



# ■ REVISTA ARGENTINA DE CIRUGÍA



Publicación de la Asociación Argentina de Cirugía  
*Scientific publication of the Asociación Argentina de Cirugía*

VOLUMEN 118  
Abril-junio 2026  
ISSN 0048-7600

NÚMERO 2

VOLUME 118  
April-June 2026  
ISSN 0048-7600

NUMBER 2



# Revista Argentina de Cirugía

FUNDADA EN 1960 | *FOUNDED IN 1960*



PREMIO APTA - FUNDACIÓN RIZZUTO, AÑO 1981  
Indización | *Indexing*  
SCOPUS;  
Núcleo Básico de Revistas Científicas  
Argentina. Res. Nº 0772/17. Caicyt - Conicet;  
SciELO; DOAJ; Catálogo Latindex Nivel 2;  
Malena  
Base de Datos | *Database*  
LILACS (BIREME-OPS); CONDOR (S.I.I.C.);  
Base de Datos Periódica, UNAM  
Participante de los Requisitos Uniformes,  
Comité Internacional de Editores  
de Revistas Médicas  
Participante del Proyecto EXTRAMED,  
Organización Mundial de la Salud (OMS)  
ISSN 0048 - 7600  
ISSN on-line 2250-639X  
Registro de la Propiedad Intelectual 687.145



Usted es libre de:

Compartir, copiar y redistribuir el material en cualquier medio o formato

Bajo las siguientes condiciones **Reconocimiento**:

Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.

**No Comercial**: No puede utilizar el material para una finalidad comercial.

**Sin Obra Derivada**: Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

*You are free to: Share, copy and redistribute the material in any medium or format under the following conditions: Acknowledgment.- You must properly acknowledge the authorship and provide a link to the license. You can do this in any reasonable way, but not in a way that suggests that you have the licensor's endorsement or receive it for your use.*

*Non-Commercial.- You cannot use the material for a commercial purpose. Without Derivative Work.- If you remix, transform or create from the material, you cannot disseminate the modified material*

Publicación Oficial de la  
Asociación Argentina de Cirugía  
*Official scientific publication of the  
Asociación Argentina de Cirugía*  
M. T. de Alvear 2415 - (1122).  
Ciudad Autónoma de Buenos Aires  
Tel. | Phone: 4822-6489 / 4822-2905-3649  
E-mail: revista@aac.org.ar

## CONSEJO EDITORIAL | *EDITORIAL BOARD*

**Director | *Director***  
Mario L. Iovaldi  
(Hospital Alemán, Argentina)

**Editor jefe | *Editor-in-chief***  
Manuel R. Montesinos  
(Hospital de Clínicas José de San Martín, Argentina)

**Editores ejecutivos | *Executive editors***  
Rodrigo A. Gasque  
(Clínica Universitaria Reina Fabiola, Argentina)

Pablo E. Huespe  
(Hospital Italiano de Buenos Aires, Argentina)

Leonardo Landi  
(Hospital José María Cullen de Santa Fe, Argentina)

Gabriel Navarta  
(Hospital Dr. Guillermo Rawson de San Juan, Argentina)

María E. Peña  
(Sanatorio Güemes, Argentina)

Lucas N. Pina  
(Hospital de Clínicas José de San Martín, Argentina)

Victoria Santa María  
(Hospital Municipal de Oncología Marie Curie, Argentina)

**Coordinadora editorial | *Editorial coordinator***  
Natalia Ingani  
(Asociación Argentina de Cirugía, Argentina)

**Correctora de estilo | *Style corrector***  
María Isabel Siracusa  
(Asociación Argentina de Cirugía, Argentina)

**Traductora | *Translator***  
Rita Tepper  
(Asociación Argentina de Cirugía, Argentina)

## EDITORES INTERNACIONALES | *INTERNATIONAL EDITORS*

Mariana Berho  
(Cleveland Clinic, EE.UU.)

Markus W. Büchler  
(Botton-Champalimaud Pancreatic Cancer Centre, Portugal)

Guillermo M. Carriquiry  
(Universidad de La República, Uruguay)

Antonio Caycedo-Marulanda  
(Queen's University, Canadá)

Claudio Cernea  
(Hospital Das Clínicas, Brasil)

Daniel I. Chu  
(University of Alabama at Birmingham, EE.UU.)

Raúl Cutait  
(Hospital Sirio-Libanés, Brasil)

José de Vinatea de Cárdenas  
(Facultad de Medicina de la Universidad Nacional Mayor de San Marcos, Perú)

Gonzalo Estapé Carriquiry  
(Centro Latinoamericano de Economía Humana, Uruguay)

Steve Eubanks  
(Florida Hospital, EE.UU.)

Owen Korn Bruzzone  
(Hospital Clínico Universidad de Chile, Chile)

Luiz P. Kowalsky  
(Hospital A. C. Camargo, Brasil)

Claudio Navarrete García  
(Clínica Santa María, Chile)

Gregg Nelson  
(University of Calgary, Canadá)

Marco Patti  
(Universidad de Virginia, EE.UU.)

Carlos A. Pellegrini  
(University of Washington, EE.UU.)

Elina Quiroga  
(University of Washington, EE.UU.)

Eduardo M. Targarona Soler  
(Hospital de la Santa Creu i Sant Pau, España)

Paula Ugalde  
(Institut Universitaire de Cardiologie et de Pneumologie de Québec, Canadá)

Steven D. Wexner  
(Cleveland Clinic, EE.UU.)

Nathan Zundel  
(Florida International University, EE.UU.)

## EDITORES HONORARIOS | *HONORARY EDITORS*

H. Pablo Curutchet  
Vicente Gutiérrez Maxwell

Enrique A. Sívori

## AUTORIDADES ASOCIACIÓN ARGENTINA DE CIRUGÍA | *BOARD OF DIRECTORS, ASOCIACIÓN ARGENTINA DE CIRUGÍA*

**Presidente | *President***  
Oscar M. Mazza  
**Vicepresidente 1° | *1st. Vicepresident***  
Martín A. Duhalde  
**Vicepresidente 2° | *2nd. Vicepresident***  
Federico A. Brahín  
**Secretaría general | *General Secretary***  
Victoria Ardiles  
**Secretaría de actas | *Recording Secretary***  
Julieta Paleari  
**Tesorero | *Treasurer***  
Pablo Cingolani  
**Protesorera | *Pro-treasurer***  
Verónica Garay

**Vocales titulares | *Ordinary Members***  
Mauricio J. Linzey  
Fernanda V. Bersano  
Carlos F. Salas  
Lucas E. Granero  
**Vocales suplentes | *Deputy Members***  
Jorge E. Ruiz  
Marcelo G. Ferrario  
M. Alejandra Geraghty  
María S. Bruzzi  
Patricio J. Gómez Constenla  
**Director general | *General Director***  
José L. Tortosa

Imagen de tapa | *Cover illustration:*

**Tríptico del carro de heno (1512-1515)**

***The Haywain Triptych (1512-1515)***

Autor | *Painter:* El Bosco

Óleo sobre tabla | *oil on panel*

147,1 x 224,3 cm

Museo del Prado, Madrid, España | *Spain*

# Contenidos

Vol 118 N°2 (Abril - junio 2026)

- Editorial** **La importancia de saber preguntar: ingeniería de prompts y modelos de lenguaje en la consulta médica**  
*Enrique Díaz Cantón*
- Exactitud, estadificación y decisión: la inteligencia artificial frente al cáncer esofagogástrico**  
*Enrique Díaz Cantón*
- Conceptos y metodología** **El tamaño muestral en la investigación quirúrgica**  
*M. Elena Peña*
- Artículo original** **Utilidad del análisis microbiológico de bilis en colecistitis grado II de la clasificación de Tokio**  
*Guadalupe Frías Mena, María C. du Plessis, Héctor F. Posada Lobo, José R. Varela, Nicolás F. Baglietto, Carlos Canullán, Enrique J. Petracchi, Hugo I. Zandalazini*
- Resultados de corto plazo del tratamiento quirúrgico de tumores retroperitoneales en un periodo de 10 años**  
*Gustavo A. Nari, José Layún, Sol Romero, Daniela Mariot, Ángel Joseph, María E. De elias*
- Sarcomas de partes blandas de tronco y extremidades. Resultados quirúrgicos y oncológicos en pacientes con enfermedad localizada**  
*Sergio D. Quildrian, Walter S. Nardi, Victoria Scasso Rebzda, Nicolás M. Rosasco, Pablo Dezano, Cintia Novas, Luciana Bella Quero, Carlos Silva, Daniel E. Pirchi*
- Chatbots basados en inteligencia artificial: la importancia de saber preguntar**  
*M. Elena Peña, Alejandro Gigena, Fernando Iglesia*
- Exactitud de la inteligencia artificial en la estadificación y toma de decisiones en cáncer esófago gástrico**  
*Adelina E. Coturel, Paula Pereyra, Rodrigo García, Agustín Diomedj, Juan J. M. Cabas Audicio, Roberto Klappenbach*
- Carta científica** **Abdomen agudo suboclusivo por intususcepción como complicación tardía en bypass gástrico en Y de Roux**  
*María C. Gribaudo, Franco W. Lara, Luciano Rollan, Gustavo G. Del Cantare*
- Peritonitis esclerosante encapsulante: informe de un caso de síndrome de cocoon, una causa inusual de obstrucción intestinal**  
*Nemecio Lizana, Xavier Sigüenza, Rocío Bermúdez, Roberto Briceño*
- Trasplante hepático en colangiocarcinoma perihiliar irresecable: informe de un caso en el marco de un protocolo de selección oncológica estricta**  
*Melisa Amondarain, Pablo E. Huespe, Federico Díaz, Alicia Verzura, Sung Ho Hyon, Martín de Santibañes*
- Fijación externa mandibular en trauma maxilofacial por arma de fuego**  
*Juan P. Berríos-Bugueño, Juan P. Aravena, Rodrigo A. Allende, María M. Padilla, Polett C. Yáñez, Diego A. Lazo*

# Contents

Vol 118 N°2 (April - June 2026)

<b>Editorial</b>	<b>The importance of asking the right questions: prompt engineering and language models in clinical practice</b> <i>Enrique Díaz Cantón</i>
	<b>Accuracy, staging, and decision-making: artificial intelligence in gastroesophageal cancer</b> <i>Enrique Díaz Cantón</i>
<b>Concepts and methodology</b>	<b>Sample size in surgical research</b> <i>M. Elena Peña</i>
<b>Original article</b>	<b>Usefulness of microbiological bile analysis in grade II cholecystitis according to the Tokyo classification</b> <i>Guadalupe Frijas Mena, María C. du Plessis, Héctor F. Posada Lobo, José R. Varela, Nicolás F. Baglietto, Carlos Canullán, Enrique J. Petracchi, Hugo I. Zandalazini</i>
	<b>Short-term outcomes of surgical treatment for retroperitoneal tumors over a 10-year period</b> <i>Gustavo A. Nari, José Layún, Sol Romero, Daniela Mariot, Ángel Joseph, María E. De elias</i>
	<b>Soft tissue sarcomas of the extremities and trunk. Surgical and oncological outcomes in patients with local disease</b> <i>Sergio D. Quildrian, Walter S. Nardi, Victoria Scasso Rebzda, Nicolás M. Rosasco, Pablo Dezanzo, Cintia Novas, Luciana Bella Quero, Carlos Silva, Daniel E. Pirchi</i>
	<b>Artificial intelligence-based chatbots: the importance of asking the right questions</b> <i>M. Elena Peña, Alejandro Gigena, Fernando Iglesia</i>
	<b>Accuracy of artificial intelligence for staging and decision-making in esophageal and gastric cancer</b> <i>Adelina E. Coturel, Paula Pereyra, Rodrigo García, Agustín Diomedi, Juan J. M. Cabas Audicio, Roberto Klappenbach</i>
<b>Scientific letter</b>	<b>Partial bowel obstruction caused by intussusception as a late complication following Roux-en-Y gastric bypass surgery</b> <i>María C. Gribaudo, Franco W. Lara, Luciano Rollan, Gustavo G. Del Cantare</i>
	<b>Encapsulating peritoneal sclerosis: a case report of cocoon syndrome, an unusual cause of bowel obstruction</b> <i>Nemecio Lizana, Xavier Sigüenza, Rocío Bermúdez, Roberto Briceño</i>
	<b>Liver Transplantation for Unresectable Perihilar Cholangiocarcinoma: A Case Report Following a Strict Oncologic Selection Protocol</b> <i>Melisa Amondarain, Pablo E. Huespe, Federico Díaz, Alicia Verzura, Sung Ho Hyon, Martín de Santibañes</i>
	<b>External mandibular fixation in maxillofacial gunshot injury</b> <i>Juan P. Berríos-Bugueño, Juan P. Aravena, Rodrigo A. Allende, María M. Padilla, Polett C. Yáñez, Diego A. Lazo</i>

# La importancia de saber preguntar: ingeniería de *prompts* y modelos de lenguaje en la consulta médica

## *The importance of asking the right questions: prompt engineering and language models in clinical practice*

Enrique Díaz Cantón\*

La incorporación de los grandes modelos de lenguaje (LLM) a la práctica clínica ha sido vertiginosa: en pocos meses, pacientes y profesionales comenzaron a consultarlos a diario para informarse, redactar y razonar sobre problemas médicos. El trabajo *Chatbots basados en inteligencia artificial: la importancia de saber preguntar*<sup>1</sup> pone el dedo en una verdad frecuentemente soslayada: la calidad de la respuesta es función de la calidad de la pregunta. Lejos de ser un detalle técnico, esta dependencia tiene consecuencias clínicas directas.

Que el resultado varíe según cómo se formula la consulta no es una intuición, sino un hallazgo reproducible. El estudio de Chen y cols. en *JAMA Oncology* mostró que, ante un mismo escenario clínico planteado con cuatro variantes de redacción, ChatGPT<sup>®</sup> modificaba sus recomendaciones de tratamiento. El 98% de las respuestas contenía al menos una opción concordante con las guías NCCN, pero el 34,3% incluía, además, alguna recomendación no concordante y el 12,5 % constituía verdaderas “alucinaciones” –terapias inexistentes o curativas para enfermedad incurable–, mezcladas entre consejos por lo demás sensatos y, por ello, difíciles de detectar<sup>2</sup>. La formulación de la pregunta, en suma, se comporta como una variable clínica más.

### Ingeniería de *prompts*: una nueva competencia clínica

La ingeniería de *prompts* es el conjunto de técnicas que orientan a un modelo hacia respuestas precisas y verificables. Las más estudiadas incluyen el *zero-shot* (instrucción directa), el *few-shot* (aportar ejemplos resueltos), la asignación de rol, el anclaje con contexto o guías –incluida la generación aumentada por recuperación–y, especialmente, el razonamiento encadenado o *chain-of-thought*, que solicita al modelo desplegar pasos intermedios antes de concluir<sup>3</sup>. Su variante *zero-shot* –el conocido “pensemos paso a paso”– mejora de manera notable el desempeño en tareas de razonamiento complejo<sup>4</sup>. No se trata de retórica: prescribir cómo debe estructurarse la respuesta cambia su

exactitud, y por eso revistas de referencia ya publican recomendaciones formales sobre cómo redactar consignas clínicas<sup>5</sup>.

### ¿Qué modelos conviene usar en medicina?

La progresión de los modelos en los bancos de prueba médicos ha sido notable. Med-PaLM 2<sup>®</sup> alcanzó alrededor del 86,5% en MedQA, un conjunto de preguntas con formato USMLE<sup>6</sup>, y Med-Gemini<sup>®</sup> elevó el estado del arte al 91,1% en 2024<sup>7</sup>. Conviene, sin embargo, una advertencia: el desempeño en exámenes de opción múltiple no equivale a competencia clínica. Las primeras versiones apenas rozaban el umbral de aprobación del USMLE<sup>8</sup>, y, como vimos, el mismo ecosistema que sobresale respondiendo preguntas cerradas produce un tercio de recomendaciones no concordantes cuando se lo interroga sobre tratamiento abierto<sup>2</sup>. La tabla 1 sintetiza, con la prudencia debida, esta evidencia heterogénea.

■ TABLA 1

Desempeño de modelos de lenguaje (de generaciones anteriores) en bancos de prueba médicos seleccionados

Modelo (desarrollador)	Banco de prueba	Resultado	Fuente
GPT-3.5 (OpenAI) <sup>8</sup>	USMLE (Pasos 1-3)	Desempeño cercano al umbral de aprobación	[8]
GPT-3.5 (OpenAI) <sup>2</sup>	Concordancia con guías NCCN en oncología (mama, próstata, pulmón)	98% con ≥ 1 opción concordante; 34,3% con ≥ 1 no concordante; 12,5 % “alucinaciones”	[2]
Med-PaLM 2 (Google) <sup>6</sup>	MedQA (estilo USMLE)	Hasta 86,5 % de exactitud	[6]
Med-Gemini (Google) <sup>7</sup>	MedQA (estilo USMLE)	91,1 % (estado del arte, 2024)	[7]

Nota: los estudios recogidos corresponden a modelos de generaciones anteriores (2023–2024) y emplean metodologías y poblaciones distintas; no constituyen una comparación directa entre modelos ni reflejan el estado del arte actual (véase el apartado siguiente). MedQA: banco de preguntas con formato USMLE; NCCN: National Comprehensive Cancer Network; USMLE: United States Medical Licensing Examination.

\*Oncólogo clínico, Instituto Universitario CEMIC, Buenos Aires, Argentina.

Codirector, Programa de Posgrado en Inteligencia Artificial y Medicina, Academia Nacional de Medicina, Buenos Aires, Argentina.

Correspondencia: ediazcanton@iuc.edu.ar

## Un panorama en rápida transformación

Conviene subrayar que los estudios reunidos en la tabla 1 emplearon modelos de generaciones anteriores y que el campo avanza a una velocidad inusual. En 2026 conviven modelos de frontera de propósito general notablemente más capaces —como GPT-5.5<sup>o</sup>, Claude Opus 4.8<sup>o</sup> o la familia Gemini 3<sup>o</sup>—y, sobre todo, herramientas especializadas en medicina que anclan cada respuesta en la literatura primaria mediante recuperación aumentada (RAG) y la citan de forma auditable. Plataformas como OpenEvidence<sup>o</sup> —utilizada por más del 40% de los médicos de los Estados Unidos y con acuerdos de contenido con el *New England Journal of Medicine* y *JAMA*—o Vera Health<sup>o</sup> —respuestas graduadas según la calidad de la evidencia, en alianza con el American College of Emergency Physicians—ejemplifican esta tendencia<sup>9,10</sup>. La evidencia revisada por pares respalda el enfoque: incorporar conocimiento clínico formal y recuperación semántica elevó el desempeño en el Paso 3 del USMLE a alrededor del 95%, por encima de un LLM nativo (90,5%)<sup>11</sup>. Aun así, ni la mayor potencia ni la especialización eximen de validar cada

salida en escenarios clínicos complejos, donde el juicio experto sigue siendo insustituible.

La lección de este trabajo trasciende la anécdota tecnológica. El profesional que incorpore estas herramientas necesitará dos alfabetizaciones complementarias: saber preguntar, esto es, dominar la ingeniería de prompts como una destreza clínica; y saber dudar, sometiendo cada respuesta al juicio experto y a las guías vigentes. Usados con esa disciplina —y con transparencia hacia el paciente—, los *chatbots* pueden ser un aliado valioso. Usados sin ella, multiplicarán los errores con una elocuencia inédita. La pregunta correcta, una vez más, sigue siendo la mitad de la respuesta.

**Declaración sobre el uso de inteligencia artificial:** El autor declara que, en la preparación de este editorial, utilizó herramientas de inteligencia artificial generativa basadas en modelos de lenguaje de gran escala como apoyo en la búsqueda y el cotejo de la bibliografía, la organización del texto y la edición de estilo. Todas las afirmaciones, las referencias y los datos cuantitativos fueron verificados por el autor contra las fuentes primarias citadas. El contenido conceptual, las interpretaciones y las conclusiones son de su exclusiva responsabilidad; la inteligencia artificial no figura como autora ni asume responsabilidad alguna sobre el contenido.

## ■ ENGLISH VERSION

The integration of large language models (LLMs) into clinical practice has been rapid and significant: within a few months, patients and healthcare professionals began to consult them daily to gather information, draft documents, and discuss medical issues. The paper ‘Artificial intelligence-based chatbots: the importance of asking the right questions’<sup>1</sup> highlights a frequently neglected reality: the quality of the answer is contingent upon the quality of the inquiry. Far from being a mere technical detail, this dependency has direct clinical implications.

The fact that the result varies depending on how the query is phrased is not just intuition, but rather a reproducible finding. The study by Chen et al. in *JAMA Oncology* demonstrated that, when presented with the same clinical scenario described in four different ways, ChatGPT modified the recommended treatment approach. In 98% of responses, at least one NCCN-concordant treatment was included<sup>2</sup>. However, 34.3% also included at least one nonconcordant recommendation, and 12.5% of outputs were hallucinations (non-existent therapies or curative treatments for an incurable disease), which were embedded in otherwise sensible recommendations and thus difficult to detect. In summary, the way in which a question is formulated is an additional clinical variable.

### Prompt engineering: a novel clinical competence

Prompt engineering is the set of techniques used to guide a model toward accurate and verifiable responses. The most widely studied techniques include zero-shot (direct instruction), few-shot (providing solved examples), role assignment, context or guideline anchoring—including retrieval-augmented generation—and, notably, chain-of-thought reasoning, which prompts the model to display intermediate steps before arriving at a conclusion<sup>3</sup>. Its zero-shot variant—the well-known “let’s think step by step”—significantly improves performance in complex reasoning tasks<sup>4</sup>. This is not a matter of rhetoric: prescribing how a response should be structured alters its accuracy, which is why leading journals are already publishing formal recommendations on drafting clinical prompts<sup>5</sup>.

### Which models are best suited for use in medicine?

The progress made with models on medical benchmarks has been remarkable. Med-PaLM 2 achieved a score of approximately 86.5% on MedQA, an evaluation dataset of medical knowledge questions from the United States Medical Licensing Examination (USMLE)<sup>6</sup>. In 2024, Med-Gemini achieved a state-of-the-art improvement of 91.1%<sup>7</sup>. Please be aware that

performance in multiple-choice exams is not necessarily an indicator of clinical competence. The early versions barely performed at or near the USMLE passing threshold<sup>8</sup>. As we have seen, the same ecosystem that excels in closed-ended questions produces one-third of non-concordant recommendations when asked about treatments using open-ended questions<sup>1</sup>. Table 1 summarizes this heterogeneous evidence with due caution.

■ TABLE 1

Performance of (previous-generation) language models on selected medical benchmarks.

Model (developer)	Benchmark	Result	Source
GPT-3.5 (OpenAI)	USMLE (Steps 1–3)	Performance at or near the passing threshold	[8]
GPT-3.5 (OpenAI)	Concordance with NCCN guidelines in oncology (breast cancer, prostate cancer, and lung cancer)	98% with ≥ 1 concordant option; 34.3% with ≥ 1 nonconcordant option; 12.5% “hallucinations”	[2]
Med-PaLM 2 (Google)	MedQA (USMLE)	Accuracy up to 86.5%	[6]
Med-Gemini (Google)	MedQA (USMLE)	91.1% (state-of-the-art, 2024)	[7]

Note: The cited studies pertain to earlier-generation models (2023–2024) and employ different methodologies and populations. They do not constitute a direct comparison between models, nor do they reflect the current state of the art (see the following section). MedQA: benchmark dataset (USMLE subset); NCCN: National Comprehensive Cancer Network; USMLE: United States Medical Licensing Examination.

## A panorama under rapid transformation

It is worth noting that the studies compiled in Table 1 utilized previous-generation models, and that the field is advancing at an unprecedented pace. As of 2026, significantly more capable general-purpose frontier models—like GPT-5.5, Claude Opus 4.8, or the

Gemini 3 family—coexist with medical-specific tools that anchor every response to primary literature via retrieval-augmented generation (RAG) and cite it in an auditable manner. Examples of this trend include platforms such as OpenEvidence, which is used by over 40% of physicians in the United States and has content partnerships with the New England Journal of Medicine and JAMA, and Vera Health, which provides graded responses based on the quality of evidence in partnership with the American College of Emergency Physicians<sup>9,10</sup>. Peer-reviewed evidence supports this approach: incorporating formal clinical knowledge and semantic retrieval increased Step 3 USMLE performance to approximately 95%, outperforming a native LLM (90.5%)<sup>11</sup>. Nevertheless, the greater power and specialization of these models do not exempt their outputs from validation in complex clinical scenarios, where expert judgment remains irreplaceable.

The lesson of this paper goes beyond the technological anecdote. Professionals integrating these tools will require two complementary literacies: knowing how to ask—that is, mastering prompt engineering as a clinical skill—and knowing how to doubt, subjecting every response to expert judgment and current guidelines. When used with discipline and transparency toward the patient, chatbots can be valuable allies. Otherwise, they will multiply errors with unprecedented eloquence. Once again, the right question is half the answer.

**Artificial intelligence disclosure:** The author declares that, during the preparation of this editorial, generative artificial intelligence tools based on large language models were utilized to assist with literature search and cross-referencing, text organization, and copyediting. All statements, references, and quantitative data were verified by the author against the primary sources cited. The conceptual content, interpretations, and conclusions are the sole responsibility of the author; the artificial intelligence is not credited as an author, nor does it assume any responsibility for the content.

## Referencias bibliográficas /References

- Peña ME, Gigena A, Iglesia F. Chatbots basados en inteligencia artificial: la importancia de saber preguntar. Rev Argent Cir. 2026;118(2):e-1961. DOI: <http://dx.doi.org/10.25132/raac.v118.n2.1961>
- Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, Bitterman DS. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. JAMA Oncol. 2023;9(10):1459-62. doi:10.1001/jamaoncol.2023.2954.
- Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst. 2022;35:24824-37.
- Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. Adv Neural Inf Process Syst. 2022;35:22199-213.
- Schulte B. Considerations for Prompting Large Language Models. JAMA Oncol. 2024;10(4):538.
- Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-80. doi:10.1038/s41586-023-06291-2.
- Saab K, Tu T, Weng WH, et al. Capabilities of Gemini models in medicine. arXiv:2404.18416 [preprint]. 2024.
- Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198.
- Elkin PL, Mehta G, LeHouillier F, et al. Semantic Clinical Artificial Intelligence vs Native Large Language Model Performance on the USMLE. JAMA Netw Open. 2025;8(4):e256359. doi:10.1001/jamanetworkopen.2025.6359.
- OpenEvidence. OpenEvidence creates the first AI in history to score 100% on the United States Medical Licensing Examination (USMLE) [comunicado]. 15 ago 2025. Disponible en: [openevidence.com](https://openevidence.com).
- Vera Health (Veracity-Health Inc.). Plataforma de soporte a la decisión clínica basada en evidencia. Disponible en: [verahealth.ai](https://verahealth.ai) (consultado en 2026).

# Exactitud, estadificación y decisión: la inteligencia artificial frente al cáncer esofagogástrico

*Accuracy, staging, and decision-making: artificial intelligence in gastroesophageal cancer*

Enrique Díaz Cantón\*

En el cáncer esofagogástrico, la estadificación no es un ejercicio descriptivo: gobierna decisiones de alto impacto y, con frecuencia, divergentes. El trabajo *Exactitud de la inteligencia artificial en la estadificación y toma de decisiones en cáncer esófago-gástrico*<sup>1</sup> formula la pregunta pertinente –cuán exacta es la inteligencia artificial (IA)– y, sobre todo, obliga a precisar contra qué patrón se mide esa exactitud.

La evidencia más directa es alentadora pero condicional. Yao y cols. compararon tres grandes modelos de lenguaje (LLM) de despliegue local con clínicos en la estadificación TNM preoperatoria del cáncer de esófago a partir de informes radiológicos en texto libre, tomando la patología posquirúrgica como patrón de referencia. Al ensayar tres estrategias de consigna –*zero-shot*, *chain-of-thought* y un método de razonamiento interpretable–, observaron que la consigna estructurada no solo elevó la exactitud, sino que aportó una traza de razonamiento auditable<sup>2</sup>. La IA puede, entonces, alcanzar una exactitud apreciable; pero esa exactitud depende de cómo se la interroga y debe contrastarse contra el oro patológico, no contra la verosimilitud de un texto bien redactado.

## Por qué el cáncer esofagogástrico exige cautela adicional

Pocos escenarios castigan tanto un error de estadificación. La conducta depende de distinciones sutiles –el epicentro tumoral respecto de la unión esofagogástrica, la clasificación de Siewert, la divergencia histológica entre adenocarcinoma y carcinoma escamoso– que orientan esquemas perioperatorios o de quimiorradioterapia diferentes. A ello se suman biomarcadores que condicionan la terapia sistémica, como HER2, PD-L1, la inestabilidad de microsátélites y CLDN18.2<sup>3</sup>. Un modelo que confunda un tumor de cardias con uno de la unión, o que omita un biomarcador, propaga el error a todo el plan terapéutico. La “exactitud” que reclama este trabajo debe medirse, por lo tanto, dentro del marco de las guías vigentes y contra el diagnóstico patológico.

## De la estadificación a la decisión

Cuando se pasa de estadificar a recomendar, la concordancia se vuelve más variable. En la predicción de conductas de junta tumoral se han descrito congruencias de línea de tratamiento de hasta el 86% con algunos modelos<sup>4</sup>, pero el mismo ecosistema produce un tercio de recomendaciones no concordantes cuando se lo interroga sobre tratamiento oncológico abierto, con un 12,5% de alucinaciones<sup>5</sup>. La tabla 1 resume esta evidencia heterogénea, que conviene leer con prudencia.

La conclusión no es desestimar la herramienta, sino encuadrarla. La inteligencia artificial puede afinar la estadificación y aportar un razonamiento explícito y revisable, sobre todo cuando se la guía con consignas estructuradas. Pero la responsabilidad de la decisión –ética, legal y humana– permanece indelegablemente en el cirujano y el oncólogo, que deben validar cada salida contra el paciente concreto, su patología y las

■ TABLA 1

Exactitud y concordancia de la IA en estadificación y decisión oncológica

Dominio evaluado	Modelo(s)	Resultado principal	Fuente
Estadificación TNM de cáncer de esófago desde informes radiológicos (referencia: patología)	INF-72B, Qwen2.5-72B, LLaMA3.1-70B	Exactitud elevada; el razonamiento estructurado (CoT / razonamiento interpretable) superó al zero-shot	[2]
Predicción de recomendaciones de junta tumoral (cabeza y cuello)	Varios LLM	Congruencia de línea de tratamiento hasta 86 %; recomendaciones justificables hasta 98 %	[4]
Recomendaciones oncológicas abiertas vs. guías NCCN (mama, próstata, pulmón)	GPT-3.5*	34,3 % con ≥ 1 recomendación no concordante; 12,5 % “alucinaciones”	[5]

Nota: los valores provienen de estudios con metodologías, idiomas y poblaciones distintas y no constituyen una comparación directa entre modelos. CoT: chain-of-thought; NCCN: National Comprehensive Cancer Network; TNM: tumor-ganglio-metástasis.

\*Oncólogo clínico, Instituto Universitario CEMIC, Buenos Aires, Argentina.

Codirector, Programa de Posgrado en Inteligencia Artificial y Medicina, Academia Nacional de Medicina, Buenos Aires, Argentina.

Correspondencia: ediazcantón@iuc.edu.ar

guías. Medida así, con rigor y contra el patrón correcto, la exactitud de la IA deja de ser una promesa para convertirse en una herramienta; medida contra su propia elocuencia, sigue siendo un espejismo.

**Declaración sobre el uso de inteligencia artificial:** El autor declara que en la preparación de este editorial utilizó

*herramientas de inteligencia artificial generativa basadas en modelos de lenguaje de gran escala como apoyo en la búsqueda y el cotejo de la bibliografía, la organización del texto y la edición de estilo. Todas las afirmaciones, las referencias y los datos cuantitativos fueron verificados por el autor contra las fuentes primarias citadas. El contenido conceptual, las interpretaciones y las conclusiones son de su exclusiva responsabilidad; la inteligencia artificial no figura como autora ni asume responsabilidad alguna sobre el contenido.*

## ENGLISH VERSION

In gastroesophageal cancer, staging is not merely a descriptive exercise: it guides high-impact, often conflicting decisions. The paper “Accuracy of artificial intelligence for staging and decision-making in esophageal and gastric cancer”<sup>1</sup> raises the pertinent question of how accurate artificial intelligence (AI) is. More importantly, it requires specifying the standard against which this accuracy is measured.

The most direct evidence is encouraging, though conditional. Yao et al. compared the performance of three locally deployed LLMs with that of clinicians in preoperative esophageal cancer staging using free-text radiology reports. The gold standard was derived from postoperative pathological staging. When the authors tested three prompting strategies—zero-shot, chain-of-thought, and an interpretable reasoning method—they observed that the structured prompt not only improved accuracy but also provided an auditable reasoning trace<sup>2</sup>. Therefore, the accuracy of AI can be contingent on the method of querying, and it should be evaluated against the pathological gold standard rather than the plausibility of a well-written text.

### Why does gastroesophageal cancer demand particular caution?

Few scenarios impose such a heavy penalty for a staging error. The management of the disease depends on subtle distinctions—tumor location relative to the gastroesophageal junction, Siewert classification, and histological differences between adenocarcinoma and squamous cell carcinoma—that guide different perioperative or chemoradiotherapy regimens. In addition, certain biomarkers, such as HER2, PD-L1, microsatellite instability, and CLDN18.2, condition systemic therapy<sup>3</sup>. A model confounding a cardia tumor with that of the gastroesophageal junction—or one that ignores a biomarker—spreads the error throughout the entire treatment plan. The “accuracy” claimed by this study must therefore be assessed within the framework of current guidelines and in relation to the pathological diagnosis.

### From staging to decision-making

When the focus shifts from staging to recommending, concordance becomes more variable. For predicting tumor board procedural recommendations, treatment line congruence of up to 86% has been reported with some models<sup>4</sup>, but the same ecosystem generates one-third of non-concordant recommendations when queried about open oncology treatment, with 12.5% of responses being hallucinations<sup>5</sup>. Table 1 summarizes this heterogeneous evidence, which should be read with caution.

The objective is not to discredit the tool, but rather to contextualize its use. Artificial intelligence can improve staging and provide explicit, verifiable reasoning, especially when guided by structured prompts. However, the ethical, legal, and human responsibility for the decision cannot be delegated; it remains with the surgeon and the oncologist, who must verify each output against the individual patient, their pathology, and current guidelines. When evaluated with such rigor

TABLE 1

Accuracy and concordance of AI in cancer staging and decision-making.

Domain	Model(s)	Primary outcome	Source
TNM staging of esophageal cancer based on radiology reports (gold standard: pathology)	INF-72B, Qwen2.5-72B, LLaMA3.1-70B	High accuracy; structured reasoning (CoT / interpretable reasoning) outperformed zero-shot	[2]
Prediction of tumor board procedural recommendations (head and neck)	Various LLMs	Treatment line congruence of up to 86%; justifiable recommendations up to 98%.	[4]
Open oncology recommendations vs. NCCN guidelines (breast cancer, prostate cancer, lung cancer)	GPT-3.5	34.3% with $\geq 1$ nonconcordant option; 12.5% “hallucinations”	[5]

Note: These values are derived from studies using different methodologies, languages, and populations and do not constitute a direct comparison between models. CoT: chain-of-thought; NCCN: National Comprehensive Cancer Network; TNM: tumor-node-metastasis

and against the proper standard, the accuracy of AI shifts from being a promise to becoming a tool; measured against its own eloquence, it remains a mirage.

**Artificial intelligence disclosure:** The author declares that, during the preparation of this editorial, generative artificial

*intelligence tools based on large language models were utilized to assist with literature search and cross-referencing, text organization, and copyediting. All statements, references, and quantitative data were verified by the author against the primary sources cited. The conceptual content, interpretations, and conclusions are the sole responsibility of the author; the artificial intelligence is not credited as an author, nor does it assume any responsibility for the content.*

#### Referencias bibliográficas /References

1. Coturel AE, Pereyra P, García R, Diomedí A, Cabas Audicio JJM, Klappenbach R. Exactitud de la inteligencia artificial en la estadificación y toma de decisiones en cáncer esófago gástrico. Rev Argent Cir. 2026;118(2):e-1963. DOI: <http://dx.doi.org/10.25132/raac.v118.n2.1963>
2. Yao Y, Cen X, Gan L, Jiang J, Wang M, Xu Y, Yuan J. Automated Esophageal Cancer Staging From Free-Text Radiology Reports: Large Language Model Evaluation Study. JMIR Med Inform. 2025;13:e75556. doi:10.2196/75556.
3. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Gastric Cancer y Esophageal and Esophagogastric Junction Cancers. NCCN; 2025-2026. Disponible en NCCN.org.
4. Aubreville M, Ganz J, Ammeling J, et al. Prediction of tumor board procedural recommendations using large language models. Eur Arch Otorhinolaryngol. 2025;282(3):1619-29. doi:10.1007/s00405-024-08947-9.
5. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, Bitterman DS. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. JAMA Oncol. 2023;9(10):1459-62. doi:10.1001/jamaoncol.2023.2954.

# El tamaño muestral en la investigación quirúrgica

## Sample size in surgical research

M. Elena Peña\*

El cálculo del tamaño muestral constituye un componente esencial del diseño metodológico en la investigación médica. Su informe en los trabajos científicos mejora la interpretabilidad de los resultados y refuerza tanto la validez interna como la legitimidad ética del estudio. En cirugía, donde muchos eventos clínicos son relativamente infrecuentes y la variabilidad técnica entre operadores puede ser considerable, esta cuestión adquiere una relevancia particular.

A pesar de su importancia, la ausencia del cálculo del tamaño muestral continúa siendo una debilidad frecuente en publicaciones quirúrgicas<sup>1,2</sup>. Este artículo propone revisar su definición, relevancia, indicaciones y los principios generales de cálculo.

### ¿Qué es el cálculo del tamaño muestral?

El cálculo del tamaño muestral consiste en la estimación del número de participantes o unidades de observación que deben incluirse en un estudio para responder de manera adecuada a la pregunta de investigación planteada (véase más adelante)<sup>3,4</sup>.

En cirugía, las unidades de observación suelen ser pacientes o procedimientos quirúrgicos, aunque también pueden ser centros, cirujanos o episodios clínicos específicos. Por ejemplo, en un estudio que analiza todas las apendicectomías realizadas durante el año 2025 en un hospital comunitario, se incluyen 120 procedimientos. En este caso, el tamaño de la muestra corresponde a 120 apendicectomías (n=120).

### ¿Cuál es la importancia de calcular el tamaño muestral?

Cuando no se realiza un cálculo previo del tamaño muestral pueden incluirse más o menos unidades de observación de las necesarias. Si se incluyen más participantes de los requeridos, se incrementan innecesariamente los costos, el uso de recursos y el tiempo de investigación. Por el contrario, si el número de casos es insuficiente, el estudio puede no tener suficiente

capacidad para detectar una diferencia real entre los grupos analizados. Esta situación se conoce como error de tipo II o falta de potencia estadística<sup>3,5</sup>. Este error es equivalente a no ver una diferencia que verdaderamente existe en la realidad. Cuando esto sucede, se acepta la hipótesis nula erróneamente, cuando en realidad debería rechazarse<sup>3</sup>.

Además, existe una dimensión ética. En los estudios de investigación, los pacientes aportan información sensible con el objetivo de contribuir al conocimiento científico. Diseñar estudios con una probabilidad muy baja de responder la pregunta de investigación puede implicar un uso inadecuado de esos datos. Por ejemplo, intentar analizar retrospectivamente una condición extremadamente infrecuente, como la apendicitis del muñón, contando únicamente con tres casos disponibles, difícilmente permitirá obtener conclusiones robustas.

### ¿Qué tipos de preguntas de investigación existen?

De forma general, las preguntas de investigación pueden clasificarse en descriptivas y analíticas. A cada tipo de pregunta corresponde un enfoque metodológico específico y diferentes estrategias de cálculo del tamaño muestral<sup>3,5</sup>.

#### **Preguntas descriptivas**

Las preguntas descriptivas buscan caracterizar o describir cómo se manifiesta un fenómeno o cómo se distribuye una variable en una población determinada, sin intentar establecer relaciones causales. Para este tipo de preguntas suelen utilizarse diseños como series de casos o estudios transversales.

Ejemplos:

- ¿Cuál es la duración promedio de la internación luego de una hemicolectomía izquierda laparoscópica?
- ¿Cuál es la mortalidad de los pacientes con diagnóstico de pancreatitis aguda grave?

En estos casos, el tamaño muestral se calcula generalmente para estimar un parámetro poblacional (media o proporción) con un determinado nivel de precisión.

### **Preguntas analíticas**

Las preguntas analíticas buscan evaluar la relación entre dos o más variables con el objetivo de determinar si existe una asociación, diferencia o posible efecto causal entre ellas.

En este tipo de estudios se formula una hipótesis que se expresa como:

- Hipótesis nula (H0): no existe diferencia o asociación entre las variables.
- Hipótesis alternativa (H1): existe una diferencia o asociación.

El diseño del estudio intenta poner a prueba la hipótesis nula mediante análisis estadístico. Si los resultados permiten rechazarla, se acepta la hipótesis alternativa, lo que sugiere la existencia de una diferencia o asociación estadísticamente significativa. Este tipo de preguntas suele abordarse mediante diseños analíticos como estudios de cohorte (prospectivos o retrospectivos), estudios de casos y controles o ensayos clínicos.

Por ejemplo: ¿Cuál es la tasa de infección del sitio quirúrgico en apendicectomías realizadas por vía abierta en comparación con las realizadas por vía laparoscópica? En este caso, la hipótesis nula establecería que no existen diferencias en la tasa de infección entre ambos abordajes. Si el análisis estadístico permite rechazar esta hipótesis, se aceptaría la hipótesis alternativa, que plantea que la tasa de infección difiere entre los grupos.

### **¿Cómo se calcula el tamaño muestral?**

En la práctica, el cálculo del tamaño muestral requiere definir varios elementos clave. Debido a su frecuencia en las publicaciones científicas en cirugía, nos referiremos al cálculo del tamaño muestral en estudios analíticos<sup>1,3,4,6,7</sup>:

#### **1. Pregunta de investigación**

El investigador debe tener clara la pregunta de investigación. Esto le permitirá en etapas previas definir el diseño de su estudio. Por ejemplo, para la pregunta "¿Cuál es la tasa de recidiva en pacientes mujeres operadas por hernia femoral por vía laparoscópica en comparación con la vía convencional?". Se trata en este caso de una pregunta analítica, que se puede responder tanto con un estudio de cohorte como con un ensayo clínico. Su elección dependerá de

la factibilidad, recursos disponibles y preferencia del investigador.

#### **2. Variable principal**

Es la variable de interés que permite responder la pregunta de investigación. El tipo de variable (categórica o continua) determina el método estadístico utilizado para el cálculo del tamaño muestral. En el ejemplo anterior, la variable de interés es la tasa de recidiva, que corresponde a una variable categórica (sí/no).

#### **3. Magnitud del efecto**

Representa la magnitud mínima de diferencia entre grupos que se consideran clínicamente relevantes. Para obtenerla se necesita evaluar la literatura y conocer cómo se comportan estos tratamientos. Es fundamental la mirada y experiencia de los investigadores vinculados con el ámbito del problema, que se encuentran familiarizados con los efectos de ellos.

Siguiendo el ejemplo, el investigador sabe que la tasa de recidiva de hernia femoral en mujeres informada en la literatura ronda el 5-10% para la vía abierta y el 1-3% para la laparoscópica. En este caso, el investigador podría establecer que una reducción absoluta del 5% en la tasa de recidiva entre la reparación abierta y la laparoscópica constituye una diferencia clínicamente significativa para su estudio.

#### **4. Nivel de significación ( $\alpha$ )**

El nivel alfa ( $\alpha$ ) representa la probabilidad de cometer un error de tipo I, es decir, rechazar la hipótesis nula cuando en realidad es verdadera. En términos prácticos, implica concluir que existe una diferencia o asociación entre las variables estudiadas cuando en realidad esta diferencia no existe.

En investigación clínica, el nivel alfa suele fijarse convencionalmente en 0,05 (5%). Esto significa que el investigador acepta una probabilidad del 5% de encontrar una diferencia estadísticamente significativa que, en realidad, sea producto del azar.

#### **5. Poder estadístico**

El poder estadístico representa la probabilidad de detectar una diferencia real entre los grupos cuando dicha diferencia realmente existe. Se relaciona directamente con el error de tipo II ( $\beta$ ), que ocurre cuando el estudio no detecta una diferencia que verdaderamente existe y, por ende, acepta erróneamente la hipótesis nula (véase anterior). Matemáticamente, el poder se expresa como  $1-\beta$ . En investigación clínica, el poder estadístico suele fijarse convencionalmente en 80-90%, lo que implica aceptar una probabilidad del 20-10% de cometer un error de tipo II.

## 6. ¿Prueba de 1 o 2 colas?

Al calcular el tamaño muestral también debe definirse si el contraste de hipótesis será de una o de dos colas. En una prueba de dos colas se evalúa la posibilidad de que exista una diferencia entre los grupos en cualquiera de las dos direcciones. Por este motivo, el nivel alfa se divide entre ambos extremos de la distribución. Este es el enfoque más utilizado en investigación clínica. En una prueba de una cola, el análisis considera solo una dirección específica de la diferencia (por ejemplo, que una técnica sea superior a otra). Aunque requiere un tamaño muestral menor, su uso debe estar bien justificado.

