido de enero a junio de 2004. Sólo se incorporaron los expedientes con información completa sobre la madre y las condiciones al nacimiento del recién nacido, que contenían un RCTG intraparto con duración mínima de 20 minutos e integridad igual o mayor del 90%.

Debido a la imposibilidad de realizar una determinación rutinaria de gases sanguíneos del cordón umbilical, la clasificación de los expedientes de acuerdo al riesgo de asfixia al nacimiento, se realizó conforme a los criterios ampliamente adoptados sobre el puntaje del test de Apgar, como: Grupo I con Apgar ≥ 7 y Grupo II con Apgar < 7 .^{3,4}

Considerando una especificidad de 90% con límites de confiabilidad del 10% por encima y por debajo del valor esperado, el tamaño muestral mínimo fue de 35 casos por grupo, los cuales fueron seleccionados por asignación al azar, mediante el empleo de una tabla de números aleatorios, después de haber construido un listado en orden progresivo con la totalidad de los expedientes.

Se obtuvieron copias de los RCTG para su interpretación por 2 grupos de observadores. El Grupo A se conformó por 2 médicos especialistas en ginecología y obstetricia con experiencia mayor de 5 años y capacitación de posgrado en cardiotocografía; el Grupo B se integró con 2 médicos residentes de cuarto año (último año de la especialización) de ginecología y obstetricia. Ambos grupos permanecieron ciegos a los antecedentes médicos de las mujeres, así como a la información relativa al resultado obstétrico y perinatal.

La decisión de incluir 2 especialistas y 2 residentes respondió a que, en su diseño, este estudio fuera lo más parecido al escenario real en términos de quienes interpretan la prueba y toman decisiones obstétricas en lo cotidiano. No se pretende establecer generalizaciones sobre el desempeño de la prueba, debido a que ya se ha estudiado en forma amplia por diferentes autores y no corresponde al objetivo del estudio, sino conocer el desempeño individual de la prueba, la concordancia que existe entre observadores habituales y las decisiones que toman en función de sus resultados. Por lo anterior, los índices de concordancia se calcularon en pares de acuerdo al nivel de experiencia y el desempeño de la prueba se midió en forma individual para cada uno de los observadores.

La totalidad de los RCTG fueron interpretados por cada uno de los médicos en sesiones presenciales bajo la supervisión de una de las investigadoras del proyecto (AGC). Posterior al análisis de las características del RCTG se les solicitó su impresión diagnóstica como normal, sospechoso o patológico y un abordaje terapéutico propuesto como

intervención (cesárea) o no intervención (trabajo de parto con monitorización o sin monitorización).

Análisis

Para resumir los resultados se emplearon medidas de tendencia central y dispersión acordes con la escala de medida de las variables en estudio.

Para identificar las diferencias entre los grupos se utilizó la prueba t de Student para grupos independientes (variables en escala cuantitativa continua) y la Prueba de Chi cuadrada (variables en escala cualitativa nominal). Se consideró una diferencia como estadísticamente significativa con un valor de $p < 0.05$.

Para medir la concordancia entre observadores se calculó el índice Kappa y sus intervalos de confianza del 95%. Para su interpretación se utilizaron los criterios de Landis y Koch: 0-0.20, pobre; 0.21-0.40, leve; 0.41-0.60, moderada; 0.61-0.80, buena; y 0.81-1.00, muy buena.⁵

Para evaluar el rendimiento de la prueba diagnóstica (RCTG) en los diferentes observadores se calculó la sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN) y razones de verisimilitud, positiva (RVP) y negativa (RVN). Para la interpretación de la RVP con la utilización del nomograma de Fagan, se consideró una probabilidad pre-prueba de 30% de fetos con hipoxia o acidosis al nacimiento cuando el RCTG intraparto es sospechoso u ominoso.¹ En este mismo sentido, se consideró una probabilidad pre-prueba para la RVN de 90% de fetos sanos cuando el RCTG se encuentra en criterios de normalidad.⁶

RESULTADOS

Se incluyeron un total de 70 expedientes de los que 35 integraron el Grupo I (Apgar a los 5 minutos ≥ 7) y 35 conformaron el Grupo II (Apgar a los 5 minutos < 7). Las características generales y obstétricas de ambos grupos, se presentan en el Cuadro 1.

