Trabajo original

Concordancia entre las recomendaciones terapéuticas de un ateneo multidisciplinario y un sistema de inteligencia artificial basado en guías clínicas internacionales en cáncer de mama: estudio comparativo de 150 presentaciones consecutivas.

Enrique Díaz Cantón[1], Priscila Giselle Holub[2], Joyce Navarro[2], Consuelo Martínez[2], María Noel Hernández[2], Cristina Prychodzko[2], Bibiana Fernanda Naso[2], Florencia Ghegin[2], Carla Ypa[2], Javier Leveroni[2], Giselle Pizarro[3], Federico Díaz[4], Gustavo Hauszpigiel[2]

Resumen

Introducción

La inteligencia artificial se ha propuesto como herramienta de apoyo a la decisión terapéutica en oncología mamaria. La evidencia disponible proviene mayoritariamente de modelos de propósito general sin guías embebidas, con concordancias del 16% al 70% frente a un ateneo multidisciplinario.

Objetivo

Evaluar la concordancia entre las recomendaciones terapéuticas de un ateneo multidisciplinario y las generadas por un sistema de inteligencia artificial configurado con guías clínicas internacionales y criterios de elegibilidad de ensayos pivotales.

Materiales y método

Estudio observacional, retrospectivo y comparativo sobre 150 presentaciones consecutivas del ateneo de una Unidad de Mastología. Las historias clínicas completas se digitalizaron y las variables se

1 Instituto de Oncología e Instituto Universitario CEMIC, Buenos Aires, Argentina
2 Unidad de Mastología, Hospital General de Agudos Dr. José M. Penna, Buenos Aires, Argentina
3 Hospital General de Agudos Dr. Juan A. Fernández, Buenos Aires, Argentina
4 Mevaterapia, Buenos Aires, Argentina

Correo electrónico: ediazcanton@iuc.edu.ar

extrajeron de la fuente primaria. Ambas conductas se descompusieron en 21 componentes terapéuticos binarios agrupados en 11 ejes. La comparación principal se restringió a los ejes que el acta del ateneo enunció explícitamente. Se calcularon índice de Jaccard con intervalos de confianza por bootstrap, índice Kappa de Cohen y pruebas de McNemar exactas con corrección de Holm.

Resultados

De 150 presentaciones, 117 resultaron analizables. El índice de Jaccard medio fue 0,925 (IC 95% 0,893–0,955), con concordancia completa en 94 pacientes (80,3%), parcial en 22 (18,8%) y nula en 1 (0,9%); el acuerdo total o parcial alcanzó 116/117 (99,1%). La comparación no restringida sobre los mismos pacientes rindió 0,615 (IC 95% 0,576–0,655). Ningún componente mostró diferencia estadísticamente significativa tras corrección por comparaciones múltiples. Las divergencias se concentraron en supresión ovárica (10 casos) e inhibidores de CDK4/6 (4 casos).

Conclusiones

Un sistema de inteligencia artificial anclado en guías reprodujo la decisión del ateneo en la gran mayoría de las presentaciones. Las divergencias residuales se concentran en ejes identificables y clínicamente interpretables. Los resultados sostienen un rol de verificación y de ampliación del acceso a recomendaciones de calidad, sin reemplazar el juicio clínico multidisciplinario.

Palabras Clave

cáncer de mama; ateneo multidisciplinario; inteligencia artificial; guías de práctica clínica; concordancia terapéutica.

Abstract

Introduction

Artificial intelligence has been proposed as a decision-support tool in breast oncology. Available evidence derives mostly from general-purpose models without embedded guidelines, reporting 16% to 70% concordance with a multidisciplinary tumour board.

Objective

To assess concordance between multidisciplinary tumour board therapeutic recommendations and those generated by an artificial intelligence system configured with international clinical guidelines and pivotal trial eligibility criteria.