## 7. Ajuste por pérdidas

En estudios prospectivos debe considerarse la posibilidad de pérdidas durante el seguimiento, por lo que el tamaño muestral inicial suele incrementarse para compensarlas. Habitualmente se incluye un 10-20% más de unidades de observación.

Una vez definidos estos parámetros, el tamaño muestral puede calcularse mediante fórmulas estadísticas específicas o utilizando software especializado, como Epi Info<sup>®</sup>, SPSS<sup>®</sup>, R<sup>®</sup>, Stata<sup>®</sup> u otras herramientas disponibles en línea<sup>5,7</sup>.

## ¿Es siempre necesario calcular el tamaño muestral?

Idealmente, el tamaño de la muestra debería calcularse siempre antes de iniciar cualquier estudio, tanto para preguntas descriptivas como analíticas<sup>3</sup>.

Este cálculo adquiere especial relevancia en los estudios analíticos y, particularmente, en los ensayos clínicos. Estos estudios experimentales y prospectivos asignan a los participantes a una o más intervenciones con el objetivo de evaluar su eficacia y seguridad. Los ensayos clínicos suelen ser complejos y costosos, ya que requieren infraestructura, seguimiento prolongado, personal capacitado, insumos y, con frecuencia, financiamiento externo. Un cálculo incorrecto del tamaño muestral puede conducir a incluir menos participantes de los necesarios y no detectar un efecto real, desperdiciando tiempo y recursos. Por el contrario, reclutar más participantes de los indispensables aumenta

innecesariamente los costos y la exposición de los pacientes a intervenciones experimentales.

En los estudios retrospectivos, aunque los datos ya existan y los costos sean menores, también es recomendable estimar el tamaño muestral requerido. Esto permite evaluar si la cantidad de casos disponibles posee suficiente potencia estadística para detectar asociaciones o diferencias relevantes. De lo contrario, resultados no significativos pueden reflejar simplemente una falta de tamaño muestral y no la ausencia de un efecto real.

## ¿Qué hacer cuando el estudio no alcanza al tamaño muestral previsto?

Cuando no es posible alcanzar el tamaño muestral calculado, pueden considerarse diversas estrategias<sup>3</sup>:

- Reformular el estudio como exploratorio o generador de hipótesis
- Declarar explícitamente la limitación de la potencia estadística
- Realizar colaboraciones multicéntricas para aumentar el tamaño muestral
- Evitar afirmaciones de equivalencia, si el estudio no fue diseñado específicamente para demostrar no inferioridad o equivalencia

Esta transparencia metodológica fortalece la credibilidad científica incluso cuando existen limitaciones inevitables.

## Conclusión

El tamaño muestral representa la traducción cuantitativa de la hipótesis clínica y constituye un pilar fundamental del diseño metodológico. En investigación quirúrgica, su cálculo riguroso resulta indispensable para generar evidencia válida, interpretable y éticamente sustentable. La calidad de nuestras publicaciones no depende únicamente de la destreza técnica en el quirófano, sino también del rigor metodológico con el que diseñamos, analizamos y comunicamos nuestra investigación.

## ■ ENGLISH VERSION

Sample size calculation is an essential component of the methodological design in medical research. In scientific studies, reporting the sample size improves the interpretability of the results, strengthening both the internal validity and ethical legitimacy of the study. This aspect is particularly

important in surgery, given that many clinical events are relatively rare and surgical techniques can vary considerably from surgeon to surgeon.

Despite its significance, failure to calculate the sample size remains a common shortcoming in surgical publications<sup>1,2</sup>. The purpose of this article is to review

the definition, significance, indications, and general principles of sample size calculation.

### What is sample size calculation?

Sample size calculation involves estimating the number of participants or observation units that must be included in a study to adequately answer the research question posed (see below)<sup>2,4</sup>.

In surgery, the units of observation are usually patients or surgical procedures, as well as hospitals, surgeons, or specific clinical episodes. For example, a study analyzing all appendectomies performed in 2025 at a community-based hospital included 120 procedures. In this case, the sample size consists of 120 appendectomies ( $n=120$ ).

### Why is it important to calculate the sample size?

When the sample size is not preliminarily calculated, the number of observation units included may be greater or fewer than necessary. When more participants than necessary are included, costs, resource usage, and research time are unnecessarily increased. Conversely, if the sample size is too small, the study may not have sufficient power to detect a real difference between the groups being analyzed. This situation is known as a type II error or lack of statistical power<sup>3-5</sup>. This error is equivalent to failing to perceive a difference that actually exists. When this happens, the null hypothesis is erroneously accepted, when in fact it should be rejected<sup>3</sup>.

Moreover, an ethical dimension is also implicated. In research studies, patients provide sensitive information with the aim of contributing to scientific knowledge. The design of studies in which the probability of answering the research question is very low may result in an inappropriate use of the data. For instance, a retrospective analysis of an extremely rare condition, such as stump appendicitis, based on only three available cases, would not likely yield robust conclusions.

### What types of research questions exist?

Generally speaking, research questions can be classified as descriptive and analytical. Each type of question requires a specific methodological approach and different strategies for calculating the sample size<sup>3-5</sup>.

#### **Descriptive questions**

The aim of descriptive questions is to characterize or describe how a phenomenon manifests

itself or how a variable is distributed within a given population, without attempting to establish causal relationships. Case series or cross-sectional studies are commonly used for this type of question.

Examples:

- What is the average length of hospital stay following a laparoscopic left hemicolectomy?
- What is the mortality rate among patients diagnosed with severe acute pancreatitis?

In these cases, the sample size is typically calculated to estimate a population parameter (mean or proportion) with a specified level of precision.

#### **Analytical questions**

Analytical questions are designed to evaluate the relationship between two or more variables to determine whether there is an association, a difference, or a possible causal effect between them.

In this type of study, a hypothesis is formulated and expressed as follows:

- Null hypothesis (H0): There is no difference or association between the variables.
- Alternative hypothesis (H1): There is a difference or association.

The study design aims to test the null hypothesis using statistical analysis. If the results reject the null hypothesis, the alternative hypothesis is accepted, suggesting the existence of a statistically significant difference or association. These types of questions are typically addressed using analytical designs such as prospective or retrospective cohort studies, case-control studies, or clinical trials.

For example, what is the surgical site infection rate for open appendectomies compared to laparoscopic appendectomies? In this case, the null hypothesis would state that there are no differences in the infection rate between the two approaches. If the statistical analysis rejects this hypothesis, the alternative hypothesis—which posits that the infection rate differs between the groups—would be accepted.

### How is the sample size calculated?

The calculation of sample size requires defining several key elements. We will refer to the calculation of sample size in analytical studies given that it is frequently mentioned in scientific publications in the field of surgery<sup>1,3,4,6,7</sup>:

## 1. Research question

The researcher must have a clear understanding of the research question in order to define the study design in previous stages. For example, for the question “What is the recurrence rate in female patients undergoing laparoscopic femoral hernia repair compared to conventional surgery?” This is an analytical question that can be answered using either a cohort study or a clinical trial. The type of study chosen will depend on the feasibility, available resources, and the researcher’s preferences.

## 2. Primary variable

The primary variable is the variable of interest that allows answering the research question. The type of variable (categorical or continuous) determines the statistical method used to calculate the sample size. In the previous example, the variable of interest is the recurrence rate, which is a categorical variable (yes/no).

## 3. Effect size

The effect size represents the smallest difference between groups that is considered clinically significant. Determining it requires reviewing the literature to understand how these treatments work. It also relies on the insight and experience of researchers in the field who are familiar with these effects.

Continuing with the example, the researcher knows that the reported recurrence rate for femoral hernias in women is approximately 5–10% for open surgery and 1–3% for laparoscopic surgery. In this case, the researcher could determine that a 5% absolute reduction in the recurrence rate between open and laparoscopic repair constitutes a clinically significant difference in this study.

## 4. Significance level ( $\alpha$ )

The alpha ( $\alpha$ ) level is defined as the probability of committing a type I error, that is, of rejecting the null hypothesis when it is in fact true. In practice, this can lead to the conclusion of difference or association between the variables under study when, in fact, no such difference exists.

In clinical research, the alpha level is conventionally set at 0.05 (5%). This means that the researcher acknowledges a 5% probability of finding a statistically significant difference that is actually due to chance.

## 5. Statistical power

Statistical power represents the probability of detecting a real difference between groups when such

a difference actually exists. It is directly related to a type II ( $\beta$ ) error, which occurs when the study fails to detect a difference that actually exists and therefore erroneously accepts the null hypothesis (see above). Mathematically, statistical power is expressed as  $1 - \beta$ . In clinical research, statistical power is typically set at 80%–90%, which implies accepting a 20%–10% probability of committing a type II error.

## 6. One-tailed test or two-tailed test?

When calculating the sample size, it is also necessary to determine whether a one-tailed or two-tailed hypothesis test should be used. A two-tailed test assesses the possibility of a difference in either direction. For this reason, it splits the alpha level across both ends of the probability distribution. This approach is most commonly used in clinical research. A one-tailed test specifies a single direction of effect or difference (for example, that one technique is superior to another). Although smaller sample sizes are required, its implementation must be well justified.

## 7. Adjusting for losses

In the context of prospective studies, the potential for losses during follow-up must be considered. To address this, the initial sample size is typically increased to compensate for these potential losses. Usually, an additional 10–20% of observation units are included.

Once these parameters have been defined, the sample size can be calculated using specific statistical formulas or specialized software, such as Epi Info, SPSS, R, Stata, or other online tools<sup>5,7</sup>.

## Is it always necessary to calculate the sample size?

Ideally, the sample size should always be calculated before initiating any study, either for descriptive or analytical questions<sup>3</sup>.

Calculating the sample size is particularly important in analytical studies, especially in clinical trials. These experimental and prospective studies assign participants to one or more interventions in order to evaluate their efficacy and safety. Clinical trials are often complex and expensive, as they require infrastructure, long-term follow-up, trained staff, supplies, and, frequently, external funding. Errors in calculating the sample size can result in the inclusion of an insufficient number of participants, thus failing to detect a genuine effect. In turn, this can result in a waste of time and resources. Conversely, recruiting more participants than strictly necessary unnecessarily increases costs and patients’ exposure to experimental interventions.

In retrospective studies, even though the data

already exist and the costs are lower, it is also advisable to estimate the required sample size. This enables the determination of whether there are enough cases to detect relevant associations or differences with sufficient statistical power. Otherwise, non-significant results may simply reflect a lack of sample size rather than the absence of a real effect.

### How to proceed when the study does not reach the planned sample size?

When it is not possible to achieve the calculated sample size, various strategies may be considered<sup>3</sup>:

- Redefine the study as exploratory or hypothesis-generating
- Explicitly state the limitation in statistical power
- Conduct multicenter collaborations to increase the sample size

- Avoid claims of equivalence if the study was not specifically designed to demonstrate non-inferiority or equivalence

This methodological transparency strengthens scientific credibility, even in the presence of unavoidable limitations.

### Conclusion

The sample size represents the quantitative translation of the clinical hypothesis and is a fundamental pillar of the methodological design. In surgical research, rigorous calculation is essential for generating valid, interpretable, and ethically sound evidence. The quality of our publications depends not only on our technical skill in the operating room, but also on the methodological rigor with which we design, analyze, and communicate our research.

### Bibliografía / References

1. Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: Are they done correctly? *Surgery*. 2003;134(2):275-9. doi:10.1067/msy.2003.235.
2. Speich B. Adequate reporting of the sample size calculation in surgical randomized controlled trials. *Surgery*. 2020;167(5):812-4. doi:10.1016/j.surg.2019.10.011.
3. Giunta DH. *Bioestadística Handbook Fundamentos Básicos*. Buenos Aires: Merck S.A; 2018.
4. García-García JA, Reding-Bernal A, López-Alvarenga JC. Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en Educación Médica*. 2013;2(8):217-24. doi:10.1016/S2007-5057(13)72715-7.
5. Althubaiti A. Sample size determination: A practical guide for health researchers. *J Gen Fam Med*. 2023;24(2):72-8. doi:10.1002/jgf2.600.
6. Sheps S. Sample Size and Power. *J Invest Surg*. 1993;6:469-75.
7. Staffa SJ, Zurakowski D. Statistical power and sample size calculations: A primer for pediatric surgeons. *J Pediatr Surg*. 2020;55(7):1173-9. doi:10.1016/j.jpedsurg.2019.05.007

# Utilidad del análisis microbiológico de bilis en colecistitis grado II de la clasificación de Tokio

## *Usefulness of microbiological bile analysis in grade II cholecystitis according to the Tokyo classification*

Guadalupe Frías Mena <sup>ID</sup>, María C. du Plessis <sup>ID</sup>, Héctor F. Posada Lobo <sup>ID</sup>, José R. Varela <sup>ID</sup>, Nicolás F. Baglietto <sup>ID</sup>, Carlos Canullán <sup>ID</sup>, Enrique J. Petracchi <sup>ID</sup>, Hugo I. Zandalazini <sup>ID</sup>

Institución: Servicio de Cirugía General, Hospital General de Agudos Dr. Cosme Argerich, Ciudad de Buenos Aires, Argentina.

Los autores declaran no tener conflictos de interés.

*Conflicts of interest*  
None declared.

Correspondencia  
Correspondence:  
Guadalupe Frías Mena.  
E-mail:  
guadafriasmena@gmail.com

### RESUMEN

**Antecedentes:** Las guías de manejo de colecistitis de Tokio 2018 (GT18) recomiendan el análisis microbiológico de bilis (AMB) intraoperatorio en casos moderados (grado II) con el fin de ajustar la terapia antibiótica.

**Objetivo:** Describir los resultados del AMB con punciones vesiculares intraoperatorias y su relación con comorbilidades, consumo de inhibidores de la bomba de protones (IBP), antibioticoterapia previa y complicaciones posoperatorias, en pacientes operados por colecistitis grado II.

**Material y métodos:** Estudio observacional, prospectivo y descriptivo en adultos con diagnóstico de colecistitis aguda grado II, según las GT18, tratadas con colecistectomía laparoscópica (CL). Se asignaron a los pacientes en 2 grupos, según el resultado del AMB fuera negativo o positivo. Se registraron las variables demográficas, los antecedentes, las características quirúrgicas y microbiológicas. Se aplicó la prueba de Chi-cuadrado ( $\chi^2$ ) para el análisis de variables categóricas. Se consideró un valor de  $p < 0,05$  como significativo.

**Resultados:** Entre 272 pacientes con CL se incluyeron 51 con colecistitis aguda grado II; 29 (57%) fueron mujeres y el promedio de edad fue 57 años (DE 18). En 14 (27%) se encontró AMB positivo. Diez presentaron índice de masa corporal (IMC) igual a 25 o mayor, 5 tuvieron comorbilidades y 7 consumieron IBP. Ninguno tuvo tratamiento previo con antibióticos. El germen más frecuente fue E. Coli (28%). Todos fueron sensibles a ceftriaxona. No hubo eventos adversos relacionados con la punción ni complicaciones posoperatorias.

**Conclusión:** Debido a que los gérmenes aislados en la población operada por colecistitis grado II fueron sensibles al tratamiento antibiótico empírico, no se justificaría realizar el AMB vesicular sistemático por punción durante la intervención.

■ **Palabras clave:** colecistitis aguda, infección biliar, colecistectomía laparoscópica, antibacterianos, microbiología.

### ABSTRACT

**Background:** The Tokyo Guidelines 2018 for the Management of Cholecystitis (TG18) recommend intraoperative microbiological bile analysis (MBA) in moderate cases (grades II) to identify microorganisms and adjust antibiotic therapy.

**Objective:** The aim of this study was to describe the results of samples obtained by intraoperative gallbladder aspiration for MBA and its association with comorbidities, PPI use, prior antibiotic therapy, and postoperative complications in patients with grade II cholecystitis.

**Material and methods:** This observational, prospective and descriptive study included adult patients (> 18 years) diagnosed with grade II acute cholecystitis according to the TG18 managed with laparoscopic cholecystectomy (LC). Patients were assigned to two groups based on a positive or negative MBA. The data analyzed included demographic variables, medical history, and surgical and microbiological characteristics. The chi-square test ( $\chi^2$ ) was used to compare categorical variables. A p-value < 0.05 was considered statistically significant.

**Results:** Of the 272 patients who underwent LC, 51 had grade II acute cholecystitis; 29 (57%) were women, and mean age was 57 years (SD 18). The MBA was positive in 14 (27%) cases. Ten patients had a body mass index (BMI) of 25 or higher, 5 had comorbidities, and 7 used PPIs. None of the patients had received antibiotic therapy prior to hospitalization. E. coli was the most common germ involved (28%). All the germs were susceptible to ceftriaxone. There were no intraoperative adverse events resulting from gallbladder aspiration or postoperative complications.

**Conclusion:** Because the bacteria isolated from patients who underwent surgery for grade II cholecystitis of GT18 were susceptible to the empirical antibiotic treatment, there is no rationale for performing routine gallbladder aspiration during the procedure for MBA.

■ **Keywords:** acute cholecystitis, bile infection, laparoscopic cholecystectomy, anti-bacterial agents, microbiology.

Recibido | Received  
13-06-25  
Aceptado | Accepted  
29-09-25

ID ORCID: Guadalupe Frías Mena ORCID: 0000-0001-6743-8566, María C. du Plessis, 0000-0001-8867-8331, Héctor F. Posada Lobo, 0000-0002-0398-7870; José R. Varela, 0000-0003-3860-235X; Nicolás F. Baglietto, 0000-0003-4906-6361; Carlos Canullán, 0000-0002-5755-0367; Enrique J. Petracchi, 0000-0002-2575-4656; Hugo I. Zandalazini, 0000-0001-7655-2416.

## Introducción

La colecistitis aguda litiasica es una enfermedad inflamatoria de la vesícula biliar que, en la mayoría de los casos, se asocia a infección por microorganismos de origen gastrointestinal, y es una causa frecuente de consulta médica. La colecistitis aguda se clasifica en leve (I), moderada (II) o grave (III) según los criterios de las guías de manejo de colecistitis de Tokyo 2018.<sup>1</sup>

En las colecistitis grado II, el tratamiento de elección es la colecistectomía laparoscópica asociada a antibioticoterapia empírica.<sup>2-4</sup> En nuestro hospital, el esquema de primera línea es ceftriaxona 1 gramo cada 12 horas asociada a metronidazol 500 mg cada 8 horas por vía intravenosa, de acuerdo con las recomendaciones del Servicio de Infectología, basadas en las guías de la Infectious Diseases Society of America (IDSA).

La elección del tratamiento antibiótico empírico deriva de la flora bacteriana gastrointestinal y los patrones de resistencia locales. La microbiota biliar puede variar en función de diversos factores tales como hospitalizaciones recientes, exposición previa a antibióticos (ATB) y la presencia de comorbilidades, entre otras.<sup>5,6</sup> A su vez, el consumo prolongado de inhibidores de la bomba de protones (IBP) predispone a la disfunción de la motilidad vesicular y a la coledocistitis, pudiendo alterar la flora bacteriana del tubo digestivo y la vía biliar.<sup>5,7</sup> Sin embargo, no existen informes de la variación de la flora bacteriana vesicular y la relación con el consumo de aquellos.

Las guías de manejo de colecistitis de Tokio (GT18) recomiendan el análisis microbiológico de bilis (AMB) intraoperatorio en casos moderados y graves (grados II y III), con el fin de identificar los microorganismos y ajustar la terapia antibiótica.<sup>1,8</sup>

El objetivo del presente estudio fue describir los resultados del AMB obtenido por punciones vesiculares intraoperatorias y su relación con las comorbilidades, el consumo de inhibidores de la bomba de protones (IBP), antibioticoterapia previa y las complicaciones posoperatorias en pacientes con colecistitis grado II.

## Material y métodos

Se realizó un estudio observacional, prospectivo y descriptivo. Se incluyeron pacientes adultos (mayores de 18 años) con diagnóstico de colecistitis aguda clasificados como grado II según las GT18, en los que se realizó colecistectomía laparoscópica en un Hospital público de la Ciudad de Buenos Aires.

Se excluyeron los pacientes tratados previamente con colecistostomía percutánea o colangiopancreatografía retrógrada endoscópica. Se reclutaron pacientes entre julio de 2023 y marzo de 2024.

Los pacientes incluidos se operaron dentro de las 48 horas del ingreso hospitalario y recibieron antibioticoterapia empírica dentro de la primera hora de diagnóstico. Al encontrarse con un esquema antibiótico desde el ingreso, no se administraron ATB en la inducción anestésica. El alta médica estuvo sujeta a la evolución clínica y no al resultado de los cultivos. Los pacientes se externaron con ATB vía oral (amoxicilina más ácido clavulánico 1 gramo) hasta completar una semana y se controlaron por consultorio externo para corroborar la sensibilidad en el AMB.

Los pacientes fueron distribuidos en 2 grupos, negativo o positivo, según el resultado del AMB y se analizaron las variables demográficas (edad, sexo), antecedentes personales (índice de masa corporal o IMC, presencia de comorbilidades, consumo regular de IBP durante las últimas 4 semanas y consumo de antibióticos previo a la internación), tiempo operatorio, eventos adversos intraoperatorios derivados de la punción, complicaciones posoperatorias para ambos grupos y variables microbiológicas solo para el grupo positivo (microbiología y antibiograma). Las comorbilidades que se tuvieron en cuenta fueron: diabetes mellitus, hipertensión arterial o insuficiencia renal.

Se tomó una muestra de líquido vesicular bajo visión laparoscópica directa, mediante punción percutánea con aguja tipo Abbocath 16° y jeringa durante la colecistectomía. Se envió la muestra al Servicio de Microbiología para AMB y se procesó según los protocolos estandarizados del laboratorio, realizándose identificación bacteriana y estudio de sensibilidad (CIM) mediante el sistema automatizado Vitek 2 (bioMérieux®).

Se recolectaron los datos en planilla de Google Excel® y se realizó análisis bivariado para determinar las posibles asociaciones y comparaciones. Se aplicó la prueba estadística (IBM SPSS Statistics 25®) de Chi-cuadrado ( $\chi^2$ ) para el análisis de variables categóricas y t de Student para las continuas. Se consideró un valor de  $p < 0,05$  como significativo.

Se mantuvo la confidencialidad en todos los casos rigiéndose bajo los principios del secreto médico establecido en la Ley de ejercicio legal de la medicina (Ley 17132) y resolución 1480/2011. No se requirió consentimiento informado para esta investigación debido a que se realizaron procedimientos según estándares de cuidados para la asistencia de esta patología. El protocolo fue aprobado por el Comité de Ética independiente de la ciudad de Buenos Aires (Priisa.BA 11900).

## Resultados

En el período analizado se realizaron 272 colecistectomías laparoscópicas, 64 de ellas por colecistitis aguda grado II de las GT18. Se incluyeron 51 pacientes.

Fueron excluidos 5 pacientes por presencia de vesícula biliar colapsada, 5 por imposibilidad técnica de acceso percutáneo atribuida a obesidad grave (IMC elevado) y 3 por antecedentes de colecistostomía percutánea. Se encontró cultivo de AMB positivo en 14 (27%).

Se analizaron en ambos grupos las variables demográficas y antecedentes personales (Tabla 1).

Ningún paciente había recibido tratamiento antibiótico previo a la internación. No hubo eventos adversos intraoperatorios derivados de la punción ni complicaciones posoperatorias. El tiempo operatorio fue de 99 minutos (DE 38).

Seis (42,8%) de los pacientes con AMB positivo presentaron resistencia. En la tabla 2 se encuentran los resultados de los cultivos y los antibiogramas.

No hubo diferencias de duración del tratamiento o del tiempo de internación entre los pacientes con cultivo positivo y negativo.

## Discusión

Los AMB en colecistitis aguda grado II, en la mayoría de los casos, no arrojaron resultados positivos. En las GT18 se menciona una tasa de positividad de cultivos en estos casos de 29 a 54%<sup>2</sup>. En nuestra experiencia el porcentaje fue menor (27%), al igual que en otros estudios<sup>9</sup>, lo que puede deberse a las diferencias en el tamaño muestral.

En aquellos casos donde el cultivo resultó positivo, los microorganismos fueron sensibles al tratamiento empírico actualmente utilizado por nuestro Servicio. La punción vesicular para análisis microbiológico no modificó la conducta con respecto al tratamiento antibiótico posoperatorio. Si bien el metronidazol se asocia a la ceftriaxona en nuestro medio, por la supuesta presencia de gérmenes anaerobios, su sensibilidad no es evaluada mediante antibiograma.

La resistencia más frecuentemente hallada fue a ampicilina, un antibiótico que no se encuentra dentro de las recomendaciones empíricas de nuestro Servicio ni de las GT18<sup>2</sup>.

Los microorganismos más frecuentemente aislados fueron *Escherichia coli* (28%), *Enterobacter cloacae* (14%) y *Streptococcus parasanguinis* (14%). Algunos estudios previos, tanto en Latinoamérica como en otras regiones, también identifican a *Escherichia coli* como el patógeno predominante, aunque con tasas de aislamiento superiores al 45%<sup>3,5,6</sup>.

En la serie aquí presentada, la edad de los pacientes con cultivos biliares positivos fue significativamente mayor que la de aquellos con cultivos negativos (57 vs. 44, p=0,03). Este hallazgo coincide con lo informado en la literatura, donde la edad avanzada constituye un factor de riesgo independiente para colonización bacteriana de la bilis<sup>9,11</sup>, probablemente relacionado con inmunosenescencia, mayor prevalencia de comorbilidades y estasis biliar asociada a cambios

■ TABLA 1

Variables demográficas y antecedentes personales según el resultado del análisis microbiológico de bilis en 51 pacientes con colecistitis grado II

VARIABLES	AMB negativo (n=37)	AMB positivo (n=14)	P
Edad (media ± DE)	44 ± 16	57 ± 18	0,031*
Sexo n (%)			
Femenino	24 (65)	8 (57)	0,63**
Masculino	13 (35)	6 (43)	
IMC n (%)			
< 25	12 (32)	4 (28)	0,943**
≥ 25 - <30	11 (30)	4 (28)	
≥ 30	14 (38)	6 (46)	
Presencia de comorbilidades n (%)	10 (27)	5 (36)	0,54**
Consumo de IPB n (%)	12 (32)	7 (50)	0,45**

\*Prueba de t de Student; \*\*prueba de chi cuadrado.

■ TABLA 2

Hallazgos microbiológicos y antibiogramas de los pacientes con análisis microbiológico de bilis positivo

Microbiología	Sensibilidad	Resistencia
	CTX* AMS† TMS‡ CLI§	AMPI¶
4 <i>Escherichia coli</i>	MS**	-
	MS	-
	MS	-
2 <i>Enterobacter cloacae</i>	CTX CIP++	AMPI AMS
	CTX* CIP	AMS PTZ##
2 <i>Streptococcus parasanguinis</i>	MS	-
	CTX* CIP CLI	ERI§§
<i>Streptococcus gordonii</i>	MS	-
<i>Streptococcus vestibularis</i>	MS	-
<i>Candida dubliniensis</i>	MS	-
<i>Staphylococcus aureus</i>	CTX CIP TMS	ERI GEN¶¶ CLI
<i>Citrobacter braaki</i>	CTX GEN	AMPI
<i>Acinetobacter baumannii</i> complex	MS	-

\*Ceftriaxona, †Ampicilina+sulbactam, ‡Trimetoprima+sulfametoxazol, §Clindamicina, ¶Ampicilina, \*\*Multisensible, ††Ciprofloxacina, ‡‡Piperacilina+tazobactam, §§Eritromicina, ¶¶Gentamicina.

anatomofuncionales de la vesícula. Futuras investigaciones podrían centrarse en el análisis de ese subgrupo de pacientes.

Aunque no se observó una asociación estadísticamente significativa entre la presencia de comorbilidades y el AMB, 10 pacientes (71%) de aquellos con cultivos positivos presentaban IMC elevado.

El 50% de los pacientes con cultivos positivos utilizó IBP previamente, sin alcanzar diferencia significativamente estadística. No obstante, esta observación plantea la necesidad de estudios adicionales que permitan esclarecer el posible vínculo entre el uso de IBP

y la alteración de la microbiota biliar en la colecistitis aguda.

Aunque no fue objetivo de este trabajo, la baja tasa de aislamiento bacteriano observada podría cuestionar la necesidad de antibióticos empíricos en estos pacientes.

Dado que los pacientes recibieron antibiotico-terapia empírica dentro de la primera hora posterior al diagnóstico, los resultados de los cultivos pudieron haberse visto alterados. Sin embargo, este factor afectó por igual a ambos grupos, y se informó al laboratorio sobre tal circunstancia para que fuera considerada durante el análisis de las muestras. A fin de minimizar este sesgo, se podría haber realizado la punción vesicular antes de la administración de antibióticos y de la intervención quirúrgica. No obstante, esta alternativa plantea cuestionamientos éticos.

El AMB no prolongó la estadía hospitalaria, ya que el alta médica dependió exclusivamente de la evolución clínica del paciente. Los resultados de los cultivos

podrían adquirir relevancia principalmente en casos de evolución posoperatoria desfavorable.

No se observaron modificaciones en la duración del tratamiento ni en el tiempo de internación en los pacientes con cultivos positivos, dado que todos los microorganismos aislados fueron sensibles al esquema empírico utilizado.

La punción vesicular es una técnica sencilla que no prolonga el tiempo operatorio. Además, en determinadas situaciones, la evacuación controlada de la bilis mediante esta técnica facilita la tracción vesicular, lo que contribuye a una colecistectomía más segura y eficaz.

La limitación de este estudio fue su carácter monocéntrico.

En conclusión, debido a que los gérmenes aislados en la población operada por colecistitis grado II de las GT18 fueron sensibles al tratamiento antibiótico empírico, no se justificaría realizar el AMB vesicular sistemático por punción durante la intervención.

## ENGLISH VERSION

### Introduction

Acute calculous cholecystitis is a gallbladder inflammation that, in most cases, is associated with infection by microorganisms from the gastrointestinal tract and is a common reason for seeking medical care. Acute cholecystitis is classified as mild (grade I), moderate (grade II), or severe (grade III) according to the criteria of the Tokyo guidelines 2018 for the management of cholecystitis<sup>1</sup>.

The standard of care for grade II cholecystitis is a laparoscopic cholecystectomy combined with an empirical antibiotic regimen<sup>2-4</sup>. In our hospital, the first-line regimen recommended by infectologists is based on the guidelines of the Infectious Diseases Society of America (IDSA): ceftriaxone 1 g intravenously every 12 hours combined with metronidazole 500 mg intravenously every 8 hours.

The choice of an empirical antibiotic regimen depends on the bacterial flora of the gastrointestinal tract and local resistance patterns. The composition of the gallbladder microbiota can vary depending on various factors, such as recent hospitalizations, prior exposure to antibiotics, and the presence of comorbidities, among others<sup>5,6</sup>. Prolonged use of proton pump inhibitors (PPIs) can lead to gallbladder motility dysfunction and cholelithiasis, which may affect the gut and gallbladder microbiota<sup>5,7</sup>. However, there are no reports associating changes in gallbladder microbiota with PPI use.

The Tokyo Guidelines 2018 for the Management of Cholecystitis (TG18) recommend intraoperative microbiological bile analysis (MBA) in

moderate and severe cases (grades II and III) to identify microorganisms and adjust antibiotic therapy<sup>1,8</sup>.

The aim of this study was to describe the results of MBA with samples obtained by intraoperative gallbladder aspiration and its association with comorbidities, PPI use, prior antibiotic therapy, and postoperative complications in patients with grade II cholecystitis.

### Material and methods

We conducted an observational, prospective and descriptive study. The study included adult patients (> 18 years) diagnosed with grade II acute cholecystitis according to the TG18 laparoscopic cholecystectomy at a public hospital in the City of Buenos Aires.

Patients who had previously undergone percutaneous cholecystostomy or endoscopic retrograde cholangiopancreatography were excluded. Enrollment occurred between July 2023 and March 2024.

The patients included in the study underwent surgery within 48 hours of hospital admission and received empirical antibiotic therapy within one hour of diagnosis. Antibiotics were not administered during induction because the patients were already on an antibiotic regimen upon admission. Hospital discharge was decided based on patients' progress and culture results. Patients were discharged with oral antibiotics (a combination of amoxicillin and clavulanic acid, 1 g) until completing one week and were followed up in the outpatient clinic to confirm susceptibility in the MBA.

Patients were divided into two groups, negative and positive, based on the results of the MBA. The following variables were analyzed: demographic characteristics (age, sex), personal history [body mass index (BMI), presence of comorbidities, regular use of PPIs during the previous 4 weeks, and antibiotic use prior to admission], operative time, intraoperative adverse events resulting from gallbladder aspiration, postoperative complications for both groups, and microbiological variables only in the positive group (microbiology and antimicrobial susceptibility testing). The comorbidities considered were diabetes mellitus, hypertension or kidney disease.

A sample of gallbladder fluid was obtained under direct laparoscopic visualization using a 16-gauge Abbocath® catheter and syringe during cholecystectomy. The sample was submitted to the Microbiology Department for MBA testing and was processed according to the laboratory’s standardized protocols. Bacterial identification and susceptibility testing (minimal inhibitory concentration) were performed using the automated Vitek 2 system (bioMérieux®).

Data were collected in a Google Excel® spreadsheet, and bivariate analysis was performed to identify potential associations and make comparisons. The chi-square test ( $\chi^2$ ) was used to compare categorical variables, and continuous variables were compared with the Student’s t-test. All calculations were performed using IBM SPSS Statistics 25® software package. A p-value < 0.05 was considered statistically significant.

Confidentiality was maintained in all cases, in accordance with the principles of medical ethics established by the Medical Practice Act (Law 17,132) and Resolution 1480/2011. Informed consent was not required for this study because the procedures were performed following the standard of care for the management of this condition. The study protocol was approved by the independent Committee on Ethics of the city of Buenos Aires (Priisa.BA 11900).

**Results**

During the study period, 272 laparoscopic cholecystectomies were performed; 64 procedures corresponded to grade II acute cholecystitis according to the GT18 criteria and 51/64 patients were included. Five patients were excluded due to a collapsed gallbladder, five due to technical impossibility for percutaneous access attributed to severe obesity (high BMI), and three due to a history of percutaneous cholecystostomy. The MBA was positive in 12 (27%) cases.

The demographic variables and personal history data were analyzed (Table 1).

None of the patients had received antibiotic therapy before admission. There were no intraoperative adverse events resulting from gallbladder aspiration or postoperative complications. Operative time was 99 minutes (SD 38).

Six (42.8%) patients with positive MBA presented antibacterial resistance. Table 2 shows the results of the cultures and antimicrobial susceptibility tests.

There were no differences in treatment duration or length of hospital stay between patients with positive and negative cultures.

**TABLE 1**

Demographic variables and personal history according to the results of microbiological bile analysis of bile in 51 patients with grade II cholecystitis

Variables	Negative MBA (n=37)	Positive MBA (n=14)	p-value
Age (mean ± SD)	44 ± 16	57 ± 18	0.031*
Sex n (%)			
Female	24 (65)	8 (57)	0.63**
Male	13 (35)	6 (43)	
BMI n (%)			
< 25	12 (32)	4 (28)	0.943**
≥ 25 - <30	11 (30)	4 (28)	
≥ 30	14 (38)	6 (46)	
Comorbidities n (%)	10 (27)	5 (36)	0.54**
Use of PPIs n (%)	12 (32)	7 (50)	0.45**

\*Student’s t test, \*\*chi-square test.

**TABLE 2**

Bacteria found and results of antimicrobial susceptibility testing in patients undergoing microbiological bile analysis

Microbiology	Susceptibility	Resistance
	CTX* AMS† TMS‡ CLI§	AMPI¶
4 Escherichia coli	MS**	-
	MS	-
	MS	-
2 Enterobacter cloacae	CTX CIP++	AMPI AMS
	CTX* CIP	AMS PTZ##
2 Streptococcus parasanguinis	MS	-
	CTX* CIP CLI	ERI§§
Streptococcus gordonii	MS	-
Streptococcus vestibularis	MS	-
Candida dubliniensis	MS	-
Staphylococcus aureus	CTX CIP TMS	ERI GEN¶¶ CLI
Citrobacter braaki	CTX GEN	AMPI
Acinetobacter baumannii complex	MS	-

\*Ceftriaxone, †Ampicillin+sulbactam, ‡Trimethoprim+sulfamethoxazole, §Clindamycin, ¶Ampicillin, \*\*Multisusceptibility, ††Ciprofloxacin, ##Piperacillin+tazobactam, §§Erythromycin, ¶¶Gentamicin.

## Discussion

In grade II acute cholecystitis, MBA did not show positive results in most cases. The TG18 mentioned positive cultures between 29% to 54%<sup>2</sup>. In our experience, the percentage was lower (27%), as was the case in other studies<sup>9</sup>, which may be due to differences in the sample size.

In cases with a positive culture, the microorganisms were susceptible to the empirical treatment currently used by us. Gallbladder aspiration for microbiological analysis did not affect the approach to postoperative antibiotic therapy. While the combination of metronidazole and ceftriaxone is commonly used in our setting due to the presumed presence of anaerobic bacteria, their susceptibility is not assessed by antimicrobial susceptibility testing.

The most common resistance was to ampicillin, an antibiotic that is not included in the empirical treatment recommendations of our department or in the GT18<sup>2</sup>.

*Escherichia coli* (28%), *Enterobacter cloacae* (14%), and *Streptococcus parasanguinis* (14%) were the bacteria most often isolated. Previous studies conducted in Latin America, as well as in other regions, have also identified *Escherichia coli* as the predominant pathogen, although with isolation rates exceeding 45%<sup>3,5,6</sup>.

In this series, patients with positive bile cultures were significantly older than those with negative cultures (57 vs. 44;  $p = 0.03$ ). This finding aligns with the established literature, which identifies advanced age as an independent risk factor for bacterial colonization of bile<sup>9-11</sup>. This is likely due to immunosenescence, a higher prevalence of comorbidities, and biliary ectasia associated with anatomical and functional changes in the gallbladder. Further research could focus on analyzing this subgroup of patients.

Although we did not find a statistically significant association between the presence of comorbidities and MBA, BMI was high in 10 (71%) patients with positive cultures.

Fifty percent of patients with positive cultures had previously used PPIs; yet this difference was not statistically significant. Nevertheless, this finding elicits the need for further studies to shed light on the possible association between the use of PPIs and changes in the gallbladder microbiota in acute cholecystitis.

While this was not the objective of the study, the low rate of bacteria isolated in bile cultures raises questions about the necessity of empirical antibiotic use in these patients.

Because the patients received empirical antibiotic therapy within the first hour after diagnosis, the culture results may have been affected. However, this issue affected both groups to the same degree, and the laboratory was informed to take this factor into account during sample analyses. To minimize this bias, gallbladder aspiration could have been performed before the antibiotics were administered and surgery was performed. However, this approach raises ethical questions.

The MBA did not prolong hospital length of stay, as patients were discharged based exclusively on their clinical progress. The culture results may be particularly relevant in cases with unfavorable postoperative course.

There were no changes in treatment duration or length of hospital stay among patients with positive cultures, as all isolated microorganisms were susceptible to the empirical regimen used.

Gallbladder aspiration is a simple procedure that does not prolong operative time. Furthermore, in certain situations, the controlled drainage of bile using this technique facilitates gallbladder retraction, which contributes to a safer and more effective cholecystectomy.

The limitation of this study is its single-center design.

In conclusion, because the bacteria isolated from patients who underwent surgery for grade II cholecystitis of GT18 were susceptible to the empirical antibiotic treatment, there is no rationale for performing routine gallbladder aspiration during the procedure for MBA.

## Referencias bibliográficas /References

- Okamoto K, Suzuki K, Takada T, Strasberg SM, Asbun HJ, Endo I, et al. Tokyo Guidelines 2018: flowchart for the management of acute cholecystitis. *J Hepatobiliary Pancreat Sci* [Internet]. 2018;25(1):55-72. Disponible en: <http://dx.doi.org/10.1002/jhbp.516>.
- Gomi H, Solomkin JS, Schlossberg D, Okamoto K, Takada T, Strasberg SM, et al. Tokyo Guidelines 2018: antimicrobial therapy for acute cholangitis and cholecystitis. *J Hepatobiliary Pancreat Sci* [Internet]. 2018;25(1):3-16. Disponible en: <http://dx.doi.org/10.1002/jhbp.518>.
- Kujiraoka M, Kuroda M, Asai K, Sekizuka T, Kato K, Watanabe M, et al. Comprehensive diagnosis of bacterial infection associated with acute cholecystitis using metagenomic approach. *Front Microbiol* [Internet]. 2017;8:685. Disponible en: <http://dx.doi.org/10.3389/fmicb.2017.00685>.
- van Dijk AH, de Reuver PR, Tasma TN, van Dieren S, Hugh TJ, Boermeester MA. Systematic review of antibiotic treatment for acute calculous cholecystitis: Antibiotic treatment for acute calculous cholecystitis. *Br J Surg* [Internet]. 2016;103(7):797-811. Disponible en: <http://dx.doi.org/10.1002/bjs.10146>.
- Granel-Villach L, Gil-Fortuño M, Fortea-Sanchis C, Gamón-Giner RL, Martínez-Ramos D, Escrig-Sos VJ. Factores que influyen en la microbiología del líquido biliar en los pacientes colecistectomizados. *Rev Gastroenterol Méx (Engl Ed)* [Internet]. 2020;85(3):257-63. Disponible en: <http://dx.doi.org/10.1016/j.rgmx.2019.07.006>.
- Gómez-Meléndez LJ, Acosta Pérez CA, Támara-Prieto JA, Castellanos-Méndez JS, Márquez-Niño AI, Fernández-Ávila DG. Caracterización de la flora bacteriana en los pacientes con colelitiasis y colecistitis atendidos en un centro hospitalario de alta complejidad. *Rev Colomb Cir* [Internet]. 2022. Disponible en: <http://dx.doi.org/10.30944/20117582.2175>.
- Yang M, Xia B, Lu Y, He Q, Lin Y, Yue P, et al. Association between regular use of gastric acid suppressants and subsequent risk

- of cholelithiasis: A prospective cohort study of 0.47 million participants. *Front Pharmacol* [Internet]. 2021;12:813587. Disponible en: <http://dx.doi.org/10.3389/fphar.2021.813587>.
8. Ocampo C, Álvarez Rodríguez J, Zandalazini J, Chiapetta Porras L, Romano M, Fernández G y col. Punción y evacuación simple de la vesícula por vía percutánea en la patología vesicular aguda. *Rev Argent Cirug.* 1996;71(6):198-205.
  9. Yun SP, Seo H-I. Clinical aspects of bile culture in patients undergoing laparoscopic cholecystectomy. *Medicine (Baltimore)* [Internet]. 2018;97(26):e11234. Disponible en: <http://dx.doi.org/10.1097/MD.00000000000011234>.
  10. Moon DK, Kang JS, Byun Y, Choi YJ, Lee HW, Jang JY, et al. Incidence of bactibilia and related factors in patients who undergo cholecystectomy. *Ann Surg Treat Res.* 2023;104(1):10-0. Disponible en: <https://doi.org/10.4174/astr.2023.104.1.10>.
  11. Tae Young Park, Jae Hyuk Do, Oh HC, Yoo Shin Choi, Seung Eun Lee, Kang H, et al. Relationship between the Tokyo Guidelines and Pathological Severity in Acute Cholecystitis. *J Pers Med* [Internet]. 2023;13(9):1335-5. Available from: <https://pubmed.ncbi.nlm.nih.gov/37763103/> DOI: <https://doi.org/10.3390/jpm13091335>.

# Resultados de corto plazo del tratamiento quirúrgico de tumores retroperitoneales en un periodo de 10 años

## Short-term outcomes of surgical treatment for retroperitoneal tumors over a 10-year period

Gustavo A. Nari<sup>1,2,3</sup> , José Layún<sup>1,2</sup> , Sol Romero<sup>2</sup> , Daniela Mariot<sup>1,2</sup> , Ángel Joseph<sup>1</sup> , María E. De elias<sup>1</sup> 

1. Servicio de Cirugía. Hospital Tránsito Cáceres de Allende. Córdoba. Argentina.
2. Unidad de Cirugía oncológica. Sanatorio La Cañada. Córdoba. Argentina.
3. Cátedra Clínica Quirúrgica II-UHC4. Universidad Nacional de Córdoba. Argentina.

Los autores declaran no tener conflictos de interés.  
*Conflicts of interest*  
None declared.

Correspondencia  
*Correspondence:*  
Gustavo Nari  
E-mail: gusnari@hotmail.com

### RESUMEN

**Antecedentes:** los tumores retroperitoneales son una patología infrecuente, cuya cirugía implica un desafío para el cirujano, ya que el espacio donde se desarrollan tiene vasos importantes que muchas veces están laminados o desplazados por el tumor. Los sarcomas son los tumores preponderantes.

**Objetivo:** describir los resultados de corto plazo obtenidos con la cirugía de pacientes con tumores retroperitoneales.

**Material y métodos:** se realizó un estudio de diseño observacional, descriptivo, retrospectivo. Fueron revisadas las historias clínicas de los pacientes operados entre enero de 2005 y diciembre de 2024. Se registraron los datos demográficos y clínico-quirúrgicos.