El análisis de la concordancia entre observadores de acuerdo a su nivel de experiencia en la interpretación del RCTG reveló que entre los especialistas el valor de Kappa fue de 0.43 (IC95% 0.25-0.60) y entre residentes el valor de Kappa fue de 0.55 (IC95% 0.36-0.73). Del análisis global de los 4 observadores se obtuvo un valor de Kappa de 0.48 (IC95% 0.36-0.73).

Los resultados de la evaluación del desempeño del RCTG como prueba diagnóstica para asfixia perinatal al ser interpretada por los distintos observadores se presentan en el Cuadro 2. Con los valores de la RVP de cada uno de los observadores se utilizó el nomograma de Fagan y se obtuvo que en forma conjunta cuando la prueba está alterada, la probabilidad de compromiso fetal aumentó de 30

(probabilidad pre-prueba) a 40-45% (probabilidad post-prueba). El mismo tratamiento con los valores de la RVN arrojó que cuando la prueba no está alterada, la probabilidad de que no exista hipoxia se redujo del 90% (probabilidad pre-prueba) a 70-85% (probabilidad post-prueba).

En el Cuadro 3 se contrastan las decisiones clínicas tomadas por los diferentes observadores de acuerdo a los RCTG correspondientes a distintos valores de Apgar, incluyendo un subgrupo de intervención justificada con Apgar ≤ 5 .

Cuadro 1. Características generales y obstétricas.

		GRUPO I (APGAR ≥ 7) N=35	GRUPO II (APGAR < 7) N=35	p
Edad materna (años)		24.34 ± 6.46	24 ± 6.94	0.83
Edad gestacional por fecha de última menstruación (semanas)		39.14 ± 1.75	38.77 ± 1.53	0.34
Edad gestacional por el método de Capurro (semanas)		38.94 ± 1.13	39 ± 1.60	0.62
Peso (gramos)		2958.1 ± 498.8	2882.8 ± 764.3	0.86
Paridad	G1	16 (45.7%)	20 (57.1%)	0.63
	G2-G3	14 (40%)	11 (31.4%)	
	$\geq G4$	5 (14.3%)	4 (11.4%)	
Control prenatal		30 (85.7%)	28 (80%)	0.52
Hipomotilidad fetal		4 (11.4%)	4 (11.4%)	1.0
Oligohidramnios		12 (34.3%)	12 (34.3%)	0.64
Meconio		7 (20%)	14 (40%)	0.06
Presentación cefálica		33 (94.3%)	32 (91.4%)	
Vía del nacimiento	Parto	14 (40%)	6 (17%)	0.03*
	Cesárea	21 (60%)	29 (83%)	
Sexo del recién nacido	Femenino	17 (48.6%)	11 (31.4%)	0.17
	Masculino	18 (51.4%)	23 (65.7%)	

* $p < 0.05$ Diferencia estadísticamente significativa. Prueba de Chi cuadrada de Pearson.

Cuadro 2. Rendimiento de la prueba RCTG entre observadores.

OBSERVADOR	SENSIBILIDAD	ESPECIFICIDAD	VPP	VPN	RVP	RVN
Especialista 1	51.4 (50-53)	65.7 (64-67)	60 (58-62)	57.5 (56-59)	1.5	0.7
Especialista 2	82.9 (81-84)	42.9 (41-44)	59.2 (58-60)	71.4 (69-74)	1.4	0.4
Residente 1	54.3 (53-56)	71.4 (70-73)	65.5 (64-67)	61 (60-62)	1.9	0.6
Residente 2	74.3 (73-76)	60 (58-62)	65 (64-66)	70 (68-72)	1.9	0.4

Cuadro 3. Relación entre las decisiones clínicas y el Apgar a los 5 minutos.

OBSERVADOR	APGAR ≥ 7 N=35		APGAR < 7 N=35		APGAR < 5 N=19	
	NI	I	NI	I	NI	I
Especialista 1	28 (80%)	7 (20%)	24 (68.6%)	11 (31.4%)	13 (68.4%)	6 (31.6%)
Especialista 2	31 (88.6%)	4 (11.4%)	25 (71.4%)	10 (28.6%)	14 (73.7%)	5 (26.3%)
Residente 1	28 (80%)	7 (20%)	22 (63%)	13 (37%)	13 (68.4%)	6 (31.6%)
Residente 2	26 (74.3%)	9 (25.7%)	19 (54.3%)	16 (45.7%)	11 (57.9%)	8 (42.1%)
PORCENTAJE INTEGRADO DEL TOTAL DE OBSERVADORES	81%	19%	64.3%	64.3%	67%	33%