Materials and method

Observational, retrospective, comparative study of 150 consecutive multidisciplinary tumour board presentations at a Breast Unit. Complete clinical records were digitised and variables extracted from primary source. Both recommendations were decomposed into 21 binary treatment components across 11 therapeutic axes. The primary comparison was restricted to axes explicitly addressed in the board minutes. Jaccard index with bootstrap confidence intervals, Cohen's kappa and exact McNemar tests with Holm correction were computed.

Results

Of 150 presentations, 117 were analysable. Mean Jaccard index was 0.925 (95% CI 0.893–0.955), with complete agreement in 94 patients (80.3%), partial in 22 (18.8%) and none in 1 (0.9%); total or partial agreement reached 116/117 (99.1%). Unrestricted comparison of the same patients yielded 0.615 (95% CI 0.576–0.655). No component reached statistical significance after correction for multiple comparisons. Divergences were concentrated in ovarian function suppression (10 cases) and CDK4/6 inhibitors (4 cases).

Conclusions

A guideline-anchored artificial intelligence system reproduced board decisions in the large majority of presentations. Residual divergences concentrate in identifiable, clinically interpretable axes. These findings support a role for artificial intelligence in decision verification and expanding access to high-quality recommendations, rather than replacing multidisciplinary clinical judgement.

Key words

breast cancer; multidisciplinary tumour board; artificial intelligence; clinical practice guidelines; treatment concordance.

Introducción

El ateneo multidisciplinario constituye el estándar de organización de la decisión terapéutica en cáncer de mama. Su implementación se ha asociado a mejores resultados, incluyendo una reducción de la mortalidad.[1] Sin embargo, su disponibilidad es desigual. En buena parte del sistema de salud argentino y de la región, sólo una fracción de las pacientes accede a que su caso sea discutido por un equipo completo; el resto recibe indicaciones decididas por un único profesional.

La aparición de modelos de lenguaje de gran escala reavivó el interés por sistemas de apoyo a la decisión. Los resultados publicados son heterogéneos: los sistemas de primera generación (Watson for Oncology) alcanzaron acuerdos del 73% al 93% según el escenario,[2] mientras que los modelos de lenguaje de propósito general (por ejemplo ChatGPT) mostraron concordancias globales del 58,8% para casos de carcinoma invasor.[3,4] En estos últimos, la limitación radica en que se evalúan modelos genéricos interrogados mediante texto libre, sin guías embebidas ni criterios de elegibilidad de ensayos clínicos.

El presente trabajo evalúa una aproximación distinta: un sistema configurado explícitamente sobre los árboles de decisión de las guías vigentes y los criterios de inclusión de los ensayos pivotales, comparado contra las conductas efectivamente registradas por el ateneo de una unidad de mastología de un hospital público de la Ciudad de Buenos Aires.

Materiales y método

Diseño y población

Estudio observacional, retrospectivo y comparativo. Se incluyeron 150 presentaciones consecutivas del ateneo multidisciplinario de la Unidad de Mastología del Hospital General de Agudos Dr. José M. Penna. Para cada paciente se digitalizó la historia clínica completa —motivo de consulta, antecedentes, examen físico, estadificación clínica, estudios por imágenes, anatomía patológica, inmunohistoquímica, estudios de extensión y, cuando correspondía, respuesta a la neoadyuvancia y anatomía patológica definitiva— y la conducta consignada en el acta. Las variables clínicas se extrajeron de la fuente primaria; no se utilizaron resúmenes ni datos agregados.

Sistema de inteligencia artificial

Las recomendaciones se generaron mediante un Proyecto de Claude Opus 5 (Anthropic, San Francisco, CA), diseñado y configurado por el

primer autor. El sistema opera sobre un motor de reglas explícito que codifica los árboles de decisión de las guías NCCN 2025[5], ESMO[6] y ASCO y los consensos CAP-SSO-ASTRO-ESTRO sobre márgenes quirúrgicos y volúmenes de irradiación, junto con los criterios de elegibilidad de los ensayos KEYNOTE-522,[7,8] KATHERINE,[9,10] monarchE,[11] NATALEE,[12] CREATE-X,[13] TAILORx,[14] RxPONDER,[15] SOFT/TEXT[16] y NSABP B-51/RTOG 1304.[17] Cada regla es trazable a su fuente y el motor es reproducible; su código se aporta como material suplementario. El sistema no accedió en ningún momento a la conducta del ateneo.