**Resultados:** fueron operados 12 pacientes, 8 con lesiones malignas, 7 de las cuales fueron sarcomas. No hubo mortalidad en la serie y las complicaciones estuvieron presentes en 4 casos (33%). El promedio de la estadía hospitalaria fue de 6,5 días. La resección R0 por anatomía patológica se logró en 9 pacientes. De aquellos pacientes que pudieron ser controlados, tres tuvieron recaídas tumorales.

**Conclusión:** los sarcomas fueron los tumores retroperitoneales más frecuentes; la resección R0 se pudo obtener en la mayoría de los casos, con tasa de complicaciones comparables con otras serie

■ **Palabras clave:** tumores retroperitoneales, sarcomas de partes blandas, cirugía general, procedimientos quirúrgicos.

### ABSTRACT

**Background:** Retroperitoneal tumors are uncommon, and their surgical removal poses a challenge for surgeons, as the space where they develop contains major blood vessels that are often compressed or displaced by the tumor. Sarcomas are the most common retroperitoneal tumors.

**Objective:** The aim of this study was to evaluate the short-term outcomes of surgical resection for retroperitoneal tumors.

**Material and methods:** We conducted an observational, retrospective and descriptive study. The medical records of patients who underwent surgery between January 2005 and December 2024 were reviewed. The demographic, clinical and operative variables were analyzed.

**Results:** A total of 12 patients underwent surgery; 8 had malignant lesions, 7 of which were sarcomas. There were no deaths in the series and 4 patients developed in 4 cases (33%). Mean length of hospital stay was 6.5 days. An R0 resection confirmed by pathological examination was achieved in 9 patients. Three patients who could be followed up developed tumor recurrences.

**Conclusion:** Sarcomas were the most common retroperitoneal tumors. An R0 resection was achieved in most cases, with complication rates comparable to those reported in other series.

■ **Keywords:** retroperitoneal tumors, soft tissue sarcomas, general surgery, surgical procedures.

Recibido | Received  
01-07-25  
Aceptado | Accepted  
06-10-25

ID ORCID: Gustavo A. Nari, 0000-0002-2559-5704; José Layún, 0000-0003-2723-0765; Sol Romero, 0009-0009-3118-0298; Daniela Mariot, 0000-0002-1470-8400; Ángel Joseph, 0009-0008-4347-4448; María E De elias, 0000-0003-4619-4149

## Introducción

Los tumores retroperitoneales son un grupo heterogéneo de lesiones de tejidos blandos que tienen origen en diferentes estirpes celulares y que pueden ser benignos o malignos mientras que algunos de ellos tienen un comportamiento intermedio<sup>1</sup>. La mayoría suelen ser malignos y, entre estos, los sarcomas son los preponderantes. Dentro de este último grupo, los liposarcomas son los más frecuentes en la mayoría de las series<sup>1-3</sup>. Los tumores retroperitoneales suelen alcanzar grandes tamaños y gran parte de la clínica que acompaña estas lesiones suele tener que ver con el efecto masa; por otra parte, la capacidad de generar metástasis a distancia es bastante pobre<sup>3</sup>. En general, el tratamiento que mejores resultados ofrece a estos pacientes es la resección quirúrgica completa ya que algunos tumores tienen un comportamiento agresivo local, lo que favorece que exista un alto porcentaje de recurrencias locales<sup>4</sup>. El lugar de la neoadyuvancia y adyuvancia con quimioterapia y/o radioterapia, excepto contadas excepciones, no suele aportar mejoras en la supervivencia.

El objetivo del presente estudio fue evaluar los resultados a corto plazo de la resección quirúrgica de tumores retroperitoneales.

## Material y método

Fue realizado un estudio de diseño observacional, descriptivo, retrospectivo. Se realizó la búsqueda de los datos de las historias clínicas de los pacientes derivados para cirugía entre enero de 2005 y diciembre de

2024 por presentar tumores retroperitoneales. Aquellos que no fueron operados o de los cuales no se poseía la información mínima necesaria fueron excluidos del estudio. También se excluyeron los tumores que se generaron en órganos retroperitoneales o en tejido linfático, primordialmente linfomas.

Se registraron los datos demográficos, las variables clínicas, patológicas y quirúrgicas. Las complicaciones fueron estratificadas según la clasificación de Dindo-Clavien<sup>5</sup>. La supervivencia fue medida en meses y evaluada hasta el fallecimiento del paciente; en aquellos pacientes que se perdieron de control se pesquisó si aún permanecían vivos en el Renaper (Registro Nacional de Personas. Argentina.gob.ar). La estadía hospitalaria fue medida en días.

Para su tratamiento estadístico los datos fueron volcados a una tabla de referencia. Las variables cualitativas fueron expresadas con rango, frecuencia y porcentaje mientras que las cuantitativas con el rango, mediana y desviación estándar (DS).

## Resultados

En el período señalado fueron atendidos 19 pacientes con tumores retroperitoneales; de estos, 7 fueron excluidos del análisis ya que 4 se consideraron irreseccables de manera preoperatoria, 2 fueron irreseccables de manera intraoperatoria y de uno no se encontraron los datos mínimos necesarios para incorporarlo al estudio (Fig. 1).

Quedaron para el análisis 12 pacientes que fueron sometidos a resección con intención curativa de sus tumores retroperitoneales. Los datos demográficos y las variables clínico-patológicas se presentan en la Tabla 1.

■ FIGURA 1

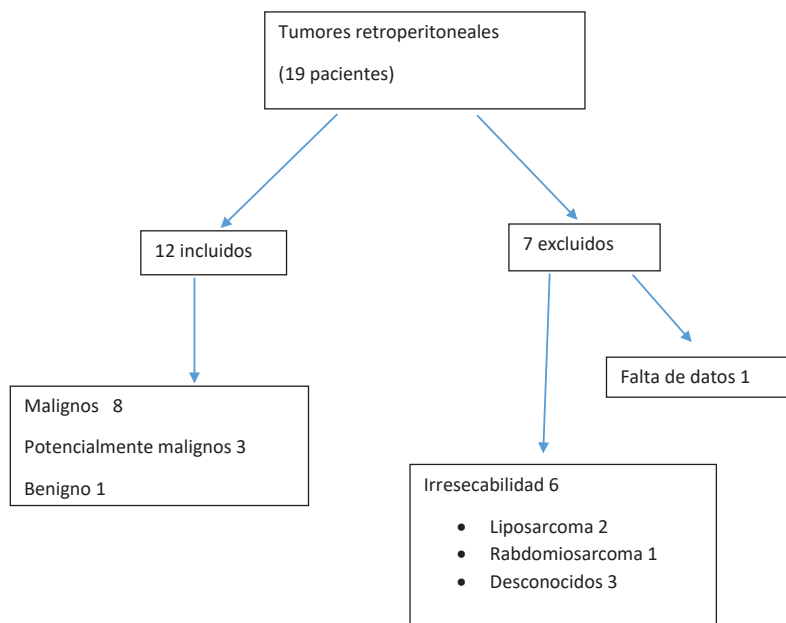


Diagrama de flujo. Tipos de tumores y criterios por los que fueron excluidos

■ TABLA 1

Variables demográficas y clínicas preoperatorias

Variable		
Género masculino (n)	8	66,6%
Edad (rango) (Mediana - DS)	19-61	42 ± 14,96
Síntomas		
Dolor abdominal (n)	12	100%
Pérdida de peso (n)	8	66,6%
Náuseas y vómitos (n)	7	58,3%
Distensión (n)	8	66,6%
Masa palpable (n)	7	58,3%
Antecedentes Familiares	7	58,3%
Laboratorio		
Albuminemia (rango) (Mediana - DS)	2,9-3,4	3,25 ± 0,16
Hemoglobina (rango) (Mediana - DS)	10-13	11,5 ± 1
Marcadores tumorales (n)	5	41,6%
Performance Status		
0 (n)	9	75%
1 (n)	3	25%
Sarcopenia (n)	4	33,3%
Estudios realizados		
Ecografía (n)	12	100%
Tomografía (n)	12	100%
RMN (n)	10	83,3%
Colonoscopia (n)	3	25%
Endoscopia alta (n)	6	50%
Punción Biopsia (n)	5	41,6%
Tamaño. Diámetro Máximo (rango)	5-30 cm	16,9 cm
Resultado Punción Biopsia	1 (+) 4 (-)	

■ TABLA 2

Variables quirúrgicas y posoperatorias

Variable		
Resección		
RO (n)	8	72,7%
R1 (n)	3	27,2%
Resección de más de un órgano (n)	1	8,33%
Dindo-Clavien		
Grado I (n)	3	25%
Grado II (n)	1	8,3%
Estadía hospitalaria (días), (Mediana - DS)	3-9	6,5 (± 1,78)
Recaída tumoral (n)	3	25%
Adyuvancia (n)	3	25%
Seguimiento en meses (rango), (Mediana - DS)	4-96	60 (± 39)

■ TABLA 3

Variables anatomopatológicas

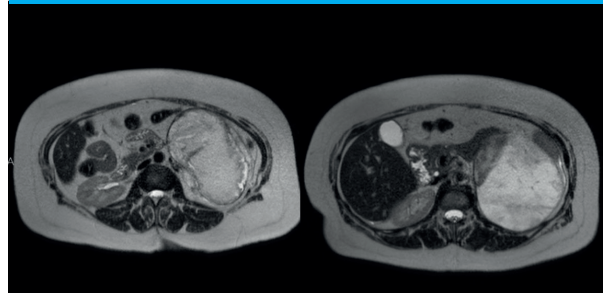
Variable	n	%
Tipos histológicos		
Liposarcoma	4	33,3
Sarcoma histiocítico	1	8,3
Sarcoma sinovial	1	8,3
Sarcoma sólido de alto grado	1	8,3
Teratocarcinoma		
Neuroblastoma	1	8,3
Schwannoma	1	8,3
Miofibroblástico	1	8,3
Inflamatorio	1	8,3
PEComa	1	8,3
Diferenciación celular		
Bajo grado	3	25
Alto Grado	1	8,3
No informa	8	66,6
Inmunoquímica	8	66,6

Ocho pacientes presentaron tumores malignos y, de estos, 7 fueron sarcomas.

Hubo complicaciones en 4 de casos: 2 infecciones del sitio quirúrgico, un serohematoma de la herida y una neumonía asociada a derrame pleural que prolongó la estadía hospitalaria para el tratamiento clínico. El resto de los datos posoperatorios pueden observarse en la tabla 2.

Los tipos anatomopatológicos se describen en la tabla 3. En las figuras 2 a 5 pueden observarse algunos de los casos más representativos.

■ FIGURA 2



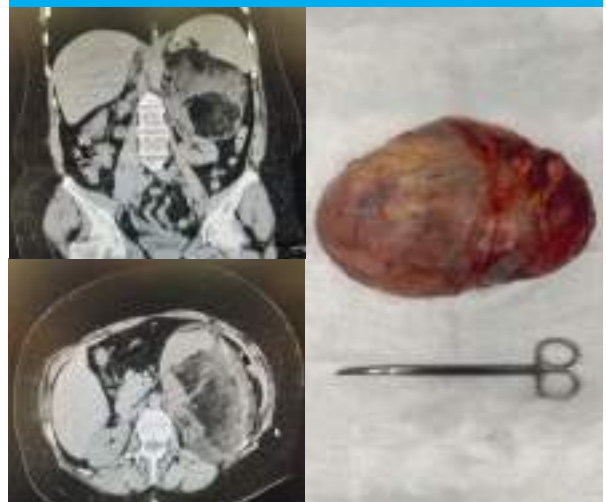
Resonancia magnética nuclear (RMN) donde se observa gran tumor miofibroblástico inflamatorio.

■ FIGURA 3



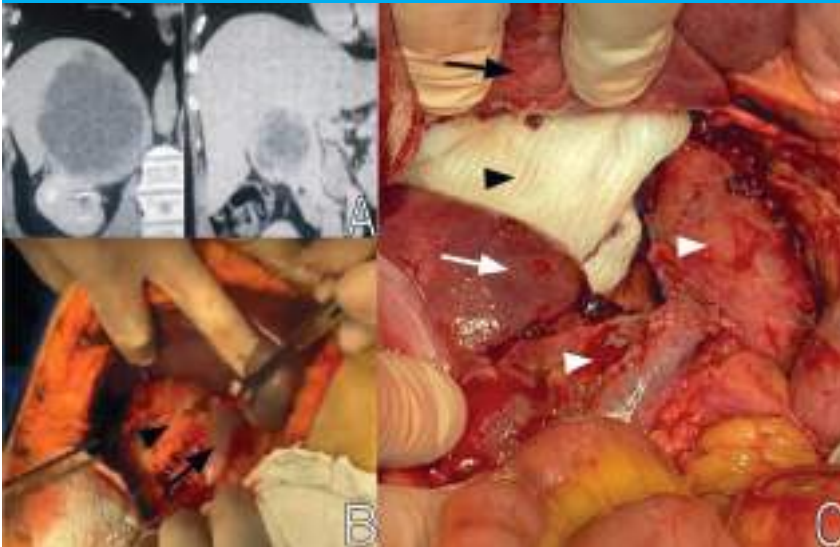
RMN donde puede observarse un liposarcoma.

■ FIGURA 4



Tomografía computarizada y pieza operatoria de liposarcoma.

■ FIGURA 5



A: imágenes de la tomografía. B: la vena cava empujada hacia adelante y desembocadura de la vena renal derecha (flecha negra), ambas empujadas hacia adelante por el tumor (cabeza de flecha negra). C: hígado hacia arriba (flecha negra) y una compresa ocupando el lugar donde se asentaba el tumor (cabeza de flecha negra), el riñón (flecha blanca), la vena renal derecha y la cava completamente liberadas (cabezas de flecha blanca).

## Discusión

Los tumores retroperitoneales comprenden un conjunto de tumores de tejidos blandos sumamente heterogéneos ubicados en el espacio retroperitoneal<sup>1,6-9</sup>. Por esta razón se han ensayado múltiples clasificaciones. A nuestro entender, aquella que toma los tumores de acuerdo con el tejido que los origina (mesenquimales, germinales, etc.), con la subdivisión de benignos y malignos, es la que mejor se acomoda a estas lesiones<sup>9</sup>. También y de una forma más amplia se utiliza la última clasificación de la Organización Mundial de la Salud (WHO) para tumores de partes blandas<sup>10</sup>. De los 12 casos que fueron intervenidos quirúrgicamente, ocho pacientes tenían tumores malignos, 7 de los cuales eran sarcomas; estos datos coinciden con la mayoría de los autores consultados, que también refieren que los liposarcomas son los más frecuentes<sup>1-3</sup>. Debido a la heterogeneidad de la serie aquí presentada, abordaremos algunos aspectos generales de importancia de los tumores malignos, y posteriormente haremos algunas consideraciones particulares de importancia con respecto a los tumores benignos.

La resección R0 en general se logró en el 8 de 11 casos potencialmente malignos (72,7% de los casos). Si bien en este trabajo hemos considerado todos los tumores retroperitoneales, estos resultados son coincidentes con los de aquellos trabajos que abordan tan solo sarcomas retroperitoneales.

La resección R0 es fundamental ya que se encuentra asociada fuertemente a un menor índice de recurrencias así como también a una prolongación de la supervivencia.

La resección multiorgánica, así como la resección vascular, en caso de ser necesaria para lograr una resección R0, debe realizarse, ya que mejora la super-

vivencia con mortalidades similares a aquellos que no implican resección de otros órganos y resección vascular, pero con una morbilidad superior<sup>1,3,4,11</sup>.

Otro dato interesante en este sentido es que, en alrededor de un 26% de los pacientes en los que no se sospecha invasión de órganos vecinos, existe de manera histológica, lo que implica que debemos insistir con una ampliación de márgenes para lograr resecciones completas<sup>12</sup>. Por su parte, Lee y cols., en referencia a los liposarcomas, sostienen que la supervivencia libre de enfermedad no ha mostrado diferencias entre resecciones R0 y R1<sup>13</sup>.

Algunos autores sugieren que, cuando compararon resecciones R2 contra manejo no operatorio, los pacientes reseccionados tuvieron mejores resultados que aquellos que no lo fueron<sup>14-16</sup>.

En este sentido es importante tener en cuenta, como expresamos anteriormente, que la sintomatología muchas veces suele estar relacionada con el efecto masa, y una resección R2 en pacientes con buen estado general y jóvenes podría estar indicada de manera paliativa para disminuir temporalmente sus síntomas.

Las complicaciones de nuestra serie fueron 4 (33,3%), cifra que —al igual que la mortalidad— se encuentra dentro de lo publicado por otros autores<sup>1,4</sup>. Es lógico esperar que aquellos pacientes en quienes el tumor tiene relaciones complejas, fundamentalmente con grandes vasos, o en los cuales la resección implica otros órganos, las complicaciones sean más altas. El estado nutricional del paciente y la sarcopenia favorecen también un aumento en la morbimortalidad, siendo este un punto para tener en cuenta a la hora de tomar la decisión de indicar la cirugía<sup>17</sup>.

La recurrencia en nuestra serie fue del 25%, pero algunos pacientes no han podido ser controlados por nosotros; una recurrencia de alrededor

del 35% ha sido informada por algunos autores<sup>18</sup>.

La supervivencia tuvo una mediana de 60 meses. Este dato en particular debe ser tomado cuidadosamente, ya que está basado en resultados de un grupo heterogéneo de tumores y algunos de ellos eran benignos. La supervivencia global informada en sarcomas retroperitoneales es similar a la comunicada por nosotros<sup>4</sup>.

La utilización de la radioterapia en tumores malignos está condicionada en primer lugar por las características espaciales que implica el retroperitoneo, aunque probablemente con técnicas como la 3D-CRT o el uso de la IMRT se podrían proteger los tejidos periféricos al sitio para irradiar<sup>19</sup>.

Diamantis y cols.<sup>19</sup> refieren que la radioterapia podría mejorar los resultados en algunos tipos especiales de lesiones. Por otro lado, el estudio STRASS<sup>20</sup> falló en demostrar algún tipo de mejoría en los resultados cuando se asocia la radioterapia. Otros autores refieren que la radioterapia tendría un papel dentro de la neoadyuvancia de sarcomas que son resecables<sup>21</sup>. La quimioterapia, por su parte, parece no tener resultados favorables desde ningún punto de vista en el tratamiento de los sarcomas retroperitoneales<sup>22,23</sup>. El tumor miofibroblástico inflamatorio retroperitoneal es extremadamente raro; su principal sitio de localización suele ser el pulmón y ocurre más frecuentemente en niños. Si bien es un tumor benigno, tiene potencialidad de convertirse en maligno y dar metástasis a distancia. La cirugía es el único tratamiento eficaz para erradicarlo; como en el resto de los tumores malignos, la quimioterapia y la radioterapia no han aportado diferencias comparadas con la cirugía<sup>24</sup>.

El schwannoma es, al igual que el anterior, un tumor benigno con potencial maligno que puede estar asociado a neurofibromatosis. También tiene la particularidad de alcanzar tamaños importantes y ejercer efecto masa. El tratamiento es siempre quirúrgico y la recurrencia es infrecuente<sup>25</sup>.

El neuroblastoma tiene su origen en las células pluripotenciales del sistema simpático, es frecuente en recién nacidos y es un tumor relativamente raro, pero el retroperitoneo suele ser el sitio de aparición más común. La cirugía incompleta de estas lesiones suele estar asociada a mal pronóstico, al igual que la edad de presentación<sup>26</sup>.

El PEComa (*perivascular epithelioid cell tumor*), anteriormente conocido como angiomiolipoma extrarenal, es un tumor muy poco frecuente, de características benignas y para el cual la cirugía de resección completa es el tratamiento definitivo<sup>8</sup>.

Finalmente, hemos utilizado la biopsia por punción histológica (*core*) en solo 5 de los pacientes resecados, y fue positiva tan solo en uno de ellos. Está demostrado que este tipo de biopsia tiene mejores resultados en los sarcomas, aunque la dificultad diagnóstica se incrementa cuando son bien diferenciados; algunos autores refieren un bajo índice de complicaciones menores que ronda el 10%<sup>27</sup>. Creemos que esta biopsia debe incorporarse dentro del arsenal diagnóstico.

En los tumores retroperitoneales malignos, la radioterapia y la quimioterapia tienen una utilidad dudosa. En sarcomas, en particular, la resección R0 que incluya, si fuera necesaria la resección visceral o vascular, el estado nutricional y el grado propuesto por la FFCCSG son las variables que más se asocian con el pronóstico.

Nuestro estudio presenta las siguientes limitaciones: su diseño retrospectivo, sobre pacientes atendidos durante varios años, un pequeño tamaño muestral y una composición heterogénea de los tipos histológicos, acorde con su baja frecuencia.

En conclusión, en la serie aquí presentada, la mayoría de los tumores retroperitoneales fueron malignos y principalmente sarcomas. La cirugía con resección R0 pudo ser obtenida en un alto porcentaje de los casos, con una tasa de complicaciones comparable a las comunicadas en otras series.

## ■ ENGLISH VERSION

### Introduction

Retroperitoneal tumors include a heterogeneous group of soft-tissue tumors originating from different cell lineages. These tumors can be benign or malignant, while some exhibit an intermediate behavior<sup>1</sup>. Most retroperitoneal tumors are malignant, with sarcomas being the most common type. Within this latter group, liposarcomas are the most common in most series<sup>1-3</sup>. Retroperitoneal tumors are generally large, and the clinical presentation associated with these lesions is often related to the mass effect. However, the incidence of distant metastasis is relatively low<sup>3</sup>. Overall, complete surgical resection offers the best outcomes

for these patients, as some tumors exhibit a locally aggressive behavior, resulting in a high percentage of local recurrences<sup>4</sup>. With a few exceptions, neoadjuvant and adjuvant chemotherapy and/or radiation therapy do not generally improve survival rates.

The aim of the present study was to evaluate the short-term outcomes of surgical resection of retroperitoneal tumors.

### Material and methods

We conducted an observational, retrospective and descriptive study. We retrieved data from the

medical records of patients referred for surgery between January 2005 and December 2024 due to retroperitoneal tumors. Participants were excluded from the study if they had not undergone surgery or for whom insufficient data was available. Tumors originating in retroperitoneal organs or lymphatic tissue—primarily lymphomas—were also excluded.

The data recorded included demographic, clinical, pathological, and operative variables. Complications were categorized using the Clavien-Dindo classification<sup>5</sup>. Survival was expressed in months and tracked until the date of death. For patients who were lost to follow-up, their survival status was verified in the Renaper database (Registro Nacional de Personas. Argentina.gov.ar). Length of hospital stay was measured in days.

Data were entered into a reference table for statistical analysis. Qualitative variables were expressed in range, frequency and percentage and quantitative variables were expressed in range, median and standard deviation (SD).

## Results

During the study period, 19 patients with retroperitoneal tumors were treated. Of these, 7 were excluded from the analysis: 4 were considered unresectable during preoperative assessment, 2 were found to be unresectable intraoperatively, and one patient did not have the minimum data required for inclusion in the study (Fig. 1).

Thus, 12 patients with retroperitoneal tumors who underwent curative intent surgery were included for the analysis. The demographic, clinical, and pathological variables are presented in Table 1.

Eight patients presented malignant

tumors, and 7 of them were sarcomas.

Complications occurred in 4 cases: 2 surgical site infections, 1 seroma near the surgical incision, and 1 case of pneumonia associated with pleural effusion, which prolonged the length of hospital stay for clinical management. Table 2 describes the postoperative findings.

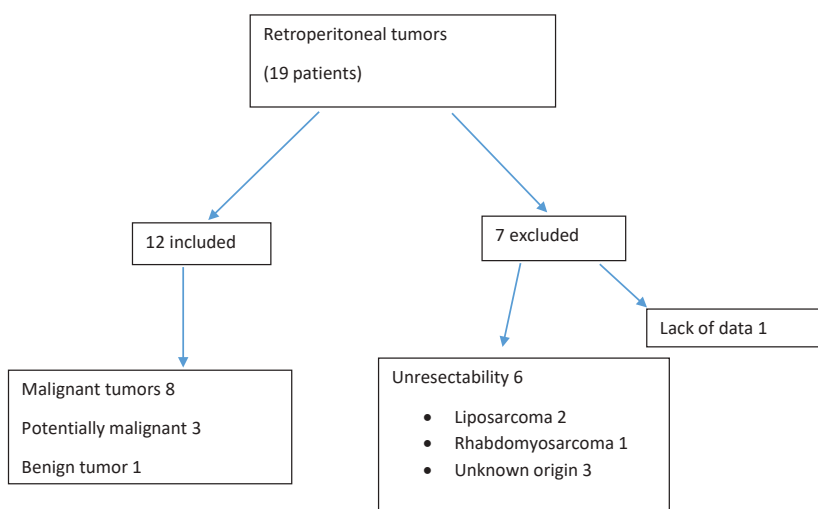
The pathological diagnoses are detailed in Table 3. The most representative cases of this series are shown in Figures 2 to 5.

■ TABLE 1

Demographic and preoperative clinical variables

Variable		
Male sex (n)	8	66.6%
Age (range) (Median - SD)	19-61	42 ± 14.96
Symptoms		
Abdominal pain (n)	12	100%
Weight loss (n)	8	66.6%
Nausea and vomiting (n)	7	58.3%
Distension (n)	8	66.6%
Palpable mass (n)	7	58.3%
Family history	7	58.3%
Laboratory tests		
Blood albumin level (range) (Median - SD)	2.9-3.4	3.25 ± 0.16
Hemoglobin (range) (Median - SD)	10-13	11.5 ± 1
Tumor markers (n)	5	41.6%
Performance Status		
0 (n)	9	75%
1 (n)	3	25%
Sarcopenia (n)	4	33.3%
Ancillary tests		
Ultrasound (n)	12	100%
Computed tomography scan (n)	12	100%
MRI (n)	10	83.3%
Colonoscopy (n)	3	25%
Upper gastrointestinal endoscopy (n)	6	50%
Needle biopsy (n)	5	41.6%
Size Maximum diameter (range)	5-30 cm	16.9 cm
Needle biopsy results	1 (+) 4 (-)	

■ FIGURE 1



Flowchart. Type of tumors and exclusion criteria

■ TABLE 2

Surgical and postoperative variables

Variable		
Resection		
R0 (n)	8	72.7%
R1 (n)	3	27.2%
Resection of more than one organ (n)	1	8.33%
Clavien-Dindo		
Grade 1 (n)	3	25%
Grade 2 (n)	1	8.3%
Length of hospital stay (days), (Median-SD)	3-9	6.5 (± 1.78)
Tumor recurrence (n)	3	25%
Adjuvant treatment (n)	3	25%
Follow-up in months (range), (Median-SD)	4-96	60 (±39)

■ TABLE 3

Pathological variables

Variable	n	%
Histological types		
Liposarcoma	4	33.3
Histiocytic sarcoma	1	8.3
Synovial sarcoma	1	8.3
High-grade solid sarcoma	1	8.3
Teratocarcinoma		
Neuroblastoma	1	8.3
Schwannoma	1	8.3
Myofibroblastic sarcoma	1	8.3
Inflammatory sarcoma	1	8.3
PEComa	1	8.3
Cell differentiation		
Low-grade	3	25
High-grade	1	8.3
Not reported	8	66.6
Immunohistochemistry	8	66.6

**Discussion**

Retroperitoneal tumors comprise a highly heterogeneous group of soft-tissue tumors located in the retroperitoneal space<sup>1,6-9</sup>. For this reason, multiple classification systems have been developed. To our understanding, the best classification for these lesions is one that categorizes tumors based on the tissue from which they originate (mesenchymal or germ cell, among others), with a further subdivision into benign and malignant tumors<sup>9</sup>. The most recent classification by the World Health Organization (WHO) for soft tissue tumors is also widely used<sup>10</sup>. Of the 12 cases that underwent surgery, 8 patients had malignant tumors, 7 of which were sarcomas. These findings are consistent with the reports of most authors, with liposarcomas being the most common tumors<sup>1-3</sup>. Due to the heterogeneity of the series presented here, we will first address some general aspects of malignant tumors and then discuss specific considerations regarding benign tumors.

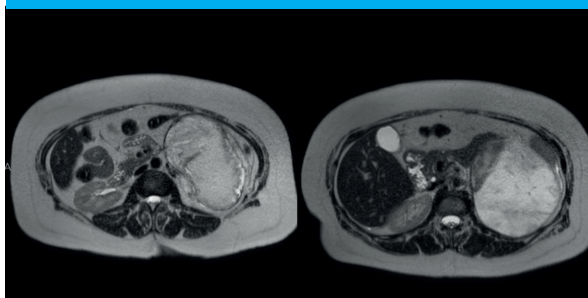
An R0 resection was achieved in 8 of 11 potentially malignant cases (72.7%). Although this study has examined all retroperitoneal tumors, these

results are consistent with those of papers that focus solely on retroperitoneal sarcomas.

An R0 resection is essential because it is strongly associated with a lower recurrence rate and improved survival.

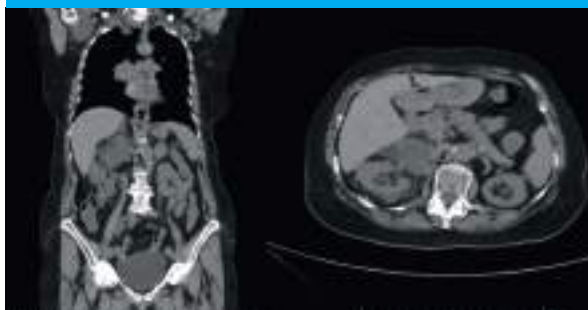
Multiorgan resection, as well as vascular resection, should be performed if necessary to achieve an R0 resection, as it improves survival with mortality rates similar to those of procedures that do not involve resection of other organs or vascular resection, but with higher morbidity rates<sup>1,3,4,11</sup>.

■ FIGURE 2



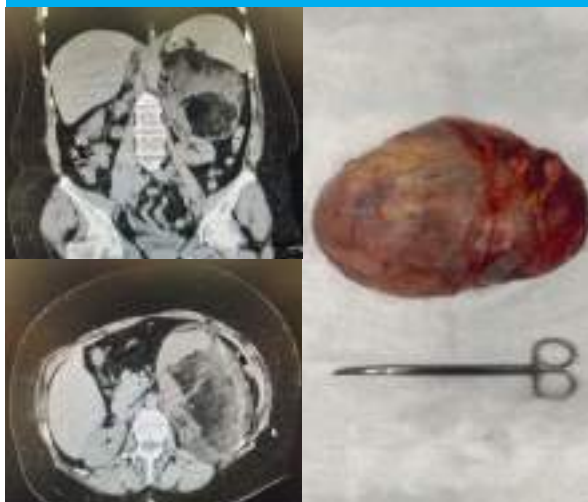
Magnetic resonance imaging (MRI) showing a large inflammatory myofibroblastic tumor

■ FIGURE 3



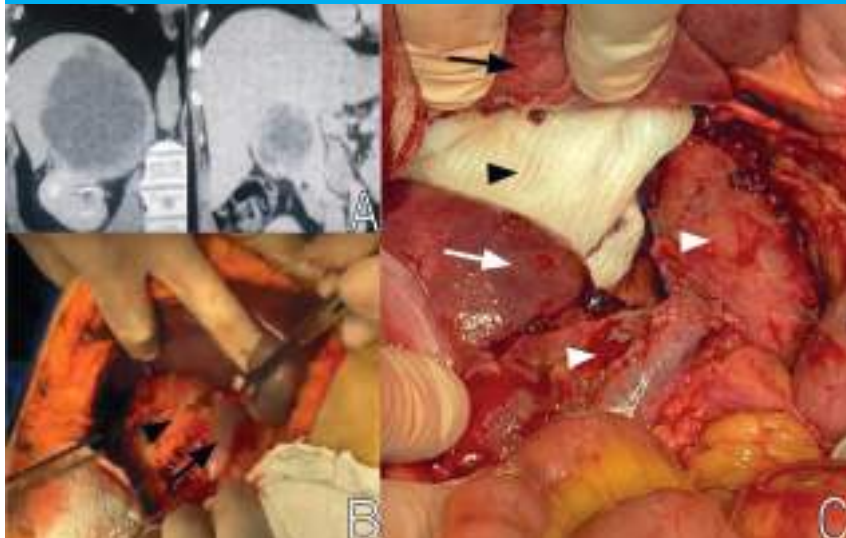
MRI showing a liposarcoma

■ FIGURE 4



Computed tomography and surgical specimen of liposarcoma

■ FIGURE 5



A. Computed tomography images. B. Inferior vena cava pushed forward and opening of the right renal vein (black arrow), both pushed forward by the tumor (black arrowhead). C. The liver is above (black arrow) and a gauze is occupying the site where the tumor was located (black arrowhead); the kidney (white arrow), right renal vein, and vena cava are completely released (white arrowheads).

A noteworthy observation in this context is that, in approximately 26% of patients where invasion of adjacent organs is not suspected, histopathologic invasion is present. This finding underscores the necessity for meticulous attention to ensure wider margins to facilitate complete resections<sup>12</sup>. Regarding liposarcomas, Lee et al. reported that disease-free survival did not differ between R0 and R1 resections<sup>13</sup>.

Some authors suggest that, when comparing R2 resections with non-surgical management, patients who underwent resection had better outcomes compared to those without resection<sup>14-16</sup>.

As previously mentioned, it is crucial to note that symptoms are frequently associated with the mass effect. Consequently, in young patients with good performance status, an R2 resection may be recommended for palliative purposes, thereby offering temporary relief from symptoms.

There were four complications in our series (33.3%), a figure that, like the mortality rate, falls within the range reported by other authors<sup>14</sup>. It is reasonable to expect that complications will be more common in patients with tumors that invade adjacent structures, especially major vessels, or in patients undergoing resections that involve other organs. Patients' nutritional conditions and sarcopenia should also be assessed when deciding to indicate surgery, as these factors contribute to greater morbidity and mortality<sup>17</sup>.

The recurrence rate in our series was 25%, but some patients were not followed up by us. Other authors have reported a recurrence rate of approximately 35%<sup>18</sup>.

Median survival was 60 months. It is essential to interpret this finding with caution, as it is based on results from a heterogeneous group of tumors, some of which were benign. The overall survival rate reported

for retroperitoneal sarcomas is similar to ours<sup>4</sup>.

The use of radiation therapy for malignant tumors is primarily determined by the anatomical characteristics of the retroperitoneum, although techniques such as 3D-CRT or IMRT could likely be used to protect the surrounding normal tissues from radiation<sup>19</sup>.

Diamantis et al.<sup>19</sup> reported that radiation therapy could improve outcomes for specific types of lesions. The STRASS study<sup>20</sup> failed to demonstrate any improvement in outcomes when combined with radiation therapy. Other authors stated that radiation therapy could play a role in the neoadjuvant treatment of resectable sarcomas<sup>21</sup>. Chemotherapy does not seem to have favorable outcomes from any perspective in the treatment of retroperitoneal sarcomas<sup>22,23</sup>. Retroperitoneal inflammatory myofibroblastic tumors is an extremely rare condition; they primarily occur in the lungs and are more common in children. Although these tumors are benign, they have the potential to become malignant and metastasize to distant sites. Surgery is the only effective treatment for eradicating the tumor. As with other malignant tumors, chemotherapy and radiation therapy have not shown any advantage over surgery<sup>24</sup>.

Schwannomas are also benign tumors with the potential to become malignant and may be associated with neurofibromatosis. They can reach large dimensions and produce mass effect. Surgery is always indicated and recurrences are rare<sup>25</sup>.

Neuroblastomas originate in pluripotent cells of the sympathetic nervous system. These relatively rare tumors occur in newborns and are most commonly found in the retroperitoneum. Incomplete surgical treatment of these lesions and the age at presentation are usually associated with a poor prognosis<sup>26</sup>.

PEComa (perivascular epithelioid cell tumor), formerly known as extrarenal angiomyolipoma, is a rare benign tumor. Complete surgical resection is the definitive treatment for this condition<sup>8</sup>.

Finally, we performed a core needle biopsy on only 5 of the patients who underwent resection, and it was positive in only one of them. It has been demonstrated that this type of biopsy yields superior results in cases of sarcoma, although the diagnosis may be complicated in well-differentiated sarcomas. Some authors report a low rate of minor complications, around 10%<sup>27</sup>. We believe that this biopsy should be included in the diagnostic workup.

The value of radiation therapy and chemotherapy in treating malignant retroperitoneal

tumors is uncertain. For sarcomas, R0 resections—including visceral or vascular resection if necessary—nutritional conditions, and the grade proposed by the FFCCSG are the variables most strongly associated with prognosis.

Our study is subject to the following limitations: it was retrospective, covering patients treated over several years; it had a small sample size; and it included a heterogeneous composition of histological types, consistent with their low prevalence.

In conclusion, most of the retroperitoneal tumors presented in this series were malignant, primarily sarcomas. R0 resections were achieved in a high percentage of cases, with a complication rate comparable to that reported in other series.

## Referencias bibliográficas /References

- Guo Q, Zhao J, Du X, Huang B. Survival outcomes of surgery for retroperitoneal Sarcomas. A systematic review and meta-analysis. *PLoS ONE*. 17(7):e0272044 <https://doi.org/10.1371/journal.pone.0272044>.
- Mack T, Purgina B. Updates in pathology for retroperitoneal soft tissue sarcomas. *Curr Oncol*. 2022;29:6400-6418. <https://doi.org/10.3390/curroncol29090504>.
- Santangelo A, Fernicola A, Santangelo D, Peluso G, Calogero A, Crocetto F, et al. Dark topics on giant retroperitoneal liposarcoma: a systematic review of 157 cases. *Cancers*. 2025; 17:740. <https://doi.org/10.3390/cancers17050740>.
- Hu H, Guo Q, Zhao J, Huang B, Du X. Aggressive surgical approach with vascular resection and reconstruction for retroperitoneal sarcoma: a systematic review. *BMC Surgery*. 2023; 23:275. <https://doi.org/10.1186/s12893-023-0218-1>.
- Clavien P, Barkun J, de Oliveira M, Vauthey J, Dindo D, Schulick R, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. 2009; 250:187-96 <https://doi.org/10.1097/SLA.0b013e3181b13ca2>.
- Ogose A, Kawashima H, Hatano H, Ariizumi T, Sasaki T, Tamagishi T, et al. The natural history of incidental retroperitoneal schwannomas. *PLoS ONE*. 2019;14(4):e0215336. <https://doi.org/10.1371/journal.pone.0215336>.
- Attili S, Chandra CR, Hemant DK, Bapsy P, Rama Rao C, Anupama G. Retroperitoneal Inflammatory Myofibroblastic tumor. *World J Surg Oncol*. 2005;3:66. <https://doi.org/10.1186/1477-7819-3-66>.
- Nari G, Brandan Recalde E, German G, Rojo S, Carreras M, Lasso Molina D. Neoplasia de células epitelioides perivasculares subtipo angiomiolipoma de localización retroperitoneal. *Oncología Clínica*. 2019;24:65-8.
- Apestegui CA. Tumores retroperitoneales. Relato oficial. 69.º Congreso Argentino de Cirugía. Rev Argent Cirug. 1998; N.º Extraordinario: 75-167.
- Choi Jh, Ro JY. The 2020 WHO classification of tumors of soft tissue: selected changes and new entities. *Adv Anat Pathol*. 2021; 28:44-58. <https://doi.org/PAP.0000000000000284>.
- Paik B, Ceo CJ, Tan J, Juan WKD, Soo KC, Ong C, et al. Asystemic review of margin status in retroperitoneal liposarcomas: does the R0 margin matter? *Front Oncol*. 2022;12:891710. <https://doi.org/10.3389/fonc.2022.891710>.
- Fairweather M, Wang J, JoV, Valdini E, Bertagnolli M, Rait C. Surgical management of primary retroperitoneal sarcomas: Rationale for selective organ resection. *Ann Surg Oncol*. 2018; 25:98-106 <https://doi.org/10.1245/s10434-017-6136-4>.
- Lee S, Goh B, Teo M, Chew M, Chow P, Wong K et al. Retroperitoneal liposarcomas: the experience of a tertiary Asian center. *World J Surg Oncol*. 2011;9:12. <https://doi.org/10.1186/1477-7819-9-12>.
- Shibata D, Lewis J, Leung D, Brennan M. Is there a role for incomplete resection in the management of retroperitoneal liposarcomas? *J Am Coll Surg*. 2001;193:373-9. [https://doi.org/10.1016/s1072-7515\(01\)01024-9](https://doi.org/10.1016/s1072-7515(01)01024-9).
- Abdelfatah E, Guzzetta A, Nagarajan N, Wolfgang C, Pawlik T, Choti M, et al. Long-term outcomes in treatment of retroperitoneal sarcomas: a 15 year single-institution evaluation of prognostic features. *J Surg Oncol*. 2016;114:56-64 <https://doi.org/10.1002/jso.24256>.
- Lehnert T, Cardona S, Hinz U, Willeke F, Mechttersheimer G, Treiber M, et al. Primary and locally recurrent retroperitoneal soft-tissue sarcoma: local control and survival. *Eur J Surg Oncol*. 2009; 35:986-993 <https://doi.org/10.1016/j.ejso.2008.11.003>.
- Nasirishargh A, Grova M, Bateni C, Judge S, Nuno M, Basmaci U, et al. Sarcopenia and frailty as predictors of surgical morbidity and oncologic outcomes in retroperitoneal sarcomas. *J Surg Oncol*. 2023;127:855-61 <https://doi.org/10.1002/jso.27199>.
- Deng H, Cao B, Cui H, Chen R, Li H, Zhao R, et al. Clinical analysis of 5-year survival and recurrence in giant retroperitoneal liposarcoma after surgery. *Clin Med J*. 2023, 136:373-5. <https://doi.org/10.1097/CM9.0000000000002405>.
- Diamantis A, Baloyiannis I, Magouliotis D, Tolia M, Symeonidis D, Bompou E, et al. Perioperative radiotherapy versus surgery alone for retroperitoneal sarcomas: asystemic review and meta-analysis. *Radiol Oncol*. 2020; .54:14-21. <https://doi.org/10.2478/raon.2020.0012>.
- Bonvalot S, Gronchi A, Le Péchoux C, Swallow C, Strauss D, Meeus P, et al. Preoperative radiotherapy plus surgery versus surgery alone for patients with primary retroperitoneal Sarcoma (EORTC-62092: STRASS): amulticentre, open-label, randomized, phase 3 trial. *Lancet Oncol*. 2020;21(10):1366-77. doi: 10.1016/S1470-2045(20)30446-0.
- Datta J, Ecker B, Neuwirth M, Geha R, Fraker D, Roses R, et al. Contemporary reappraisal of the efficacy of adjuvant chemotherapy in resected retroperitoneal sarcoma: Evidence from nationwide clinical oncology database and review of the literature. *Surg Oncol*. 2017; 26:117-24 <https://doi.org/10.1016/j.suronc.2017.01.008>.
- Zhou D, Connolly E, Mar J, Lazarakis S, Grimison P, Connor J, et al. A systematic review of the role of chemotherapy in retroperitoneal sarcoma by the Australian and New Zealand Sarcoma association/clinical practice guidelines working party. *Cancer Treatment Rev*. 2024; 122:102663 <https://doi.org/10.1016/j.ctrv.2023.102663>.
- Attili S, Chandra R, Hemant D, Bapsy P, Rama Rao C, Anupama G. Retroperitoneal inflammatory myofibroblastic tumor. *World J Surg Oncol*. 2005; 3:66. <https://doi.org/www.wjso.com/content/3/1/66>.
- Alventosa Mateu C, Castillo López G, Alcort Antequera C. Retroperitoneal Schwannoma. *Rev Esp Enferm Digest*. 2018; 110:597 <https://doi.org/10.17235/reed.2018.5569/2018>.
- Liu X, Zhao J, Zhang L, Huang J, Ma Y. The effect of age distribution on the prognosis of retroperitoneal neuroblastoma. *Medicine*. 2020;99:23(e20639). <https://dx.doi.org/10.1097/MD00000000000020639>.
- Nardi W, Nyla N, El Zein S, Tzanis D, Buhadiba T, Helfre S, et al. Diagnostic Accuracy and safety of percutaneous core needle biopsy of retroperitoneal tumors. *Eur J Surg Oncol*. 2024;50(1):107298. doi:10.1016/j.ejso.2023.107298.
- Coindre J, Terrier P, Bui N, Bonichon F, Collin, Le Doussal V F. et al. Prognostic factors in adult patients with locally controlled soft tissue sarcoma. A study of 546 patients from the French Federations of Cancer Centers Sarcoma Group. *J Clin Oncol*. 1996;14:869-77. <https://doi.org/10.1200/JCO.1996.14.3.869>.

# Sarcomas de partes blandas de tronco y extremidades. Resultados quirúrgicos y oncológicos en pacientes con enfermedad localizada

## Soft tissue sarcomas of the extremities and trunk. Surgical and oncological outcomes in patients with local disease

Sergio D. Quildrian<sup>1,2</sup>, Walter S. Nardi<sup>1,2</sup>, Victoria Scasso Rebzda<sup>1</sup>, Nicolás M. Rosasco<sup>1</sup>, Pablo Dezanzo<sup>3</sup>, Cintia Novas<sup>4</sup>, Luciana Bella Quero<sup>4</sup>, Carlos Silva<sup>4</sup>, Daniel E. Pirchi<sup>1</sup>

1. Servicio de Cirugía General.  
2. Unidad de Sarcomas y Melanoma.  
3. Servicio de Anatomía Patológica.  
4. Servicio de Oncología.

