



<https://doi.org/10.24245/gom.v90i1.6992>

## Comentario y respuesta al artículo: "Comparación de cuatro tablas de crecimiento fetal para la predicción de desenlaces perinatales adversos en un hospital de tercer nivel de México".

Gineol Obstet Mex 2021; 89 (9): 704-14. <https://doi.org/10.24245/gom.v89i9.5817>

### Comparison of four growth charts fetal for the prediction of perinatal outcomes adverse events in a tertiary hospital in Mexico.

Héctor Oviedo-Cruz,<sup>1</sup> Erika Ruth Carrasco-Blancas,<sup>1</sup> Marco Antonio Cortes-Martínez<sup>2</sup>

#### Sr. Editor:

El trabajo de Mendoza-Carrera *et al*<sup>1</sup> es único y valioso porque describe el rendimiento de cuatro curvas de crecimiento fetal para predecir eventos perinatales adversos en una población mexicana; hecho sin precedentes. Sin embargo, no se realizó la comparación estadística entre los indicadores de rendimiento de las pruebas. De acuerdo con el objetivo del estudio, "comparar el rendimiento", esta es una deficiencia metodológica que afecta la conclusión por privarle de sustento formal.

El desempeño de las pruebas de predicción se evalúa de forma similar al de las pruebas diagnósticas.<sup>2</sup> Para comparar el desempeño entre dos pruebas diagnósticas en el mismo grupo de pacientes, el análisis estadístico puede basarse en la diferencia entre las sensibilidades y especificidades<sup>3</sup> o en la comparación pareada de las áreas bajo las curvas de característica de receptor-operador (AUC-ROC).<sup>4</sup>

<sup>1</sup> Departamento de Medicina Materno-Fetal, Centro Médico para Atención Especializada. Hospital Español. Ciudad de México.

<sup>2</sup> Departamento de Bioquímica y Calidad, Laboratorio CEMAFE, Ciudad de México.

**Recibido:** octubre 2021

**Aceptado:** diciembre 2021

#### Correspondencia

Héctor Oviedo Cruz  
[hector.oviedo@antesdelparto.com](mailto:hector.oviedo@antesdelparto.com)

#### Este artículo debe citarse como:

Comentario y respuesta al artículo:  
"Comparación de cuatro tablas de crecimiento fetal para la predicción de desenlaces perinatales adversos en un hospital de tercer nivel de México".  
Ginecol Obstet Mex. 2022; 90 (1): 115-118.

El estadístico más recomendable para comparar las sensibilidades y especificidades entre dos pruebas es el de McNemar,<sup>5</sup> que utiliza la distribución de  $\chi^2$ ;<sup>6</sup> ésta también sirve de base para los estadísticos que comparan tres o más pruebas.<sup>7</sup>

La comparación entre más de dos curvas ROC se hace mediante estadísticos de diseño propio,<sup>8-11</sup> preferentemente codificados en programas de cómputo;<sup>12-14</sup> pero el análisis también puede basarse solo en los intervalos de confianza de las AUC-ROC.<sup>15</sup>

El documento de Mendoza-Carrera *et al*<sup>1</sup> omitió explicar cómo se llegó al tamaño de muestra, incumpliendo las recomendaciones STROBE<sup>16</sup> declaradas por los autores; es menester reportarlo incluso si fue por conveniencia, como en estudios piloto o cuando no hay datos previos, aunque suele basarse en los índices de desempeño de la prueba en cuestión.<sup>17</sup>

Por lo anterior, el trabajo de Mendoza-Carrera *et al*<sup>1</sup> es único y valioso, pero su conclusión debió ponderarse formalmente con alguna prueba estadística.