Descomposición y comparación

Ambas conductas se descompusieron en 21 componentes terapéuticos binarios agrupados en 11 ejes: cirugía mamaria, cirugía axilar, radioterapia, quimioterapia neoadyuvante, quimioterapia adyuvante, bloqueo anti-HER2, inmunoterapia, hormonoterapia, inhibidores de CDK4/6, escalamiento post-neoadyuvante e inhibidores de PARP, y estratificación y estudios complementarios.

Las actas del ateneo registran únicamente el eje sometido a discusión, mientras que el sistema emite siempre el plan completo. En esta serie el ateneo se pronunció en promedio sobre 2,17 de los 11 ejes por paciente. Comparar planes completos propuestos por el sistema de inteligencia artificial (IA) contra actas sintéticas del ateneo mide la exhaustividad del registro y no el acuerdo clínico. Por ello la comparación principal se enfocó en los componentes en los que el ateneo multidisciplinario se refirió explícitamente. Para el eje de radioterapia (RT), cuando el acta consignó únicamente "RT" sin detallar áreas a irradiar, se omitió la evaluación de la irradiación ganglionar regional. Como referencia se reporta también la comparación no restringida sobre los mismos pacientes.

Se consideraron no analizables las presentaciones cuya conducta fue condicional o incompleta: enunciados del tipo "si extensión negativa, quimioterapia neoadyuvante y luego cirugía", reestadificaciones pendientes, cambios de línea sin especificación de esquema y un caso con estudios de extensión no interpretables en el registro fuente.

Análisis estadístico

El acuerdo se cuantificó mediante el índice de Jaccard por paciente, con intervalos de confianza del 95% obtenidos por bootstrap de 10.000 remuestreos. Por componente se calcularon el acuerdo observado, el índice κ (kappa) de Cohen y pruebas de McNemar exactas, con corrección de Holm para comparaciones múltiples. Se realizó un análisis de sensibilidad tratando las formulaciones condicionales del acta ("evaluar X", "ofrecer X") como no indicadas.

Aspectos éticos

El protocolo fue aprobado por el comité de ética institucional excluyendo la necesidad de consentimiento informado por tratarse de un análisis retrospectivo de registros. Los datos se anonimizaron antes del procesamiento.

Resultados

De las 150 presentaciones, 33 se excluyeron por conducta condicional o incompleta (28 enunciados condicionales dependientes de la estadificación, 2 cambios de línea, 1 reestadificación pendiente, 1 esquema no consignado y 1 con estudios de extensión no interpretables), quedando 117 pacientes analizables.

La mediana de edad de la cohorte fue de 57 años (rango: 28–85), de las cuales 37 pacientes (31,6%) eran premenopáusicas. Según el subtipo tumoral, la distribución fue: luminal B (n=46; 39,3%), luminal A (n=30; 25,6%), triple negativo (n=11; 9,4%), HER2 positivo con receptores hormonales positivos (n=9; 7,7%), HER2 positivo con receptores negativos (n=9; 7,7%) y carcinoma ductal in situ (n=9;7,7%); 3 casos corresponden a lesiones de riesgo. Un total de 43 pacientes (36,8%) recibieron tratamiento neoadyuvante, de las cuales 11 lograron respuesta patológica completa.

La concordancia global media fue 0,925 (IC 95% 0,893–0,955). El acuerdo fue completo en 94 casos (80,3%), parcial en 22 (18,8%) y nulo en 1 (0,9%), alcanzando un acuerdo completo o parcial en 116 de 117 (99,1%) (Tabla 1).