Hospital Británico de Buenos Aires. Argentina.

Los autores declaran no tener conflictos de interés.  
Conflicts of interest  
None declared.

Correspondencia  
Correspondence:  
Sergio D. Quildrian  
E-mail: squildrian@intramed.net

### RESUMEN

**Antecedentes:** los sarcomas de partes blandas (SPB) localizados de tronco superficial y extremidades (TyE) requieren tratamiento multimodal para alcanzar un control local mayor del 80%.

**Objetivo:** describir los resultados quirúrgicos y oncológicos alejados en pacientes con SPB localizados de TyE.

**Material y métodos:** estudio observacional descriptivo retrospectivo de pacientes con SPB de TyE operados entre 2008 y 2024. Se agruparon en: tumores primarios (G1), recaídas locales (G2), ampliaciones tras resección inadecuada (G3). Se registraron las características clínico-patológicas, tipo de resección y reconstrucción, morbilidad y tratamientos complementarios. Se estimó la sobrevida libre de recurrencia local (SLRL) y la sobrevida global (SG).

**Resultados:** se incluyeron 124 pacientes: G1=61 (49%), G2=37 (30%) y G3=26 (21%). La mediana de tamaño fue 8,5 cm (RIC 3,25-12). Se lograron resecciones completas en todos los casos, con conservación de miembro en el 98,8%. Se realizaron procedimientos reconstructivos en 36 casos (29%), más frecuentes en G2-G3. En G1, 22 pacientes recibieron radioterapia, 6 en forma neoadyuvante y 16 adyuvante. A cinco años, la SLRL, metástasis a distancia y SG fueron 82,8%, 63,5% y 79,3%, respectivamente. En G2-G3, las tasas fueron 61,6%, 75% y 85,3%. Con una mediana de seguimiento de 34 meses (RIC 15-67), la SG a 5 años de la serie fue 82,5% y no hubo diferencias entre los subgrupos analizados ( $p=0,64$ ).

**Conclusión:** se logró un control local superior al 80% con alto porcentaje de conservación en SPB de extremidades. Se necesitaron más procedimientos reconstructivos en pacientes G2-G3, pero esto no impactó negativamente en la SLRL ni en la SG.

■ **Palabras clave:** sarcomas de partes blandas, sarcomas de tronco superficial, sarcomas de extremidades, resección oncológica, cirugía con conservación de miembro, cirugía reconstructiva, radioterapia.

### ABSTRACT

**Background:** Local soft tissue sarcomas (STS) of the extremities and superficial trunk require multimodal treatment to achieve local control rates > 80%.

**Objective:** The aim of this study was to describe the long-term surgical and oncological outcomes of patients with STS of the extremities and trunk with local disease.

**Materials and methods:** We conducted a retrospective, descriptive and observational study of patients with STS of the extremities and trunk who underwent surgery between 2008 and 2024. Patients were divided into 3 groups: primary tumors (G1), local recurrences (G2) and margin widening following inadequate resections (G3). The clinical and pathological characteristics, type of resection and reconstruction, complications, and adjunctive treatment therapy were recorded. Local recurrence-free survival (LRFS) and overall survival (OS) were evaluated in each group.

**Results:** A total of 124 patients were included: 61 (49%) in group 1, 37 (30% in group 2, and 26 (21%) in group 3. Median tumor size was 8.5 cm (IQR 3.25-12). All the tumors were completely resected with limb preservation in 98.8%. Reconstructive procedures were performed in 36 cases (29%), with a higher incidence in groups G2 and G3. Radiation therapy was administered to 22 patients in G1, 6 as neoadjuvant therapy and 16 as adjuvant therapy. In G1, 5-year LRFS, distant metastases and OS were 82.8%, 63.5%, and 79.3%, respectively. In G2 and G3, the rates were 61.6%, 75%, and 85.3%. After a median follow-up of 34 months (IQR 15-67), the 5-year OS rate for the cohort was 82.5%, and there were no differences among the subgroups analyzed ( $p = 0.64$ ).

**Conclusion:** Local control rates exceeded 80%, with a high rate of limb preservation in cases of extremity STS. More reconstructive procedures were required in G2-G3 patients, but this did not have a negative impact on LRFS or OS.

■ **Keywords:** soft tissue sarcoma, trunk sarcoma, extremity sarcoma, oncologic resection, limb-sparing surgery, reconstructive surgery, radiotherapy.

Recibido | Received  
22-10-25  
Aceptado | Accepted  
28-01-26

ID ORCID: Sergio D. Quildrian, 0000-0001-5794-7084; Walter S. Nardi, 0000-0002-2106-4738; Victoria Scasso Rebzda, 0000-0001-8389-2144; Nicolás M. Rosasco, 0009-0005-7558-9598; Pablo Dezanzo, 0009-0000-4333-4383; Cintia Novas, 0009-0001-1680-2652; Luciana Bella Quero, 0009-0000-6710-876X; Daniel E. Pirchi, 0000-0002-7353-0470

## Introducción

Los sarcomas de partes blandas (SPB) constituyen un grupo heterogéneo de tumores con más de 100 subtipos histológicos. Corresponden al 1% de los tumores malignos del adulto y cerca del 15% de los tumores malignos de la niñez, con una prevalencia de 2-4/100 000 habitantes/año. En adultos, la edad promedio de aparición es de 50-60 años. Existen diferencias en la prevalencia de los distintos tipos histológicos según el grupo etario y la localización anatómica, lo que enfatiza la importancia de personalizar su manejo<sup>1</sup>.

El tratamiento de elección en pacientes con enfermedad localizada es la resección completa con márgenes libres luego de una evaluación en un comité multidisciplinario especializado en sarcomas<sup>2,3</sup>. Las guías de manejo de SPB resaltan la importancia de que estos pacientes sean tratados en centros especializados que cuentan con equipos multidisciplinarios con experiencia en sarcomas.

Los SPB de tronco superficial y extremidades (TyE) corresponden a cerca del 50% del total y comparten el enfoque terapéutico multimodal basado en cirugía, radioterapia y eventual terapia sistémica<sup>1</sup>. La combinación y secuenciación de estos tratamientos dependerá de las características tumorales (tipo histológico, grado de diferenciación), anatómicas y tipo de resección realizada. Esta conducta presenta tasas de control local superiores al 80% en los informes de centros de referencia<sup>4-7</sup>.

Debido a la baja incidencia de los SPB, la gran variedad histológica y las características anatómicas resulta difícil para el cirujano general adquirir experiencia en su manejo, lo que puede impactar negativamente en el pronóstico. Esto se ve reflejado en el escaso número de publicaciones en nuestro medio, las que a menudo incluyen localizaciones en las que el tratamiento no es homogéneo (retroperitoneo, cabeza y cuello) o pacientes con enfermedad sistémica<sup>9-19</sup>. Dado que no existe en nuestro país un sistema de derivación de pacientes con SPB, muchos son tratados inicialmente de manera subóptima, lo que dificulta el control local. De este modo, surge el interrogante mayor: ¿Es factible en pacientes con recaídas obtener sobrevida libre de enfermedad similar a la de aquellos tratados inicialmente en centros especializados?

El objetivo del presente estudio fue describir los resultados quirúrgicos y oncológicos alejados en pacientes con SPB localizados de TyE tratados quirúrgicamente con intención curativa, en una Unidad de Sarcomas.

## Material y métodos

Se llevó a cabo un estudio de diseño observacional descriptivo y retrospectivo sobre los pacientes operados por SPB de TyE en la Unidad de Sarcomas y

Melanoma entre 2008 y 2024, ya que estos requieren el mismo tipo de tratamiento multimodal.

Se incluyeron y agruparon los pacientes de acuerdo con la forma de presentación: tumores primarios (Grupo 1), tumores recidivados (Grupo 2) y pacientes con resecciones incompletas previas (Grupo 3). Se excluyeron los pacientes con sarcomas cabeza y cuello, retroperitoneo/pelvis, viscerales, tumores desmoides y dermatofibrosarcoma protuberans, así como aquellos pacientes con enfermedad sistémica. Todos los pacientes fueron discutidos en el Comité Multidisciplinario de Sarcomas semanal sin excepción.

Se registraron las características demográficas y del tumor, la localización, los tratamientos previos, el tipo de cirugía y de resección según la clasificación R de tumor residual (R0=sin tumor residual, R1=tumor residual microscópico en margen de resección y R2=tumor residual macroscópico), la necesidad de cirugía reconstructiva por grupo, la morbilidad y los tratamientos complementarios. Se evaluó la sobrevida libre de recurrencia local y la sobrevida global de cada grupo. Para el estudio comparativo se utilizó la prueba de Fisher y para la evaluación de la sobrevida método de Kaplan-Meier y Log-Rank.

## Resultados

Durante el período de análisis, sobre 175 pacientes operados por SPB, 124 (70,8%) pacientes lo fueron por SPB en TyE, y conforman la cohorte de estudio luego de aplicar los criterios de exclusión.

Del total, 83 (67%) sarcomas fueron de extremidades y 41 (33%) de tronco superficial. En 61 (49%) pacientes se realizó resección primaria (Grupo 1), 37 (30%) pacientes fueron operados por recaída local (Grupo 2) y en 26 (21%) se realizó ampliación de márgenes luego de resección inicial inadecuada sin haberse efectuado una biopsia preoperatoria (Whoops procedure) (Grupo 3).

El tipo histológico más frecuente fue el liposarcoma (38/124, 30,5%). La mediana de tamaño del tumor primario fue de 8,75 (RIC, 3,25- 12) centímetros, mientras que la información del grado tumoral se encontraba disponible para su análisis en 106 (85,5%) lesiones y 66% (70/106) fueron de alto grado (G2-3). La mediana de seguimiento para toda la serie fue de 34 (RIC, 15-67) meses. Las características demográficas y del tumor primario se pueden ver en la tabla 1 según la forma de presentación.

A los pacientes del Grupo 1 se les realizó resección macroscópica completa en bloque, y 53 tuvieron resecciones R0 y 8 resecciones R1. Se planificó una cirugía con la resección R1 acorde con el subtipo histológico de la biopsia preoperatoria y/o las relaciones anatómicas del tumor: así, en 3 pacientes por presentar liposarcoma bien diferenciado, 2 con disección periadventicial, 2 resecciones de periostio y un paciente con

disección perineural. En todos los sarcomas de extremidades se realizó cirugía conservadora de miembros (100% de conservación del miembro) (Tabla 2).

Se les administró radioterapia a 22 pacientes (36%): 6 en forma neoadyuvante y 16 adyuvante. Las razones por las que 39 (65,6%) pacientes no recibieron RT fueron: 10 sarcomas de bajo riesgo, 7 liposarcomas bien diferenciados, 4 por rápida progresión sistémica, 12 por ubicación anatómica, 2 por complicaciones loca-

les que requirieron amputación de miembro, 1 sarcoma radioinducido, 2 por edad avanzada y uno por decisión del paciente.

Con una mediana de seguimiento de 41 meses (RIC, 15-77), se registraron 7 recurrencias locales, 15 progresiones sistémicas y 4 pacientes presentaron ambas. En este grupo de pacientes, el control local se obtuvo en el 82% de los casos (50/61). La supervivencia libre de recurrencia local, de metástasis a distancia y global estimada a 5 años fue de 82,8% (IC 95%: 72,1-95,0), 63,5 (IC 95%: 50,7-79,4) y 79,3% (IC 95%: 68,4-92,1), respectivamente.

En los 37 pacientes del Grupo 2, que consultaron por recaída local, se pudo realizar resección macroscópica completa: 33 resecciones R0 y 4 resecciones R1. En todos los sarcomas de extremidades se realizó conservación de miembro.

En 26 pacientes del Grupo 3 se realizó ampliación de márgenes luego de exéresis inadecuada y se encontró enfermedad microscópica en 7 (27%).

Con una mediana de seguimiento de 31 meses (RIC, 15-67), se registraron 11 recurrencias locales, 7 progresiones sistémicas y 5 pacientes presentaron ambas. La supervivencia libre de recurrencia local, de metástasis a distancia y global estimada a 5 años fue de 61,6% (IC 95%: 45,9-82,6), 75% (IC 95%: 63,4-88,8) y 85,3% (IC 95%: 74,7-97,4), respectivamente.

Fue indicada quimioterapia preoperatoria en 7 pacientes (7,1%) de los 98 que fueron tratados por presencia de tumor (Grupos 1 y 2).

Asociados a la resección quirúrgica se debieron realizar 37 procedimientos reconstructivos en 36 pacientes (36/124, 29%) luego de la resección oncológica en bloque: 19 en extremidades y 17 en tronco. Los procedimientos más frecuentes fueron: 9 colgajos libres microquirúrgicos (8 colgajos anterolaterales de muslo y 1 colgajo dorsal libre), 3 colgajos pediculados (2 dorsales y 1 oblicuo mayor) y 7 colgajos locales fasciocutáneos. Al evaluar procedimientos reconstructivos complejos (colgajos pediculados y libres) según cada grupo, observamos mayor necesidad en pacientes de los Grupos 2 y 3 en comparación con el Grupo 1 (12% vs. 6,5%; p=0,49). De los sarcomas de extremidades (n = 84), en 9 (10,7%) pacientes (5 sarcomas primarios, 2 recidivados y 2 luego de resección inadecuada) se realizaron grandes resecciones que requirieron reconstrucciones complejas para evitar la amputación del miembro afectado: 6 pacientes presentaron exposición ósea y/o de elementos neurovasculares que se cubrieron con colgajos libres, mientras que 3 pacientes necesitaron reconstrucción vascular con bypass arterial. Dos pacientes presentaron pérdida del colgajo libre: uno de ellos requirió un nuevo colgajo libre y evitó la amputación del miembro y en otro paciente se realizó un procedimiento menor con conservación de miembro. A destacar, un paciente presentó en el posoperatorio

■ TABLA 1

VARIABLES DEMOGRÁFICAS Y CARACTERÍSTICAS TUMORALES

	Resección Primaria (Grupo 1) n = 61 (49%)	Resección Recurrencia (Grupo 2) n = 37 (30%)	Ampliación de Márgenes (Grupo 3) n = 26 (21%)
Sexo, n (%)			
Masculino	33 (54)	19 (51,5)	21 (80,7)
Femenino	28 (46)	18 (48,5)	5 (19,3)
Edad, mediana (RIC)	57,5 (46-70)	57 (39-69)	51 (41-73)
Tamaño del tumor primario en cm, mediana (RIC)	10 (5,8-14,2)	4 (2,5-8,5)	3,5 (2,3-6)
Localización, n (%)			
Tronco	18 (30)	14 (38)	9 (35)
Extremidad Superficial/Profundo, n (%)	43 (70)	23 (62)	17(65)
Superficial	20 (33)	24 (65)	9 (35)
Profundo	41 (67)	13 (35)	17 (65)
Histopatología, n (%)			
WDLPS	8 (13)	3 (8)	3 (11,5)
LPS-DD	10 (16,5)	4 (11)	1 (4)
LPS-M	8 (13)	2 (5,5)	1 (4)
LMS	9 (15)	7 (19)	7 (27)
UPS	3 (5)	9 (24)	4 (15,5)
MFS	4 (6,5)	0	5 (19)
TMVNP	4 (6,5)	0	0
TFS	3 (5)	0	0
Otros*	12 (19,5)	12 (32,8)	5 (19)
Grado histológico FNCLCC, n (%)			
1	22 (36)	8 (21,5)	6 (23)
2	10 (16,5)	7 (19)	10 (38,5)
3	22 (36)	15 (40,5)	6 (23)
No informados	7 (11,5)	7 (19)	4 (15,5)

WDLPS = Liposarcoma bien diferenciado; LPS-DD = Liposarcoma desdiferenciado; LPS-M = Liposarcoma mixoide; LMS = Leiomiomasarcoma; UPS = Sarcoma pleomorfo indiferenciado; MFS = Mixofibrosarcoma; TMVNP = Tumor maligno de la vaina periférica del nervio; TFS = Tumor fibroso solitario; FNCLCC = French Federation of Cancer Centers Sarcoma Group; RT = Radioterapia.

\*Otros = sarcoma fibromixoide (3), sarcoma sinovial (3), angiosarcoma (2), hemangiosarcoma (2), rhabdomiomasarcoma (1), sarcoma epitelioides (1), sarcoma miofibroblástico (1), sarcoma fibroblástico mixoinflamatorio (1), dermatofibrosarcoma (1), sarcoma de células claras (1), sarcomas inclassificados (2).

■ TABLA 2

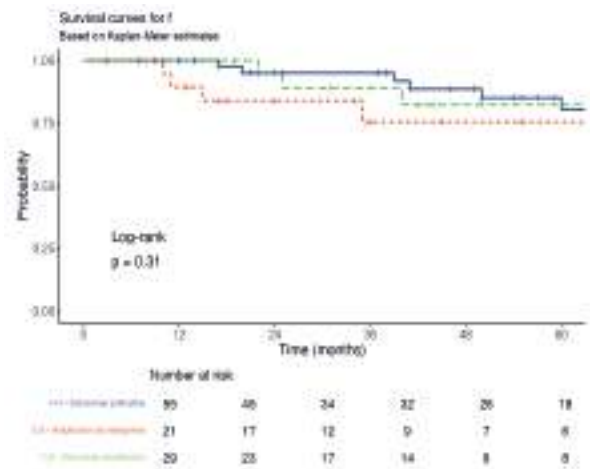
Tratamiento y morbilidad asociada			
	Resección Primaria (Grupo 1) n = 61	Resección Recurrencia (Grupo 2) n = 37	Ampliación de Márgenes (Grupo 3) n = 26
Márgenes quirúrgicos, n (%)			
R0	53 (87)	4 (11)	2 (7,6)
R1	8 (13)	8 (21,5)	9 (35)
R2	0	11 (9,5)	5 (19)
No informados	0	14 (58)	10 (38,5)
Reconstrucción, n (%)			
Cierre primario	44 (72)	25 (67,5)	19
Injerto de piel	3 (5)	2 (5,5)	0
Colgajo local de avance	4 (6,5)	2 (5,5)	3
Colgajo pediculado	1 (1,5)	2 (5,5)	0
Colgajo libre microquirúrgico	3 (5)	3 (8)	3
Malla	3 (5)	3 (8)	1
Vascular	3 (5)	0	0
Radioterapia perioperatoria, n (%)			
Neoadyuvante	6 (10)	0	0
Adyuvante	16 (26)	13 (35)	5 (19)
Complicaciones (Clavien-Dindo), n (%)			
Total	14 (23)	9 (24)	3 (11,5)
IIIa	0	2 (5,5)	0
IIIb	7 (11,5)	2 (5,5)	2 (7,6)
IV	0	1 (2,7)	0
Seguimiento (meses), mediana (IQR)	41 (16,5-76)	31 (13-69,5)	31,5 (13-57)

sangrado masivo por diabrosis de bypass arterial a los 20 días y requirió una amputación de urgencia. La tasa global de conservación de miembro primaria fue de 98,8% (82/83). Durante el seguimiento se debieron realizar 3 amputaciones con criterio curativo en pacientes con recurrencias locales no pasibles de cirugía conservadora (tasa de conservación de miembro secundaria de 97,5% (80/82).

En total se registraron 26 complicaciones (21%) de las cuales 14 (14/124, 11,3%) correspondieron a complicaciones mayores (Clavien-Dindo  $\geq 3$ ): infección de sitio quirúrgico que requirió toilette y antibioticoterapia (n = 10), hematoma (n = 2) y pérdida de colgajo libre (n = 2). No se registró mortalidad operatoria en esta serie (véase Tabla 2).

La sobrevida libre de recurrencia local para los Grupos 1, 2 y 3 no presentó una diferencia estadísticamente significativa ( $p = 0,31$ ) (Fig. 1). La sobrevida global (SG) de la serie a 5 años fue de 82,5% (IC 95%: 74-91,9). La SG a 5 años para los Grupos 1, 2 y 3 fue de 79,3%, 79,6% y 94,7%;  $p = 0,64$ ) (Figs. 2 y 3).

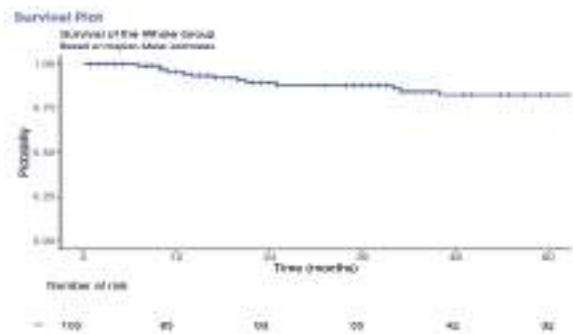
■ FIGURA 1



Intervalo libre de recurrencia local: grupo 1 azul, grupo 2 verde, grupo 3 rojo.

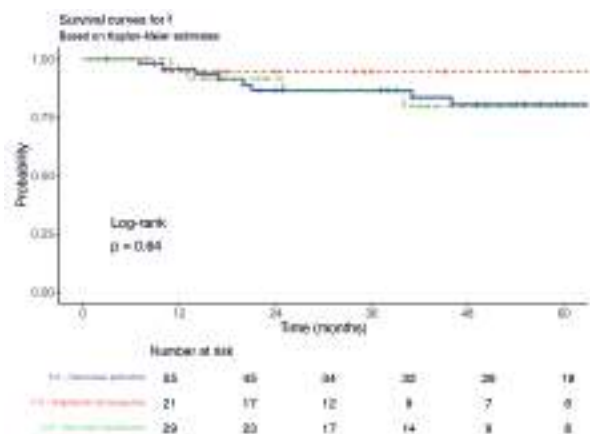
■ FIGURA 2

Site	Number at Risk	Number of Events	Survival	95% Confidence Interval	Lower	Upper
12	83	4	95,2%	91,8%	98,6%	
24	66	3	95,6%	91,1%	99,1%	
48	58	3	96,6%	94,3%	98,9%	



Supervivencia global de toda la serie.

■ FIGURA 3



Supervivencia global por grupo: grupo 1 azul, grupo 2 verde, grupo 3 rojo.

## Discusión

En pacientes con SPB localizados de tronco y extremidades, el tratamiento multimodal con eje en la resección completa con márgenes negativos continúa siendo el tratamiento estándar y el único tratamiento potencialmente curativo<sup>2,3,20,21</sup>. Es de destacar que se observaron mejores resultados oncológicos en centros de referencia en comparación con centros no especializados y esto se debe en gran parte a una mayor adherencia a las guías de práctica clínica (estudios de imágenes, biopsia preoperatoria, diferentes técnicas de anatomía patológica) y una evaluación multidisciplinaria inicial<sup>22</sup>. En nuestra serie y en línea con este concepto, todos los pacientes fueron evaluados en un comité multidisciplinario y con biopsia confirmatoria previa a la conducta quirúrgica.

Históricamente, el tratamiento de los pacientes con SPB de extremidades consistía en la amputación en casi el 50% de los casos, con un aceptable control local (5% de recurrencias locales)<sup>23,24</sup>. No obstante, el pronóstico estaba dado por la progresión sistémica de la enfermedad. Durante la década del 70 se produjo un cambio trascendental en el manejo de los SPB, a partir de estudios prospectivos aleatorizados que establecieron las bases del tratamiento actual con cirugía conservadora asociada a radioterapia<sup>25-27</sup>. Aunque el control local es menor que el obtenido con la amputación, la sobrevida es similar y con mejor calidad de vida al conservar un miembro funcional. En el caso de los sarcomas de tronco superficial, su tratamiento se asimila al de los SPB de extremidades con el mismo esquema multimodal de tratamiento. En nuestra serie, la tasa de conservación de miembro fue mayor del 98%, en línea con resultados actuales de centros de referencia<sup>28</sup>.

Danieli y cols.<sup>4</sup> informan los resultados a lo largo de 30 años con este tratamiento multimodal observando una mayor indicación tanto de quimioterapia (de 18 a 29%) como de radioterapia preoperatoria (de 1,6 a 26%), mientras que el uso concurrente de quimioterapia y radioterapia varió de 0,8 a 14,3%. En la presente serie, luego de la evaluación multidisciplinaria, el 36% de los pacientes con tumores primarios (Grupo 1) recibieron radioterapia (preoperatoria o posoperatoria). En el caso de los pacientes con recurrencia o resecciones incompletas (Grupos 2 y 3), la decisión es mucho más compleja, ya que existen condicionantes (cirugía y/o radioterapia y/o tratamiento sistémico previos) que pueden incidir en la indicación de tratamientos complementarios a la cirugía. Nosotros indicamos quimioterapia en el 7% de los pacientes que consultaron por tumor primario o recurrente.

El trabajo de O'Sullivan y cols.<sup>29</sup> determinó que la radioterapia tanto preoperatoria como posoperatoria tiene igual efecto en el control local de la enfermedad, con más complicaciones tempranas (dehiscencia de he-

rida, infección) durante el posoperatorio inmediato en pacientes que recibieron radioterapia preoperatoria y más complicaciones tardías (linfedema, rigidez) en los pacientes que recibieron radioterapia en el posoperatorio. La tendencia actual parece priorizar el tratamiento preoperatorio, sobre todo en aquellos pacientes en los que se esperan márgenes cercanos en zonas de contacto del tumor con estructuras no invadidas directamente (paquete vasculonervioso, hueso, etc.) o cuando –por la magnitud de la resección– se espera una mayor morbilidad de herida que retrase o impida la indicación de radioterapia adyuvante<sup>2,30</sup>.

El trabajo de Danieli y cols.<sup>4</sup> informa 17,9% de recurrencia local, 72,2% de metástasis a distancia y ambas en el 9,3% como primer evento antes de la muerte asociada al sarcoma. En nuestra serie observamos 14,5% de recurrencias locales, 17,7% de recurrencias sistémicas y 7,2% de recurrencias locales y sistémicas asociadas. El menor porcentaje de recurrencias sistémicas observadas podría estar relacionado con el menor período de tiempo de nuestra serie.

Ante la falta de estudios prospectivos aleatorizados que evalúen el margen quirúrgico óptimo, las guías de consenso recomiendan realizar resecciones con margen negativo, y se acepta una resección marginal en cercanías de estructuras vitales (neurovasculares y óseas) en ausencia de invasión directa para evitar una mayor morbilidad. Se sugiere un margen a 1-2 cm de la pseudocápsula tumoral, aunque valores menores en el material fijado se consideran apropiados<sup>31</sup>. Para lograr esto, a menudo se deben realizar grandes resecciones con exposición de estructuras neurovasculares u óseas. Un aspecto para tener en cuenta en estos pacientes es la posibilidad de necesitar algún tipo de reconstrucción.

En la presente serie, casi un tercio de los pacientes requirieron un procedimiento reconstructivo. Si bien se utilizaron injertos y colgajos locales en algunos casos, en el 7% se debieron realizar colgajos libres para lograr una cobertura satisfactoria, y los colgajos fasciocutáneos fueron los más frecuentemente utilizados. La decisión de optar por un procedimiento u otro depende de factores locales, del paciente y del tumor. En el caso de grandes resecciones por tumores distales de las extremidades o de algunos sectores de la pared torácica, la única opción quirúrgica reconstructiva son los colgajos libres, por lo que una evaluación preoperatoria correcta es primordial para asegurar una cirugía satisfactoria<sup>17</sup>.

Otro ejemplo de manejo quirúrgico multidisciplinario es la evaluación del compromiso vascular. En nuestra serie, 3 pacientes requirieron resección vascular con reemplazo. En la mayoría de los casos se puede realizar una disección por el plano adventicial si no hay compromiso vascular directo. Este tipo de disección también se utiliza con estructuras nerviosas y hueso cuando no hay invasión directa. Esto lleva al concep-

to de “margen positivo planeado”, que hace referencia a las zonas donde no se logran grandes márgenes por cercanía de estructuras vitales no invadidas. Gerrand y cols.<sup>30</sup> demostraron que este tipo de resección presenta un buen control local, cuando se asocia a radioterapia, y evita una morbilidad mayor cuando se planea desde el preoperatorio.

En esta serie se observó una sobrevida a 5 años de los pacientes tratados por tumor primario cercana al 80%, similar a lo informado en centros de referencia. A su vez, la sobrevida de los pacientes tratados fuera de nuestro centro se alineó con la obtenida en pacientes tratados inicialmente en nuestra Unidad. No obstante, estos pacientes requirieron procedimientos más complejos para lograr una cirugía óptima. Por ello, es

necesaria una rápida derivación de ellos a centros especializados.

Si bien nuestra serie presenta las limitaciones propias de todo estudio retrospectivo, consideramos que tiene como fortaleza el hecho de ser una serie de pacientes tratados en un mismo centro y por una Unidad Especializada en sarcomas. Creemos que nuestros resultados se deben al abordaje multidisciplinario desde el inicio y la adherencia a guías de manejo. Resaltamos la importancia de contar con un equipo especializado en sarcomas y también con la posibilidad de interactuar con otras especialidades quirúrgicas (cirugía reconstructiva y vascular) para asegurar una resección oncológica y maximizar la posibilidad de conservación del miembro en tumores de extremidades.

## ■ ENGLISH VERSION

### Introduction

Soft tissue sarcomas (STS) constitute a heterogeneous group of tumors with more than 100 histological subtypes. They account for 1% of malignant tumors in adults and about 15% of malignant tumors in children, with a prevalence of 2–4 per 100,000 people-years. In adults, STS occur at an average age of 50–60 years. There are differences in the prevalence of the various histological types depending on the age group and anatomical location, which underscores the importance of tailored management<sup>1</sup>.

Complete en bloc resection with clear margins is the treatment of choice after discussing the case in a multidisciplinary tumor board specializing in sarcomas<sup>2,3</sup>. The guidelines for the management of STS emphasize the importance of treating these patients at specialized centers with multidisciplinary teams that have expertise in sarcomas.

Soft tissue sarcomas of the extremities and superficial trunk account for about 50% of all cases and are treated using a multimodal approach that combines surgery, radiation therapy, and, if necessary, systemic therapy<sup>1</sup>. The combination and sequencing of these treatments will depend on the tumor characteristics (histological type, degree of differentiation), anatomic factors, and the type of resection performed. This course of action has achieved local control rates exceeding 80% in reports from referral centers<sup>4-7</sup>.

Given the low incidence of STS, their histological diversity, and anatomical characteristics, general surgeons often find it difficult to gain experience in managing these tumors, which can negatively impact the prognosis. This is reflected in the limited number of publications in our setting, which often involve sites where treatment is not uniform (such as the retroperitoneum and head and neck) or patients with systemic disease<sup>9-19</sup>. Since our country lacks a referral

system for patients with STS, many of them receive suboptimal initial treatment, which hinders local control. This raises the key question: Is it feasible for patients with recurrent disease to achieve disease-free survival comparable to that of patients initially treated at specialized centers?

The aim of this study was to describe the long-term surgical and oncological outcomes of patients with localized STS who underwent curative intent surgery at a Sarcoma Unit.

### Material and methods

We conducted a descriptive, retrospective and observational study of patients undergoing surgery for STS of the trunk and extremity in the Sarcoma and Melanoma Unit between 2008 and 2024, as these patients require the same type of multimodal treatment.

Patients were included and divided into groups based on their presentation: primary tumors (group 1), recurrent tumors (group 2), and patients with prior incomplete resection (group 3). Patients with head and neck sarcomas, retroperitoneal/pelvic sarcomas, visceral sarcomas, desmoid tumors, dermatofibrosarcoma protuberans and systemic disease were excluded. The weekly Multidisciplinary Sarcoma Committee reviewed all patients without exception.

The following variables were recorded: demographic data, tumor characteristics and location, prior treatments, type of surgery, and resection type base on the R classification of residual tumor (R0 = no residual tumor, R1 = microscopic residual tumor at the resection margin, and R2 = macroscopic residual tumor). The need for reconstructive surgery by group, morbidity, and adjunctive therapy was also recorded.

Local recurrence-free survival and overall survival were evaluated in each group. Fisher’s exact test was used for comparative analysis, and the Kaplan-Meier method and the log-rank test were used to assess survival.

**Results**

During the study period, of the 175 patients who underwent surgery for STS, 124 (70.8%) corresponded to STS of the extremities and trunk and constitute the study cohort after applying the exclusion criteria.

Eighty-three (67%) were located in the extremities and 41 (33%) in the superficial trunk. Primary resection was performed in 61 (49%) patients (group 1), 37 (30%) patients underwent surgery for local relapse (group 2), and 26 (21%) underwent margin widening following an inadequate initial resection without a preoperative biopsy (Whoops procedure) (group 3).

Liposarcoma was the most common histological type (38/124, 30.5%). The median size of the primary tumor was 8.75 cm (IQR, 3.25–12). Data on tumor grade were available for 106 (85.5%) lesions, 66% (70/106) of which were high-grade tumors (G2–3). Median follow-up for the entire series was 34 months (IQR, 15-67). The demographic data and characteristics of the primary tumor according to the type of presentation are shown in Table 1.

Patients in group 1 underwent complete en bloc resection; R0 resections were performed in 53 patients and R1 in 8 patients. Surgery was planned with R1 resection based on the histological subtype of the preoperative biopsy and/or the anatomical relationships of the tumor. Thus, in 3 patients with well-differentiated liposarcomas, 2 underwent periadventitial dissection, 2 underwent periosteal resection, and 1 underwent perineural dissection. Limb-sparing surgery was performed in all cases of extremity sarcoma (limb preservation in 100% of cases) (Table 2).

Radiation therapy was administered to 22 patients (36%), 6 as neoadjuvant therapy and 16 as adjuvant therapy. The reasons for not recommending radiation therapy to 39 patients (65.6%) were as follows: 10 patients had low-risk sarcomas; 7 had well-differentiated liposarcomas; 4 cases developed rapid systemic progression; the anatomical location of the tumor was an issue for 12 patients; 2 cases involved local complications requiring limb amputation; 1 case involved a radiation-induced sarcoma; 2 patients were elderly; and one patient declined treatment.

After a median follow-up of 41 months (IQR, 15–77), 7 patients experienced local recurrence, 15 developed systemic disease, and both events occurred in 4 patients. In this group of patients, local control was achieved in 82% of cases (50/61). The estimated 5-year rates of local recurrence-free survival, distant metastasis, and overall survival were 82.8% (95% CI,

72.1–95); 63.5 (95% CI, 50.7–79.4); and 79.3% (95% CI, 68.4–92.1), respectively.

The 37 patients in group 2 who presented with local recurrence underwent complete macroscopic resection; R0 resections were achieved in 33 cases and R1 resections in 4. Limb-sparing surgery was performed in all cases of extremity sarcomas.

In 26 patients in group 3, margin widening was performed following inadequate resection, and 7 patients presented microscopic disease (27%).

After a median follow-up of 31 months (IQR, 15–67), 11 patients experienced local recurrence, 7 developed systemic disease, while both events

**TABLE 1**

Descriptive variables and tumor characteristics

	Primary resection (Group 1) n = 61 (49%)	Recurrence resection (Group 2) n = 37 (30%)	Margin widening (Group 3) n = 26 (21%)
Sex, n (%)			
Male	33 (54)	19 (51.5)	21 (80.7)
Female	28 (46)	18 (48.5)	5 (19.3)
Age, median (IQR)	57.5 (46-70)	57 (39-69)	51 (41-73)
Primary tumor size in cm, median (IQR)	10 (5.8-14.2)	4 (2.5-8.5)	3.5 (2.3-6)
Location, n (%)			
Trunk	18 (30)	14 (38)	9 (35)
Extremity	43 (70)	23 (62)	17(65)
Superficial/Deep, n (%)			
Superficial	20 (33)	24 (65)	9 (35)
Deep	41 (67)	13 (35)	17 (65)
Histopathology, n (%)			
WDLPS	8 (13)	3 (8)	3 (11.5)
DDLPS	10 (16.5)	4 (11)	1 (4)
MLPS	8 (13)	2 (5.5)	1 (4)
LMS	9 (15)	7 (19)	7 (27)
UPS	3 (5)	9 (24)	4 (15.5)
MFS	4 (6.5)	0	5 (19)
MPNST	4 (6.5)	0	0
SFT	3 (5)	0	0
Other*	12 (19.5)	12 (32.8)	5 (19)
FNCLCC histological grade, n (%)			
1	22 (36)	8 (21.5)	6 (23)
2	10 (16.5)	7 (19)	10 (38.5)
3	22 (36)	15 (40.5)	6 (23)
Not reported	7 (11.5)	7 (19)	4 (15.5)

WDLPS = well-differentiated liposarcoma; DDLPS = dedifferentiated liposarcoma; MLPS = myxoid liposarcoma; LMS = leiomyosarcoma; UPS = undifferentiated pleomorphic sarcoma; MFS = myxofibrosarcoma; MPNST = malignant peripheral nerve sheath tumor; SFT = solitary fibrous tumor; FNCLCC = Fédération Nationale des Centres de Lutte Contre le Cancer; RT = radiation therapy.

\*Other = fibromyxoid sarcoma (3), synovial sarcoma (3), angiosarcoma (2), hemangiosarcoma (2), rhabdomyosarcoma (1), epithelioid sarcoma (1), myofibroblastic sarcoma (1), myofibroblastic sarcoma (1), dermatofibrosarcoma (1), clear cell sarcoma (1), unclassified sarcomas (2).

■ TABLE 2

Treatment and associated mortality			
	Primary resection (Group 1) n = 61	Recurrence resection (Group 2) n = 37	Margin widening (Group 3) n = 26
Surgical margins, n (%)			
R0	53 (87)	4 (11)	2 (7.6)
R1	8 (13)	8 (21.5)	9 (35)
R2	0	11 (9.5)	5 (19)
Not reported	0	14 (58)	10 (38.5)
Reconstruction, n (%)			
Primary closure	44 (72)	25 (67.5)	19
Skin graft	3 (5)	2 (5.5)	0
Advancement flap	4 (6.5)	2 (5.5)	3
Pedicled flap	1 (1.5)	2 (5.5)	0
Microsurgical free flap	3 (5)	3 (8)	3
Mesh	3 (5)	3 (8)	1
Vascular	3 (5)	0	0
Perioperative radiation therapy, n (%)			
Neoadjuvant	6 (10)	0	0
Adjuvant	16 (26)	13 (35)	5 (19)
Clavien-Dindo complications, n (%)			
Total	14 (23)	9 (24)	3 (11.5)
3a	0	2 (5.5)	0
3b	7 (11.5)	2 (5.5)	2 (7.6)
4	0	1 (2.7)	0
Follow-up (months), median (IQR)	41 (16.5-76)	31 (13-69.5)	31.5 (13-57)

occurred in 5 patients. The estimated 5-year survival rates free of local recurrence, distant metastasis, and overall survival were 61.6% (95% CI: 45.9–82.6), 75% (95% CI: 63.4–88.8), and 85.3% (95% CI: 74.7–97.4), respectively.

Preoperative chemotherapy was indicated for 7 patients (7.1%) of the 98 who were treated due to tumor presence (Groups 1 and 2).

Along with surgical resection, 37 reconstructive procedures were performed on 36 patients (36/124, 29%) following en bloc oncological resection: 19 on the extremities and 17 on the trunk. The most common surgical procedure involved 9 microsurgical free flaps (8 anterolateral thigh flaps and 1 dorsalis free flap), three pedicled flaps (2 dorsalis flaps and 1 external oblique flap), and 7 local fasciocutaneous flaps. When evaluating complex reconstructive procedures (pedicled and free flaps) by group, patients in groups 2 and 3 were found to have a greater need for such procedures compared to group 1 (12% vs. 6.5%;  $p = 0.49$ ). Of the 84 cases of extremity sarcomas, 9 patients (10.7%) (5 with primary sarcomas, 2 with recurrent sarcomas, and 2 cases following inadequate resection)

underwent extensive resections that required complex reconstructions to avoid amputation of the affected limb. Six patients with bone and/or neurovascular exposure were reconstructed with free flaps, while 3 patients required vascular reconstruction with arterial bypass. Two patients experienced free flap failure: one required a new free flap and avoided limb amputation, while the other underwent a minor limb-preserving procedure. Notably, one patient experienced massive postoperative bleeding due to erosions of the arterial bypass 20 days following surgery and required an emergency amputation. The overall primary limb preservation rate was 98.8% (82/83). During follow-up, three curative amputations were performed in patients with local recurrences that were not eligible for conservative surgery. The secondary limb preservation rate was 97.5% (80/82).

A total of 26 complications (21%) were recorded, of which 14 (14/124, 11.3%) were major complications (Clavien-Dindo grade  $\geq 3$ ): surgical site infection requiring lavage and antibiotic therapy ( $n = 10$ ), hematoma ( $n = 2$ ), and free flap failure ( $n = 2$ ). There were no deaths in this series (Table 2).

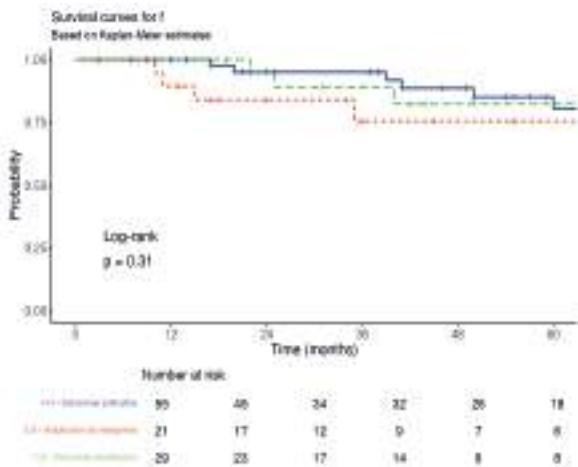
There were no statistically significant differences in local recurrence-free survival for groups 1, 2, and 3 ( $p = 0.31$ ) (Fig. 1). The 5-year overall survival (OS) rate for the entire cohort was 82.5% (95% CI: 74–91.9): 79.3% for group 1, 79.6% for group 2 and 94.7% for group 3;  $p = 0.64$ ) (Figures 2 and 3)

## Discussion

In patients with STS of the extremities and local trunk, multimodal treatment focused on complete resection with clear margins remains the standard of care and the only potentially curative treatment.<sup>2,3,20,21</sup> It is worth noting that better oncological outcomes were observed in referral centers compared to non-specialized centers, largely due to greater adherence to clinical practice guidelines (imaging tests, preoperative biopsy, expert pathology review) and an initial multidisciplinary evaluation<sup>22</sup>. In our series, and in line with this approach, all patients were evaluated by a multidisciplinary committee and underwent a confirmatory biopsy prior to surgery.

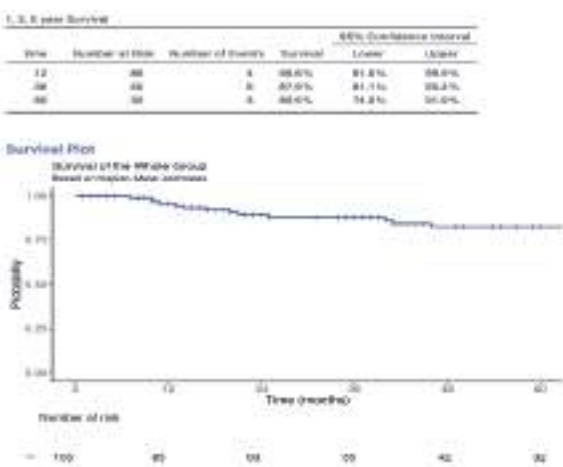
Historically, management of patients with STS of the extremities involved amputation in nearly 50% of cases, with acceptable local control (5% of local recurrence)<sup>23,24</sup>. However, the prognosis was determined by the systemic progression of the disease. During the 1970s, a significant shift occurred in the management of STS following prospective randomized trials that laid the foundation for the current treatment approach combining conservative surgery with radiation therapy<sup>25–27</sup>. Although local control is inferior to that of amputation, survival rates remain similar, offering a better quality of life through the preservation

■ FIGURE 1



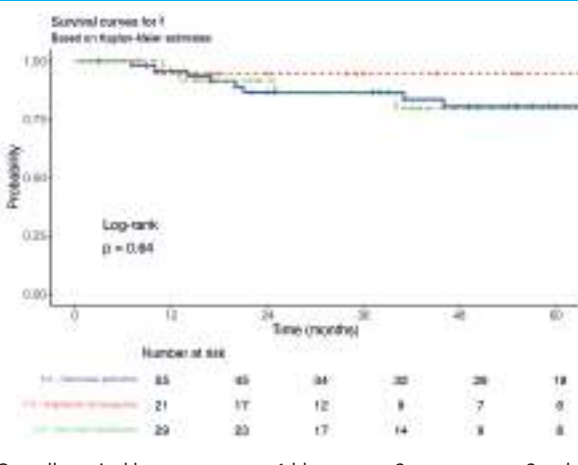
Local recurrence-free survival interval: group 1 blue, group 2 green, group 3 red

■ FIGURE 2



Overall survival for the entire series

■ FIGURE 3



Overall survival by group: group 1 blue, group 2 green, group 3 red

of a functional limb. For superficial trunk sarcomas, treatment is similar to that of STS of the extremities, following the same multimodal treatment regimen. In our series, the limb preservation rate was > 98%, consistent with current results from referral centers<sup>28</sup>.

Danieli et al.<sup>4</sup> reported 30-year results for this multimodal treatment, noting an increase in the use of both chemotherapy (from 18% to 29%) and preoperative radiation therapy (from 1.6% to 26%), while the use of combined chemoradiotherapy ranged from 0.8% to 14.3%. In this series, following multidisciplinary evaluation, 36% of patients with primary tumors (group 1) received radiation therapy (preoperative or postoperative). For patients with recurrence or incomplete resection (groups 2 and 3), the decision is much more complex, as certain factors, such as previous surgery, radiation therapy, and/or systemic therapy, may influence the decision regarding adjunctive treatments following surgery. We prescribed chemotherapy for 7% of the patients who sought medical advice for a primary or recurrent tumor.