## REFERENCIAS

- Mendoza-Carrera CE, Acevedo-Gallegos S, M. L-M, Gallardo-Gaona JM and Copado-Mendoza DY. Comparación de cuatro tablas de crecimiento fetal para la predicción de desenlaces perinatales adversos en un hospital de tercer nivel de México. *Ginecol Obstet México* 2021;89:704-714. <https://doi.org/10.24245/gom.v89i9.5817>
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
- Hawass NE. Comparing the sensitivities and specificities of two diagnostic procedures performed on the same group of patients. *Br J Radiol* 1997;70:360-6. <https://doi.org/10.1259/bjr.70.832.9166071>
- Bandos AI, Rockette HE and Gur D. A conditional nonparametric test for comparing two areas under the ROC curves from a paired design. *Acad Radiol* 2005;12:291-7. <https://doi.org/10.1016/j.acra.2004.08.013>
- Kim S and Lee W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? *Stat Methods Med Res* 2017;26:142-154. <https://doi.org/10.1177/0962280214541852>
- Traiman A and Luiz RR. McNemar chi2 test revisited: comparing sensitivity and specificity of diagnostic examinations. *Scand J Clin Lab Invest* 2008;68:77-80. <https://doi.org/10.1080/00365510701666031>
- Yu Q, Tang W, Ma Y, Gamble SA and Tu XM. Comparing Multiple Sensitivities and Specificities with Different Diagnostic Criteria: Applications to Sexual Abuse and Sexual Health Research. *Comput Stat Data Anal* 2008;53:27-37. <https://doi.org/10.1016/j.csda.2008.05.031>
- Hanley JA and McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-43. <https://doi.org/10.1148/radiology.148.3.6878708>
- Bandos AI, Rockette HE and Gur D. A permutation test for comparing ROC curves in multireader studies a multireader ROC, permutation test. *Acad Radiol* 2006;13:414-20. <https://doi.org/10.1016/j.acra.2005.12.012>
- Tang L, Du P and Wu C. Compare diagnostic tests using transformation-invariant smoothed ROC curves(). *J Stat Plan Inference* 2010;140:3540-3551. <https://doi.org/10.1016/j.jspi.2010.05.026>
- Blanche P, Dartigues JF and Riou J. A closed max-t test for multiple comparisons of areas under the ROC curve. *Biometrics* 2020; <https://doi.org/10.1111/biom.13401>
- Vergara IA, Norambuena T, Ferrada E, Slater AW and Melo F. StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics* 2008;9:265. <https://doi.org/10.1186/1471-2105-9-265>
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77. <https://doi.org/10.1186/1471-2105-12-77>
- Novoselova N, Della Beffa C, Wang J, Li J, Pessier F, *et al*. HUM calculator and HUM package for R: easy-to-use software tools for multicategory receiver operating characteristic analysis. *Bioinformatics* 2014;30:1635-6. <https://doi.org/10.1093/bioinformatics/btu086>
- Zou GY and Yue L. Using confidence intervals to compare several correlated areas under the receiver operating characteristic curves. *Stat Med* 2013;32:5077-5090. <https://doi.org/10.1002/sim.5889>
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, *et al*. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-9. <https://doi.org/10.1016/j.jclinepi.2007.11.008>
- Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193-204. <https://doi.org/10.1016/j.jbi.2014.02.013>



### Respuesta al comentario anterior

Agradecemos los comentarios y retroalimentación constructiva por parte de los doctores Oviedo-Cruz, Carrasco-Blancas y Cortés-Martínez. Dado que en el estudio de Mendoza-Carrera *et al* (1) evaluó el rendimiento predictivo global de cuatro tablas de crecimiento fetal en relación con un desenlace perinatal adverso compuesto utilizando diferentes parámetros (es decir, riesgo relativo, sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, sensibilidad ajustada a una tasa de falsos positivos del 10%, y área bajo la curva [AUC, *c-statistic*]), la comparación formal de cada uno de esos parámetros entre las cuatro tablas utilizando pruebas estadísticas como se sugiere, incrementaría de manera significativa la tasa de falsos positivos<sup>2</sup> y sería necesario utilizar algún método para corregir dicha multiplicidad (por ejemplo, la prueba propuesta por Bonferroni). Igualmente, se requeriría realizar una comparación por pares para cada parámetro ya que no se estarían evaluando dos tablas, sino cuatro, lo que requeriría un tamaño de muestra mucho mayor.<sup>3</sup> También es importante mencionar que en el estudio de Mendoza-Carrera *et al* se realizaron varios análisis de subgrupos (por ejemplo, resultado respiratorio compuesto, ingreso a unidad de cuidados intensivos neonatales, hiperbilirrubinemia) lo que incrementaría aún más la tasa de falsos positivos si se llevara a cabo una comparación estadística formal.