Tabla 1. *Concordancia global entre el ateneo multidisciplinario y el sistema de IA (n = 117).*

Medida	Resultado
Índice de Jaccard medio (comparación eje-restringida)	0,925 (IC 95% 0,893–0,955)
Concordancia completa	94 (80,3%)
Concordancia parcial	22 (18,8%)
Concordancia nula	1 (0,9%)
Acuerdo total o parcial	116/117 (99,1%)
Comparación no restringida (mismos pacientes)	0,615 (IC 95% 0,576–0,655)
Análisis de sensibilidad (condicionales = no indicado)	0,919 (IC 95% 0,886–0,950)

El acuerdo fue homogéneo entre ejes terapéuticos (Tabla 2). Los ejes de mayor volumen —hormonoterapia y RT, enunciados en el 73,5% y el 70,9% de las presentaciones respectivamente— alcanzaron índices de 0,918 y 0,932. Los ejes de quimioterapia neoadyuvante, inmunoterapia y escalamiento post-neoadyuvante mostraron concordancia completa en todas las presentaciones en que fueron enunciados, aunque con escaso número de casos.

Tabla 2. *Concordancia por eje terapéutico.*

Eje terapéutico	n	Jaccard	Concordancia completa
Hormonoterapia	86	0,918	65 (75,6%)
Radioterapia	83	0,932	65 (78,3%)
Inhibidores de CDK4/6	21	0,855	12 (57,1%)
Bloqueo anti-HER2	16	0,939	12 (75,0%)
Cirugía mamaria	11	0,848	9 (81,8%)
Estratificación y estudios	10	0,942	8 (80,0%)
Quimioterapia adyuvante	9	0,863	4 (44,4%)
Cirugía axilar	8	0,917	7 (87,5%)
Quimioterapia neoadyuvante	4	1,000	4 (100%)
Inmunoterapia	3	1,000	3 (100%)
Escalamiento post-neoadyuvante	3	1,000	3 (100%)

Ningún componente alcanzó diferencia estadísticamente significativa tras la corrección de Holm. La mayor diferencia nominal correspondió a la supresión ovárica: en 10 pacientes el sistema la indicó y el acta no la consignó, frente a 1 caso en sentido inverso (acuerdo 87,2%; kappa 0,641; p = 0,012 sin corregir; p = 0,188 corregido). El componente de hormonoterapia mostró acuerdo completo en las 86 presentaciones en que el eje fue enunciado (kappa 1,000), y el de indicación de RT en 82 de 83 (98,8%).

Se registraron 23 pacientes con acuerdo incompleto. Cuatro de estas divergencias fueron sustantivas y merecen especial comentario en cuanto a la indicación adyuvante de inhibidores de CDK4/6. En uno de los casos, el ateneo multidisciplinario indicó el tratamiento a pesar de que la paciente no cumplía con los criterios de elegibilidad de los ensayos monarchE y NATALEE. En los tres casos restantes, se observó que la IA no consideró el estadio clínico inicial y, por ende, omitió la indicación de inhibidores CDK4/6 en pacientes que contaban con criterio para recibirlo.

En dos pacientes el ateneo indicó quimioterapia adyuvante directa en escenarios en que las guías sugieren una firma genómica previa para decidir.

Uno de los casos de discordancia consistió en una paciente con un tumor bilateral con fenotipos distintos. Obtuvo pCR en el tumor fenotipo HER2 y quedó enfermedad residual en el tumor luminal, la IA recomendó TDM-1 interpretando incorrectamente la lateralidad.

Finalmente, en una paciente con axila clínicamente positiva convertida a ypN0, el ateneo indicó radioterapia post mastectomia, mientras que el sistema la omitió conforme a los resultados del estudio NSABP B-51.

Como hallazgo descriptivo adicional, 6 pacientes cumplían criterios de CALGB 9343 y PRIME II para la omisión opcional de radioterapia tras

cirugía conservadora bajo hormonoterapia y fueron igualmente irradiadas. El motor señala esa opción sin indicarla, por lo que no computó como discordancia.