O’Sullivan et al.<sup>29</sup> found that both preoperative and postoperative radiation therapy have the same effect on local disease control, with higher rates of early complications (wound dehiscence, infection) during the immediate postoperative period in patients who received preoperative radiation therapy and higher rates of late complications (lymphedema, stiffness) in patients who received postoperative radiation therapy. The current trend appears to prioritize preoperative treatment, especially for patients in whom tumor margins are expected to be adjacent to structures not directly invaded, such as neurovascular bundles or bone. Preoperative treatment is also recommended when more wound complications are expected due to the extent of the resection, which could delay or preclude the use of adjuvant radiation therapy<sup>2,30</sup>.

The study by Danieli et al.<sup>4</sup> reported a 17.9% incidence of local recurrence, 72.2% of distant metastasis, and 9.3% of cases in which both occurred as the first event prior to sarcoma-related death. In our series, local recurrence was observed in 14.5% of cases, systemic recurrence in 17.7%, and combined local and systemic recurrence in 7.2%. The lower rate of systemic recurrences observed may be related to the shorter duration of our study.

In the absence of randomized prospective studies evaluating the optimal surgical margin, the guidelines recommend performing resections with negative margins. However, a marginal resection near vital structures (such as neurovascular bundles and bone) is accepted in the absence of direct invasion to minimize morbidity. A margin of 1–2 cm from the tumor pseudocapsule is recommended, although smaller margins in the fixed material are considered appropriate<sup>31</sup>. To achieve this, it is often necessary to

perform extensive resections that expose neurovascular bundles or bones. An aspect to consider in these patients is the possibility that they may require some kind of reconstruction.

In this series, nearly one-third of the patients required a reconstructive procedure. Although local grafts and flaps were utilized in certain cases, free flaps were necessary in 7% of cases to achieve adequate coverage, with fasciocutaneous flaps being the most frequently employed. The decision of which procedure to use depends on local factors, the patient, and the tumor. In cases of major resections due to tumors in the distal extremities or certain areas of the chest wall, free flaps are the only reconstructive surgical option. Thus, a thorough preoperative evaluation is essential to ensure a successful surgical outcome<sup>17</sup>.

Another example of multidisciplinary surgical management is the assessment of vascular involvement. In our series, 3 patients required vascular resection with reconstruction. In most cases, an adventitial dissection can be performed if there is no direct vascular involvement. This type of dissection is also used on nerve structures and bone when there is no direct invasion. This leads to the concept of a "planned positive margin," which refers to areas where wide margins cannot be achieved due to the proximity

of vital structures that have not been invaded. Gerrand et al.<sup>30</sup> demonstrated that this type of resection, when combined with radiation therapy, provides good local control and avoids greater morbidity when planned preoperatively.

In this series, the 5-year survival rate for patients treated for primary tumors was approximately 80%, similar to rates reported by referral centers. In turn, the survival rates of patients treated outside our center were comparable to those of patients initially treated at our unit. However, these patients required more complex procedures to ensure optimal surgical outcomes. For this reason, it is necessary to refer them promptly to specialized centers.

While our study has limitations common to retrospective studies, its strength lies in the fact that it consists of patients treated at a single center by a unit with expertise in sarcoma management. We believe that our results can be attributed to our initial multidisciplinary approach and adherence to clinical guidelines. We emphasize the importance of having a team specialized in the management of sarcomas, as well as the ability to collaborate with other surgical specialties, such as reconstructive and vascular surgery, to ensure oncological resection and maximize the likelihood of limb preservation in cases of extremity tumors.

## Referencias bibliográficas /References

- Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. *Ann Surg.* 2014;260(3):416-21; discussion 421-2. doi:10.1097/SLA.0000000000000869.
- Gronchi A, Miah AB, Dei Tos AP, Abecassis N, Bajpai J, Bauer S, et al. Soft tissue and visceral sarcomas: ESMO-EURACAN-GEN-TURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2021;32(11):1348-65. doi:10.1016/j.annonc.2021.07.006.
- NCCN Guidelines® in Soft Tissue Sarcoma. Version 1.2025.
- Danieli M, Barretta F, Fiore M, Radaelli S, Sangalli C, Barisella M, et al. Refining the approach to patients with primary soft tissue sarcoma of the extremities and trunk wall: outcome improvement over time at a single institution. *Ann Surg Oncol.* 2022;29(5):3274-86. doi:10.1245/s10434-021-11189-2.
- Gervais MK, Basile G, Dulude JP, Mottard S, Gronchi A. Histology-tailored approach to soft tissue sarcoma. *Ann Surg Oncol.* 2024;31(12):7915-29. doi:10.1245/s10434-024-15981-8.
- Gronchi A, Miceli R, Colombo C, Collini P, Stacchiotti S, Olmi P, et al. Primary extremity soft tissue sarcomas: outcome improvement over time at a single institution. *Ann Oncol.* 2011;22(7):1675-81. doi:10.1093/annonc/mdq643.
- Roland CL, van Houdt W, Gronchi A. The Landmark Series: multimodality treatment of extremity sarcoma. *Ann Surg Oncol.* 2020;27(10):3672-82. doi:10.1245/s10434-020-08872-1.
- Califano L, Saco P, González A, Adan R, Voogd A, Urrutia G y col. Sarcomas de cabeza y cuello. Experiencia del Instituto Roffo. *Rev Argent Cirug.* 2002;82(3-4):141-8.
- Loccisano MH, Montesinos MF, Brécoli PA, Paradedda E, Avagnina A, Montesinos MR. Resultados del tratamiento quirúrgico de los sarcomas de partes blandas en adultos. *Rev Argent Cirug.* 2021;111(3):143-61.
- Montesinos MR, Falco JE, Sinagra DL, Mezzadri NA, Curutchet HP. Sarcomas retroperitoneales. *Rev Argent Cirug.* 2000;78:1-5.
- Nardi W, Aragona L, Quildrian S. Primary treated versus referred trunk and extremities soft tissue sarcomas: comparative analysis of initial treatment impact on disease control. *ecancer.* 2025;19:1933. doi:10.3332/ecancer.2025.1933.
- Novillo M, Albergo JJ, Huespe I, Latallade V, Farfalli GL, Ayerza MA y cols.. Resultados oncológicos y factores pronósticos en pacientes con sarcoma sinovial tratados quirúrgicamente. *Medicina (B Aires).* 2023;83(5):737-43. PMID:37870331.
- Norte MF, Recanatti M, Juárez M de los Á, Montesinos MR, Debonis DL. Experiencia en el tratamiento de sarcomas de partes blandas de cabeza y cuello en el adulto. *Rev Argent Cirug.* 2016;108(3):125-9.
- Pradier RN, González AR, Adan RS, Saco PA, Califano LL, Loria D, et al. Sarcomas de partes blandas. *Rev Argent Cirug.* 1993;64:26-35.
- Quildrian SD, Biscochea J, Calónico N, Chapela J, Sokol G, Cortés Bellomo A y col. Rbdomiosarcoma de partes blandas del adulto. *Rev Argent Cirug.* 2007;93(1-2):8-12.
- Quildrian SD, Sokol G, Biscochea JC, López Moris E, Calónico N, Chapela J y col. Liposarcoma de tronco y extremidades. *Rev Argent Cirug.* 2009;96(5-6):181-5.
- Quildrian SD, Nardi WS, Vega MG, Chapela JA. The role of free flap reconstruction after resection of extremity and trunk soft tissue sarcomas. *Clin Surg Oncol.* 2024;3(2):10042. doi:10.1016/j.cson.2024.100042.
- Sylvestre Begnis G, Sylvestre Begnis D, Viú Armengol R, Grunfeld P, Rastelli C. Sarcomas de partes blandas: resultados de tratamiento y factores pronósticos. *Rev Argent Cirug.* 2002;82(5-6):199-206.
- Yanzon A, Gómez NL, Picco P, Boccalatte L, Cayol F, Larrañaga J, et al. Head and neck sarcomas: treatment outcomes in a tertiary referral center in Argentina. *Oral Maxillofac Surg.* 2021;25(4):509-18. doi:10.1007/s10006-021-00944-0.
- Pasquali S, Palmerini E, Quagliuolo V, Martin-Broto J, López-Pousa A, Grignani G, et al. Neoadjuvant chemotherapy in high-risk soft tissue sarcomas: a Sarculator-based risk stratification analysis of the ISG-ST5 1001 randomized trial. *Cancer.* 2022;128(1):85-93. doi:10.1002/cncr.33895.
- Gronchi A, Ferrari S, Quagliuolo V, Broto JM, Pousa AL, Grignani G, et al. Histotype-tailored neoadjuvant chemotherapy versus standard chemotherapy in patients with high-risk soft-tissue sarcomas (ISG-ST5 1001): an international, open-label, randomised, controlled, phase 3, multicentre trial. *Lancet Oncol.* 2017;18(6):812-22. doi:10.1016/S1470-2045(17)30334-0.
- Blay JY, Honoré C, Stoeckle E, Meeus P, Jafari M, Gouin F, et al. Surgery in reference centers improves survival of sarcoma patients: a nation-wide study. *Ann Oncol.* 2019;30(7):1143-53. doi:10.1093/annonc/mdz124.
- Abbas JS, Holyoke ED, Moore R, Karakousis CP. The surgical treatment

- and outcome of soft-tissue sarcoma. Arch Surg. 1981;116(6):765-9. doi:10.1001/archsurg.1981.01380180025006.
24. Shiu MH, Castro EB, Hajdu SI, Fortner JG. Surgical treatment of 297 soft tissue sarcomas of the lower extremity. Ann Surg. 1975;182(5):597-602. doi:10.1097/00000658-197511000-00011.
  25. Beane JD, Yang JC, White D, Steinberg SM, Rosenberg SA, Rudloff U. Efficacy of adjuvant radiation therapy in the treatment of soft tissue sarcoma of the extremity: 20-year follow-up of a randomized prospective trial. Ann Surg Oncol. 2014;21(8):2484-9.
  26. Rosenberg SA, Tepper J, Glatstein E, Costa J, Baker A, Brennan M, et al. The treatment of soft-tissue sarcomas of the extremities: prospective randomized evaluations of (1) limb-sparing surgery plus radiation therapy compared with amputation and (2) the role of adjuvant chemotherapy. Ann Surg. 1982;196(3):305-15.
  27. Yang JC, Chang AE, Baker AR, Sindelar WF, Danforth DN, Topalian SL, et al. Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity. J Clin Oncol. 1998;16(1):197-203.
  28. Conti L, Buriro F, Baia M, Pasquali S, Miceli R, De Rosa L, et al. Contemporary role of amputation for patients with extremity soft tissue sarcoma. Eur J Surg Oncol. 2023;49(5):934-40. doi:10.1016/j.ejso.2022.12.002.
  29. O'Sullivan B, Davis AM, Turcotte R, Bell R, Catton C, Chabot P, et al. Preoperative versus postoperative radiotherapy in soft-tissue sarcoma of the limbs: a randomized trial. Lancet. 2002;359(9325):2235-41. doi:10.1016/S0140-6736(02)09292-9.
  30. Gerrand CH, Wunder JS, Kandel RA, O'Sullivan B, Catton CN, Bell RS, et al. Classification of positive margins after resection of soft-tissue sarcoma of the limb predicts the risk of local recurrence. J Bone Joint Surg Br. 2001;83(8):1149-55. doi:10.1302/0301-620X.83B8.12028.
  31. Borghi A, Gronchi A. Extremity and truncal soft tissue sarcoma: risk assessment and multidisciplinary management. Semin Radiat Oncol. 2024;34(2):147-63. doi:10.1016/j.semradonc.2023.12.001.

# Chatbots basados en inteligencia artificial: la importancia de saber preguntar

## Artificial intelligence-based chatbots: the importance of asking the right questions

M. Elena Peña , Alejandro Gigena , Fernando Iglesia 

Sanatorio Güemes,  
Servicio de Cirugía  
General. Buenos Aires.  
Argentina

Los autores declaran no  
tener conflictos  
de interés.

Conflicts of interest  
None declared.

Correspondencia  
Correspondence:  
M. Elena Peña  
E-mail: mpena@  
sg.com.ar

### RESUMEN

**Antecedentes:** los *chatbots* basados en inteligencia artificial (CBIA) son una fuente muy utilizada de información médica. La ingeniería de prompts (IP) está orientada a formular y optimizar las preguntas realizadas a los CBIA para mejorar sus respuestas.

**Objetivo:** evaluar la calidad de los *prompts* y las respuestas generadas por un CBIA en la resolución de casos clínicos de cirugía general, antes y después de una capacitación en IP.

**Materiales y métodos:** se elaboraron tres casos clínicos ficticios, que debieron ser resueltos por residentes de Cirugía General utilizando ChatGPT-4®. Luego de recibir capacitación en IP, los participantes resolvieron nuevamente los casos. La calidad de los *prompts* se evaluó mediante una escala (5-15 puntos) que consideró completitud, contexto, datos de entrada, formato de salida e instrucción. Las respuestas del *chatbot* se valoraron con una escala (3-15 puntos) que incluyó precisión, completitud y relevancia. Se compararon los resultados antes y después de la capacitación en IP.

**Resultados:** dieciséis residentes de primero a cuarto año participaron del estudio. La calidad de los *prompts* mejoró significativamente luego de la capacitación en IP, en puntaje total [7,9(1,8) vs. 10,4(2,1),  $p < 0,01$ ], completitud, contexto, datos de entrada y formato de salida. También mejoraron las respuestas del *chatbot* en todas las categorías y puntaje total [10,2(2) vs. 11,9(1,8),  $p < 0,01$ ].

**Conclusión:** la capacitación en IP mejoró significativamente la calidad de los *prompts* y las respuestas del CBIA en la resolución de casos clínicos de cirugía general.

■ **Palabras clave:** inteligencia artificial, chatbot, ChatGPT, prompt, Cirugía General.

### ABSTRACT

**Background:** Artificial intelligence-based chatbots (CBIA) are a widely used source of medical information. Prompt engineering (PE) focuses on designing and optimizing the questions asked of AI/CBs to improve responses.

**Objective:** The aim of this study was to compare the quality of prompts and the responses provided by an AiCB for clinical case resolution in general surgery, before and following PE training.

**Materials and methods:** Three fictional clinical cases were developed for residents in general surgery to solve using ChatGPT-4®. After they were trained in PE, the participants solved the cases again. The quality of the prompts was evaluated using a scale (5-15 points) that explored completeness, context, input data, output format, and instructions. The chatbot's answers were assessed using a scale (3-15 points) that included accuracy, completeness, and relevance. The results obtained before and following PE training were compared.

**Results:** Sixteen postgraduate year 1 to 4 residents participated in the study. The quality of prompts improved significantly following PE training, as assessed by total score [7.9 (1.8) vs. 10.4 (2.1),  $p < 0.01$ ] for completeness, context, data input, and output format categories. Chatbot's responses also improved across the categories and total score [10.2 (2) vs. 11.9 (1.8),  $p < 0.01$ ].

**Conclusion:** Training in PE significantly improved the quality of prompts and AiCB's responses for solving general surgery clinical cases.

■ **Keywords:** artificial intelligence, chatbot, ChatGPT, prompt, general surgery.

Recibido | Received  
25-12-25  
Aceptado | Accepted  
01-04-26

ID ORCID: M. Elena Peña, 0000-0001-7298-895X; Alejandro Gigena ORCID: 0009-0004-9518-1533; Fernando Iglesia, 0009-0000-5073-2726

## Introducción

La inteligencia artificial (IA) ha mostrado un crecimiento notable en los últimos años, especialmente en los grandes modelos de lenguaje (LLM). Estas son herramientas de procesamiento de lenguaje natural (NLP) entrenadas mediante aprendizaje automático en grandes cantidades de datos no estructurados. De esta forma, los LLM se utilizan hoy en día para la comprensión de texto, reconocimiento de voz, generación de lenguaje, traducción, etc.<sup>1</sup>.

ChatGPT (de Chat-Generative Pre-Trained-Transformer) es un *chatbot* entrenado en el LLM GPT y se destaca por ser capaz de simular conversaciones humanas mediante una interfaz amigable de preguntas y respuestas. Desde su lanzamiento en noviembre de 2022 por OpenAI, ha ganado popularidad en diversas áreas, entre ellas el ámbito médico. Ha demostrado ser útil en la educación de profesionales de la salud<sup>2,3</sup> y como asistente en la toma de decisiones médicas<sup>4,6</sup>. Actualmente, es ampliamente utilizado tanto por profesionales, como pacientes e instituciones de salud<sup>7</sup>. Sin embargo, existen controversias en cuanto a la validez y seguridad de la información que brindan los *chatbots* basados en inteligencia artificial (CBIA) en el ámbito médico<sup>8,9</sup>.

La pregunta, instrucción o palabras que el usuario introduce en el CBIA para obtener una respuesta se denomina *prompt*. Diversos estudios previos han demostrado que la manera de formular los *prompts* influye significativamente en la calidad de las respuestas que este brinda<sup>10</sup>. La ingeniería de *prompts* (IP) es un campo de investigación dedicado a diseñar y refinar los *prompts* con el objetivo de interactuar más eficazmente con los CBIA y obtener respuestas más adecuadas<sup>1</sup>.

Por ser esta un área de conocimiento relativamente nueva y poco conocida por gran parte de los usuarios en el ámbito médico y educativo, surge la pregunta de si el entrenamiento en IP podría influir en la calidad de la información obtenida con un CBIA.

El objetivo de este estudio fue comparar la calidad de los *prompts* generados por residentes y las respuestas brindadas por un CBIA para la resolución de casos clínicos en Cirugía General, antes y después de recibir capacitación en IP. La hipótesis fue que la adquisición de conocimientos en IP mejoraría tanto la calidad de los *prompts* como de las respuestas obtenidas.

## Materiales y métodos

Se realizó un estudio de diseño prospectivo, analítico, cuasi-experimental. Como ámbito de trabajo se eligió ChatGPT-4 (última versión al momento del diseño). La población de estudio estuvo conformada por los residentes de Cirugía General y los casos clínicos en Cirugía General diseñados a tal fin. La unidad de análisis

fue el conjunto de *prompts* elaborados por los residentes y las respuestas provistas por ChatGPT. El estudio se desarrolló en el período entre junio y julio de 2024.

Los criterios de inclusión de la población de estudio fueron: residentes de todos los años de Cirugía General de un sanatorio de tercer nivel, y casos clínicos sobre patologías frecuentes de cirugía general. No hubo criterios de exclusión.

## Procedimientos

Se elaboraron tres casos clínicos ficticios sobre patologías frecuentes en Cirugía General (Anexo 1). El caso clínico 1 se refirió a una paciente joven de sexo femenino que consultó a la Guardia por un cuadro de abdomen agudo quirúrgico. El objetivo de este caso fue evaluar los posibles diagnósticos diferenciales. El caso clínico 2 se trató de un paciente internado en la institución por colecistitis aguda. El objetivo de este caso fue evaluar el tratamiento de la colecistitis de acuerdo con su gravedad y basado en evidencia científica. El caso clínico 3 se trató de un paciente atendido en consultorio ambulatorio por una eventración compleja. El objetivo de este caso fue evaluar la descripción de los pasos de la técnica quirúrgica.

Se convocó a todos los residentes de Cirugía General de un sanatorio de tercer nivel a participar voluntariamente del estudio. Se realizó un único encuentro presencial de 3 horas de duración que se desarrolló en cuatro etapas.

### *Etapas 1: Evaluación de la experiencia previa de los participantes en el uso de CBIA e IP*

Se preguntó a los participantes si habían utilizado alguna vez un CBIA para cualquier actividad, si lo habían utilizado para su actividad profesional médica y si poseían conocimientos sobre IP. Se les envió un formulario de Google para recopilar las respuestas. A continuación, se brindó una breve demostración sobre el uso y los componentes básicos de ChatGPT-4.

### *Etapas 2: Resolución de casos clínicos*

Se entregaron a los participantes los tres casos clínicos ya descriptos. Se les solicitó que resolvieran los casos interactuando con ChatGPT mediante la generación de *prompts*. Se indicó que iniciaran una nueva sesión para cada caso con el fin de evitar retención de memoria por parte del chat y que no modificasen los hiperparámetros del modelo. Se permitió la reformulación del *prompt* (iteración). Se solicitó que enviaran el *prompt* generado para cada caso clínico con la respectiva respuesta de ChatGPT mediante un formulario de Google, recopilando un total de tres *prompts* y tres respuestas por cada residente.

### Etapa 3: Capacitación en IP

Por último, se brindó capacitación a los residentes sobre IP basándose en la publicación de Meskó<sup>11</sup>. Esto incluyó la definición de IP y su impacto en la obtención de mejores respuestas del CBIA en el ámbito médico. Luego se enumeraron y definieron los elementos básicos de un *prompt* (instrucción, contexto, datos de entrada y de salida). Finalmente se brindaron estrategias para refinar los *prompts*, por ejemplo, realizar iteraciones, juegos de roles, etcétera.

### Etapa 4: Resolución de casos clínicos luego de la capacitación en IP

Se solicitó a los participantes que resolvieran nuevamente los casos clínicos aplicando los conocimientos aprendidos sobre IP. Se solicitó que enviaran el *prompt* generado para cada caso clínico con la respectiva respuesta de ChatGPT mediante un formulario de Google idéntico al anterior; se recopiló un total de tres *prompts* y tres respuestas adicionales por cada residente.

### Análisis de variables

Se evaluó la calidad de los *prompts* elaborados por los residentes y las respuestas provistas por ChatGPT. Luego se compararon los resultados antes y después de recibir la capacitación en IP.

Para el análisis de la calidad de los *prompts* se consideraron cinco categorías: completitud, contexto, datos de entrada, formato de salida e instrucción (Tabla 1). Cada categoría fue evaluada con escalas de Likert y con la sumatoria se calculó el puntaje total de 5 a 15 puntos. La categoría "completitud" recibió el mayor peso relativo en el puntaje final debido a que constituye un requisito fundamental para el razonamiento clínico válido. Esto es particularmente relevante en el ámbito médico, donde la toma de decisiones depende de la disponibilidad de información clínica suficiente y pertinente.

Para el análisis de la calidad de las respuestas se consideraron tres categorías: precisión, completitud y relevancia (Tabla 2). Cada categoría fue evaluada con escalas de Likert y con la sumatoria se calculó el puntaje total de 3 a 15 puntos. En ambas escalas, un puntaje

■ TABLA 1

Descripción de las variables relacionadas con la calidad de los *prompts*

Variable	Definición	Evaluación
Completitud	Inclusión de todos los datos relevantes del caso clínico en el <i>prompt</i> (Por ejemplo: "Paciente mujer de 35 años que consulta por dolor abdominal de 6 horas de evolución")	Escala de Likert. Puntaje 1 al 5 (1: Muy incompleto, 2: Incompleto, 3: Moderadamente completo, 4: Completo, 5: Muy completo)
Contexto	Inclusión de datos contextuales del caso clínico en el <i>prompt</i> (Por ejemplo: "Soy residente de primer año")	Escala de Likert. Puntaje 1 al 3 (1: No incluye, 2: Incluye parcialmente, 3: Incluye totalmente)
Datos de entrada	Inclusión de información externa sobre la cual basarse en el <i>prompt</i> (Por ejemplo: "Responder usando las guías de la Sociedad europea de hernias")	Escala de Likert. Puntaje 1 al 3 (1: No incluye, 2: Incluye con poco detalle, 3: Incluye con detalle)
Formato de salida	Inclusión del formato de salida para formular el <i>prompt</i> (Por ejemplo: "Responder en 200 palabras")	Escala de Likert. Puntaje 1 al 2 (1: No incluye, 2: Incluye)
Instrucción	Inclusión de instrucciones claras y precisas en el <i>prompt</i> (Por ejemplo: "Elaborar una lista de diagnósticos diferenciales")	Escala de Likert. Puntaje 1 al 2 (1: No incluye, 2: Incluye)
Puntaje <i>prompt</i> total	Sumatoria de los puntajes obtenidos en completitud, contexto, datos de salida, datos de entrada e instrucción	Puntaje 5 al 15

■ TABLA 2

Descripción de las variables relacionadas con la calidad de las respuestas

Variable	Definición	Evaluación
Precisión	Veracidad de la información presentada en la respuesta	Escala de Likert. Puntaje 1 al 5 (1: Totalmente incorrecta, 2: Más incorrecta que correcta, 3: Iguales de correcta que incorrecta, 4: Más correcta que incorrecta, 5: Totalmente correcta)
Completitud	Presencia de toda la información necesaria para responder el caso clínico en la respuesta	Escala de Likert. Puntaje 1 al 5 (1: Muy incompleta, 2: Incompleta, 3: Moderadamente completa, 4: Completa, 5: Muy completa)
Relevancia.	Pertinencia de la información presente en la respuesta para responder el caso clínico	Escala de Likert. Puntaje 1 al 5 (1: Totalmente irrelevante, 2: Poco relevante, 3: Moderadamente relevante, 4: Relevante, 5: Totalmente relevante)
Puntaje total respuesta	Sumatoria de los puntajes obtenidos en precisión, completitud y relevancia	Puntaje 3 al 15

superior indica mejor calidad de respuesta o *prompt*.

La evaluación de los *prompts* y de las respuestas fue realizada por dos evaluadores (AG y FI), médicos especialistas en Cirugía General. Ambos fueron previamente entrenados en ingeniería de *prompts* con la misma modalidad que los residentes. Ambos recibieron el mismo set completo de *prompts* y respuestas codificadas, sin información sobre los participantes ni la intervención. Se solicitó que evaluaran los *prompts* y respuestas utilizando las escalas ya mencionadas. Los resultados de ambos evaluadores fueron promediados obteniendo un resultado final.

Para el tratamiento estadístico, las variables cuantitativas se informaron como promedio y desvío estándar para facilitar su interpretación. Las variables cualitativas se expresaron como frecuencias absolutas y relativas. Para el análisis comparativo de los *prompts* y respuestas antes y después de la capacitación se utilizó el test de Wilcoxon para muestras dependientes.

La concordancia interevaluador se evaluó mediante el coeficiente kappa de Cohen ponderado, debido al carácter ordinal de las escalas utilizadas. Se analizaron tanto la puntuación total de los *prompts* y las respuestas como las dimensiones específicas antes y después de la intervención. La magnitud del acuerdo se interpretó según los criterios de Landis y Koch<sup>12</sup>.

Se realizó un análisis multivariado mediante regresión ordinal para evaluar la asociación entre la experiencia previa en el uso de CBIA (tanto para actividades generales como para uso médico) y el año de residencia, con la calidad de los *prompts* y de las respuestas generadas por ChatGPT. Las variables dependientes fueron los puntajes totales de calidad de los *prompts* y de las respuestas, evaluados antes y después de la capacitación en ingeniería de *prompts*. El año de residencia se modeló como una variable ordinal, evaluándose su efecto como una tendencia lineal. Se estimaron odds ratios (OR) con sus correspondientes intervalos de confianza y valores de p. Se verificaron los supuestos del modelo, incluyendo la ausencia de colinealidad entre variables independientes mediante el factor de inflación de la varianza (VIF) y el supuesto de regresión paralela mediante el test de Brant. Se consideró la  $p < 0,05$  como significativa. El análisis estadístico fue realizado con el programa SPSS v.25<sup>®</sup> y RStudio<sup>®</sup>.

## Resultados

De los 18 residentes pertenecientes al Servicio de Cirugía General, 16 participaron del estudio. El 75% había utilizado algún CBIA previamente y el 63% lo utilizó para su actividad profesional. Ningún residente poseía conocimientos previos sobre IP (Tabla 3). Todos los residentes resolvieron los tres casos clínicos antes y después de la capacitación en IP, recopilando un total de 96 *prompts* y 96 respuestas.

La calidad de los *prompts* mejoró significativamente luego de la capacitación en IP [Puntaje total antes: 7,9 (1,8); puntaje total después: 10,4 (2,1);  $p < 0,01$ ]. Esta mejoría se observó en las categorías de completitud, contexto, formato de entrada y formato de salida (Tabla 4). En el caso clínico 1 se observó un incremento significativo del puntaje total, completitud, contexto y formato de salida (Anexo 2, Tabla A1). En el caso clínico 2 se observó un incremento significativo del puntaje total, completitud y contexto (Anexo 2, Tabla A3). En el caso clínico 3 se observó un incremento significativo del puntaje total, completitud, contexto y datos de entrada (Anexo 2, Tabla A5).

La calidad de las respuestas mejoró significativamente luego de la capacitación en IP [Puntaje total antes: 10,2 (2); puntaje total después: 11,9 (1,8);  $p < 0,01$ ]. Esta mejoría se observó en las categorías de precisión, completitud y relevancia (Tabla 5). En los casos

■ TABLA 3

Cantidad de residentes según el año de residencia

Variable	n (%)
Participantes	16
Año de residencia	
R1	6 (37%)
R2	3 (19%)
R3	4 (25%)
R4	3 (19%)
Uso previo de CBIA	12 (75%)
Uso previo de CBIA para actividad profesional	10 (63%)
Conocimientos de IP	0

R1: residente de primer año, R2: residente de segundo año, R3: residente de tercer año, R4: residente de cuarto año, CBIA: chatbots basados en inteligencia artificial, IP: ingeniería de prompts.

■ TABLA 4

Variables relacionadas con los prompts antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Completitud	2,2 (1,1)	3,2 (1,3)	<0,01
Contexto	1,3 (0,6)	2,2 (0,7)	<0,01
Datos de entrada	1,3 (0,5)	1,6 (0,8)	<0,01
Formato de salida	1,3 (0,5)	1,6 (0,5)	<0,01
Instrucción	1,8 (0,4)	1,9 (0,3)	0,4
Puntaje total de prompt	7,9 (1,8)	10,4 (2,1)	<0,01

■ TABLA 5

Variables relacionadas con las respuestas antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Precisión	3,6 (0,6)	4,1 (0,7)	<0,01
Completitud	3,6 (0,9)	4,1 (0,8)	<0,01
Relevancia	3,1 (0,9)	3,7 (0,8)	<0,01
Puntaje total respuesta	10,2 (2)	11,9 (1,8)	<0,01

clínicos 1 y 2 se observó un incremento significativo del puntaje total, precisión, completitud y relevancia (Anexo 2, Tablas A2 y A4). En el caso clínico 3 se observó un incremento significativo del puntaje total, precisión y completitud (Anexo 3, Tabla A6).

En el análisis de la concordancia interobservador se encontró que, para el puntaje total de los *prompts*, la concordancia entre evaluadores fue leve antes de la capacitación ( $\kappa = 0,04$ ) y aceptable luego de ella ( $\kappa = 0,25$ ). Al analizar las dimensiones específicas se observó que la completitud, contexto, datos de entrada e instrucción presentaron niveles de concordancia moderados-sustanciales antes y después de la intervención, mientras que para el formato de salida fue pobre-aceptable (Tabla A7).

En cuanto al puntaje total de las respuestas, la concordancia fue leve antes de la capacitación ( $\kappa = 0,04$ ) y pobre luego de esta ( $\kappa = -0,11$ ). Al analizar las dimensiones específicas, la concordancia fue pobre para las tres categorías antes y después de la intervención (Tabla A8).

Para el análisis de predictores de calidad se encontró que, en el modelo de regresión ordinal, no se observó asociación entre el año de residencia y el puntaje total de los *prompts* ni de las respuestas generadas por ChatGPT, tanto antes como después de la capacitación en IP.

En cuanto a la experiencia previa en el uso de CBIA para actividades generales, se observó una asociación negativa con el puntaje total de los *prompts* posteriores a la capacitación [OR 0,7 (IC 95%: 0,01-0,57)]. No se observó asociación con el resto de los puntajes para esta variable.

Asimismo, la experiencia previa en el uso de ChatGPT con fines médicos no mostró asociación significativa con el puntaje total de los *prompts* ni de las respuestas en ninguno de los escenarios evaluados (Tablas A9 y A10).

## Discusión

En el presente estudio analizamos la calidad de los *prompts* generados por residentes y las respuestas proporcionadas por ChatGPT-4 en la resolución de casos clínicos de Cirugía General, comparando los resultados antes y después de una capacitación en IP. Los hallazgos principales fueron: a) La calidad de los *prompts* mejoró significativamente luego de la capacitación; b) La calidad de las respuestas mejoró significativamente luego de la capacitación; c) El año de residencia y la experiencia previa de los residentes en el uso de CBIA sin capacitación específica en IP no se asociaron a mejor calidad de *prompts* y respuestas.

La IP es un campo de investigación emergente que se basa en el diseño y refinamiento de los *prompts*. El objetivo es interactuar de forma más eficaz con los CBIA y obtener respuestas más adecuadas<sup>1</sup>. Existen

múltiples estrategias para lograrlo. Por ejemplo, la técnica zero-shot donde el usuario solicita una tarea al CBIA sin brindarle ejemplos previos, en contraste con la técnica few-shot, donde antes de generar el prompt se brindan ejemplos de cómo debería responder<sup>1,13</sup>. Diversos estudios han puesto en evidencia que el refinamiento de los *prompts* mejora el desempeño de los modelos de lenguaje<sup>14,15</sup>. Sin embargo, en el ámbito médico, especialmente en Cirugía General, este tipo de información es muy escasa y no existen recomendaciones estandarizadas ni guías sobre la interacción óptima con los CBIA<sup>5,11,16</sup>. En nuestro trabajo observamos un impacto positivo de la IP en la calidad de *prompts* creados por los residentes, destacando las categorías de completitud, contexto, datos de entrada, formato de salida e instrucción, las cuales podrían guiar la elaboración de *prompts* médicos en el futuro.

ChatGPT es uno de los CBIA más conocidos y utilizados en la actualidad. Su rendimiento en distintas tareas mejoró significativamente con el avance de los LLM<sup>6,7,17</sup>. Algunos estudios previos analizaron la utilidad de ChatGPT 3.5<sup>o</sup> en el ámbito de la educación y medicina. En el examen de licencia médica de Estados Unidos (USMLE), por ejemplo, ChatGPT logró resolver más del 50% de las preguntas, equiparando al conocimiento de un estudiante de Medicina de tercer año<sup>18,19</sup>. La versión utilizada en nuestro estudio, ChatGPT-4, ofrece mejoras en precisión y contextualización de respuestas<sup>20-22</sup>. Nuestros resultados sugieren que, aunque esta versión es más precisa, la calidad de sus respuestas depende aún en gran medida de la formulación del prompt, especialmente en entornos médicos donde la precisión y ética de la información son cruciales para evitar consecuencias negativas graves en los pacientes<sup>22</sup>.

Es importante destacar que los casos clínicos diseñados abordaron distintos aspectos de la evaluación médica, tales como la elaboración de diagnósticos diferenciales, tratamiento según la evidencia científica y descripción de la técnica quirúrgica. En todos los casos hubo una mejoría de los *prompts* en las categorías de completitud y contexto, lo cual se tradujo en mejores respuestas del *chatbot*. Este hallazgo resalta la relevancia de incorporar información completa y contextualizada en los *prompts*, independientemente del tipo de caso clínico planteado. En contraste, las otras tres categorías mostraron mejoras no significativas tras la capacitación en la mayoría de los escenarios. Particularmente, la categoría instrucción no evidenció cambios en ninguno de los casos. Esto podría explicarse porque los puntajes basales ya eran elevados antes de la capacitación, lo que habría limitado el margen de mejora y reducido la necesidad de enfatizar este aspecto durante el entrenamiento en IP.

Otro aspecto relevante es que, aunque la mayoría de los residentes había utilizado un CBIA en el pasado para fines personales o profesionales, ninguno poseía conocimientos de IP antes de la capacitación. Esto pone de manifiesto la falta de difusión de la IP, pese a

la creciente popularización de estas herramientas. Por otro lado, ni el año de residencia ni la experiencia previa en el uso de ChatGPT, tanto para actividades generales como médicas, se asociaron de forma consistente con una mejor calidad de *prompts* y de respuestas antes y después de la capacitación. Esto sugiere que un nivel formativo mayor y el uso casual de los CBIA no garantizan una interacción más eficaz con estos modelos de lenguaje. En conjunto, tales hallazgos destacan la importancia de incluir conocimientos sobre el uso de CBIA e IP en los programas de formación médica.

Este estudio presenta algunas limitaciones. En primer lugar, se analizó únicamente ChatGPT-4, por lo que los resultados pueden no ser generalizables a otras versiones y modelos de IA. Por otro lado, el uso de casos clínicos ficticios puede limitar la aplicabilidad de estos hallazgos a escenarios de la vida real, con mayor complejidad y variabilidad. Otra limitación fue la baja concordancia interevaluador para las respuestas del CBIA, incluso después de la capacitación. Esto sugiere que su valoración podría estar influida por com-

ponentes interpretativos y subjetivos del evaluador. Asimismo, los resultados de la regresión ordinal deben interpretarse con cautela, dado que la amplitud de los intervalos de confianza reflejan una menor precisión en la estimación de los efectos. Finalmente, la muestra de participantes fue pequeña y proveniente de una sola institución, lo cual podría afectar la generalización de los resultados a otros contextos clínicos y formativos.

En conclusión, en este estudio, la capacitación en IP a residentes de Cirugía General mejoró significativamente la calidad de los *prompts* y de las respuestas generadas por ChatGPT. El año de residencia y el uso previo de CBIA sin capacitación específica en IP no se asoció de manera significativa a mejor calidad de *prompts* y respuestas. Todo esto reafirma la importancia de obtener conocimientos en IP para mejorar la eficacia de los *chatbots* de IA en la toma de decisiones médicas. Se recomienda incorporar entrenamiento en IP en los programas de formación médica para maximizar el potencial de estas herramientas, promoviendo su uso seguro y efectivo.

## ■ ANEXOS

### Anexo 1

#### Caso clínico 1

Consulta a la Guardia una paciente mujer de 35 años por presentar dolor abdominal en fosa ilíaca derecha de 6 horas de evolución (Escala visual analógica: 8/10), sin otros síntomas asociados.

- Antecedentes personales: apendicectomía a los 25 años, colecistectomía a los 30 años.
- Examen físico: tensión arterial 120/80 mm Hg, frecuencia cardíaca 90 lpm, frecuencia respiratoria 18 rpm, temperatura 37 °C. Abdomen blando, depresible, doloroso en fosa ilíaca derecha sin defensa ni dolor a la descompresión. Ruidos hidroaéreos normales. Resto del examen físico sin particularidades.

#### Exámenes complementarios

- Laboratorio (hemograma, función renal, hepatograma, ionograma): leucocitos 11 000/mm<sup>3</sup>, resto sin particularidades.
- Ecografía de abdomen: se observa líquido libre en el cuadrante inferior derecho. Resto sin particularidades.

La paciente le pregunta cuál es el diagnóstico y la conducta para seguir.

#### Caso clínico 2

Usted recibe una interconsulta por un paciente

varón de 60 años internado en sala general por dolor abdominal en hipocondrio derecho de 5 días de evolución. Al examen físico se encuentra lúcido, febril (38 °C), taquicárdico (105 lpm), normotenso. Abdomen blando, depresible, doloroso en hipocondrio derecho con defensa y descompresión. En el laboratorio presenta leucocitosis (25 000 glóbulos blancos), sin otra particularidad. En la ecografía de abdomen se observa vesícula biliar con paredes de 5 mm, lito enclavado en bacinete y líquido perivesicular.

Su compañero de guardia le pregunta cuál es el mejor tratamiento para este paciente de acuerdo con la evidencia científica.

#### Caso clínico 3

Usted se encuentra en el consultorio de cirugía de pared abdominal. Consulta un paciente varón de 58 años por un bulto doloroso en centro abdominal de 2 años de evolución. Presenta episodios frecuentes de dolor que alteran su calidad de vida.

- Antecedentes personales: diabetes, hipertensión arterial, hemicolectomía izquierda por diverticulitis aguda hace 5 años.
- Examen físico: cicatriz mediana infraumbilical. Eventración de línea media con saco de 20 × 20 centímetros.
- Tomografía computarizada (TC) de abdomen y pelvis: eventración centroabdominal con saco de 20 × 20 cm con contenido intestinal y anillo de 12 cm de ancho.

Se diagnostica una eventración compleja. Se le propone al paciente realizar una eventroplastia convencional, pero usted no recuerda del todo los pasos de la técnica quirúrgica.

## Anexo 2

■ TABLA A1

Caso clínico 1: variables relacionadas con los *prompts* antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Complejidad	2,2 (1,2)	2,9 (1,2)	<0,01
Contexto	1,3 (0,7)	2,2 (0,8)	<0,01
Datos de entrada	1,2 (0,4)	1,4 (0,6)	0,08
Formato de salida	1,3 (0,5)	1,6 (0,5)	0,03
Instrucción	1,8 (0,4)	1,9 (0,3)	0,3
Puntaje total <i>prompt</i>	7,75 (2)	10 (1,9)	<0,01

■ TABLA A2

Caso clínico 1: variables relacionadas con las respuestas antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Precisión	3,3 (0,6)	3,9 (0,7)	<0,01
Complejidad	3,3 (0,9)	3,8 (0,8)	0,01
Relevancia	2,7 (0,9)	3,5 (0,8)	<0,01
Puntaje total respuesta	9,3 (2,2)	11,3 (1,7)	<0,01

■ TABLA A3

Caso clínico 2: variables relacionadas con los *prompts* antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Complejidad	2,2 (1,1)	3,4 (1,4)	<0,01
Contexto	1,3 (0,6)	2,2 (0,7)	<0,01
Datos de entrada	1,4 (0,5)	1,7 (0,9)	0,13
Formato de salida	1,1 (0,3)	1,4 (0,5)	0,06
Instrucción	1,8 (0,4)	1,8 (0,4)	1
Puntaje total <i>prompt</i>	7,8 (2)	10,6 (2,7)	<0,01

■ TABLA A4

Caso clínico 2: variables relacionadas con las respuestas antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Precisión	3,8 (0,7)	4,4 (0,7)	<0,01
Complejidad	3,8 (0,8)	4,1 (0,6)	0,2
Relevancia	3,3 (0,8)	4 (0,7)	0,03
Puntaje total respuesta	10,8 (2)	12,4 (1,7)	0,01

■ TABLA A5

Caso clínico 3: variables relacionadas con los *prompts* antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Complejidad	2,3 (1,1)	3,5 (1,3)	<0,01
Contexto	1,3 (0,5)	2,2 (0,8)	<0,01
Datos de entrada	1,4 (0,5)	1,7 (0,8)	0,03
Formato de salida	1,4 (0,5)	1,7 (0,5)	0,09
Instrucción	1,8 (0,4)	1,9 (0,3)	0,7
Puntaje total <i>prompt</i>	8,2 (1,5)	10,9 (1,7)	<0,01

■ TABLA A6

Caso clínico 3: variables relacionadas con las respuestas antes y después de recibir capacitación en IP

	Antes de IP	Después de IP	p
Precisión	3,7 (0,5)	4 (0,7)	0,03
Complejidad	3,7 (0,7)	4,3 (0,9)	0,03
Relevancia	3,3 (0,9)	3,7 (0,8)	0,3
Puntaje total respuesta	10,7 (1,9)	11,9 (1,9)	<0,01

■ TABLA A7

Concordancia interevaluador en el análisis de los *prompts*

	Intervención	Kappa	Interpretación
Complejidad	Antes	0,59	Moderada
Complejidad	Después	0,67	Sustancial
Contexto	Antes	0,52	Moderada
Contexto	Después	0,59	Moderada
Datos de entrada	Antes	0,7	Sustancial
Datos de entrada	Después	0,77	Sustancial
Formato de salida	Antes	0,23	Aceptable
Formato de salida	Después	-0,12	Pobre
Instrucción	Antes	0,57	Moderada
Instrucción	Después	0,49	Moderada
Puntaje total	Antes	0,04	Leve
Puntaje total	Después	0,25	Aceptable

■ TABLA A8

Concordancia interevaluador en el análisis de las respuestas

	Intervención	Kappa	Interpretación
Precisión	Antes	0,14	Leve
Precisión	Después	0,06	Leve
Complejidad	Antes	0,13	Leve
Complejidad	Después	-0,07	Pobre
Relevancia	Antes	0,2	Leve
Relevancia	Después	0,04	Leve
Puntaje total	Antes	0,04	Leve
Puntaje total	Después	-0,11	Pobre

■ TABLA A9

Asociación entre el año de residencia, la experiencia previa en el uso de CBIA (general y ámbito médico) y el puntaje total de los prompts antes y después de la capacitación en IP

	Intervención	OR (IC 95%)	p
Año de residencia	Antes	1,34 (0,43–4,15)	0,61
Año de residencia	Después	1,95 (0,66–5,81)	0,23
Uso previo CBIA	Antes	1,49 (0,22–10,24)	0,69
Uso previo CBIA	Después	0,07 (0,01–0,57)	0,01
Uso previo CBIA médico	Antes	0,47 (0,08–2,97)	0,42
Uso previo CBIA médico	Después	1,91 (0,31–11,71)	0,46

■ TABLA A10

Asociación entre el año de residencia, la experiencia previa en el uso de CBIA (general y ámbito médico) y el puntaje total de las respuestas antes y después de la capacitación en IP

	Intervención	OR (IC 95%)	p
Año de residencia	Antes	1,76 (0,65–4,73)	0,27
Año de residencia	Después	1,23 (0,43–3,48)	0,7
Uso previo CBIA	Antes	0,59 (0,1–3,61)	0,57
Uso previo CBIA	Después	3,08 (0,41–23,37)	0,28
Uso previo CBIA médico	Antes	1,41 (0,28–6,98)	0,68
Uso previo CBIA médico	Después	0,72 (0,11–4,68)	0,73

## ■ ENGLISH VERSION

### Introduction

Artificial intelligence (AI) has grown remarkably in recent years, especially in the area of large language models (LLMs). A large language model is a natural language processing (NLP) tool that has been trained on vast amounts of data using machine learning. Consequently, LLMs are currently employed for text understanding, speech recognition, language generation, translation, and other related tasks<sup>1</sup>.