Asimismo, agradecemos la oportunidad para aclarar que en el estudio de Mendoza-Carrera *et al* se utilizó un método de muestreo por conveniencia,<sup>4</sup> incluyendo todos los datos disponibles en el periodo de estudio. En la discusión del artículo de Mendoza-Carrera *et al* se destaca el tamaño de muestra relativamente pequeño y la necesidad de realizar más proyectos de investigación (idealmente multicéntricos) en población mexicana y en otros países de América

Latina para confirmar los resultados reportados en dicho estudio.

La información presentada en el estudio de Mendoza-Carrera *et al* podría ser de utilidad para realizar el análisis de poder para estudios de investigación similares en los que se plantee una hipótesis específica definida *a priori*<sup>5</sup> (por ejemplo, se plantea la hipótesis que la tabla de crecimiento fetal de la Fundación de Medicina Fetal Barcelona<sup>6</sup> tiene un mejor rendimiento predictivo [tomando el AUC como referencia] que la tabla propuesta por INTERGROWTH-21<sup>st</sup> en población mexicana). Por otra parte, con una única comparación e hipótesis correspondiente, se evitaría el problema de multiplicidad antes descrito.

Finalmente, estamos de acuerdo con los doctores Oviedo-Cruz, Carrasco-Blancas, y Cortés-Martínez en cuanto a diseñar y analizar estudios de investigación con el mayor rigor metodológico posible, siguiendo siempre las recomendaciones propuestas por la red *Enhancing the QUALity and Transparency Of health Research (EQUATOR)*<sup>8</sup> y por otras organizaciones internacionales especializadas en ginecología y obstetricia.

Claudia Elvira Mendoza-Carrera, Sandra Acevedo-Gallegos, Mario Lumbreras-Márquez, Juan M Gallardo-Gaona, Diana Yazmín Copado-Mendoza, María J Rodríguez-Sibaja

### REFERENCIAS

1. Mendoza-Carrera CE, Acevedo-Gallegos S, Lumbreras-Márquez M, Gallardo-Gaona JM, Copado-Mendoza DY, Rodríguez-Sibaja MJ. Comparison of four fetal growth charts in the prediction of adverse perinatal outcomes in a tertiary hospital in Mexico. *Ginecología y Obstetricia de México*. 2021 Sep 1;89(9):704–14.
2. Dmitrienko A, D'Agostino RB. Multiplicity Considerations in Clinical Trials. *New England Journal of Medicine*. 2018 May 31;378(22):2115–22.

3. Li G, Taljaard M, van den Heuvel ER, Levine MAH, Cook DJ, Wells GA, et al. An introduction to multiplicity issues in clinical trials: The what, why, when and how. *International Journal of Epidemiology*. 2017;46(2):746–56.
4. Battaglia M. Convenience sampling. In: Lavrakas P, editor. *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America : Sage Publications, Inc.; 2011. p. 149.
5. Harrington D, D'Agostino RB Sr, Gatsonis C. New guidelines for statistical reporting in the Journal. *New England Journal of Medicine*. 2019 Oct 17;381(16):1595–7.
6. Figueras F, Meler E, Iraola A, Eixarch E, Coll O, Figueras J, et al. Customized birthweight standards for a Spanish population. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2008;136(1):20–4.
7. Papageorghiou AT, Ohuma EO, Altman DG, Todros T, Ismail LC, Lambert A, et al. International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *The Lancet*. 2014;384(9946):869–79.
8. EQUATOR Network Group.[Internet] [Consultado 15 dic 2021]. [Disponible en: <http://www.equator-network.org/>]

### CITACIÓN ACTUAL

De acuerdo con las principales bases de datos y repositorios internacionales, la nueva forma de citación para publicaciones periódicas, digitales (revistas en línea), libros o cualquier tipo de referencia que incluya número doi (por sus siglas en inglés: Digital Object Identifier) será de la siguiente forma:

#### REFERENCIAS

1. Yang M, Gou, ZW, Deng CJ, Liang X, et al.\* A comparative study of three different forecasting methods for trial of labor after cesarean section. *J Obstet Gynaecol Res*. 2017;25(11):239-42. <https://doi.org/10.1016/j.gyobfe.2015.04..0015>.\*\*
- \* Cuando la referencia contiene hasta tres autores, éstos se colocarán de forma completa. En caso de 5 autores o más, solo se colocan cuatro, seguidos de la palabra en latín “et al”.
- \*\* El registro Doi deberá colocarse con el link completo (como se indica en el ejemplo).