DISCUSIÓN

Tradicionalmente se ha considerado que si una prueba diagnóstica se aplica en una población de individuos semejante y los observadores tienen la misma experiencia, su desempeño no variará sustancialmente cuando se utilice en otros escenarios que cumplan con condiciones semejantes. Sin embargo, en todas las pruebas médicas basadas en observaciones, como la interpretación de radiografías, electrocardiogramas, angiogramas, cardiotocogramas y especímenes de patología, subyace una fuente de heterogeneidad dada por la variabilidad entre

distintos observadores y aún por el mismo observador en ocasiones separadas entre sí por un lapso de tiempo. Aun en grupos altamente especializados se ha evidenciado que dos expertos pueden tener desacuerdos hasta en el 20% de los casos.⁷

Beaulieu et al., en 1982 demostraron una marcada variabilidad en la interpretación de cardiotocogramas intraparto de un revisor a otro, aun tratándose de gineco-obstetras experimentados en la prueba⁸ y desde entonces diferentes autores la han corroborado valiéndose de diversos métodos estadísticos para medir la variación interobservador.⁹ De entre los diferentes factores que

contribuyen a la baja eficacia de la cardiotocografía intraparto la pobre reproducibilidad es esencial debido a su impacto directo en la toma de decisiones. En el presente estudio, la interpretación del RCTG intraparto como único elemento para la toma de decisiones conduciría a que el 19% de los recién nacidos sin asfixia naciera vía cesárea. A este valor habría que agregar el correspondiente a las cesáreas justificadas por una indicación materna u obstétrica no asociada a asfixia perinatal, para entender el impacto, en la proporción de nacimientos atendidos por cesárea, que tiene la monitorización electrónica rutinaria de la frecuencia cardíaca fetal intraparto en la población de bajo riesgo. Por otra parte, en el caso de los recién nacidos con una alta probabilidad de hipoxia fetal, la decisión de no intervenir alcanzó en conjunto el 67%, lo cual nos lleva a concluir que, en las condiciones reales de uso del RCTG intraparto por este grupo hospitalario, la prueba diagnóstica no sólo no tiene ninguna utilidad clínica, sino que incluso se constituye en una práctica de riesgo, tanto por la alta tasa de falsos positivos ampliamente descrita para la prueba, como por la elevada tasa de falsos negativos que varió entre 25-49% en este grupo de observadores, independientemente de su nivel de experiencia. De modo tal que, desde la perspectiva de la toma de decisiones clínicas, la aplicación rutinaria de la prueba en forma aislada resultaría en un incremento en la tasa de cesáreas y un número elevado de recién nacidos asfixiados no detectados durante el trabajo de parto. Sin embargo, cuando contrastamos las decisiones reales con las propuestas de intervención de los observadores que participaron en esta investigación, encontramos que la frecuencia de la operación cesárea en el Grupo II (Apgar a los 5 minutos <7) fue de 83% en comparación al 35.7% del estudio, lo que probablemente represente la participación en la toma de decisiones obstétricas de otros elementos que conforman el escenario clínico real y que no están incluidos en la interpretación aislada del RCTG. Lo mismo aplica para el 60% de cesáreas en la población con Apgar a los 5 minutos ≥ 7 , en comparación al 19% del estudio. De lo anterior se desprende que, en la indicación de la cesárea por este grupo participan tanto factores médicos como extra médicos que provocan que la frecuencia de la intervención quirúrgica sea elevada más allá de la interpretación del RCTG.

Como se pudo comprobar en este trabajo, las inconsistencias en los datos recogidos a través de este método de medición pueden tener impactos definitivos en decisiones cruciales que coloquen en riesgo la seguridad de la paciente. En este contexto, Young et al., encontraron como causas frecuentes del cuidado intraparto subóptimo, los errores en la actuación, el retraso en la respuesta y las fallas en la identificación de

anormalidades del RCTG.¹⁰

Por ello, diversos colegios, asociaciones y agencias internacionales se han dado a la tarea, desde los inicios de la cardiotocografía, de elaborar guías de aceptación internacional para la interpretación y manejo clínico de los diferentes hallazgos en el RCTG ante e intraparto.¹¹