Discusión

El presente estudio demuestra que un sistema de IA basado en guías clínicas internacionales alcanza un nivel de concordancia sustancial con las decisiones del ateneo multidisciplinario en el manejo del cáncer de mama. El acuerdo observado —con un índice de Jaccard de 0,925 y una concordancia total o parcial en el 99,1% de las presentaciones— supera de manera sustancial al reportado para modelos de lenguaje de propósito general (16%-70%).[2-4] Esta diferencia radica en el diseño del sistema: en lugar de interrogar a un modelo genérico en lenguaje natural, se ejecuta un motor de reglas cuyos nodos reproducen estrictamente los árboles de decisión de las guías y los criterios de inclusión de los ensayos clínicos. Así, el sistema no razona sobre el caso de forma probabilística, sino que lo clasifica.

La comparación no restringida sobre los mismos 117 pacientes presentó un acuerdo de 0,615 (en comparación al índice de Jaccard de 0,925). La caída de 0.31 puntos sin que cambiara una sola conducta cuantifica el sesgo de registro: el acta del ateneo consigna los ejes bajo discusión y omite los que ya están resueltos o son de indicación evidente. Este fenómeno explica buena parte de la distancia entre nuestro resultado y las cifras publicadas, y sugiere que las series que comparan planes completos contra actas clínicas subestiman sistemáticamente el acuerdo real.

Las divergencias residuales tienen un patrón interpretable. La supresión ovárica concentra la mayor diferencia nominal: en 10 pacientes premenopáusicas que recibieron quimioterapia y/o presentaban compromiso ganglionar —población de SOFT/TEXT— el sistema la indicó y el acta no la mencionó. Esto puede deberse a que los profesionales del ateneo indicaron supresión de la función ovárica con criterios más estrictos, sólo a los casos de muy alto riesgo en pacientes muy jóvenes.

Por otro lado, se evidenciaron situaciones donde el ateneo multidisciplinario optó por escalar el tratamiento adyuvante más allá de los criterios estrictos de las guías o ensayos clínicos. Esto incluyó tanto la indicación de CDK4/6 en una paciente que no cumplía los criterios

de elegibilidad de los estudios monarchE o NATALEE, como la prescripción directa de quimioterapia adyuvante en escenarios donde la normativa sugiere realizar previamente una firma genómica. Este segundo hallazgo merece una lectura contextual: la ausencia de acceso oportuno a plataformas genómicas en el sistema público obliga a decidir la quimioterapia con criterios clinicopatológicos. No es un error, sino la traducción de una restricción de recursos a la práctica. Estas discrepancias reflejan la brecha entre la adherencia estricta a un algoritmo y la flexibilidad del juicio clínico ante la complejidad individual del paciente.

Otra discordancia consistió en una paciente con un tumor bilateral con fenotipos distintos que obtuvo pCR en el tumor fenotipo HER2 y quedó enfermedad residual en el tumor luminal, en la que el sistema de IA recomendó TDM-1 interpretando incorrectamente la lateralidad. Esta situación subraya la importancia de la supervisión con criterio clínico sobre los resultados de la IA.

El caso de irradiación regional tras conversión a ypN0 ilustra otra dimensión del problema: la incorporación de evidencia reciente. El estudio NSABP B-51 mostró que la irradiación ganglionar no reduce la recurrencia en pacientes que convierten a ypN0 tras neoadyuvancia.[17] Un motor de reglas actualizado aplica ese resultado de manera inmediata y uniforme; un ateneo lo incorpora con la latencia propia de la deliberación clínica y de la prudencia ante un cambio de práctica establecida. El presente diseño metodológico no permite determinar cuál de las dos conductas es preferible.