ChatGPT (Chat-Generative Pre-Trained Transformer) is a chatbot trained on the GPT LLM and is notable for its ability to simulate human conversations through a user-friendly question-and-answer interface. Since its launch in November 2022 by OpenAI, it has gained popularity in various settings, including the medical field. ChatGPT has proven useful for educating healthcare professionals<sup>2,3</sup> and assisting in medical decision-making<sup>4–6</sup>. It is currently widely used by professionals, patients, and healthcare institutions<sup>7</sup>. However, the validity and reliability of the information provided by artificial intelligence-based chatbots (AiBCs) in the medical field remain controversial<sup>8,9</sup>.

The question, instruction, or words that users enter into the AiCB to obtain a response are called a prompt. Several previous studies have shown that the way prompts are formulated has a significant impact on the quality of the responses they generate<sup>10</sup>. Prompt engineering (PE) is a field of research dedicated to the design and refinement of prompts to interact more effectively with AiCBs and obtain more appropriate responses<sup>1</sup>.

As this is a relatively new field of knowledge with which many users in medical and educational settings are unfamiliar, the question arises whether PE training could influence the quality of information obtained through an AiCB.

The aim of this study was to compare the quality of prompts generated by residents and the responses provided by an AiCB for clinical case

resolution in general surgery, before and following PE training. The hypothesis was that gaining knowledge in PE would improve both the quality of prompts and responses obtained.

### Materials and methods

We conducted a prospective, analytic and quasi-experimental study. ChatGPT-4 (the most recent version at the time of design) was selected as the platform for this project. The study population included general surgery residents and clinical cases designed for this purpose. The unit of analysis was the set of prompts created by the residents and the responses provided by ChatGPT. The study took place between June and July 2024.

The inclusion criteria for the study population included all the residents in the general surgery program at a tertiary care hospital, and clinical cases involving common general surgery conditions. There were no exclusion criteria.

### Procedures

Three fictional clinical cases were developed based on common conditions in general surgery (Appendix 1). Clinical case 1 corresponded to a young female patient who attended the emergency department for acute abdomen requiring surgery. The aim of this case was to evaluate the possible differential diagnoses. Clinical case 2 corresponded to a patient hospitalized for acute cholecystitis. The aim of this case was to assess evidence-based management of cholecystitis according to its severity. Clinical case 3 was a patient attending the outpatient clinic for a complex incisional hernia. The aim of this case was to evaluate the description of the surgical technique steps.

All residents in the general surgery department

at a tertiary care hospital were invited to participate in the study voluntarily. A single 3-hour in-person meeting was held, divided into four stages.

#### *Stage 1: Assessment of participants' prior experience with AiCB and PE*

Participants were asked whether they had ever used an AiCB for any activity, whether they had used it in their medical practice, and whether they were aware of PE. A Google form was distributed to collect their responses. Next, the use and basic components of ChatGPT-4 were briefly introduced.

#### *Stage 2: Solving clinical cases.*

The three clinical cases described above were distributed to the participants. They were asked to solve the cases by interacting with ChatGPT using prompts. Participants were instructed to start a new session for each case, to prevent the chat system from retaining memory, and to avoid modifying the model's hyperparameters. The residents were allowed to reformulate (iterate) the prompt. They were asked to submit the prompt generated for each clinical case along with ChatGPT's corresponding response via a Google form, resulting in a total of three prompts and three responses per resident.

#### *Step 3: PE training.*

Finally, residents were trained on PE based on the publication by Meskó<sup>11</sup>. This included the definition of PE and its impact on obtaining better responses from the AiCB in the medical field. The basic elements of prompts (instructions, context, input data, and output

format) were then listed and defined. Finally, strategies for refining the prompts were provided, including iteration and role-playing.

#### *Step 4: Solving clinical cases following PE training*

Participants were asked to revisit the clinical cases and apply the knowledge they had gained about PE to resolve them. They were asked to submit the prompt generated for each clinical case, along with ChatGPT's corresponding response, via a Google form identical to the previously submitted, resulting in a total of three prompts and three responses per resident.

### **Analysis of variables**

The quality of the prompts elaborated by the residents and of the responses provided by ChatGPT was evaluated. Then, the results obtained before and following PE training were compared.

Prompt quality was analyzed considering five categories: completeness, context, input data, output format, and instructions (Table 1). Each category was assessed using Likert scales, and the total score — ranging from 5 to 15 points — was calculated by adding up the individual scores. The “completeness” category was assigned the highest relative weight in the final score because it is a fundamental requirement for valid clinical reasoning. This is particularly relevant in the medical field, where decision-making depends on the availability of sufficient and relevant clinical information.

The quality of the responses was analyzed using three categories: accuracy, completeness, and relevance (Table 2). Each category was assessed using

■ TABLE 1

Description of the variables related to prompt quality

Variable	Definition	Evaluation
Completeness	Inclusion of all relevant clinical case data within the prompt (e.g., “35-year-old female patient presenting with abdominal pain that began 6 hours ago...”)	Likert scale. Score from 1 to 5 (1: Highly deficient, 2: Incomplete, 3: Partially complete, 4: Substantially complete, 5: Fully comprehensive).
Context	Inclusion of contextual data regarding the clinical case within the prompt (e.g., “I am a postgraduate year-1 resident”).	Likert scale. Score from 1 to 3 (1: Does not include, 2: Partially includes, 3: Fully includes).
Input Data	Inclusion of external information upon which to base the prompt (e.g., “Respond using the guidelines from the European Hernia Society”).	Likert scale. Score from 1 to 3 (1: Does not include, 2: Includes with minor detail, 3: Includes with detail).
Output Format	Inclusion of the required output format when formulating the prompt (e.g., “Respond in 200 words”).	Likert scale. Score from 1 to 2 (1: Does not include, 2: Includes).
Instruction	Inclusion of clear and precise instructions within the prompt (e.g., “Develop a list of differential diagnoses”).	Likert scale. Score from 1 to 2 (1: Does not include, 2: Includes).
Total Prompt Score	Sum of the scores obtained for completeness, context, input data, output format, and instruction.	Score from 5 to 15.

■ TABLE 2

Description of the variables related to response quality

Variable	Definition	Assessment
Accuracy	Accuracy of the information provided in the response	Likert scale Score from 1 to 5 (1: Completely incorrect, 2: Mostly incorrect, 3: Equally correct and incorrect, 4: Mostly correct, 5: Completely correct).
Completeness	Presence of all necessary information within the response to address the clinical case.	Likert scale. Score from 1 to 5 (1: Highly deficient, 2: Incomplete, 3: Partially complete, 4: Substantially complete, 5: Fully comprehensive)
Relevance	Relevance of the information provided by the response to address the clinical case	Likert scale Score from 1 to 5 (1: Completely irrelevant, 2: Mostly irrelevant, 3: Partially relevant, 4: Mostly relevant, 5: Completely relevant).
Total response score	Sum of the scores obtained for accuracy, completeness, and relevance.	Score from 3 to 15

Likert scales, and the total score — ranging from 3 to 15 points — was calculated by adding up the individual scores. For both scales, a higher score indicates greater prompt or response quality.

The prompts and responses were assessed by two evaluators (AG and FI), who are general surgeons. Both surgeons had received training in prompt engineering using the same method as that employed with residents. They were administered the same set of prompts and encoded responses, with no information about the participants or the intervention provided. They were asked to assess the prompts and responses using the mentioned scales. The scores calculated by both evaluators were averaged to obtain a final score.

For statistical analysis, quantitative variables were reported as mean and standard deviation to facilitate their interpretation. Qualitative variables were expressed as absolute and relative frequencies. The Wilcoxon test for dependent samples was used to compare prompts and responses before and following the training intervention.

Inter-rater agreement was assessed using Cohen's weighted kappa coefficient, given the ordinal nature of the scales used. The total scores obtained from prompts and responses, as well as the specific dimensions, were analyzed both before and following the intervention. The magnitude of the agreement was interpreted according to the criteria of Landis and Koch<sup>12</sup>.

The association between prior experience of AiCB (for both general and medical purposes) and year in the residency program with the quality of prompts and responses generated by ChatGPT was evaluated using ordinal multivariate regression. The dependent variables were the total prompt and response quality scores, assessed before and following prompt engineering training. The year of residency was modeled as an ordinal variable, and its effect was assessed as a linear trend. Odds ratios (OR) were estimated, along with their corresponding confidence intervals and p-values. The model assumptions were verified, including the absence of multicollinearity among independent variables, using the variance inflation factor (VIF), and the assumption of parallel

regression using the Brant test. A p-value < 0.05 was considered statistically significant. All the statistical calculations were performed using SPSS v.25<sup>®</sup> and RStudio<sup>®</sup> software packages.

## Results

Of the 18 residents in the department of general surgery, 16 participated in the study. Seventy-five percent had used an AiCB before, with 63% having used it in their professional activities. None of the residents had any prior knowledge of PE (Table 3). All residents solved the three clinical cases before and following PE training, providing a total of 96 prompts and 96 responses.

Prompts' quality improved significantly following PE training [total score before training: 7.9 (1.8); total score following training: 10.4 (2.1);  $p < 0.01$ ]. This improvement was noted in completeness, context, input data and output format categories (Table 4). In clinical case 1, a significant increase was observed in the total score, completeness, context, and output format categories (Appendix 2, Table A1). In clinical case 2, a significant increase was observed in the total score, completeness and context categories (Appendix 2, Table A3). In clinical case 3, the total score, completeness, context, and input data categories increased significantly (Appendix 2, Table A5).

The quality of responses improved significantly following PE training [total score before training: 10.2 (2); total score following training: 11.9 (1.8);  $p < 0.01$ ]. This improvement was noted in the accuracy, completeness, and relevance categories (Table 5). In clinical cases 1 and 2, the total score, accuracy, completeness, and relevance categories increased significantly (Appendix 2, Tables A2 and A4). In clinical case 3, there was a significant increase in the total score, accuracy, and completeness categories (Appendix 3, Table A6).

The analysis of inter-rater agreement revealed that agreement was low before training ( $\kappa = 0.04$ ) and acceptable following training for the total prompt score ( $\kappa = 0.25$ ). When analyzing the specific dimensions, moderate and substantial levels of agreement were

■ TABLE 3

Number of residents by year of residency

Variable	n (%)
Participants	16
Year of residency	
PGY-1	6 (37%)
PGY-2	3 (19%)
PGY-3	4 (25%)
PGY-4	3 (19%)
Previous use of AiCB	12 (75%)
Previous use of AiCB for professional activities	10 (63%)
PE knowledge	0

PGY-1: post-graduate year 1 resident; PGY-2: post-graduate year 2 resident; PGY-3: post-graduate year 3 resident; PGY-4: post-graduate year 4 resident; AiCB: artificial intelligence-based chatbot; PE: prompt engineering.

■ TABLE 4

Variables related to prompts before and following PE training

	Before PE	Following PE	p-value
Completeness	2.2 (1.1)	3.2 (1.3)	<0.01
Context	1.3 (0.6)	2.2 (0.7)	<0.01
Input data	1.3 (0.5)	1.6 (0.8)	<0.01
Output format	1.3 (0.5)	1.6 (0.5)	<0.01
Instruction	1.8 (0.4)	1.9 (0.3)	0.4
Total prompt score	7.9 (1.8)	10.4 (2.1)	<0.01

■ TABLA 5

Variables related to response before and following PE training

	Before PE	Following PE	p-value
Accuracy	3.6 (0.6)	4.1 (0.7)	<0.01
Completeness	3.6 (0.9)	4.1 (0.8)	<0.01
Relevance	3.1 (0.9)	3.7 (0.8)	<0.01
Total response score	10.2 (2)	11.9 (1.8)	<0.01

found for completeness, context, input data and instructions before and following the intervention, respectively. For the output format, poor and acceptable levels of agreement were found, respectively (Table A7).

The analysis of inter-rater agreement revealed that agreement was low before training ( $\kappa = 0.04$ ) and poor following training for the total response score ( $\kappa = 0.11$ ). When analyzing the specific dimensions, agreement was poor for all three categories, both before and following the intervention (Table A8).

The analysis of quality predictors revealed no association between the year of residency and the total prompt score or total response score, either before or following PE training in the ordinal regression model.

Regarding previous experience in using AiCB for general activities, there was a negative association with the total prompt score following training [OR 0.7

(95% CI: 0.01–0.57)]. No association occurred with the rest of the scores for this variable.

Similarly, previous experience with ChatGPT for medical purposes was not significantly associated with total prompt or response scores in any scenario (Tables A9 and A10).

## Discussion

In this study, we analyzed the quality of prompts generated by residents and the responses provided by ChatGPT-4 for clinical case resolution in general surgery, before and following PE training. The main findings were: a) the prompt quality significantly improved following training; b) response quality significantly improved following the training; c) the year of residency and prior experience of residents in the use of AiCB without specific PE training were not associated with better prompt quality and response quality.

Prompt engineering is an emerging field of research that focuses on designing and refining prompts. The goal is to interact with the AiCB more effectively and obtain more appropriate responses<sup>1</sup>. There are several strategies to achieve this goal. For example, in the zero-shot technique, the user asks the AiCB to perform a task without providing any examples, whereas in the few-shot technique, the user provides examples of how the model should respond before generating the prompt<sup>1,13</sup>. Several studies have shown that refining prompts improves the performance of language models<sup>14,15</sup>. However, in the medical field, particularly in general surgery, such information is scarce, and there are no standardized recommendations or guidelines about optimal interaction with AiCBs<sup>5,11,16</sup>. In our study, we observed that PE produced a positive impact on the quality of prompts created by residents, particularly in the categories completeness, context, input data, output format, and instructions, which could guide the development of medical prompts in the future.

ChatGPT is one of the most widely known and used AiCBs at present. Its performance on various tasks improved significantly with the development of LLMs<sup>6,7,17</sup>. Some previous studies have examined the usefulness of ChatGPT 3.5<sup>®</sup> in the field of education and medicine. In the United States Medical Licensing Examination (USMLE), for instance, ChatGPT demonstrated capabilities to answer over 50% of the questions, exhibiting a level of knowledge comparable to that of a third-year medical student<sup>18,19</sup>. The version used in our study, ChatGPT-4, offers improved accuracy and contextual relevance of responses<sup>20–22</sup>. Our results suggest that, while this version demonstrates enhanced accuracy, response quality still largely depends on prompt formulation, particularly within medical contexts where the accuracy and ethical integrity of

the information are crucial to avoiding serious adverse outcomes for patients<sup>22</sup>.

It is worth noting that the clinical cases presented addressed various aspects of medical evaluation, such as the formulation of differential diagnoses, evidence-based treatments, and descriptions of surgical techniques. The prompts improved in both completeness and context in all cases, resulting in better chatbot responses. This finding underscores the importance of including comprehensive and contextualized information in prompts, regardless of the type of clinical case presented. Conversely, the other three categories showed no significant improvement in most scenarios following training. In particular, the instruction category showed no changes in any of the cases. This could be explained by the fact that baseline scores were already high before training. This may have limited the potential for improvement and reduced the emphasis placed on this aspect during PE training.

Another relevant aspect is that although most residents had already used an AiCB in their personal or professional activities, none had prior knowledge of PE before training. This highlights the PE is not well disseminated, despite the growing popularity of these tools. Conversely, the year of residency and prior experience using ChatGPT, whether for general or medical purposes, were not consistently associated with higher-quality prompts and responses before or following training. This suggests that a higher level of training and casual use of AiCBs do not ensure more effective interaction with these language models.

Altogether, these findings underscore the importance of incorporating knowledge on the use of AiCBs and PE into medical training programs.

This study has some limitations. First, we analyzed only ChatGPT-4, so the results cannot be generalized to other versions or other AI models. The use of fictional clinical cases may limit the applicability of these findings to real-world scenarios, which are more complex and variable. Another limitation was the low inter-rater agreement about the responses provided by the AiCB, even following training. This suggests that their assessment might be influenced by their own subjective interpretations. Furthermore, the results of the ordinal regression should be interpreted with caution, as the wide confidence intervals indicate lower accuracy in estimating the effects. Finally, the sample of participants was small and drawn from a single institution, which could limit the generalizability of the results to other clinical and educational settings.

In conclusion, this study found that training general surgery residents in PE significantly improved the quality of prompts and ChatGPT responses. Year of residency and prior AiCB use without specific PE training were not significantly associated with higher-quality prompts or responses. These findings underscore the importance of leveraging PE expertise to enhance the effectiveness of AI-based chatbots in medical decision-making. It is recommended that PE training be incorporated into medical education programs to maximize the potential of these tools and promote their safe and effective use.

## ■ APPENDICES

### Appendix 1

#### **Clinical case 1**

A 35-year-old female patient presents to the emergency department with abdominal pain in the right iliac fossa (visual analogue scale: 8/10) that started 6 hours ago, with no other associated symptoms.

- Personal history: appendectomy when she was 25 years old and cholecystectomy at the age of 30.
- Physical examination: blood pressure 12/80 mm Hg, heart rate 90 bpm, respiratory rate 18 bpm, temperature 37 °C. The abdomen is soft, depressible, and tender on palpation in the right iliac fossa, without guarding or rebound tenderness. The bowel sounds are normal. There are no other relevant signs on physical examination.

#### *Ancillary tests:*

- Laboratory tests: complete blood count, kidney

function, electrolytes: white cell count 11 000/mm<sup>3</sup>, with no other abnormal results.

- Abdominal ultrasound: free peritoneal fluid in the right lower quadrant. There were no other abnormalities.

The patient asks about the diagnosis and course of action.

#### **Clinical case 2**

You are required for a consultation on a 60-year-old male patient admitted to the general ward with abdominal pain in the right hypochondriac region that has persisted for 5 days. On physical examination, the patient is alert; his body temperature is 38 °C, his blood pressure is normal, and he presents tachycardia (105 bpm). The abdomen is soft, depressible, and tender on palpation in the right hypochondriac region, with guarding and rebound tenderness. The laboratory tests show high white cell count (25,000/mm<sup>3</sup>) with no other abnormal results. On abdominal ultrasound, the gallbladder walls measure 5 mm, a stone is lodged in

the gallbladder infundibulum, and pericholecystic fluid is visible.

Your on-call colleague asks you about the best treatment for this patient, based on scientific evidence.

### Clinical case 3

You are in the abdominal wall surgery clinic. A 58-year-old male patient presents with a painful bulge in the mid-abdominal region that has been present for two years. He complains of frequent episodes of pain that affect his quality of life.

- Personal history: diabetes, hypertension, left hemicolectomy for acute diverticulitis 5 years ago.
- Physical examination: infraumbilical median scar. Midline incisional hernia with a sac measuring 20 × 20 cm.
- Computed tomography scan of the abdomen and pelvis: midline incisional hernia with a 20 × 20 cm sac containing intestinal contents and a 12 cm wide ring.

The diagnosis is complex incisional hernia. The decision is to perform conventional hernia repair, but you don't fully recall the surgical steps.

## Appendix 2

■ TABLE A1

Clinical case 1:  
prompt-related variables before and following PE training

	Before PE	Following PE	p-value
Completeness	2.2 (1.2)	2.9 (1.2)	<0.01
Context	1.3 (0.7)	2.2 (0.8)	<0.01
Input data	1.2 (0.4)	1.4 (0.6)	0.08
Output format	1.3 (0.5)	1.6 (0.5)	0.03
Instruction	1.8 (0.4)	1.9 (0.3)	0.3
Total prompt score	7.75 (2)	10 (1.9)	<0.01

■ TABLE A2

Clinical case 1:  
response-related variables before and following PE training

	Before PE	Following PE	p-value
Accuracy	3.3 (0.6)	3.9 (0.7)	<0.01
Completeness	3.3 (0.9)	3.8 (0.8)	0.01
Relevance	2.7 (0.9)	3.5 (0.8)	<0.01
Total response score	9.3 (2.2)	11.3 (1.7)	<0.01

■ TABLE A3

Clinical case 2:  
prompt-related variables before and following PE training

	Before PE	Following PE	p-value
Completeness	2.2 (1.1)	3.4 (1.4)	<0.01
Context	1.3 (0.6)	2.2 (0.7)	<0.01
Input data	1.4 (0.5)	1.7 (0.9)	0.13
Output format	1.1 (0.3)	1.4 (0.5)	0.06
Instruction	1.8 (0.4)	1.8 (0.4)	1
Total prompt score	7.8 (2)	10.6 (2.7)	<0.01

■ TABLE A4

Clinical case 2:  
response-related variables before and following PE training

	Before PE	Following PE	p-value
Accuracy	3.8 (0.7)	4.4 (0.7)	<0.01
Completeness	3.8 (0.8)	4.1 (0.6)	0.2
Relevance	3.3 (0.8)	4 (0.7)	0.03
Total response score	10.8 (2)	12.4 (1.7)	0.01

■ TABLE A5

Clinical case 3:  
prompt-related variables before and following PE training

	Before PE	Following PE	p-value
Completeness	2.3 (1.1)	3.5 (1.3)	<0.01
Context	1.3 (0.5)	2.2 (0.8)	<0.01
Input data	1.4 (0.5)	1.7 (0.8)	0.03
Output format	1.4 (0.5)	1.7 (0.5)	0.09
Instruction	1.8 (0.4)	1.9 (0.3)	0.7
Total prompt score	8.2 (1.5)	10.9 (1.7)	<0.01

■ TABLE A6

Clinical case 3:  
response-related variables before and following PE training

	Before PE	Following PE	p-value
Accuracy	3.7 (0.5)	4 (0.7)	0.03
Completeness	3.7 (0.7)	4.3 (0.9)	0.03
Relevance	3.3 (0.9)	3.7 (0.8)	0.3
Total response score	10.7 (1.9)	11.9 (1.9)	<0.01

■ TABLE A7

Inter-rater agreement in prompt analysis

	Intervention	Kappa	Interpretation
Completeness	Before	0.59	Moderate
Completeness	Following	0.67	Significant
Context	Before	0.52	Moderate
Context	Following	0.59	Moderate
Input data	Before	0.7	Significant
Input data	Following	0.77	Significant
Output format	Before	0.23	Acceptable
Output format	Following	-0.12	Poor
Instruction	Before	0.57	Moderate
Instruction	Following	0.49	Moderate
Total score	Before	0.04	Mild
Total score	Following	0.25	Acceptable

■ TABLE A8

Inter-rater agreement in response analysis

	Intervention	Kappa	Interpretation
Accuracy	Before	0.14	Mild
Accuracy	Following	0.06	Mild
Completeness	Before	0.13	Mild
Completeness	Following	-0.07	Poor
Relevance	Before	0.2	Mild
Relevance	Following	0.04	Mild
Total score	Before	0.04	Mild
Total score	Following	-0.11	Poor

■ TABLE A9

Association between year of residency, prior experience with AiCB (general and medical purposes), and total prompt scores before and following PE training

	Intervention	OR (95% CI)	p-value
Year of residency	Before	1.34 (0.43–4.15)	0,61
Year of residency	Following	1.95 (0.66–5.81)	0,23
Previous use of AiCB	Before	1.49 (0.22–10.24)	0,69
Previous use of AiCB	Following	0.07 (0.01–0.57)	0,01
Previous use of AiCB for medical purposes	Before	0.47 (0.08–2.97)	0,42
Previous use of AiCB for medical purposes	Following	1.91 (0.31–11.71)	0,46

■ TABLE A10

Association between year of residency, prior experience with AiCB (general and medical purposes), and total prompt scores before and following PE training

	Intervention	OR (95% CI)	p-value
Year of residency	Prior	1.76 (0.65–4.73)	0.27
Year of residency	Following	1.23 (0.43–3.48)	0.7
Previous use of AiCB	Prior	0.59 (0.1–3.61)	0.57
Previous use of AiCB	Following	3.08 (0.41–23.37)	0.28
Previous use of AiCB for medical purposes	Prior	1.41 (0.28–6.98)	0.68
Previous use of AiCB for medical purposes	Following	0.72 (0.11–4.68)	0.73

## Referencias bibliográficas /References

- Wang J, Enze S, Sigang Y, Zihao W, Chong M, Haixing D, et al. Prompt Engineering for Healthcare: Methodologies and Applications. *ArXiv*. 2023; abs/2304.14670. doi:10.48550/arXiv.2304.14670.
- Goodman RS, Patrinely JR, Stone CA Jr, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and Reliability of Chatbot Responses to Physician Questions. *JAMA Netw Open*. 2023;6(10):e2336483. doi:10.1001/jamanetworkopen.2023.36483.
- Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)*. 2023;11(6):887. doi: 10.3390/healthcare11060887.
- Sorin V, Klang E, Sklair-Levy M, Cohen I, Zippel DB, Balint Lahat N, et al. Large language model (ChatGPT) as a support tool for breast tumor board. *NPJ Breast Cancer*. 2023;9(1):44. doi:10.1038/s41523-023-00557-8.
- Kuşçu O, Pamuk AE, SütaySüslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? *Front Oncol*. 2023;13:1256459. doi:10.3389/fonc.2023.1256459.
- Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. *Int J Environ Res Public Health*. 2023;20(4):3378. doi:10.3390/ijerph20043378.
- Cascella M, Semeraro F, Montomoli J, Bellini V, Piazza O, Bignami E. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. *J Med Syst*. 2024;48(1):22. doi:10.1007/s10916-024-02045-3.
- Birkun AA, Gautam A. Large Language Model (LLM)-Powered Chatbots Fail to Generate Guideline-Consistent Content on Resuscitation and May Provide Potentially Harmful Advice. *Prehosp Disaster Med*. 2023;38(6):757–63. doi: https://doi.org/10.1017/S1049023X23006568.
- Zúñiga Salazar G, Zúñiga D, Vindel CL, Yoong AM, Hincapie S, Zúñiga AB, et al. Efficacy of AI Chats to Determine an Emergency: A Comparison Between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. *Cureus*. 2023;15(9):e45473. doi: https://doi.org/10.7759/cureus.45473.
- Liu M, Okuhara T, Chang X, Shirabe R, Nishiie Y, Okada H, et al. Performance of ChatGPT Across Different Versions in Medical Licensing Examinations Worldwide: Systematic Review and Meta-Analysis. *J Med Internet Res*. 2024;26:e60807. doi:10.2196/60807.
- Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *J Med Internet Res*. 2023;25:e50638. doi:10.2196/50638.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
- Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-80. doi:10.1038/s41586-023-06291-2.
- Wang L, Chen X, Deng X, Wen H, You M, Liu W, et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *NPJ Digit Med*.

- 2024;7(1):41. doi:10.1038/s41746-024-01029-4.
15. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. ArXiv. 2023:2302.11382v1. <https://doi.org/10.48550/arXiv.2302.11382>
  16. Sorin V, Klang E, Sklair-Levy M, Cohen I, Zippel DB, Balint Lahat N, et al. Large language model (ChatGPT) as a support tool for breast tumor board. NPJ Breast Cancer. 2023;9(1):44. doi:10.1038/s41523-023-00557-8.
  17. Open AI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 Technical Report ArXiv. 2023: 2303.08774v6. doi: [10.48550/arXiv.2303.08774](https://doi.org/10.48550/arXiv.2303.08774).
  18. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ. 2023;9:e45312. doi:10.2196/45312.
  19. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198.
  20. Rojas M, Rojas M, Burgess V, Toro-Pérez J, Salehi S. Exploring the Performance of ChatGPT Versions 3.5, 4, and 4 With Vision in the Chilean Medical Licensing Examination: Observational Study. JMIR Med Educ. 2024;10:e55048. doi: 10.2196/55048.
  21. Shieh A, Tran B, He G, Kumar M, Freed JA, Majety P. Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK and clinical case reports. Sci Rep. 2024;14(1):9330. doi: 10.1038/s41598-024-58760-x.
  22. Guillen-Grima F, Guillen-Aguinaga S, Guillen-Aguinaga L, Alas-Brun R, Onambele L, Ortega W, et al. Evaluating the Efficacy of ChatGPT in Navigating the Spanish Medical Residency Entrance Examination (MIR): Promising Horizons for AI in Clinical Medicine. Clin Pract. 2023;13(6):1460-87. doi: 10.3390/clinpract13060130.

# Exactitud de la inteligencia artificial en la estadificación y toma de decisiones en cáncer esófago gástrico

## Accuracy of artificial intelligence for staging and decision-making in esophageal and gastric cancer

Adelina E. Coturel <sup>1</sup>, Paula Pereyra <sup>1</sup>, Rodrigo García <sup>1</sup>, Agustín Diomedi <sup>1</sup>, Juan J. M. Cabas Audicio <sup>1</sup>, Roberto Klappenbach <sup>1</sup>

Institución:  
Hospital de Alta  
Complejidad del  
Bicentenario de Esteban  
Echeverría. Buenos  
Aires, Argentina.

Los autores declaran no  
tener conflictos  
de interés.  
Conflicts of interest  
None declared.

Correspondencia  
Correspondence:  
Adelina E. Coturel  
E-mail:  
adelinacoturel@gmail.com

### RESUMEN

**Antecedentes:** la inteligencia artificial (IA) ha mostrado un creciente potencial en el ámbito de la oncología, no solo en el análisis de imágenes sino también como herramienta de apoyo en la toma de decisiones clínicas.

**Objetivo:** evaluar un modelo conversacional basado en IA, entrenado con guías clínicas, para asistir en la estadificación y planificación terapéutica inicial en cáncer esofagogástrico.

**Material y métodos:** se realizó un estudio retrospectivo en pacientes con cáncer de esófago, estómago y de la unión esofagogástrica, en un hospital de tercer nivel de atención, tratados entre diciembre de 2023 y mayo de 2025. Se desarrolló un asistente virtual o chatbot con ChatGPT-4.5<sup>®</sup> (OpenAI), configurado para interpretar los estudios según guías NCCN y ESMO. El chatbot debía determinar el estadio clínico (cTNM), sugerir la conducta terapéutica y detectar hallazgos relevantes. Sus respuestas se compararon con las decisiones del equipo médico tratante.

**Resultados:** de 53 pacientes evaluados, 34 cumplieron los criterios de inclusión. La concordancia entre el cTNM y el estadio clínico asignados por el equipo médico y el chatbot fue del 94%. La coincidencia en la conducta terapéutica sugerida fue del 85%. Las discrepancias fueron por diferencias en la interpretación del estadio, consideraciones clínicas no disponibles para la IA y aspectos sociales o logísticos que influyeron en la decisión médica.

**Conclusión:** la IA aplicada mediante un chatbot mostró utilidad para asistir en la estadificación y propuesta terapéutica inicial del cáncer esofagogástrico, con alta concordancia con las decisiones clínicas.

■ **Palabras clave:** inteligencia artificial, neoplasias esofágicas, neoplasias gástricas, estadificación de tumores, sistemas de apoyo a la toma de decisiones clínicas.

### ABSTRACT

**Background:** Artificial intelligence (AI) has demonstrated increasing potential in oncology, including in imaging analysis and as a tool for supporting clinical decision-making.

**Objective:** The aim of this study was to evaluate an AI conversational model, trained using up-to-date clinical guidelines, to assist with staging and initial treatment planning in patients with gastroesophageal cancer.

**Materials and methods:** We conducted a retrospective study of patients with esophageal, gastric, and gastroesophageal junction cancer treated at a tertiary care hospital between December 2023 and May 2025. A virtual clinical assistant or chatbot was developed using ChatGPT-4.5<sup>®</sup> (OpenAI), configured to interpret test results in accordance with the NCCN and ESMO guidelines. The chatbot was designed to determine the clinical stage (cTNM), suggest a treatment plan, and identify relevant findings. The answers were compared with the decisions made by the treating physician.

**Results:** Of the 53 patients evaluated, 34 met the inclusion criteria. The accuracy between the cTNM staging determined by the medical team and that of the chatbot was 94%. The concordance in the suggested therapeutic approach was 85%. The discrepancies were due to differences in the interpretation of the stage, clinical considerations not available to the AI, and social or logistical factors that influenced the medical decision.

**Conclusion:** AI, implemented through a chatbot trained using clinical guidelines, proved useful in supporting staging and initial treatment planning for gastroesophageal cancer, with high concordance with clinical decisions.

■ **Keywords:** artificial intelligence, esophageal neoplasms, stomach neoplasms, neoplasm staging, decision support systems, clinical.

Recibido | Received  
01-01-26  
Aceptado | Accepted  
01-04-26

ID ORCID: Adelina E. Coturel, 0000-0001-5379-6872; Paula Pereyra, 0009-0007-2682-586X; Rodrigo García, 0000-0001-9442-1989; Agustín Diomedi, 0009-0009-0961-2461; Juan J. M. Cabas Audicio, 0009-0007-4379-6941; Roberto Klappenbach, 0000-0002-0069-0035.

## Introducción

La inteligencia artificial (IA) se ha consolidado como una herramienta transformadora en medicina, especialmente en oncología, donde ha demostrado mejorar la precisión diagnóstica, optimizar la planificación terapéutica y apoyar la toma de decisiones personalizadas<sup>1</sup>. En términos generales, la IA comprende sistemas capaces de reproducir funciones cognitivas humanas, como el aprendizaje y el razonamiento<sup>2</sup>. Dentro de este campo, los modelos de lenguaje natural utilizados en asistentes conversacionales, *chatbots*, permiten interpretar e interactuar con lenguaje clínico complejo.

El avance del *deep learning* ha ampliado la capacidad de estas tecnologías para analizar información heterogénea, imágenes, texto libre y datos clínicos estructurados, y vincularla con guías de práctica clínica. En oncología digestiva se han informado aplicaciones exitosas en detección temprana de cáncer esofagogástrico, identificación automatizada de lesiones endoscópicas y predicción de respuesta terapéutica<sup>3</sup>. Sin embargo, hay poca evidencia sobre el uso de modelos conversacionales para el desarrollo de asistentes que permitan colaborar con la estadificación y planificación inicial del tratamiento.

En la Argentina, la evidencia sobre el uso de la IA en medicina aún es limitada. Se han publicado experiencias en reconocimiento de imágenes para la visión crítica de seguridad en colecistectomías en el Hospital Argerich<sup>4</sup>, en la evaluación del uso de ChatGPT para recomendaciones de *screening* de cáncer colorrectal<sup>5</sup>, y en la implementación de un copiloto de IA integrado a historias clínicas electrónicas en el Hospital Italiano de Buenos Aires<sup>6</sup>, orientado a mejorar el acceso a información relevante del paciente.

En este contexto, el objetivo principal de nuestro estudio fue evaluar la precisión de un modelo conversacional de IA para predecir el estadio clínico (cTNM y estadio global) a partir de los estudios disponibles en la primera consulta oncológica de pacientes con cáncer esofagogástrico. Como objetivo secundario, analizar la concordancia entre las conductas terapéuticas sugeridas por el *chatbot* y las indicadas por el comité de tumores o el equipo tratante.

## Material y métodos

Se realizó un estudio retrospectivo sobre la base de datos de pacientes que consultaron por primera vez, por patología oncológica, en el consultorio de cirugía esofagogástrica de la institución, entre diciembre de 2023 y mayo de 2025.

Fueron incluidos aquellos pacientes de los cuales se contaba con estudios preoperatorios completos, definidos como: pacientes con diagnóstico de tumor de esófago, unión esofagogástrica (UEG) o estómago,

informe de endoscopia digestiva alta, informe anatomopatológico de biopsia, tomografía computarizada, tomografía por emisión de positrones (PET-TC) en pacientes con cáncer de esófago o tumores de la UEG.

Fueron excluidos los pacientes sin la totalidad de los estudios preoperatorios mencionados, o con histología diferente de adenocarcinoma o carcinoma epidermoide.

Las decisiones terapéuticas se tomaron en forma multidisciplinaria entre cirujanos especialistas en Cirugía Esofagogástrica oncológica, oncólogos clínicos, anestesiólogos, médicos clínicos y nutricionistas.

## Desarrollo y uso del asistente virtual

Para el análisis automatizado se desarrolló un asistente clínico virtual utilizando ChatGPT (OpenAI), versión GPT-4.5<sup>®</sup>, a través de una suscripción paga ChatGPT Plus<sup>®</sup>. El desarrollo y configuración del asistente fue realizado por cirujanos con formación en Cirugía Esofagogástrica y experiencia en estadificación tumoral. Si bien no cuentan con formación formal en programación ni en desarrollo algorítmico, poseen capacitación en IA aplicada al ámbito sanitario, lo que permitió estructurar el uso del modelo bajo criterios clínicos y metodológicos adecuados.

El *chatbot* fue configurado específicamente para esta tarea mediante la carga y lectura contextual de las guías clínicas actualizadas de la Red Nacional del Cáncer de Estados Unidos (NCCN, por sus siglas del inglés National Comprehensive Cancer Network)<sup>7,8</sup> y de la Sociedad Europea de Oncología Clínica (ESMO, del inglés European Society for Medical Oncology)<sup>9,10</sup>, correspondientes a cáncer de esófago y cáncer gástrico.

El modelo fue instruido para procesar, en lenguaje natural, la información clínica de cada paciente simulando una consulta oncológica inicial. Cabe destacar que, para ello, no fue reentrenado ni modificado a nivel algorítmico. Se trabajó mediante ingeniería de *prompts*, definiendo instrucciones estructuradas basadas en las guías internacionales mencionadas. A partir de los datos aportados (edad, sexo, comorbilidades, motivo de consulta, informes de endoscopia, anatomía patológica e imágenes), el asistente debía: a) determinar el estadio clínico del tumor según la clasificación TNM 8ª edición (AJCC)<sup>11</sup>; b) sugerir el tratamiento oncológico inicial adecuado según guías internacionales.

Previamente al análisis definitivo, se realizaron pruebas piloto con casos históricos no incluidos en la cohorte final, con el objetivo de estandarizar el formato de ingreso de datos, ajustar la redacción de las instrucciones y homogeneizar la estructura de las respuestas. No se realizaron modificaciones sustanciales en las instrucciones durante el análisis de los casos incluidos en el estudio.

Cada caso de estudio fue procesado de manera

individual, cargando los datos clínicos en la plataforma y registrando las respuestas generadas por la IA en una tabla estructurada. Posteriormente, las propuestas del asistente fueron comparadas con las decisiones tomadas por el comité de tumores o el equipo tratante en la práctica clínica habitual.

En aquellos casos en los que se identificaron discrepancias en la estadificación o en la conducta terapéutica sugerida, se realizó una interacción dirigida con el *chatbot*, con el objetivo de explorar las causas de la diferencia y determinar si se trataba de una interpretación alternativa válida, una omisión de datos relevantes, o una limitación en la comprensión contextual de los estudios.

### Análisis estadístico

Las variables analizadas fueron:

1) demográficas: edad, sexo, comorbilidades, motivo de consulta; 2) del tumor: topografía, histopatología, informes de endoscopia, tomografía y PET-TC, cTNM, estadio clínico; 3) del tratamiento: realización de tratamiento endoscópico, tratamiento quirúrgico sin neoadyuvancia, tratamiento quirúrgico con neoadyuvancia o quimioterapia perioperatoria, tratamiento paliativo.

Se realizó un análisis descriptivo de los datos recolectados. Las variables categóricas se expresaron como frecuencias absolutas y porcentajes, mientras que las variables continuas se informaron como media y rango.

La concordancia entre la estadificación clínica (cTNM y estadio global) propuesta por el *chatbot* y la realizada por el equipo médico fue evaluada mediante comparación directa caso por caso. Del mismo modo, se analizó la concordancia en la conducta terapéutica sugerida.

El estudio se diseñó bajo la hipótesis de que un modelo conversacional de IA, entrenado con guías clínicas actualizadas, podría predecir correctamente la estadificación clínica en al menos el 80% de los casos y sugerir tratamientos acordes con las recomendaciones internacionales. Dado el carácter exploratorio del estudio y el tamaño de la muestra, no se aplicaron pruebas estadísticas inferenciales.

La exactitud del modelo se calculó como la proporción de casos en los que la respuesta del *chatbot* coincidió con la estadificación o la conducta propuesta por el equipo médico, sobre el total de casos evaluados. Se utilizó la siguiente fórmula:

$$\text{Exactitud (\%)} = (\text{Casos concordantes} / \text{Total de casos}) \times 100$$

Se calcularon por separado la exactitud para cTNM, estadio clínico y tratamiento sugerido.

Consideraciones éticas: El presente estudio fue realizado de acuerdo con los principios éticos de la Declaración de Helsinki. Por tratarse de un estudio retrospectivo, observacional y basado en el análisis de datos clínicos anonimizados, no se requirió consentimiento informado.

### Resultados

De 53 pacientes que consultaron por tumores de esófago, estómago y de la UEG en el período indicado, se excluyeron 19 por falta de datos relevantes en la historia clínica, y se incluyeron 34 casos que cumplían los criterios de inclusión. De ellos, 18 (51%) eran de sexo masculino y la edad promedio fue de 67 años (rango 51-80). En 29 pacientes se encontraba detallado el motivo de consulta, y eso se analizó en la tabla 1, junto al resto de los datos demográficos y la ubicación del tumor.

■ TABLA 1

Variables demográficas y de los tumores	
Variables	n (%)
Comorbilidades	
Diabetes mellitus	11 (32%)
HTA	19 (56%)
EPOC	6 (18%)
Tabaquista o extabaquista	9 (27%)
Enfermedad cardiovascular	9 (27%)
Cáncer de mama	6 (18%)
Otros cánceres	4 (11%)
Motivo de consulta	
Hemorragia digestiva	6 (18%)
Disfagia	16 (47%)
Dispepsia o dolor epigástrico	14 (41%)
Síndrome pilórico o vómitos	9 (26%)
Descenso de peso	26 (76%)
Ubicación del tumor	
Esófago	10 (29%)
Estómago	13 (38%)
Unión esofagogástrica	11 (32%)
- Siewert I	2 (18%)
- Siewert II	6 (54%)
- Siewert III	3 (27%)
Anatomía patológica	
Células en anillo de sello	
Esófago	1 (10%)
Unión esofagogástrica	3 (27%)
Estómago	4 (31%)
Carcinoma escamoso*	5 (50%)

\*Del total de los cánceres de esófago.

HTA: hipertensión arterial; EPOC: enfermedad pulmonar obstructiva crónica.

Según el estadio, 3 pacientes presentaron estadio I, 7 estadio II, 17 estadio III y 7 estadio IV. El resto de las características se describen en la tabla 2.

La exactitud entre el cTNM estipulado por el equipo médico y el del *chatbot* fue del 94%. Dentro del tratamiento propuesto, la concordancia fue del 85%.

Los casos discordantes se presentan en la tabla 3.

Al dividir la cohorte en tres grupos secuenciales (n = 12, 11 y 11 pacientes), encontramos una mejoría en la precisión en el cTNM y el estadio del *chat-*

*bot*: 83% en el grupo 1, 100% en el grupo 2 y también 100% en el grupo 3.

## Discusión

El presente estudio evaluó la utilidad de un modelo conversacional de inteligencia artificial, entrenado con guías clínicas actualizadas, para asistir en la estadificación y planificación terapéutica inicial en pacientes con cáncer de esófago, unión esofagogástrica y estómago. Los resultados muestran una alta concordancia entre la estadificación clínica propuesta por el *chatbot* y la realizada por el equipo médico tratante (94%), tanto a nivel de cTNM como de estadio global. Si bien la concordancia en la conducta terapéutica fue algo menor (85%), en la mayoría de los casos las diferencias se explicaron por consideraciones clínicas contextuales no siempre disponibles para el modelo, como limitaciones sociales.