En el caso específico del RCTG, se ha procurado incrementar el rendimiento de la prueba añadiéndole otras, como las estimulación vibroacústica, la determinación del pH del cuero cabelludo fetal, la oximetría de pulso fetal y el análisis del segmento ST del electrocardiograma fetal, así como reducir la variabilidad inter e intraobservador a través de sistemas de análisis computarizado, elaboración de guías de interpretación de la FCF y diseño de metodologías adecuadas de capacitación, entre las que destacan la instrucción en las bases fisiológicas del cardiotocograma, el entrenamiento en interpretación basado en computadora, toma de decisiones basada en los resultados de la monitorización y trabajo en equipo frente a la emergencia obstétrica. Diferentes autores recomiendan que el entrenamiento incluya a todo el personal involucrado en la interpretación de la prueba y toma de decisiones y que este se repita con una periodicidad no mayor a 6 meses.¹⁰⁻¹²

Finalmente, las investigaciones orientadas a identificar el nivel y las causas de la variabilidad en distintas actividades clínicas, así como identificar los eventos adversos graves que se derivan de ella, pueden contribuir decisivamente a la mejora de la calidad y seguridad de la atención médica.¹³

REFERENCIAS

1. Barrena MN, Carvajal CJ. Evaluación fetal intraparto. Análisis crítico de la evidencia. *Rev Chil Obstet Ginecol* 2006; 71:63-68.
2. Alfirevic Z, Devane D, Gyte GML. Cardiotocografía continua (CTG) como una forma de monitoreo electrónico fetal (MEF) para evaluación durante el trabajo de parto. Base de datos Cochrane de Revisiones Sistemáticas 2006, Issue 3. No. CD006066.
3. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Eng J Med* 2001; 344:467-71.
4. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstet Gynecol* 2001; 98:65-70.
5. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-74.
6. Schiermeier S, Pildner von Steinburg S, Thieme A, Reinhard J, Daumer M et al. Sensitivity and specificity

of intrapartum computerized FIGO criteria for cardiotocography and fetal scalp pH during labour: multi-centre, observational study. *BJOG* 2008; 115:1557-1563.

7. Leape LL, Berwick DM, Bates DW. What practices will most improve safety? Evidence-based medicine meets patient safety. *JAMA* 2002; 288:501-7.
8. Beaulieu MD, Fabia J, Leduc B, Brisson J, Bastide A et al. The reproducibility of intrapartum cardiotocogram assessments. *CMA Journal* 1982; 127: 214-6.
9. Palomäki O, Luukkaala T, Luoto R, Tuimala R. Intrapartum cardiotocography – the dilemma of interpretational variation. *J Perinat Med* 2006; 34: 298-302.
10. Young P, Hamilton R, Hodgett S, Moss M, Rigby C et al. Reducing risk by improving standards of intrapartum fetal care. *J R Soc Med* 2001; 94:226-31.
11. Santo S, Ayres de Campos D. Human factors affecting the interpretation of fetal heart rate tracings: an update. *Curr Opin Obstet Gynecol* 2012; 24:84-8.
12. Pehrson C, Sorensen JL, Amer-Wahlin I. Evaluation and impact of cardiotocography training programmes: a systematic review. *BJOG* 2011; 118:926-35.
13. Grimes DA, Peipert JF. Electronic fetal monitoring as a public Elath screening program, the arithmetic of failure. *Obstet Gynecol* 2010; 116:1397-1400.



Centro de Investigación Materno Infantil Gen

Encuentro de la Red de Unidades de Investigación en Enfermería ENEO - UNAM Instituciones e Institutos Nacionales de Salud

Tema Central
**"La Investigación y la Práctica del Cuidado
Contribuyendo a la Salud"**



Dirigido a Profesionales de la Salud **25 y 26 de Octubre 2012**
08:00 hrs a 14:00 hrs

Lugar: Auditorio "Alfonso Caso"
Unidad de Posgrado
junto a la Torre de Humanidades C. U. **talleres**

Costo: \$200.00 m/n sólo evento \$300.00 m/n evento y taller	Consultoría Prenatal* 25 de Octubre 2012 Horario: 16:00 a 20:00 hrs.	Introducción al SPSS* 25 de Octubre 2012 Horario: 16:00 a 20:00 hrs.
------------------------------------------------------------------------	--------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------

Informes: Enseñanza CIMIGen
Tel: 5012370454
ensenanza@cimigen.org.mx
cimigen1@cimigen.org.mx
www.cimigen.org.mx

*** Los Talleres se realizarán en las instalaciones de la E.N.E.O. - U.N.A.M. Camino Viejo a Xochimilco y Viaducto Tlalpan s/n Col. San Lorenzo Huipulco, Delegación Tlalpan, México, D. F. C. P. 14370.**