Implicancias para el acceso a la atención

El hallazgo con mayor proyección no es el porcentaje de acuerdo sino lo que ese acuerdo habilita. Sólo una fracción de las pacientes con cáncer de mama en nuestro medio accede a que su caso sea discutido en un ateneo multidisciplinario completo. La mayoría es tratada en centros sin oncólogo, radioterapeuta, patólogo y cirujano mamario disponibles de manera simultánea, o donde el ateneo existe pero sesiona con frecuencia insuficiente para el volumen asistencial. La consecuencia es una desigualdad silenciosa: dos pacientes con la misma enfermedad reciben tratamientos distintos según dónde consulten. Datos de la Sociedad Argentina de Mastología muestran que en 2019, de los nuevos casos con diagnóstico de cáncer de mama a nivel país, sólo el 21% fue tratado en unidades de mastología. Esto contrasta con cifras de otros países tales como Alemania, donde el porcentaje de atención en unidades era del 77%.[18]

Un sistema que reproduce la decisión de un ateneo de referencia en el 99,1% de las presentaciones permite, en principio, que la totalidad de las pacientes de un servicio —y no la fracción que alcanza a ser presentada— reciba una recomendación alineada con las guías vigentes y con la práctica de un centro de primer nivel. No sustituye la discusión multidisciplinaria: la complementa allí donde no existe, y la verifica donde existe. En un sistema de salud con recursos desiguales, esa función de nivelación es probablemente el aporte más concreto de la herramienta.

Sobre la ausencia de una comparación de superioridad

Este diseño no permite establecer qué estrategia es mejor, por tres razones. Primera: no existe patrón de referencia de resultado; ninguna de las dos conductas fue contrastada contra supervivencia, recaída o toxicidad observadas. Segunda: la comparación de adherencia a guías sería circular, porque el sistema fue construido a partir de esas mismas guías. Tercera: las desviaciones del ateneo responden mayoritariamente a disponibilidad de fármacos, cobertura, comorbilidad y preferencia de la paciente, variables que el motor de reglas no observa. Lo que la serie sostiene es concordancia, y un rol de verificación y acceso derivado de ella.

Limitaciones

La primera limitación es que se trata de una serie retrospectiva unicéntrica. Por otro lado, la variable de exposición no es "la inteligencia artificial" sino una configuración específica de un modelo determinado, cuyo comportamiento puede variar entre versiones. Ilustrando este hecho, un análisis previo de esta misma serie, realizado con una versión anterior del modelo y una versión previa del motor de reglas, arrojó un índice de 0,909 (en comparación con el resultado actual de 0,925) con diferencias en la identificación de casos discordantes. La descomposición de la conducta del ateneo a partir de texto libre introduce una capa de interpretación, mitigada mediante reglas explícitas y auditables pero no eliminada del todo. Los 33 casos excluidos podrían diferir sistemáticamente de los incluidos: concentran presentaciones localmente avanzadas pendientes de estadificación, es decir, escenarios de mayor complejidad, por lo que el acuerdo informado no debe extrapolarse a ellos. El análisis de sensibilidad sobre formulaciones condicionales no modificó el resultado.

Conclusión

Un sistema de inteligencia artificial configurado sobre guías clínicas internacionales y criterios de elegibilidad de ensayos pivotales reprodujo la conducta de un ateneo multidisciplinario experto en el 99,1% de las presentaciones analizables, con un índice de Jaccard de 0,925. Las divergencias residuales se concentran en ejes identificables —supresión ovárica, escalamiento adyuvante con inhibidores de CDK4/6 e incorporación de evidencia reciente— y son clínicamente interpretables.

Los resultados no sostienen la superioridad de ninguna de las dos estrategias, la cual no puede establecerse con este diseño. Sostienen, en cambio, un rol de verificación de la decisión donde el ateneo existe, y de ampliación del acceso a una recomendación de calidad donde no existe. Las perspectivas incluyen diseños prospectivos y multicéntricos, con adjudicación ciega por panel externo y evaluación de resultados oncológicos a largo plazo.