Los *chatbots* son programas que procesan y simulan conversaciones humanas mediante lenguaje natural<sup>12</sup>. Su uso en salud creció durante la pandemia de COVID-19, aunque aún permanece subutilizado y se espera que aumente en la próxima década. En cáncer gástrico, distintos modelos de inteligencia artificial han mostrado precisión prometedora para identificar individuos de alto riesgo, predecir agresividad tumoral, recurrencia y supervivencia<sup>13,14</sup>. El deep learning ha demostrado capacidad para analizar imágenes endoscópicas, datos genómicos e histopatológicos, mejorando diagnóstico y pronóstico. La integración multimodal de datos clínicos, genéticos e imagenológicos permite obtener una visión más completa de la enfermedad y

■ TABLA 2

Comparación entre la estadificación clínica realizada por el equipo médico y el *chatbot*

cTNM**	Equipo médico	Chatbot
cT		
cT1b	1(3%)	1 (3%)
cT1b-2	1(3%)	1 (3%)
cT2-3	3(9%)	3 (9%)
cT3	21(62%)	21 (62%)
cT3-4	4(12%)	4 (12%)
cT4	4(12%)	4 (12%)
cN		
cN0-1	1(3%)	1 (3%)
cN+	20(59%)	21 (62%)
cM		
cM1	7(21%)	6(18 %)
Estadios		
I	3(9%)	3 (9%)
II	7(20%)	6 (18%)
III	17(50%)	19 (56%)
IV	7(20%)	6 (18%)

■ TABLA 3

Casos discordantes entre el Equipo Médico y el *chatbot*

Caso	Categoría de la discrepancia	Motivo	Equipo Médico	chatbot
1	cTNM y estadio clínico	Adenopatías lateroaórticas en estudios por imágenes	Considera M1, estadio IV	Considera adenopatías regionales, M0, estadio III
1	Tratamiento	Discrepancia de estadio clínico (IV vs. III)	Tratamiento paliativo	Neoadyuvancia y cirugía
2	cTNM y estadio clínico	Tumor de la UEG, sin adenopatías hiper captantes en la PET. El chatBot lo considera N + probable por la longitud del tumor (5,8 cm) y SUV 13,1	Estadificación clínica: cT-3N0M0 (EII)	Estadificación clínica: y el Chatbot cT3N1M0 (EIII)
3	Tratamiento	Sexo femenino, 63 años, HTA y trasplante renal Carcinoma escamoso de tercio medio de esófago	CROSS y esofagectomía	Quimioterapia Radioterapia definitiva
4	Tratamiento	Sexo femenino, 80 años, ingresa por Guardia por hemorragia digestiva y disnea. Hallazgo de adenocarcinoma gástrico cT3N+M0	Prehabilitación 3 semanas y gastrectomía	Tratamiento paliativo
5	Tratamiento	Masculino, 65 años, cáncer gástrico difuso (lininitis), mala tolerancia a la vía oral. Sin cobertura social	Prehabilitación 3 semanas internado (nutrición parenteral y enteral) y gastrectomía	FLOT y gastrectomía
6	Tratamiento	Sexo femenino, 51 años, adenocarcinoma de la UEG cT3N+MODiscordancia entre Siewert	FLOT y gastrectomía	CROSS y esofagectomía
7*	Tipo de cirugía	Paciente con adenocarcinoma de cuerpo gástrico	FLOT y gastrectomía total	FLOT y gastrectomía distal

\*Este caso no se considera discrepancia formal (el plan terapéutico es correcto), pero se menciona ya que varía el tipo de gastrectomía, lo que es relevante revisar para asegurar una resección oncológica adecuada.

UEG: unión gastroesofágica; PET: tomografía por emisión de positrones (PositronEmissionTomography); SUV: StandardizedUptakeValue; HTA: hipertensión arterial; CROSS: ChemoRadiotherapy for Oesophageal cancer followed by Surgery; FLOT: quimioterapia perioperatoria con 5-Fluorouracilo, leucovorina, oxaliplatino y tetraxetaxel (docetaxel).

podría contribuir a estrategias terapéuticas personalizadas.

Un estudio reciente de China<sup>15</sup> comparó el rendimiento de ChatGPT-4o y Gemini Advanced® en la toma de decisiones para cáncer gástrico avanzado, utilizando tres enfoques complementarios: preguntas clínicas clave, pacientes evaluados en reuniones de equipo multidisciplinario y casos raros publicados en PubMed. En todos los escenarios, ChatGPT-4o mostró un desempeño superior, con mayor precisión y completitud de las recomendaciones, mejor alineamiento con guías clínicas y mayor capacidad para adaptarse a perfiles complejos de pacientes. La tasa de alucinaciones también fue menor en ChatGPT-4o (5,25%) en comparación con Gemini (8,42%), sin implicar riesgos clínicos inmediatos. Si bien estos resultados refuerzan el potencial de los modelos conversacionales como apoyo a la toma de decisiones, el propio estudio destaca que, debido a las limitaciones inherentes de las IA y la ausencia de guías formales de uso, estas herramientas deben integrarse únicamente bajo supervisión experta y en el marco de contextos clínicos reales.

Es importante señalar que ChatGPT no es un sistema de IA diseñado específicamente para uso médico, ni cuenta con certificación como dispositivo médico. Se trata de un modelo de lenguaje generalista, cuya aplicación en el ámbito sanitario debe realizarse bajo supervisión profesional y no sustituye el juicio clínico especializado. Sin embargo, su amplia disponibilidad, fácil acceso y utilización, lo convierten en una herramienta atractiva para explorar su potencial como asistente en la práctica clínica real, especialmente en contextos donde no se dispone de sistemas de IA médica específicos.

La evaluación clínica inicial del cáncer gástrico o esofágico permite clasificar a los pacientes en cuatro grupos<sup>7,8</sup>.

- 1) tumores tempranos, resecables por endoscopia;
- 2) tumores pequeños no endoscópicamente resecables, sin adenopatías ni metástasis, candidatos a cirugía primaria;
- 3) enfermedad localmente avanzada, con invasión profunda y/o adenopatías locorregionales, que se beneficia de tratamiento neoadyuvante o perioperatorio previo a la cirugía, y
- 4) enfermedad metastásica o con adenopatías fuera del territorio regional (estadio IV), cuya conducta es paliativa.

Una correcta clasificación, junto con el estado general y las comorbilidades, orienta la selección del tratamiento más adecuado.

En nuestra serie, tanto el equipo médico como el *chatbot* enfrentaron dificultades para discriminar entre T2-T3 y T3-T4a en algunos casos, dado que la interpretación dependió exclusivamente de los informes de TC. Esto se debe a que, si bien la TC es el método más utilizado en la práctica cotidiana para estadificar

cáncer gástrico<sup>16</sup>, como los tumores T2 y T3/4 se reconocen por signos transmurales, la diferenciación puede ser difícil, particularmente entre T3 y T4a, donde la serosa no se visualiza directamente y el tejido graso perigástrico varía entre pacientes. Algunos hallazgos como irregularidad de la superficie externa, opacidad de la grasa perigástrica o hiperatenuación serosa sugieren T4a, mientras que la extensión franca a órganos adyacentes orienta a T4b<sup>17,18</sup>.

La evaluación del compromiso ganglionar es clave para definir la necesidad de terapia neoadyuvante, aunque no existe un método de referencia (gold standard) en cáncer gástrico y de la UEG. La ecoendoscopia presenta una exactitud variable (30-90%), con baja sensibilidad en estaciones 7-12 y marcada dependencia del operador, lo que limita su disponibilidad<sup>19</sup>. La tomografía computarizada tiene una tasa considerable de falsos negativos, especialmente en tumores difusos o mixtos, y utiliza criterios morfológicos (tamaño, forma y realce), siendo el tamaño el más relevante.

La PET-TC aporta poco en tumores gástricos difusos debido a la baja avidéz por FDG, dificultando aún más la estadificación ganglionar<sup>20</sup>. Para metástasis a distancia, la TC continúa siendo el estudio de elección, aunque su sensibilidad y especificidad para enfermedad peritoneal son moderadas (66% y 77%). Los hallazgos sugestivos incluyen ascitis, nódulos o placas peritoneales, engrosamiento sobre intestino delgado o hiperrealce del peritoneo.

En nuestra práctica, la estadificación se realiza con TC con contraste oral e intravenoso. No contamos con ecoendoscopia y la PET-TC no se solicita de rutina en tumores gástricos, aunque algunos pacientes la aportan desde otras instituciones; en contraste, sí la utilizamos con mayor frecuencia en cáncer de esófago y de la UEG.

Respecto a la estadificación clínica del cáncer de esófago, la ecoendoscopia parece ser precisa para el T, con una sensibilidad y especificidad del 85 y 87% para predecir el T1a. En cuanto a los ganglios linfáticos, la precisión es más variable, pero puede mejorarse realizando punción con aguja fina de los ganglios linfáticos<sup>21</sup>.

La tomografía computarizada es el estudio más utilizado para evaluar el cáncer de esófago, pero tiene una baja sensibilidad en los estadios tempranos, y una baja sensibilidad y especificidad para evaluar los ganglios linfáticos, con falsos negativos en casos de micrometástasis, y de falsos positivos en presencia de infecciones o patologías inflamatorias. La PET-TC es más sensible para identificar los tumores primarios, la diseminación ganglionar y metástasis a distancia en este tipo de cáncer, que la tomografía. Sin embargo, la baja resolución espacial genera que esta no sea fiable para evaluar la extensión y profundidad de la lesión (T). Respecto del N (ganglios linfáticos), puede tener pobre identificación en los ganglios adyacentes al tumor, sin

diferenciarlos de él. Además, las micrometástasis pueden no captar el FDG. Mientras que, para el M (metástasis), la PET-TC tiene la mejor performance, con una sensibilidad del 71% y una especificidad del 93%.

En nuestra serie, la estadificación se basó principalmente en TC y PET-TC, dado que la ecoendoscopia no estaba sistemáticamente disponible al momento del diagnóstico. Aun así, el *chatbot* logró una concordancia adecuada con el comité de tumores, lo que sugiere que estas herramientas pueden integrarse a la práctica real incluso cuando no se cuenta con todos los métodos diagnósticos ideales.

Este estudio presenta varias limitaciones. Su diseño retrospectivo implica heterogeneidad y posible pérdida de datos, sumado a la variabilidad en los informes de estudios realizados en distintos centros, lo que dificulta la estandarización. Además, el análisis se basó en informes escritos y no en imágenes originales, lo que impide una reevaluación objetiva por parte del modelo conversacional.

Ningún paciente contó con ecoendoscopia (EUS), método de referencia para evaluar profundidad

tumoral y compromiso ganglionar, lo que podría haber limitado la precisión del cTNM en algunos casos. Sin embargo, la baja disponibilidad de esta tecnología en la Argentina, hace que los resultados sean pragmáticos y extrapolables a la mayoría de los centros locales. Asimismo, al tratarse de un estudio de un único centro de tercer nivel, los resultados no son generalizables sin validación externa, requisito indispensable para modelos de IA<sup>22</sup>. Finalmente, el tamaño muestral reducido limita el poder estadístico y la detección de patrones infrecuentes.

En conclusión, sobre la base de los resultados presentados, es posible afirmar que la IA, aplicada mediante un *chatbot* entrenado con guías clínicas, demostró utilidad para apoyar la estadificación y la planificación terapéutica inicial del cáncer esofagogástrico. Aun con las limitaciones del estudio, se observó una concordancia elevada con las decisiones del equipo tratante. Estos resultados posicionan a los modelos conversacionales como herramientas prometedoras y accesibles, cuya integración debe continuar evaluándose en estudios prospectivos y multicéntricos.

## ■ ENGLISH VERSION

### Introduction

Artificial Intelligence (AI) has emerged as a transformative medical tool, particularly in oncology, demonstrating its ability to improve diagnostic accuracy, optimize treatment planning, and support personalized decision-making<sup>1</sup>. Broadly speaking, AI refers to systems capable of replicating human cognitive functions, such as learning and reasoning<sup>2</sup>. In this field, natural language models enable conversational assistants—chatbots—to interpret and interact with complex clinical language.

Advances in deep learning have expanded the ability of these technologies to analyze diverse information—including images, free-text data, and structured clinical data—and link it to clinical practice guidelines. In the field of gastrointestinal oncology, notable successes have been made in the early detection of gastroesophageal cancer, the automated identification of endoscopic lesions, and the prediction of treatment response<sup>3</sup>. However, little evidence exists regarding the use of conversational models to develop assistants that can help with staging and initial treatment planning.

In Argentina, evidence regarding the use of AI in medicine is still limited. Studies have been published on the use of AI to identify critical view of safety during cholecystectomy procedures at Hospital Argerich<sup>4</sup>, of ChatGPT for colorectal cancer screening recommendations<sup>5</sup>, and on implementing an AI co-pilot integrated into electronic medical records at Hospital Italiano de Buenos Aires<sup>6</sup> to improve access to relevant patient information.

In this context, the primary objective of our study was to assess the accuracy of an AI-driven conversational model in predicting clinical staging (cTNM and overall stage) using diagnostic data available at the initial oncology consultation for patients with gastroesophageal cancer. The secondary objective was to analyze the concordance between the treatment recommendations suggested by the chatbot and those provided by the tumor board or the treating team.

### Material and methods

We conducted a retrospective study using the database of patients who had their first consultation for a cancer-related condition at the institution's gastroesophageal surgery clinic between December 2023 and May 2025.

The study included patients with complete preoperative workups, defined as patients diagnosed with esophageal, gastroesophageal junction (GEJ), or gastric tumors, an upper gastrointestinal endoscopy report, a biopsy with pathology report, computed tomography (CT) scan, and positron emission tomography-computed tomography (PET-CT) in patients with esophageal cancer or tumors of the GEJ.

Patients who did not undergo all the preoperative tests mentioned or whose histology was other than adenocarcinoma or squamous cell carcinoma were excluded.

Treatment decisions were made by a multidisciplinary team of esophageal cancer surgeons,

clinical oncologists, anesthesiologists, general practitioners, and nutritionists.

### **Development and use of the virtual assistant**

For automated analysis, a virtual clinical assistant was developed using ChatGPT (OpenAI), version GPT-4.5<sup>®</sup>, via a subscription to ChatGPT Plus<sup>®</sup>. The development and configuration of the assistant were carried out by surgeons trained in gastroesophageal surgery and with expertise in tumor staging. Although these surgeons lack formal training in programming or algorithm development, they have received training in AI related to the healthcare field. Therefore, they were able to structure the use of the model in accordance with appropriate clinical and methodological criteria.

The chatbot was specifically configured for this task through data ingestion and contextual parsing of the updated clinical guidelines from the National Comprehensive Cancer Network (NCCN)<sup>7,8</sup> and the European Society for Medical Oncology (ESMO)<sup>9,10</sup> on esophageal and gastric cancer.

The model was trained to process each patient's clinical information in natural language, simulating an initial oncology consultation. It should be noted that, for this purpose, the algorithm was not retrained or modified. We used prompt engineering to define structured instructions based on the aforementioned international guidelines. Based on the provided data (age, sex, comorbidities, reason for consultation, endoscopy reports, pathology reports, and imaging results), the assistant was required to: a) determine the clinical stage of the tumor according to the 8th edition of the AJCC staging manual<sup>11</sup>; and b) suggest the appropriate initial cancer treatment in accordance with international guidelines.

Prior to the final analysis, pilot tests were conducted using historical cases excluded from the final cohort to standardize data entry formats, refine instructional prompting, and ensure response consistency. There were no substantial changes to the instructions during the analysis of the cases included in the study.

Each case study was processed individually, uploading the clinical data to the platform and recording the AI-generated responses in a structured table. The recommendations made by the assistant were then compared with the decisions made by the tumor board or the treating team in routine clinical practice.

In cases where discrepancies in staging or treatment were identified, targeted interactions with the chatbot were performed to identify their cause. We assessed whether these arose from a valid alternative interpretation, an omission of relevant data, or a limitation in the model's contextual understanding of the test results.

### **Statistical analysis**

The variables analyzed included:

1) demographic data: age, sex, comorbidities and reasons for consultation; 2) tumor data: location, histopathologic characteristics, endoscopy reports, CT-scan and PET-CT results, cTNM, clinical stage; 3) treatment: endoscopic treatment, surgery without neoadjuvant therapy, surgery with neoadjuvant therapy or perioperative chemotherapy, palliative treatment.

A descriptive analysis was conducted using the data collected. Categorical variables were expressed as absolute frequencies and percentages and continuous variables as mean and range.

Concordance between the clinical staging (cTNM and overall stage) proposed by the chatbot and that determined by the medical team was assessed via direct case-by-case comparison. Similarly, the concordance in the suggested therapeutic approach was analyzed.

The study was designed based on the hypothesis that an AI conversational model, trained using updated clinical guidelines, could correctly predict clinical staging in at least 80% of cases and suggest treatments in line with international recommendations. Given the exploratory nature of the study and the sample size, inferential statistical tests were not performed.

The accuracy of the model was calculated by dividing the number of cases where the chatbot's response matched the medical team staging or therapeutic strategy by the total number of evaluated cases. The following formula was used:

$$\text{Accuracy (\%)} = (\text{Concordant cases} / \text{Total cases}) \times 100$$

Accuracy was calculated separately for cTNM, clinical stage, and recommended treatment.

Ethical considerations: The study was conducted following the recommendations of the Declaration of Helsinki. As the study was observational and based on the analysis of retrospective anonymous clinical data, an informed consent was not required

### **Results**

Of the 53 patients who sought medical care for esophageal, gastric, or GEJ tumors during the specified period, 19 were excluded due to an absence of relevant data in their medical records, resulting in 34 cases that met the inclusion criteria. Mean age was 67 years (range 51-80) and 51% (n = 18) were men. The reason for consultation was documented in 29 patients and is presented in Table 1, along with the other demographic data and tumor location.

Three patients were in stage I, 7 in stage II, 17 in

■ TABLE 1

Demographic data and tumor characteristics	
Variables	n (%)
Comorbidities:	
Diabetes mellitus	11 (32%)
HTN	19 (56%)
COPD	6 (18%)
Current or former smoker	9 (27%)
Cardiovascular disease	9 (27%)
Breast cancer	6 (18%)
Other cancers	4 (11%)
Reason for consultation	
Gastrointestinal bleeding	6 (18%)
Dysphagia	16 (47%)
Dyspepsia or epigastric pain	14 (41%)
Pyloric syndrome or vomiting	9 (26%)
Weight loss	
Tumor location	
Esophagus	10 (29%)
Stomach	13 (38%)
Gastroesophageal junction	11 (32%)
- Siewert I	2 (18%)
- Siewert II	6 (54%)
- Siewert III	3 (27%)
Anatomical pathology	
Signet ring cells	
Esophagus	1 (10%)
Gastroesophageal junction	3 (27%)
Stomach	4 (31%)
Squamous cell carcinoma*	5 (50%)

\*Of total esophageal cancers.

COPD: chronic obstructive pulmonary disease; HTN: hypertension.

■ TABLE 2

Comparison between clinical staging determined by the medical team and the chatbot

cTNM**	Medical team	Chatbot
cT		
cT1b	1(3%)	1 (3%)
cT1b-2	1(3%)	1 (3%)
cT2-3	3(9%)	3 (9%)
cT3	21(62%)	21 (62%)
cT3-4	4(12%)	4 (12%)
cT4	4(12%)	4 (12%)
cN		
cN0-1	1(3%)	1 (3%)
cN+	20(59%)	21 (62%)
cM		
cM1	7(21%)	6(18 %)
Stage		
I	3(9%)	3 (9%)
II	7(20%)	6 (18%)
III	17(50%)	19 (56%)
IV	7(20%)	6 (18%)

stage III, and 7 in stage IV. The rest of the characteristics are described in Table 2.

The accuracy between the cTNM staging determined by the medical team and that of the chatbot was 94%. Concordance of the treatment proposed was 85%.

Discordant cases are presented in Table 3.

When the cohort was divided into three sequential groups (n = 12, 11, and 11 patients, respectively), the accuracy of the chatbot to predict cTNM and staging improved to 83% in group 1, 100% in group 2, and 100% in group 3.

## Discussion

The present study evaluated the usefulness of an AI conversational model, trained using up-to-date clinical guidelines, to assist with staging and initial treatment planning in patients with esophageal, gastroesophageal junction, and gastric cancer. The results show a high level of concordance (94%) between the clinical staging proposed by the chatbot and that determined by the treating medical team, both in terms of cTNM and overall stage. Although concordance in therapeutic management was somewhat lower (85%), the differences were mostly explained by contextual clinical considerations that were not always available to the model, such as social constraints.

Chatbots are programs that process and simulate human conversations using natural language<sup>12</sup>. The use of chatbots in healthcare increased during the COVID-19 pandemic, although it remains underutilized and is expected to increase over the next decade. In the field of gastric cancer, various artificial intelligence models have demonstrated promising accuracy in identifying high-risk individuals and predicting tumor aggressiveness, recurrence, and survival<sup>13,14</sup>. Deep learning has demonstrated the ability to analyze endoscopic images, genomic data, and histopathological data, thereby improving diagnosis and prognosis. Integrating multimodal clinical, genetic, and imaging data provides a comprehensive disease profile, facilitating the implementation of personalized treatment strategies.

A recent study from China<sup>15</sup> compared the performance of ChatGPT-4o and Gemini Advanced® in decision-making for advanced gastric cancer. The study utilized three complementary approaches: key clinical questions, patients evaluated in multidisciplinary team meetings, and rare cases published in PubMed. ChatGPT-4o demonstrated superior performance in all scenarios, providing more accurate and comprehensive recommendations, better aligning with clinical guidelines, and demonstrating a greater ability to adapt to complex patient profiles. The hallucination rate was also lower in ChatGPT-4o (5.25%) compared to Gemini (8.42%), without implying immediate clinical risks. While

■ TABLE 3

Discordant cases between the medical team and the chatbot

Case	Discrepancy category	Reason	Medical team	chatbot
1	cTNM and clinical stage	Lateral aortic lymph nodes on imaging studies	Considers M1, stage IV	Considers regional lymph nodes, M0, stage III
1	Treatment	Discrepancy of clinical stage (IV vs. III)	Palliative treatment	Neoadjuvant therapy and surgery
2	cTNM and clinical stage	GEJ tumor, without uptake in lymph nodes on PET-CT. The chatbot considers it to be N+ based on the tumor's length (5.8 cm) and SUV of 13.1	Clinical stage: cT3N0M0 (EII)	Clinical stage: and chatbot cT3N1M0 (EIII)
3	Treatment	Female, 63 years old, HTN, and kidney transplant; squamous cell carcinoma of the middle third of the esophagus	CROSS and esophagectomy	Chemotherapy Definitive radiation therapy
4	Treatment	Female, 80 years old, admitted to the emergency department with gastrointestinal bleeding and dyspnea. Gastric adenocarcinoma cT3N+M0	Prehabilitation for 3 weeks and gastrectomy	Palliative treatment
5	Treatment	Male, 65 years old, diffuse gastric cancer (linitis plastica). Intolerance to oral feeding. Absence of social security coverage.	Admission for prehabilitation for 3 weeks (parenteral and enteral nutrition) and gastrectomy	FLOT and gastrectomy
6	Treatment	Female, 51 years old, adenocarcinoma of the GEJ cT3N+M0 Discrepancy with Siewert	FLOT and gastrectomy	CROSS and esophagectomy
7*	Type of surgery	Gastric body adenocarcinoma	FLOT and total gastrectomy	FLOT and distal gastrectomy

\*This case is not considered a formal discrepancy (the treatment plan is correct), but it is mentioned because the type of gastrectomy differs, which is important to review to ensure adequate oncological resection.

GEJ: gastroesophageal junction; PET: positron emission tomography; SUV: standardized uptake value; HTM: hypertension; CROSS: chemoradiotherapy for esophageal cancer followed by surgery; FLOT: perioperative chemotherapy with 5-fluorouracil, leucovorin, oxaliplatin, and docetaxel.

these findings underscore the potential of conversational models to support decision-making, the study also highlights that, due to the inherent limitations of AI and the lack of formal guidelines for their use, these tools should be implemented only under expert supervision and within the context of real-world clinical settings.

It is noteworthy that ChatGPT is not an AI system designed specifically for medical use, nor is it certified as a medical device. It is a general-purpose language model; its use in the healthcare setting must be carried out under professional supervision and does not replace specialized clinical judgment. However, its high availability, accessibility, and ease of use make it an attractive tool for exploring its potential in real-world clinical practice, particularly where specialized medical AI systems are unavailable.

Initial clinical evaluation of gastric or esophageal cancer allows for patient classification into four groups<sup>7,8</sup>:

- 1) early-stage tumors, amenable to endoscopic resection;
- 2) localized tumors not suitable for endoscopic resection, without lymph node involvement or metastasis, eligible for primary surgery;
- 3) locally advanced disease, characterized by deep invasion and/or locoregional lymph node involvement, benefiting from neoadjuvant or perioperative treatment prior to surgery; and
- 4) metastatic disease or lymph node involvement extending beyond the regional territory (stage IV), where management is primarily palliative.

An accurate staging, along with the patient's performance status and comorbidities, guides the selection of the most appropriate treatment.

In our series, both the medical team and the chatbot encountered difficulties in distinguishing between T2-T3 and T3-T4a in some cases, since the interpretation relied solely on CT reports. Although CT is the most commonly used method for staging gastric cancer<sup>16</sup>, T2 and T3/4 tumors are identified by transmural features. Consequently, differentiation can be challenging, especially between T3 and T4a tumors, as the serosa is not directly visualized and the amount of perigastric fat varies among patients. Certain findings, such as an irregular outer layer of the gastric wall, haziness of the perigastric fat and the hyperattenuating serosa sign suggest T4a, whereas clear invasion of adjacent organs suggests T4b<sup>17,18</sup>.

Assessment of lymph node involvement is key to determining the need for neoadjuvant therapy, although there is no gold standard for gastric and EGJ cancer. The accuracy of endoscopic ultrasound is variable (30–90%), with low sensitivity for stations 7–12 and significant operator dependence, which limits its availability<sup>19</sup>. Computed tomography has a significant false-negative rate, particularly in diffuse or mixed-type tumors. It relies on morphological criteria (size, shape, and enhancement), with size remaining the primary diagnostic factor.

PET-CT provides little information in diffuse gastric tumors due to their low FDG uptake, thereby further complicating lymph node staging<sup>20</sup>. For

detecting distant metastases, CT remains the standard imaging modality, although its sensitivity and specificity for peritoneal involvement are moderate (66% and 77%, respectively). Findings suggestive of peritoneal seeding include ascites, peritoneal nodules or plaques, bowel wall thickening, or peritoneal hyperenhancement.

In our practice, we used oral and intravenous contrast-enhanced CT scans for staging. Endoscopic ultrasound is not available in our setting, and we do not routinely request PET-CT scans for gastric tumors, although some patients bring results from other institutions. We do request PET-CT scans more frequently for esophageal and GEJ cancers.

Endoscopic ultrasound demonstrates high accuracy for staging esophageal cancer T-status, with a sensitivity and specificity of 85% and 87%, respectively, for identifying T1a lesions. The accuracy of endoscopic ultrasound for detecting lymph node involvement is more variable; however, it can be significantly enhanced by incorporating fine-needle aspiration<sup>21</sup>.

While CT is the most widely used imaging modality for esophageal cancer, it lacks sensitivity for early-stage disease and demonstrates poor diagnostic performance for lymph node assessment, with high false-negative rates in the presence of micrometastases and false-positive rates due to concurrent infection or inflammation. PET-CT is more sensitive for identifying primary tumors, lymph node involvement, and distant metastases in this type of cancer. However, the low spatial resolution makes it unreliable for assessing the extent and depth of the lesion (T). Regarding N-staging (lymph nodes), PET-CT may face challenges in detecting lymph nodes in close proximity to the primary tumor, as it often fails to distinguish them from the primary lesion. In addition, micrometastases may not uptake FDG. PET-CT scans perform best for M-staging (metastasis), with a sensitivity of 71% and a specificity of 93%.

In our series, staging was primarily based on CT and PET-CT scans, since endoscopic ultrasound was not routinely available at the time of diagnosis. Even so, the chatbot achieved adequate concordance with the tumor board, suggesting that these tools can be incorporated into real-world practice even when not all the ideal diagnostic methods are available.

This study has some limitations. Its retrospective design suggests heterogeneity and the possibility of data loss, along with variability in the reporting of studies conducted at different centers, which hinders standardization. Furthermore, the analysis was based on written reports rather than original images, which prevents the conversational model from conducting an objective reassessment.

No patients underwent endoscopic ultrasound (EUS), the gold standard for assessing tumor depth and lymph node involvement, which may have limited the accuracy of the cTNM staging in some cases. However, the limited availability of this technology in Argentina means that the results are pragmatic and can be extrapolated to most local centers. Furthermore, since this is a single-center tertiary-level study, the results cannot be generalized without external validation, which is an essential requirement for AI models<sup>22</sup>. Finally, the small sample size limits the statistical power and the ability to detect rare patterns.

In conclusion, based on the results presented, it can be stated that AI, implemented through a chatbot trained using clinical guidelines, proved to be useful in supporting staging and initial treatment planning for gastroesophageal cancer. Despite the mentioned limitations, the study demonstrated high concordance with the treatment decisions made by the treating medical team. These findings position conversational models as promising and accessible tools, and their integration should be further evaluated in prospective, multicenter studies.

## Referencias bibliográficas /References

1. Marra A, Morganti S, Pareja F, Campanella G, Bibeau F, Fuchs T, et al. Artificial intelligence entering the pathology arena in oncology: current applications and future perspectives. *Ann Oncol*. 2025;36(7):712-25. doi:10.1016/j.annonc.2025.03.006.
2. Chong PL, Vaigeshwari V, Mohammed Reyasudin BK, Noor Hidayah BRA, Tatchanaamoorti P, Yeow JA, et al. Integrating artificial intelligence in healthcare: applications, challenges, and future directions. *Future Sci OA*. 2025;11(1):2527505. doi:10.1080/20565623.2025.2527505.
3. Chen ZL, Wang C, Wang F. Revolutionizing gastroenterology and hepatology with artificial intelligence: From precision diagnosis to equitable healthcare through interdisciplinary practice. *World J Gastroenterol*. 2025;31(24):108021. doi:10.3748/wjg.v31.i24.108021.
4. Petracchi EJ, Olivieri SE, Varela J, Canullán CM, Zandalazini H, Ocampo C, et al. Use of artificial intelligence in the detection of the critical view of safety during laparoscopic cholecystectomy. *J Gastrointest Surg*. 2024;28(6):877-9. doi:10.1016/j.gasur.2024.03.018.
5. Pereyra L, Schlottmann F, Steinberg L, Lasa J. Colorectal Cancer Prevention: Is Chat Generative Pretrained Transformer (Chat GPT) ready to Assist Physicians in Determining Appropriate Screening and Surveillance Recommendations?. *J Clin Gastroenterol*. 2024;58(10):1022-7. doi:10.1097/MCG.0000000000001979.

# Abdomen agudo suboclusivo por intususcepción como complicación tardía en *bypass* gástrico en Y de Roux

## Partial bowel obstruction caused by intussusception as a late complication following Roux-en-Y gastric bypass surgery

María C. Gribaudo , Franco W. Lara , Luciano Rollan , Gustavo G. Del Cantare 

Clínica Regional del Sud,  
Río Cuarto, Córdoba,  
Argentina

### RESUMEN

La intususcepción es más frecuente en pacientes pediátricos, y su presencia en adultos es de baja incidencia, principalmente relacionada con divertículos, pólipos, tumores, cirugías previas, etc. La resolución quirúrgica temprana se plantea como tratamiento de elección. Se presenta una mujer de 44 años, con antecedentes de *bypass* gástrico en Y de Roux, que ingresa por sospecha clínica de abdomen agudo suboclusivo, sugestivo de intususcepción por tomografía computarizada. Se realiza una laparoscopia exploradora, diagnóstica y terapéutica, en la cual se reduce el segmento del asa alimentaria invaginada a diez centímetros de pie de Y de Roux, con vitalidad y motilidad conservada. En el posquirúrgico presenta buena evolución, con alta a las cuarenta y ocho horas sin complicaciones alejadas. Se considera la laparoscopia como la mejor vía de abordaje por presentar buena evolución posquirúrgica, disminución de íleo, rápida inserción laboral, menor dolor y mejores resultados estéticos, con respecto a la vía convencional.

■ **Palabras clave:** intususcepción, *bypass* gástrico, laparoscopia.

Los autores declaran no tener conflictos de interés.  
*Conflicts of interest*  
None declared.

Correspondencia  
*Correspondence:*  
María C. Gribaudo.  
E-mail:  
gribaudoconsuelo@gmail.com

### ABSTRACT

Intussusception is more common in pediatric patients, and its incidence in adults is low, typically associated with diverticula, polyps, tumors, or a history of previous surgeries. Early surgical intervention is considered the treatment of choice. A 44-year-old female patient with a history of Roux-en-Y gastric bypass is admitted with a clinical diagnosis of partial bowel obstruction, suggestive of intussusception based on computed tomography scan findings. An exploratory laparoscopy was performed for diagnostic and therapeutic purposes. An invaginated segment of the alimentary loop was found 10 centimeters from the base of the Roux-en-Y anastomosis, with preserved blood supply and motility, and was reduced. The postoperative period was uneventful, and the patient was discharged 48 hours after the procedure. There were no late complications. Laparoscopy is considered the preferred approach because it offers better postoperative outcomes, reduced postoperative ileus, a faster return to work, reduced pain, and enhanced cosmetic results compared to the conventional approach.

■ **Keywords:** intussusception, gastric bypass, laparoscopy

Recibido | Received  
06-06-25  
Aceptado | Accepted  
16-10-25

ID ORCID: María C. Gribaudo, 0009-0007-8166-0893; Franco W. Lara, 0009-0003-8159-9456; Luciano Rollan, 0009-0006-9101-1531; Gustavo G. Del Cantare, 0009-0002-9312-5138.

La obesidad es un problema de salud importante, con una prevalencia mundial en adultos del 13% y un estimado de 4 millones de muertes anuales relacionadas con el sobrepeso y la obesidad<sup>1,2</sup>. Existen diferentes técnicas quirúrgicas entre las cuales se encuentra el *bypass* gástrico en Y de Roux, y la invaginación intestinal o intususcepción como una complicación poco frecuente. Se define como la introducción de un segmento intestinal en otro, semejando un telescopio (ya sea anterógrada o retrógrada, la primera la más frecuente).

Si bien la intususcepción en adultos es poco frecuente, con una incidencia estimada del 0,1 y el 1,2%<sup>3</sup> respecto de la población pediátrica, a menudo es patológica debido a la presencia de puntos de partida tales como pólipos, divertículo de Meckel y

neoplasias<sup>4,5</sup>, sumada a un meso complaciente que permita el deslizamiento de un asa sobre la otra. Curiosamente, en pacientes posquirúrgicos de *bypass* gástrico en Y de Roux, se encuentra con mayor frecuencia la forma retrógrada<sup>6</sup>.

El diagnóstico clínico es limitado debido a la inespecificidad de los síntomas, los hallazgos de la exploración física y los resultados de laboratorio. Se diagnostica mediante radiografía, tomografía computarizada con contraste (TCC) o como hallazgo intraquirúrgico<sup>5</sup>.

Presentamos el caso de una mujer de 44 años que consultó por Guardia por presentar vómitos incoercibles de características fecaloides de seis horas de evolución acompañados de dolor abdominal generalizado.

Como antecedentes patológicos presentaba

hipotiroidismo, síndrome anémico y constipación crónica, antecedentes quirúrgicos de bypass gástrico en Y de Roux ocho años antes, colecistectomía convencional con instrumentación de la vía biliar, dermoplastia, mastoplastia de aumento y hernioplastia hiatal.

En el examen físico presentaba dolor abdominal generalizado a la palpación superficial con mayor intensidad en hipocondrio izquierdo y epigastrio.

Se colocó una sonda nasogástrica y se solicitó análisis de laboratorio que no presentaron alteraciones, y tomografía computarizada con contraste intravenoso de abdomen y pelvis, que informó signos de oclusión intestinal a punto de partida de invaginación intestinal versus hernia interna (Fig. 1).

Se realizó una laparoscopia exploradora, en donde se evidencia intususcepción yeyuno yeyunal retrógrada de asa alimentaria a diez centímetros de pie de anastomosis en Y de Roux (Fig. 2A). Se liberó la intususcepción donde se observa yeyuno indemne (Fig. 2B).

La paciente presentó buena evolución posquirúrgica en sala de cuidados generales, con alimentación vía oral temprana a las veinticuatro horas y alta sanatoria a las cuarenta y ocho horas con evolución favorable hasta la actualidad.

La intususcepción, como causa de abdomen agudo oclusivo, es una patología poco frecuente en la población adulta. Cuando esta se presenta, se debe sospechar a punto de partida de neoplasias, pólipos y divertículos<sup>4</sup>. En pacientes sometidos a cirugías *bypass*

gástrico en Y de Roux, se debe tener en cuenta la línea de sutura como punto de partida para ocasionar la invaginación, así como también la dirección en la que se anastomosan las asas (isoperistáltico y anisoperistáltico). Otro factor para considerar es el tamaño de boca de anastomosis, debido a que a menor tamaño el intestino debe realizar mayor intensidad de propulsión para vencer dicha disminución de calibre y generar alteración en la motilidad, y favorece la intususcepción según se realice con sutura mecánica o manual. Por todo lo antes mencionado y lo descrito en la evidencia actual se deberían contemplar dichos aspectos en la primera cirugía sugiriendo realizar sutura mecánica anisoperistáltica para disminuir la incidencia de esta complicación y una boca anastomótica que no genere cambio abrupto de calibre. No hay evidencia que indique modificar el largo de asas ni preferencia en cuanto a realizar sutura mecánica versus manual<sup>5</sup>.

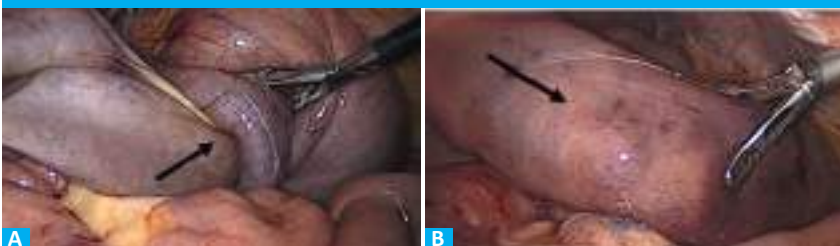
Debido a la gran variedad de causas que generan cuadros de abdomen agudo oclusivo, clínicamente similares, se sugiere realizar TCC para confirmar el diagnóstico ya que se considera el procedimiento de referencia (gold standard).<sup>5</sup> En la mayoría de los casos, la respuesta al tratamiento médico es insuficiente, por lo cual la cirugía es el tratamiento definitivo. Se prefiere, en este caso, el abordaje laparoscópico debido a la temprana recuperación de la motilidad intestinal (lo cual es importante en la etiopatogenia de esta patología) que se asocia a menor incidencia o gravedad del íleo posquirúrgico y más rápido reinicio de la alimentación vía oral; menor dolor que favorece la movilización

■ FIGURA 1



Tomografía computarizada con contraste intravenoso evidenciando "signo de la escarapela" en hemiabdomen izquierdo compatible con invaginación intestinal (señalado con flecha blanca).

■ FIGURA 2



A: intususcepción en laparoscopia exploradora (señalada con flecha negra). B: liberación de invaginación (señalada con flecha negra).

temprana del paciente; mejores resultados estéticos y reinserción laboral rápida; siempre teniendo en cuenta la experiencia del equipo quirúrgico y la disponibilidad de esta vía de abordaje. De ser posible, según la vitalidad intestinal del asa comprometida, se prefiere conservarla ya que la resección y anastomosis generarían un nuevo punto de partida para intususcepción y sus posibles complicaciones, como fugas o sangrados, además de una prolongación posquirúrgica. En caso de presentar compromiso de vitalidad del asa se debería optar por resecar el segmento para evitar perforaciones o isquemias. Cuando el equipo quirúrgico presente mayor experiencia en cirugía convencional, se sugiere abordaje por dicha vía para mayor seguridad y evitar complicaciones tempranas o tardías<sup>6</sup>.

En conclusión, el caso aquí presentado expone la intususcepción retrógrada como una complicación poco frecuente en el posquirúrgico tardío de *bypass* gástrico en Y de Roux, que pudo resultar satisfactoriamente por laparoscopia.

Es importante sospecharla durante el diagnóstico diferencial de abdomen agudo oclusivo en cirugías digestivas. Las diferentes alternativas terapéuticas incluyen tratamiento médico y expectante o resolución quirúrgica, ya sea por vía laparoscópica o convencional. Según lo expuesto en la bibliografía científica disponible, la vía laparoscópica se sugiere como la mejor alternativa dado que presenta múltiples beneficios para el paciente con respecto a la vía convencional, siempre teniendo en cuenta la experiencia del equipo tratante.

## ■ ENGLISH VERSION

Obesity is a major health problem. It has a global prevalence of 13% among adults and is estimated to cause 4 million deaths annually related to being overweight or obese<sup>1,2</sup>. Among the various surgical techniques, the Roux-en-Y gastric bypass is associated with intestinal invagination or intussusception, a rare complication. Intussusception is defined as the invagination of one segment of the bowel into an immediately adjacent segment of the bowel, resembling a telescoping motion. It can be antegrade or retrograde, with antegrade being more common.

Adult intussusception is rare, with an estimated incidence of 0.1% to 1.2%<sup>3</sup> compared to the pediatric population. It is often associated with lead points such as polyps, Meckel's diverticulum, and neoplasms<sup>4,5</sup>, combined with a compliant mesentery that allows one loop to slide over another. Interestingly, retrograde intussusception is more commonly found in patients who have undergone Roux-en-Y gastric bypass surgery<sup>6</sup>.

The clinical diagnosis is challenging because symptoms, physical examination signs, and laboratory tests are nonspecific. The diagnosis is made by X-ray, computed tomography (CT) scan with contrast agents, or as an intraoperative finding.

We present the case of a 44-year-old female patient who arrived at the emergency department with six hours of uncontrollable fecal vomiting accompanied by generalized abdominal pain.

She reported a history of hypothyroidism, anemia, and chronic constipation. She also reported undergoing Roux-en-Y gastric bypass surgery eight years prior, as well as conventional cholecystectomy with common bile duct exploration, dermoplasty, augmentation mammoplasty, and hiatal hernia repair.

The abdomen was tender on superficial palpation, particularly in the left hypochondriac and epigastric regions.

A nasogastric tube was inserted. Laboratory

test results were normal, and a CT scan of the abdomen and pelvis with intravenous contrast agent revealed signs of bowel obstruction caused by either intussusception or an internal hernia (Fig. 1).

An exploratory laparoscopy was performed, revealing a retrograde jejunojunal intussusception of the alimentary limb, ten centimeters from the foot of the Roux-en-Y anastomosis (Fig. 2A). The intussusception was reduced, revealing an intact jejunum (Fig. 2B).

The patient had a favorable postoperative course in the general ward, initiated early oral feeding within 24 hours, and was discharged 48 hours later, with a satisfactory outcome to date.

Intussusception is a rare cause of acute abdomen in adults, and its presence should raise suspicion of neoplasms, polyps or diverticula as lead points<sup>4</sup>. In patients with a history of surgery with enteroenteric anastomoses, such as Roux-en-Y gastric bypass, the suture line and the direction of the anastomosis (isoperistaltic or anisoperistaltic) must be considered lead points for intussusception. Another factor to consider is the size of the anastomotic orifice. A smaller size forces the intestine to exert greater propulsion intensity to overcome this luminal narrowing. This can lead to motility alterations and favor intussusception, depending on whether mechanical stapler or hand-sewn anastomosis is used. Consequently, based on these considerations and the current evidence, it is recommended that these factors be taken into account during the initial surgery employing anisoperistaltic mechanical stapling to minimize the occurrence of this complication and incorporating an anastomotic orifice that does not result in an abrupt change in diameter. There is no evidence to suggest that loop length should be modified or that mechanical stapler is better than hand-sewn sutures<sup>5</sup>.

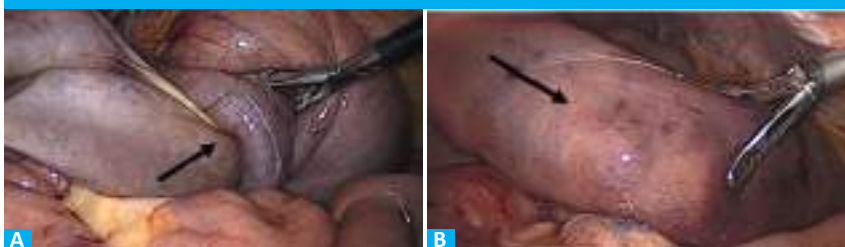
Given the variety of causes that produce clinically similar presentations of acute intestinal

■ FIGURE 1



Computed tomography with intravenous contrast showing the “target sign” in the left lower abdomen, consistent with intussusception (indicated by the white arrow).

■ FIGURE 2



A: Intussusception in exploratory laparoscopy. B: Invagination reduced (black arrow).

obstruction, a CT scan is recommended to confirm the diagnosis, as it is considered the gold standard<sup>5</sup>. In most cases, response to medical treatment is inadequate, making surgery the definitive treatment. In this case, the laparoscopic approach is preferred because it promote early recovery of bowel motility -an important factor in the pathogenesis of this condition- which is associated with a lower incidence of and less severe postoperative ileus, as well as a faster resumption to oral feeding. Other benefits include less pain, which promotes early patient mobilization; better cosmetic results; and a faster return to work. The surgical team expertise and the availability of this approach must always be taken into account. It is preferable to preserve the affected loop if its viability is not compromised, since resection and anastomosis would create a new site for intussusception and its potential complications, such as leaks or bleeding, as well as a prolonged postoperative recovery. If loop viability is compromised, the segment

should be resected to prevent perforation or ischemia. It is recommended that surgical teams with extensive experience in conventional surgery use this approach to ensure greater safety and avoid early or late complications<sup>6</sup>.

In conclusion, this case describes retrograde intussusception as a rare complication in the late postoperative period following Roux-en-Y gastric bypass, which was successfully treated via laparoscopy.

It is important to consider this condition in the differential diagnosis of acute obstructive abdomen in gastrointestinal surgery. Treatment options include medical or expectant management and surgical procedures, such as laparoscopy or conventional surgery. According to the available scientific literature, the laparoscopic approach is the preferred option because it offers patients numerous benefits compared with the conventional approach, always taking into account the expertise of the treatment team.

