

LA APLICACIÓN DEL ANÁLISIS POR SUBGRUPOS Y MACHINE LEARNING PARA MEJORAR EL ESTUDIO DEL CÁNCER DE MAMA

Julianna Ribera, doctora en Medicina y Biología, directora de Operaciones y cofundadora de Amalfi Analytics.
Manuel David Gil, Farmacéutico hospitalario del Hospital Universitario de Puerto Real, Cádiz.

CONTEXTO

El cáncer de mama es una de las neoplasias más frecuentes en las mujeres. El análisis por subgrupos permite valorar la eficacia y seguridad de los tratamientos en pacientes con determinadas características en este contexto clínico. Por otra parte, para profundizar en la variabilidad de casos, sus comorbilidades asociadas y en la evolución de las pacientes puede usarse la tecnología *machine learning*.

La mayoría de las pacientes con diagnóstico de **cáncer de mama metastásico** tienen un perfil hormonal positivo y reciben terapia endocrina. Los autores del ensayo MONARCH 2¹, en el que a este tratamiento se le añade un inhibidor de las quinasas dependientes de ciclinas 4 y 6 (CDK4 y CDK6), sugirieron un mayor beneficio en pacientes con metástasis viscerales y resistencia primaria al tratamiento endocrino. Sin embargo, esta afirmación no se basa en ningún método riguroso.

Para valorar si esta afirmación era cierta, se empleó la metodología de Gil-Sierra et al.², que es una herramienta de análisis de subgrupos en ensayos clínicos. **La misma permite valorar si existen diferencias estadísticamente significativas entre los subgrupos; la plausibilidad biológica de las hipotéticas diferencias en los efectos de los tratamientos; y la consistencia de los resultados de los subgrupos entre estudios.** Tal y como ha señalado la Agencia Europea del Medicamento³, el análisis por subgrupos cuenta con un impacto creciente en el posicionamiento terapéutico de los fármacos.



El análisis por subgrupos permite valorar la eficacia y seguridad de los tratamientos

TECNOLOGÍA ANÁLISIS DE DATOS



Los ensayos clínicos pueden apoyarse en el análisis de subgrupos con *machine learning*

RETO

Optimizar el manejo de los pacientes con cáncer de mama mediante el uso de herramientas para la interpretación de datos de ensayos clínicos como el análisis por subgrupos, así como el uso de la tecnología *machine learning* en datos de la vida real.

¹ Sledge GW Jr, Toi M, Neven P, et al. **The Effect of Abemaciclib Plus Fulvestrant on Overall Survival in Hormone Receptor-Positive, ERBB2-Negative Breast Cancer That Progressed on Endocrine Therapy-MONARCH 2: A Randomized Clinical Trial.** JAMA Oncol. 2019. doi:10.1001/jamaoncol.2019.4782.

² Gil-Sierra MD, Fénix-Caballero S, Abdel Kader-Martin L, Fraga-Fuentes MD, Sánchez-Hidalgo M, Alarcón de la Lastra-Romero C, et al. **Checklist for clinical applicability of subgroup analysis.** J Clin Pharm Ther. 2020;45(3):530-8. Doi: 10.1111/jcpt.13102

³ Sun X, Briel M, Walter SD, Guyatt GH. **Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses.** BMJ 2010;340:c117. DOI:10.1136/bmj.c117.

SOLUCIÓN

Se aplicó la **herramienta de Gil-Sierra et al.**, con la que se descartó asumir una diferencia en la eficacia del inhibidor de las quinasas dependientes de ciclinas 4 y 6 en los subgrupos con metástasis viscerales y resistencia primaria al tratamiento endocrino con respecto a sus subgrupos complementarios, al no observarse diferencia estadísticamente significativa entre las subpoblaciones.

Por otra parte, se propone utilizar **la tecnología machine learning** con datos reales de pacientes con cáncer de mama con los siguientes objetivos:

1. Mejorar la personalización por grupos de protocolos a través del estudio de la variabilidad de casos.
2. Analizar las comorbilidades asociadas a los diferentes grupos para observar si existen asociaciones y problemas con exceso de presencia entre las pacientes con cáncer de mama.
3. Examinar la evolución de estas pacientes y las transiciones entre sus estados de salud.

DESARROLLO

Al aplicar la herramienta sobre análisis por subgrupos a los resultados del ensayo MONARCH2, se consiguió evitar una interpretación inadecuada del análisis por subgrupos, en aras de una mejor eficiencia.

Para la aplicación del *machine learning*, **se procesaron los datos** rutinarios de pacientes con cáncer de mama atendidas en diversos hospitales de una región en un periodo de un año. El objetivo era analizar la variabilidad de casos para mejorar la personalización por grupos de los protocolos.

Con el fin de estudiar los **patrones de comorbilidad** de las pacientes se utilizó la técnica **Naïve Bayes Clustering**⁴. Se definieron cinco clústeres distintos, en función de la edad y comorbilidades (problemas metabólicos y depresión). A través de este análisis se detectó que el grupo de más alto coste y con una mortalidad superior era el de elevada presencia de metástasis.

Para saber qué comorbilidades se asociaban a qué clústeres se utilizó la técnica de **frequent itemsets**⁵. Las pacientes con cáncer de mama presentaban asociaciones y problemas con exceso de presencia comparativamente con el resto de la población base de todos los pacientes atendidos. Por ejemplo, se detectaron el triple de casos de accidente isquémico transitorio.

Por último, se empleó la metodología **Continuous-Time Markov Chain**⁶ para visualizar la evolución de los pacientes. La misma identifica los estados temporales de las pacientes, y también las transiciones entre ellos. De este modo, se pueden estudiar estados previos a determinadas complicaciones, y también estados de alto riesgo para prevenir una evolución no deseada.

CONCLUSIÓN

Este tipo de análisis ofrece una visión completa de las pacientes y tiene en cuenta, además del cáncer de mama, otros problemas de salud, tratamientos y circunstancias sociales. Por eso, permite representar la población real afectada, lo que es útil para centrar los recursos en los grupos de alto riesgo.



Este tipo de análisis ofrece una visión completa de las pacientes

⁴ J. Bordas-Martínez, R. Gavalda, J.G. Shull, V. Vicens-Zygmunt, L. Planas-Cerezales, G. Bermudo-Peloché, S. Santos, N. Salord, C. Monasterio, M. Molina-Molina, G. Suarez-Cuartin. **IPF cluster analysis highlights diagnostic delay and cardiovascular comorbidities association with outcome.** *ERJ Open Research, a journal of the European Respiratory Society*. Jan 2021. 00897-2020; DOI: 10.1183/23120541.00897-2020. Disponible en: <https://openresersjournals.com/content/7/2/00897-2020>

⁵ Marti Zamora, Manel Baradad, Ester Amado, Silvia Cordoní, Esther Limón, Juliana Ribera, Marta Arias, Ricard Gavalda. **Characterizing Chronic Disease and Polymedication Prescription Patterns from Electronic Health Records.** 2nd IEEE Intl. Conf. on Data Science and Advanced Analytics (IEEE DSAA 2015). Disponible en: <https://www.cs.upc.edu/~gavalda/papers/dsaa2015-techrep.pdf>

⁶ Marti Zamora and Ricard Gavalda. **Interpretable Patient Trajectories from Temporally Annotated Health Records.** Poster at 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).