Referencias

1. Kesson EM, Allardice GM, George WD, Burns HJG, Morrison DS. Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women. BMJ. 2012;344:e2718. doi:10.1136/bmj.e2718 ◄

2. Somashekhar SP, Sepúlveda MJ, Puglielli S, et al. Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. Ann Oncol. 2018;29(2):418-423. doi:10.1093/annonc/mdx781 ◄ ◄

3. Griewing S, Gremke N, Wagner U, Lingenfelder M, Kuhn S, Boekhoff J. Challenging ChatGPT 3.5 in senology—an assessment of concordance with breast cancer tumor board decision making. J Pers Med. 2023;13(10):1502. doi:10.3390/jpm13101502 ◄ ◄

4. Lukac S, Dayan D, Fink V, et al. Evaluating ChatGPT as an adjunct for the multidisciplinary tumor board decision-making in primary breast cancer cases. Arch Gynecol Obstet. 2023;308(6):1831-1844. doi:10.1007/s00404-023-07130-5 ◄ ◄

5. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer. Version 4.2025. Plymouth Meeting (PA): NCCN; 2025. Disponible en: https://www.nccn.org ◄

6. Loibl S, André F, Bachelot T, et al. Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2024;35(2):159-182. doi:10.1016/j.annonc.2023.11.016 ◄

7. Schmid P, Cortes J, Pusztai L, et al; KEYNOTE-522 Investigators. Pembrolizumab for early triple-negative breast cancer. N Engl J Med. 2020;382(9):810-821. doi:10.1056/NEJMoa1910549 ◄

8. Schmid P, Cortes J, Dent R, et al; KEYNOTE-522 Investigators. Event-free survival with pembrolizumab in early triple-negative breast cancer. N Engl J Med. 2022;386(6):556-567. doi:10.1056/NEJMoa2112651 ◄

9. Gunter von Minckwitz G, Huang CS, Mano MS, et al; KATHERINE Investigators. Trastuzumab emtansine for residual invasive HER2-positive breast cancer. N Engl J Med. 2019;380(7):617-628. doi:10.1056/NEJMoa1814017 ◄

10. Geyer CE Jr, Untch M, Huang CS, et al; KATHERINE Study Group. Survival with trastuzumab emtansine in residual HER2-positive breast cancer. N Engl J Med. 2025;392(3):249-257. doi:10.1056/NEJMoa2406070 ◄

11. Johnston SRD, Harbeck N, Hegg R, et al; monarchE Committee Members and Investigators. Abemaciclib combined with endocrine therapy for the adjuvant treatment of HR+, HER2-, node-positive, high-risk, early breast cancer (monarchE). J Clin Oncol. 2020;38(34):3987-3998. doi:10.1200/JCO.20.02514 ◄

12. Slamon D, Lipatov O, Nowecki Z, et al. Ribociclib plus endocrine therapy in early breast cancer. N Engl J Med. 2024;390(12):1080-1091. doi:10.1056/NEJMoa2305488 ◄

13. Masuda N, Lee SJ, Ohtani S, et al. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. N Engl J Med. 2017;376(22):2147-2159. doi:10.1056/NEJMoa1612645 ◄

14. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. N Engl J Med. 2018;379(2):111-121. doi:10.1056/NEJMoa1804710 ◄

15. Kalinsky K, Barlow WE, Gralow JR, et al. 21-gene assay to inform chemotherapy benefit in node-positive breast cancer. N Engl J Med. 2021;385(25):2336-2347. doi:10.1056/NEJMoa2108873 ◄

16. Francis PA, Regan MM, Fleming GF, et al. Adjuvant ovarian suppression in premenopausal breast cancer. N Engl J Med. 2015;372(5):436-446. doi:10.1056/NEJMoa1412379 ◄

17. Mamounas EP, White JR, Bandos H, et al. Omitting regional nodal irradiation after response to neoadjuvant chemotherapy. N Engl J Med. 2025;392(21):2113-2124. doi:10.1056/NEJMoa2414859 ◄ ◄

18. Terrier F. Unidades de Mastología: actualidad y perspectivas en Argentina. Pautas de la SAM. En: XII Foro de Expertos en Mastología; 2020; Buenos Aires, Argentina. Sociedad Argentina de Mastología. ◄