#### Referencias bibliográficas /References

- Hedberg S, Thorell A, Österberg J, Peltonen M, Andersson E, Näslund E, et al. Comparison of Sleeve Gastrectomy vs Roux-en-Y Gastric Bypass. *JAMA netw open*. 2024; 7 (1:e2353141). doi: 10.1001/jamanetworkopen.2023.53141.
- Adams TD, Davidson LE, Litwin SE, Kim J, Kolotkin RL, Nanjee MN, et al. Weight and metabolic outcomes 12 years after gastric bypass. *N Engl J Med*. 2017;377(12):1143-55. doi: 10.1056/NEJMoa1700459.
- Simper SC, Erzinger JM, McKinlay RD, Smith SC. Retrograde (reverse) jejunal intussusception might not be such a rare problem: a single group's experience of 23 cases. *Surg Obes Relat Dis*. 2008;4:77-83. doi: 10.1016/j.soard.2007.12.004
- Wong CSL, Ramadan M, Recurrent retrograde intussusception in patient with previous gastric bypass surgery. *J Surg Case Rep*. 2022;2022 (9:rjac414). doi: 10.1093/jscr/rjac414.
- Ghattas Y, Perry BL, Lester D, Charles A, Ang D. Small Bowel Recurrent Intussusception Status Post Gastric Bypass. *Cureus*. 2024;16 (10:e71230). doi: 10.7759/cureus.71230.
- Daellenbach L, Suter M. Jejunojejunal intussusception after Roux-en-Y gastric bypass: a review. *Obes Surg*. 2011;21:253-63. doi: 10.1007/s11695-010-0298-5.

# Peritonitis esclerosante encapsulante: informe de un caso de síndrome de cocoon, una causa inusual de obstrucción intestinal

## Encapsulating peritoneal sclerosis: a case report of cocoon syndrome, an unusual cause of bowel obstruction

Nemecio Lizana G.<sup>1</sup> , Xavier Sigüenza R.<sup>2</sup> , Rocío Bermúdez D.<sup>2</sup> , Roberto Briceño R.<sup>2</sup> 

1. Escuela de Medicina, Universidad de O'Higgins. Chile  
2. Servicio de Cirugía, Hospital Franco Ravera Zunino. Chile

Los autores declaran no tener conflictos de interés.

Conflicts of interest  
None declared.

Correspondencia  
Correspondence:  
Roberto Briceño  
E-mail:  
roberto.bri.re@gmail.com

### RESUMEN

La peritonitis esclerosante encapsulante (PEE), también llamada síndrome de cocoon, es una causa poco frecuente de obstrucción intestinal. Se caracteriza por encapsulamiento del intestino delgado por una membrana fibrosa. Presentamos el caso de un varón de 76 años, sin antecedentes de cirugía abdominal, que consultó por dolor abdominal de tipo cólico, vómitos y distensión abdominal. La tomografía axial computarizada (TAC) reveló obstrucción intestinal de alto grado. En la cirugía se hallaron asas de íleon envueltas en tejido fibroso, lo que confirmó una PEE tipo II. Se realizó liberación de adherencias y reparación de perforaciones intestinales intraoperatorias. El paciente evolucionó favorablemente. El diagnóstico suele ser complejo y con frecuencia se confirma durante la cirugía. La TAC permite orientar el tratamiento al mostrar asas dilatadas y membranas engrosadas. No existen guías estandarizadas. Sin embargo, en casos obstructivos, la cirugía es la intervención de elección. Este caso resalta la importancia de considerar PEE en pacientes sin antecedentes quirúrgicos.

■ **Palabras clave:** síndrome de cocoon, peritonitis encapsulante esclerosante, obstrucción intestinal, fibrosis peritoneal.

### ABSTRACT

Encapsulating peritoneal sclerosis (EPS), also known as cocoon syndrome, is a rare cause of bowel obstruction in which the small intestine is encased by a thin fibrous membrane. We report the case of a 76-year-old male patient with no history of abdominal surgery who presented with colicky abdominal pain, vomiting, and abdominal distension. A computed tomography (CT) scan revealed severe bowel obstruction. During surgery, loops of the ileum were found to be encased in fibrous tissue, confirming a type II EPS. The adhesions were released and intraoperative intestinal perforations were repaired. The patient evolved with favorable outcome. The diagnosis is often complex and is frequently confirmed during surgery. A CT scan helps guide treatment by revealing dilated loops and thickened membranes. There are no standardized international guidelines. In cases of obstruction, surgery is the treatment of choice. This case highlights the importance of considering EPS in patients with no history of surgery.

■ **Keywords:** cocoon syndrome, encapsulating peritoneal sclerosis, bowel obstruction, peritoneal fibrosis.

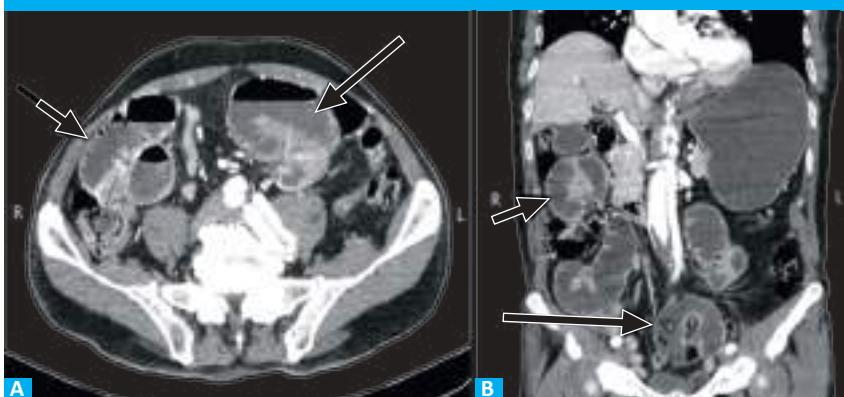
Recibido | Received 16-07-25  
Aceptado | Accepted 16-10-25  
ID ORCID: Nemecio Lizana G., 0000-0002-3824-4129; Xavier Sigüenza, 0009-0001-9067-3940; Rocío Bermúdez, 0000-0001-7488-4484; Roberto Briceño R. 0009-0006-8581-5373

La peritonitis esclerosante encapsulante (PEE) o síndrome de cocoon es una patología inflamatoria crónica que posee una etiología desconocida, cuyos síntomas incluyen el dolor abdominal intermitente y la obstrucción intestinal, ya sea en yeyuno o íleon<sup>1</sup>. Esto es provocado por tejido fibroso que envuelve el intestino.

Se presenta el caso de un varón de 76 años, con antecedentes mórbidos de hipertensión arterial, disección aórtica toracoabdominal y de arterias ilíacas comunes conocida y en tratamiento médico desde 2008, y antecedente quirúrgico de hernioplastia inguinal abierta izquierda en 2024. Acudió al Servicio de Urgencias por un cuadro de dolor abdominal de 24 horas de evolución, de tipo cólico, y progresivo en región epigástrica y periumbilical, asociado a vómitos profusos con características de retención, ausencia de deposiciones y gases por ano. En el examen físico se destaca un abdomen distendido, timpánico, con disminución de ruidos hidroaéreos, blando y depresible, sin signos de

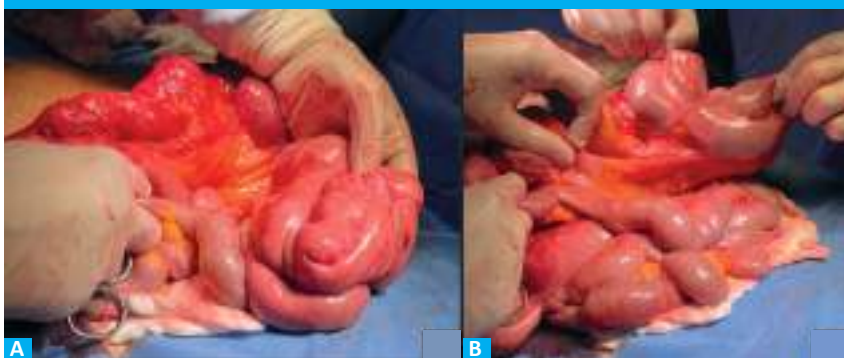
irritación peritoneal. Ante la sospecha de obstrucción intestinal, se coloca una sonda nasogástrica (SNG) que da salida a 800 mL de contenido gástrico. En los exámenes de laboratorio de urgencia se destacan PCR 1,47 mg/dL, leucocitos 10 200 mg/dL, creatinina sérica 1,25 mg/dL, BUN 29,1 mg/dL, sin trastorno hidroelectrolítico, y pruebas de coagulación sin alteraciones. La tomografía axial computarizada (TAC) de abdomen y pelvis con contraste (Fig. 1 a y b) destaca signos de oclusión intestinal mecánica de alto grado a la altura del hipogastrio sin signos de sufrimiento de asa, además de la presencia de colgajo de disección Stanford B sin cambios con relación a estudio previo realizado en 2017. Dados los hallazgos y sin antecedentes de cirugía abdominal, se decide ingresar en quirófano para una laparotomía exploradora. En la cirugía se evidencia un yeyuno vital pero dilatado, secundario a obstrucción distal en relación con 6 ovillos de íleon cubiertos por adherencias firmes blanquecinas (Fig. 2, a y b), que son liberadas con tijera y técnica roma. Sin embargo, a 150 cm

■ FIGURA 1



TAC preoperatoria donde se evidencian, en corte axial y coronal, signos de oclusión intestinal mecánica de alto grado, con cambio de calibre en la región hipogástrica en íleon medio sin signos de isquemia transmural de asas al momento del estudio; por otro lado, flap de disección aorto-iliaco crónico tipo B de Stanford. (En flechas negras se aprecian niveles de obstrucción intestinal debido a bandas fibrosas).

■ FIGURA 2



Imágenes durante la cirugía donde se aprecian ovillos de íleon distal cubiertos por adherencias firmes blanquecinas interasas.

del ángulo de Treitz durante estas maniobras, se produce una perforación intestinal que da salida a una moderada cantidad de líquido intestinal y es reparada con Monocryl 3-0<sup>®</sup>, (monofilamento absorbible), puntos sero-musculares. A 200 cm se evidencia otra perforación puntiforme que se repara de la misma forma. Durante el posoperatorio, el paciente evoluciona con íleo adinámico, pese a lo cual se logra retirar la SNG a los 4 días. Posteriormente tolera realimentación progresiva vía oral, presenta deposiciones y gases por ano. Es dado de alta a los 5 días del posoperatorio. En el control a los 18 días, presenta buen estado general, tolera adecuadamente régimen oral y realiza actividades diarias normalmente.

La PEE es una patología inflamatoria crónica de etiología desconocida que se cree que surge de cuadros recurrentes de peritonitis subclínica o de bajo grado que se presentan sin signos abdominales específicos<sup>1</sup>. Además, es una causa poco frecuente de obstrucción intestinal<sup>2</sup>. En el caso referido, la ausencia de signos clínicos específicos y la evolución hacia una obstrucción intestinal aguda coincidieron con esta presentación clínica poco frecuente.

Por otro lado, esta patología se caracteriza por presentar un intestino delgado cubierto por una membrana fibrosa o un saco en forma de capullo formando ovillos en el intestino<sup>2</sup>. Debido a su rareza y a la diversidad de sus causas, la incidencia y prevalencia globales

de la PEE no están bien documentadas al día de hoy<sup>1</sup>. Durante la cirugía realizada, se identificaron múltiples ovillos en el íleon distal cubiertos por una cápsula fibrosa, hallazgo característico que confirmó retrospectivamente el diagnóstico del PEE.

La PEE puede dividirse en dos tipos: la membrana del capullo de tipo I es fina, el tracto intestinal puede verse bajo la cápsula y no se forma una segunda cavidad abdominal; por otro lado, la membrana del capullo de tipo II es gruesa y forma la segunda cavidad abdominal, y el intestino bajo la cápsula no se ve claramente<sup>3</sup>. En este caso, la presencia de una cápsula fibrosa gruesa y transparente sugiere una presentación tipo II.

En general, el diagnóstico preoperatorio de esta patología es difícil, y en la mayoría de los casos, solo se establece durante la realización de la laparotomía exploradora<sup>4</sup>. La laparotomía fue determinante en este paciente, ya que permitió evidenciar el origen real de la obstrucción, que no fue sospechado clínicamente por los antecedentes del paciente.

Los síntomas más comunes dentro de esta patología son el dolor abdominal intermitente, que aparece en el 33,1% de los casos, y la obstrucción del intestino delgado en un 57,9%; estos síntomas se producen cuando el tejido fibroso comprime o envuelve el intestino. Tal compresión provoca síntomas similares al íleo, ya sean dolor abdominal intermitente o tipo

cólico, náuseas, vómitos, anorexia y pérdida de peso<sup>1,5</sup>. En nuestro caso, el paciente consultó por un cuadro de dolor abdominal de tipo cólico, vómitos y distensión abdominal, manifestaciones clínicas comunes en la PEE con compromiso obstructivo.

Es importante recalcar que el estudio de imágenes desempeña un papel crucial en el proceso diagnóstico de esta patología<sup>1</sup>. Es aquí donde la TAC adquiere un papel fundamental, al evidenciar hallazgos característicos como asas intestinales dilatadas y enredadas, envueltas por una membrana peritoneal engrosada (> 2 mm), presencia de adherencias y líquido ascítico interasas, proporcionando al equipo quirúrgico información valiosa respecto de la complejidad y duración esperada del procedimiento<sup>2</sup>. En este caso, la TAC permitió evidenciar signos de obstrucción de alto grado, orientando hacia la necesidad de la realización de una laparotomía exploradora.

La ausencia de guías estandarizadas para el tratamiento de la PEE, genera que los procedimientos quirúrgicos se realicen principalmente en casos complicados por obstrucción intestinal. Entre los principales objetivos de la cirugía se incluye la liberación completa de las bandas fibróticas a lo largo del intestino delgado y la raíz mesentérica. La resección intestinal se reserva

únicamente para segmentos no vitales, a fin de evitar agregar morbilidad al procedimiento con el fin de prevenir complicaciones como el síndrome de intestino corto<sup>5</sup>. La cirugía en este paciente se centró en liberar las adherencias firmes encontradas interasas, y evitar la resección intestinal.

La PEE representa un desafío diagnóstico y terapéutico debido a su baja incidencia, etiología idiopática y presentación clínica inespecífica. Debe ser descartado además en casos de dolor abdominal que no tengan un origen mecánico. En la presentación del caso clínico se evidenció cómo un cuadro de obstrucción intestinal en un paciente sin antecedentes de cirugías abdominales previas orientó, a través de la TAC y la confirmación quirúrgica, al diagnóstico de PEE tipo II. El estudio imagenológico resultó clave en la planificación quirúrgica, al sugerir la presencia de una obstrucción mecánica de alto grado.

La laparotomía exploradora realizada permitió tanto el diagnóstico definitivo como el tratamiento resolutorio. Este caso resalta la importancia de considerar la PEE o síndrome de cocoon dentro del diagnóstico diferencial de obstrucción intestinal, especialmente en ausencia de cirugía previa, lo que destaca la relevancia del enfoque quirúrgico oportuno para una evolución clínica favorable.

## ■ ENGLISH VERSION

Encapsulating peritoneal sclerosis (EPS), also known as cocoon syndrome, is a chronic inflammatory condition of unknown origin characterized by intermittent abdominal pain and bowel obstruction in either the jejunum or the ileum<sup>1</sup>. In this condition, the bowel is wrapped with a thick fibrotic membrane.

We report the case of a 76-year-old male patient with a history of hypertension, chronic dissection of the thoracoabdominal aorta and common iliac artery on medical treatment since 2008, and open left inguinal hernia repair in 2024. The patient presented in the emergency department with progressive colicky abdominal pain in the epigastric and periumbilical region that had started 24 hours before, associated with profuse retention vomiting, and inability to pass stool and gas. On physical examination, the abdomen was distended, tympanic, and soft with no rigidity or rebound tenderness. The bowel sounds were decreased. A nasogastric tube (NGT) was inserted due to suspected intestinal obstruction, and 800 mL of gastric contents were drained. The laboratory tests requested in the emergency department showed PCR 1.47 of mg/dL, white blood cell count of 10,200/mm<sup>3</sup>, serum creatinine levels of 1.25 mg/dL and BUN of 29.1 mg/dL. The electrolytes and coagulation tests were within normal ranges. A computed tomography (CT) scan of the abdomen and pelvis with contrast agent

(Fig. 1 a and b) revealed severe mechanical bowel obstruction in the hypogastrium, with no evidence of loop ischemia. A Stanford type B aortic dissection flap was also visible with no changes compared to a 2017 scan. Given these findings and the absence of a history of abdominal surgery, it was decided to proceed to the operating room for an exploratory laparotomy. During surgery, the jejunum was found to be vital, though dilated due to a distal obstruction involving six clusters of ileal loops covered by firm, whitish adhesions (Fig. 2a and 2b), which were released using a blunt technique with scissors. However, a bowel perforation occurred 150 cm from the angle of Treitz during these maneuvers, resulting in the leakage of a moderate amount of intestinal fluid. The perforation was repaired with absorbable monofilament Monocryl® sutures with seromuscular bites. Another punctiform perforation was found 200 cm from the angle, which was repaired in the same fashion. During the postoperative period, and despite the patient presented adynamic ileus, the NGT could be removed 4 days later. The patient tolerated the gradual reintroduction of oral feeding and began passing stool and gas through the anus. He was discharged on postoperative day 5. Eighteen days later, during the follow-up checkup, his performance status was good. He tolerated oral feeding normally and was able to perform daily activities as usual.

■ FIGURE 1



Preoperative CT scans in axial and coronal sections showing signs of severe mechanical bowel obstruction, with a diameter change in the mid-ileum at the level of the hypogastrium. There were no signs of transmural ischemia of the bowel loops when the scan was performed. A Stanford type B aorto-iliac chronic dissection flap is also evident.

■ FIGURE 2



Operative images of the distal ileum with clusters of ileal loops covered by firm, whitish adhesions interposed between bowel loops.

Encapsulating peritoneal sclerosis (EPS) is a chronic inflammatory condition of unknown origin. It is believed to develop from repeated episodes of subclinical or low-grade peritonitis without specific abdominal symptoms<sup>1</sup>. It is also an unusual cause of bowel obstruction<sup>2</sup>. In the case described, the absence of specific clinical signs and the progression to acute bowel obstruction were consistent with this rare clinical presentation.

In EPS, the small bowel is covered by a fibrous membrane or a cocoon-like sac, forming clusters in the intestine<sup>2</sup>. The overall incidence and prevalence of EPS have not been well documented yet due to its rarity and the variety of causes<sup>1</sup>. In our case, the surgical exploration identified multiple masses in the distal ileum covered by a fibrotic capsule—a characteristic that retrospectively confirmed the diagnosis of EPS.

Encapsulating peritoneal sclerosis can be categorized into two types: types I and II. Type I is characterized by a thin cocoon membrane, the presence of an intestinal tube beneath the capsule, and the absence of a second abdominal cavity. In contrast, type II EPS is distinguished by a thick cocoon membrane, which gives rise to a second abdominal cavity, and the intestinal tube beneath the capsule is not clearly visible<sup>3</sup>. In this case, the presence of a thick, transparent fibrous capsule suggests a type II presentation.

In general, the preoperative diagnosis of this condition is difficult, and in most cases, it is only confirmed during exploratory laparotomy<sup>4</sup>. In this patient, laparotomy was crucial, as it revealed the true origin of the obstruction, which had not been clinically suspected based on the patient's history.

The most common symptoms of this condition are intermittent abdominal pain, which occurs in 33.1% of cases, and small bowel obstruction in 57.9%. These symptoms occur when fibrous tissue compresses or encases the intestine. This compression causes symptoms similar to those observed in ileus, such as intermittent or colicky abdominal pain, nausea, vomiting, loss of appetite, and weight loss<sup>4,5</sup>. In this case, the patient presented with colicky abdominal pain, vomiting, and abdominal distension—clinical manifestations commonly seen in EPS with bowel obstruction.

It is important to emphasize that imaging tests play a crucial role in the workup process of this condition<sup>1</sup>. This is where CT scan plays a crucial role, as it reveals characteristic findings such as clusters of dilated and obstructed bowel loops, surrounded by a thickened peritoneal membrane (> 2 mm), adhesions, and interloop fluid. These findings provide the surgical team with valuable information regarding the complexity and expected duration of the procedure<sup>2</sup>. In this case, the

CT scan revealed signs of severe obstruction, indicating the need for an exploratory laparotomy.

The lack of guidelines for the standard of care of EPS results in surgical procedures being performed primarily in cases complicated by bowel obstruction. The primary goal of surgery is to completely release fibrotic bands along the small intestine and the mesenteric root. Bowel resection is restricted to non-vital segments only to minimize the risk of complications, such as short bowel syndrome<sup>5</sup>. The surgery for this patient focused on releasing the firm adhesions interposed between bowel loops and avoiding intestinal resection.

EPS poses a diagnostic and therapeutic challenge due to its low incidence, idiopathic etiology, and nonspecific clinical presentation. The diagnosis

should be ruled out in cases of abdominal pain that do not have a mechanical cause. The clinical case presentation demonstrated how CT and surgery led to the diagnosis of type II EPS in a patient with no history of previous abdominal surgery and symptoms of intestinal obstruction. The imaging test was key to surgical planning, as it suggested the presence of a severe mechanical obstruction.

The exploratory laparotomy facilitated both a definitive diagnosis and curative treatment. This case highlights the importance of considering EPS or cocoon syndrome in the differential diagnosis of bowel obstruction, especially in the absence of prior surgery, and the relevance of a timely surgical approach for a favorable clinical outcome.

#### Referencias bibliográficas /References

1. Vipudhamorn W, Juthasilaparut T, Sutharat P, Sanmee S, Supatrakul E, Chowsilpa S, et al. Unraveling the enigma of sclerosing encapsulating peritonitis: a comprehensive review. *Ann Coloproctol.* 2025;41(3):175-89. doi: 10.3393/ac.2024.00486.0069.
2. Mohammed F, Abdulkarim M, Ibn Yasir A, Taleballah O, Shani D, Salih N. Abdominal cocoon syndrome, a case report of a rare disease entity causing intestinal obstruction. *Int J Surg Case Rep.* 2021;87:106401. doi:10.1016/j.ijscr.2021.106401.
3. Ye Y, Cao L, Zhang Z, He M. Abdominal cocoon syndrome: An extremely rare cause of intestinal obstruction. *Asian J Surg.* 2022;45(10):1868-9. doi:10.1016/j.asjsur.2022.03.115.
4. Frazão J, Martins AR, Calado J, Godinho A. Abdominal Cocoon Syndrome: A Rare Cause of Intestinal Obstruction. *Cureus.* 2022;14(3):e22929. doi:10.7759/cureus.22929.
5. Chorti A, Panidis S, Konstantinidis D, Cheva A, Papavramidis T, Michalopoulos A, et al. Abdominal cocoon syndrome: Rare cause of intestinal obstruction-Case report and systematic review of literature. *Medicine (Baltimore).* 2022;101(27):e29837. doi: 10.1097/MD.00000000000029837.
6. Vipudhamorn W, Juthasilaparut T, Sutharat P, Sanmee S, Supatrakul E. Abdominal cocoon syndrome-a rare culprit behind small bowel ischemia and obstruction: Three case reports. *World J Gastrointest Surg.* 2024;16(3):955-65. doi:10.4240/wjgs.v16.i3.955.

# Trasplante hepático en colangiocarcinoma perihiliar irresecable: informe de un caso en el marco de un protocolo de selección oncológica estricta

## *Liver Transplantation for Unresectable Perihilar Cholangiocarcinoma: A Case Report Following a Strict Oncologic Selection Protocol*

Melisa Amondarain , Pablo E. Huespe , Federico Díaz , Alicia Verzura , Sung Ho Hyon , Martín de Santibañes 

Hospital Italiano de Buenos Aires. Argentina

### RESUMEN

Los autores declaran no tener conflictos de interés.  
*Conflicts of interest None declared.*

Correspondencia  
*Correspondence:*  
Melisa Amondarain.  
E-mail:  
amondarainmelisa@gmail.com

El colangiocarcinoma perihiliar (CCpH) representa uno de los tumores hepatobiliares más desafiantes por su complejidad quirúrgica y mal pronóstico en casos irresecables. En pacientes seleccionados sin diseminación metastásica, el trasplante hepático (TH) bajo protocolo con terapia neoadyuvante ha surgido como una alternativa curativa aceptada. Presentamos el caso de un varón de 52 años con CCpH irresecable por compromiso vascular bilateral, tratado mediante un protocolo institucional estricto que incluyó quimiorradioterapia neoadyuvante y trasplante hepático ortotópico. Pese a colangitis recurrente durante el seguimiento, los marcadores tumorales permanecieron normales y la laparoscopia de estadificación descartó enfermedad diseminada. El TH se realizó exitosamente con injerto cadavérico. Este caso demuestra la factibilidad y los buenos resultados del trasplante protocolizado para CCpH irresecable, apoyando su implementación en centros especializados con pacientes cuidadosamente seleccionados. El TH bajo protocolo representa una posibilidad de supervivencia prolongada en casos históricamente sin opciones curativas.

■ **Palabras clave:** colangiocarcinoma perihiliar, trasplante hepático, terapia neoadyuvante.

### ABSTRACT

Perihepatic cholangiocarcinoma (pCCA) is one of the most challenging hepatobiliary tumors to treat in terms of surgical complexity and poor prognosis in unresectable cases. In selected patients without metastatic disease, liver transplantation (LT) performed according to a protocol that includes neoadjuvant therapy has emerged as an accepted curative option. We present the case of a 52-year-old man with unresectable pCCA due to bilateral vascular involvement, who was treated according to a strict institutional protocol that included neoadjuvant chemoradiotherapy and orthotopic liver transplantation. Despite recurrent episodes of cholangitis during follow-up, tumor markers remained normal, and a staging laparoscopy ruled out distant metastases. Cadaveric liver transplantation was successfully performed without significant intraoperative complications. This case demonstrates the feasibility and favorable outcomes of standardized transplantation for unresectable pCCA, supporting its implementation in specialized centers with carefully selected patients. Liver transplantation following a protocol offers the possibility of prolonged survival in cases where curative options have historically been unavailable.

■ **Keywords:** perihilar cholangiocarcinoma, liver transplantation, neoadjuvant therapy.

Recibido | Received  
15-09-25  
Aceptado | Accepted  
12-02-26

ID ORCID: Melisa Amondarain, 0000-0001-7941-123X; Pablo E. Huespe, 0000-0003-2644-5449; Federico Díaz, 0009-0008-5117-8446; Alicia Verzura, 0009-0004-1531-9812; Sung Ho Hyon, 0000-0002-0189-753X; Martín de Santibañes, 0000-0002-0163-1248

El colangiocarcinoma perihiliar (CCpH) representa la forma más frecuente de cáncer de la vía biliar extrahepática. Su localización en la confluencia de los conductos hepáticos dificulta el diagnóstico temprano y, en la mayoría de los casos, compromete la posibilidad de resección quirúrgica con márgenes negativos. La resección curativa sigue siendo el tratamiento estándar cuando es factible, pero en pacientes con enfermedad localmente avanzada, sin diseminación metastásica, el pronóstico continúa siendo pobre con terapias convencionales<sup>1</sup>. En las últimas décadas, el trasplante hepático ha emergido como una alternativa terapéutica en pacientes con CCpH irresecable, especialmente en el contexto de protocolos que combinan terapia neoadyuvante y criterios de selección estrictos, como el desarrollado por la Mayo Clinic. Este enfoque ha demostrado tasas de supervivencia a largo plazo significativamente superiores a las obtenidas con resección en casos seleccionados.

Las tasas de supervivencia a 5 años oscilan entre 65 y 75%. Estos resultados superan ampliamente los obtenidos con resección en casos avanzados, especialmente cuando los márgenes son difíciles y la enfermedad está localmente avanzada<sup>2</sup>.

Se presenta el caso de un varón de 52 años, sin comorbilidades relevantes, que consultó en julio de 2023 por ictericia obstructiva de 48 horas de evolución. La resonancia magnética informó dilatación de la vía biliar intrahepática con estenosis en la confluencia biliar. Mediante colangiografía transhepática se colocaron drenajes biliares percutáneos y *stents* (endoprótesis) metálicos biliares. El diagnóstico fue CCpH irresecable por compromiso bilateral de estructuras vasculares hepáticas.

Se realizó una ecoendoscopia lineal que evidenció tres ganglios peripancreáticos (12-17 mm) y dos ganglios hiliares (12-14 mm), sin lesiones pancreáticas ni vasculares adicionales. Se efectuó punción con aguja

ProCore 25G, obteniéndose material citológico e histológico compatible con inflamación reactiva sin hallazgos de malignidad.

Durante el seguimiento presentó cuatro episodios de colangitis, tratados con antibióticos intravenosos y recambio de drenajes percutáneos. Los marcadores tumorales (CEA, CA 19-9) se mantuvieron dentro de rangos normales. La tomografía por emisión de positrones (PET/TC) mostró captación metabólica inespecífica en la confluencia biliar, atribuible a inflamación, sin signos de enfermedad metastásica.

El caso fue evaluado en ateneo multidisciplinario. Se propuso su inclusión en el protocolo institucional de trasplante hepático para CCpH irresecable, que incluye quimiorradioterapia neoadyuvante y eventual trasplante hepático con duodenopancreatectomía céfálica (DPC) en caso de confirmarse compromiso tumoral de la vía biliar distal o estructuras pancreatoduodenales en la laparoscopia de estadificación.

Recibió radioterapia externa (50 Gy) (Figs. 1 y 2) y quimioterapia con 5-FU según protocolo.

En marzo de 2024 se realizó una laparoscopia exploradora con biopsia peritoneal y linfadenectomía hepática, sin hallazgos de diseminación.

En diciembre de 2024 se realizó un trasplante hepático ortotópico con injerto cadavérico, sin complicaciones intraoperatorias mayores. Se realizó la sección de vía biliar, y se la envió a biopsia por congelación, que informó margen libre de lesión, por lo que no se encontró necesidad de DPC, y se confirmó la viabilidad de continuar hacia trasplante hepático aislado.

Se practicó anastomosis biliar término-lateral en Y de Roux. El paciente evolucionó favorablemente. El análisis anatomopatológico confirmó CCpH moderadamente diferenciado sin invasión vascular ni ganglionar (ypT1a ypN0 ypMx).

El trasplante hepático posterior a terapia neoadyuvante ha modificado de manera sustancial el abordaje del CCpH irresecable. El protocolo desarrollado por la Mayo Clinic demostró resultados superiores a la resección quirúrgica en pacientes estrictamente seleccionados, con tasas de supervivencia a 5 años del 65-75%, en contraste con supervivencias inferiores al 30% informadas tras resección en enfermedad localmente avanzada, incluso con márgenes negativos<sup>2</sup>.

Los criterios de inclusión: tumor  $\leq 3$  cm, ausencia de metástasis, ganglios negativos y ausencia de procedimientos transperitoneales previos buscan minimizar el riesgo de diseminación tumoral y explicarían los resultados favorables observados en las distintas series. En estudios multicéntricos norteamericanos, los pacientes que completaron la terapia neoadyuvante y accedieron al trasplante presentaron supervivencias significativamente mayores que aquellos tratados con resección, aun cuando estos últimos lograran resecciones R0<sup>2,3</sup>.

El caso presentado comparte características con las cohortes publicadas: irresecabilidad por compromiso vascular bilateral, marcadores tumorales persistentemente normales y ausencia de enfermedad metastásica en estudios de estadificación. A pesar de complicaciones infecciosas recurrentes durante la

■ FIGURA 1



Reconstrucción 3D: Plano Sagital (DLI)  
Flecha corta negra: en verde columna, cabeza de flecha negra: azul hígado, flecha larga negra: rojo corazón, flecha larga blanca: riñones y cabeza de flecha blanca: isodosis correspondiente a dosis del 100% en región delimitada para tratar

■ FIGURA 2



Plano axial: flecha corta negra: curva en rojo con isodosis correspondiente a PTV con cobertura del 100% de la dosis (51 Gy). Cabeza de flecha negra: en azul el hígado

espera, principalmente colangitis asociada a drenajes biliares, el control tumoral se mantuvo, situación también descrita en hasta el 40% de los pacientes incluidos en protocolos de trasplante para CCpH, sin impacto negativo en la supervivencia cuando el protocolo se completa adecuadamente. La laparoscopia de estadiación desempeñó un papel central, permitiendo descartar diseminación peritoneal oculta, hallazgo informado en hasta el 20% de los pacientes evaluados inicialmente como candidatos a trasplante en series contemporáneas<sup>4</sup>. Asimismo, la ausencia de compromiso distal de la vía biliar permitió evitar una duodenopancreatometomía cefálica asociada, reduciendo la morbilidad operatoria sin comprometer el control oncológico.

Algunos metanálisis recientes confirman la solidez de esta estrategia, con supervivencias globales a 5

años superiores al 60% en pacientes trasplantados bajo protocolo, posicionando al trasplante hepático como la mejor alternativa curativa disponible para CCpH irresecable en centros especializados<sup>5</sup>. En línea con estos resultados, el paciente permanece libre de enfermedad al año del trasplante, con seguimiento trimestral mediante marcadores tumorales, resonancia magnética abdominal y tomografía computarizada de tórax, sin evidencia de recurrencia, y sin tratamiento adyuvante.

En conclusión, el trasplante hepático bajo protocolo representa una opción curativa real para pacientes seleccionados con CCpH irresecable. Este caso ejemplifica su éxito cuando se aplica conforme a la evidencia vigente y destaca la importancia de centros especializados, selección rigurosa y abordaje multidisciplinario<sup>6</sup>.

## ■ ENGLISH VERSION

Perihilar cholangiocarcinoma (pCCA) is the most common type of extrahepatic bile duct cancer. The location of the tumor at the confluence of the hepatic ducts complicates early diagnosis and, in most cases, precludes the possibility of surgical resection with negative margins. Curative resection remains the standard of care when feasible, but in patients with locally advanced disease and absence of distant metastasis, the prognosis remains poor with conventional therapies<sup>1</sup>.

In recent decades, liver transplantation has emerged as a therapeutic option for patients with unresectable pCCA, particularly within the context of protocols that combine neoadjuvant therapy and strict selection criteria, such as the one developed by the Mayo Clinic. This approach has demonstrated long-term survival rates that are significantly higher than those achieved with resection in selected cases. Five-year survival rates range from 65% to 75%. These results far exceed those obtained with resection in advanced cases, especially when achieving clear margins is difficult, and the disease is locally advanced<sup>2</sup>.

We report the case of a 52-year-old otherwise healthy male patient who presented with obstructive jaundice lasting 48 hours in July 2023. On magnetic resonance imaging, the intrahepatic bile ducts were dilated and presented a stricture at the confluence. Percutaneous biliary drains and metallic stents were inserted using transhepatic cholangiography. The diagnosis was unresectable pCCA due to bilateral involvement of hepatic vascular structures.

Linear endoscopic ultrasound identified three peripancreatic lymph nodes (12–17 mm) and two hilar lymph nodes (12–14 mm). No secondary pancreatic or vascular involvement was observed. Samples were obtained with a 25G ProCore needle. The cytological

and histological analysis showed findings consistent with reactive inflammation, with no evidence of malignancy.

During follow-up, the patient developed four episodes of cholangitis which were treated with intravenous antibiotics and replacement of percutaneous drains. The tumor markers (CEA, CA 19-9) were within normal ranges. Positron emission tomography/computed tomography (PET/CT) scan demonstrated nonspecific uptake at the biliary confluence suggestive of inflammation, with no evidence of metastatic disease.

The case was evaluated in a multidisciplinary case conference. The patient was proposed for inclusion in the institutional liver transplantation protocol for unresectable pCCA. The protocol consists of neoadjuvant chemoradiotherapy followed by liver transplantation, with concomitant cephalic pancreaticoduodenectomy (CPD) if staging laparoscopy confirms tumor involvement of the distal bile duct or pancreatoduodenal structures.

He received external radiation therapy (50 Gy) (Figs. 1 and 2), concurrent with 5-FU chemotherapy according to protocol.

An exploratory laparoscopy, peritoneal biopsy, and hepatic lymphadenectomy were performed in March 2024. There was no evidence of metastasis.

In December 2024, the patient underwent orthotopic cadaveric liver transplantation without significant intraoperative complications. The biliary tract was sectioned and submitted for frozen section. Since the margins were clear, CPD was not necessary, and the decision was made to proceed with the liver transplant.

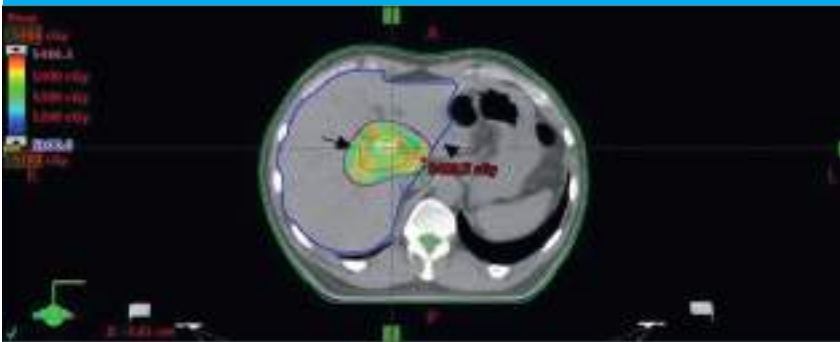
An end-to-side Roux-en-Y biliary anastomosis was performed. The patient evolved with favorable outcome. The pathological examination revealed

■ FIGURE 1



3D reconstruction, sagittal view (prescribed dose at isocenter). Short black arrow: spine (green); black arrowhead: liver (blue); long black arrow: heart (red); long white arrow: kidneys; and white arrowhead: radiation isodose line corresponding to 100% of the prescribed dose within the target treatment volume.

■ FIGURE 2



Axial view; short black arrow: red curve showing the isodose corresponding to the planning target volume with 100% dose coverage (51 Gy). Black arrowhead: liver (blue)

moderately differentiated pCCA without vascular or lymph node invasion (ypT1a, ypN0, ypMx).

Liver transplantation following neoadjuvant therapy has substantially changed the approach to unresectable pCCA. The Mayo Clinic protocol has demonstrated better outcomes compared with surgical resection in carefully selected patients, with 5-year survival rates of 65–75% versus less than 30% following resection for locally advanced disease, even with negative margins.

The inclusion criteria—tumor size  $\leq 3$  cm, absence of metastasis, negative lymph nodes, and no prior transperitoneal procedures—are intended to minimize the risk of tumor seeding and may explain the favorable outcomes observed in various series. Multicenter studies conducted in the United States have shown that patients who have undergone neoadjuvant therapy and transplantation have significantly higher survival rates than those treated with resection, even in cases of R0 resection<sup>2,3</sup>.

This case exhibits characteristics consistent with previously published cohorts, including unresectability due to bilateral vascular involvement, persistently normal tumor markers, and the absence of metastatic disease on staging examinations. Despite recurrent infections during the waiting period, primarily cholangitis associated with biliary drains, the tumor remained under control—a situation also reported in up to 40% of patients enrolled in transplant protocols for pCCA—with no negative impact on survival when

the protocol is properly completed. Staging laparoscopy played a key role in ruling out occult peritoneal seeding, a finding reported in up to 20% of patients initially evaluated as transplant candidates in contemporary series<sup>4</sup>. Moreover, the absence of distal bile duct involvement enabled the avoidance of a concomitant cephalic pancreaticoduodenectomy, thereby reducing surgical morbidity without compromising cancer control.

Several recent meta-analyses have confirmed the effectiveness of this strategy. Overall, 5-year survival rates exceed 60% for patients who undergo transplantation according to protocol. This establishes liver transplantation as the best curative option for unresectable pCCA in specialized centers<sup>5</sup>. Consistent with these results, the patient has remained disease-free for a period of one year following transplantation. He is subject to quarterly monitoring, which includes tumor marker assessments, abdominal MRI scans, and chest CT scans. No evidence of recurrence has been observed, and no adjuvant treatment has been administered.

In conclusion, liver transplantation under protocol represents a genuine curative option for selected patients with unresectable pCCA. This case illustrates the success of this approach when implemented in accordance with current evidence, underscoring the importance of specialized centers, rigorous patient selection, and multidisciplinary approach<sup>6</sup>.

## Referencias bibliográficas /References

1. Ernani L, Martino RB, Andraus W, Fernandes ESM, Mello FPT, Andrade R, et al. Protocol for liver transplantation in hilar cholangiocarcinoma. *Arq Bras Cir Dig.* 2022;34(3):e1618. doi: 10.1590/0102-672020210002e1618. PMID: 35019130; PMCID: PMC8735267.
2. Heimbach JK, Gores GJ, Haddock MG, Alberts SR, Nyberg SL, Ishitani MB, et al. Transplantation for hilar cholangiocarcinoma: The Mayo Clinic experience. *Hepatology.* 2006;43(4):872-80. DOI: 10.1002/lt.20266.
3. Ethun CG, López-Aguilar AG, Anderson DJ, Adams AB, Fields RC, Doyle MBM, et al. Transplantation versus resection for hilar cholangiocarcinoma: an argument for shifting treatment paradigms. *Ann Surg.* 2018;267(5):797-805. DOI: 10.1097/SLA.0000000000002574
4. Rea DJ, Heimbach JK, Rosen CB, Haddock MG, Alberts SR, Kremers WK, et al. Liver transplantation with neoadjuvant chemoradiation is more effective than resection for hilar cholangiocarcinoma. *Ann Surg.* 2005;242(3):451-8.
5. Darwish Murad S, Kim WR, Harnois DM, Douglas DD, Burton J, Kulik LM, et al. Efficacy of neoadjuvant chemoradiation followed by liver transplantation for perihilar cholangiocarcinoma at 12 US centers. *Gastroenterology.* 2012;143(1):88-98. <https://pubmed.ncbi.nlm.nih.gov/22504095/>
6. Cambridge WA, Fairfield C, Powell JJ, Harrison EM, Wigmore SJ, Garden OJ, et al. Meta-analysis of survival after liver transplantation for unresectable perihilar cholangiocarcinoma. *Ann Surg.* 2021;273(2):240-50. DOI:10.1097/SLA.0000000000003801

## Fijación externa mandibular en trauma maxilofacial por arma de fuego

### External mandibular fixation in maxillofacial gunshot injury

Juan P. Berríos-Bugueño<sup>1</sup> , Juan P. Aravena<sup>1,2</sup> , Rodrigo A. Allende<sup>1</sup> , María M. Padilla<sup>1</sup> , Polett C. Yáñez<sup>1</sup> , Diego A. Lazo<sup>1,2</sup> 

1. Departamento de Cirugía Maxilofacial, Hospital San José. Chile  
2. Departamento de Cirugía y Traumatología Bucal y Maxilofacial. Facultad de Odontología, Universidad de Chile. Chile.

Los autores declaran no tener conflictos de interés.

Conflicts of interest  
None declared.

Correspondencia  
Correspondence:  
Juan P. Berríos-Bugueño  
E-mail: [juan.berrios.b@ug.uchile.cl](mailto:juan.berrios.b@ug.uchile.cl)

#### RESUMEN

El trauma maxilofacial por arma de fuego representa un desafío clínico por la complejidad anatómica y el alto riesgo de morbilidad y mortalidad. Las fracturas mandibulares conminutas secundarias a impactos balísticos suelen asociarse a pérdida ósea, contaminación bacteriana y daño extenso de tejidos blandos, lo que limita la aplicación de la reducción abierta con fijación interna rígida. Se presenta el caso de un varón de 19 años con fractura mandibular conminuta por proyectil balístico, tratado mediante fijación externa con clavos de Kirschner bicorticales, tubo de silicona y acrílico autopolimerizable. La técnica permitió estabilizar los fragmentos óseos sin abordajes extensos, preservando la vascularización periosteal y reduciendo el riesgo de complicaciones. La fijación externa constituye una alternativa eficaz y de bajo costo para el manejo de fracturas mandibulares complejas, especialmente en pacientes politraumatizados o con tejidos comprometidos, que funcionan como tratamiento definitivo o puente hacia una reconstrucción futura.

■ **Palabras clave:** herida arma de fuego, mandíbula, fijadores externos.

#### ABSTRACT

Maxillofacial gunshot injuries constitute a clinical challenge due to the complex anatomy and high risk of morbidity and mortality. Comminuted mandibular fractures resulting from ballistic trauma are often associated with bone loss, bacterial infection, and extensive soft tissue damage, limiting the use of open reduction with rigid internal fixation. We present the case of a 19-year-old male patient with a comminuted mandibular fracture caused by a bullet, managed with external fixation with bicortical Kirschner wires, a silicone tube, and self-curing acrylic resin. The technique enabled the stabilization of the bone fragments without the necessity for extensive approaches, thereby preserving the periosteal blood supply and reducing the risk of complications.

External fixation constitutes an effective and low-cost alternative for managing complex mandibular fractures, particularly in patients with polytrauma or involvement of other tissues. This method serves as both a definitive treatment and a bridging treatment for future reconstruction.

■ **Keywords:** Wounds, Gunshot, Mandible/surgery, External Fixators.

Recibido | Received  
02-10-25  
Aceptado | Accepted  
10-02-26

ID ORCID: Juan P. Berríos-Bugueño, 0009-0005-5428-4701; Juan P. Aravena, 0000-0002-1089-8446; Rodrigo A. Allende, 0009-0003-2754-4665; María M. Padilla, 0009-0004-7014-3506; Polett C. Yáñez, 0000-0003-2177-653X; Diego A. Lazo, 0000-0003-0016-4075.

El trauma maxilofacial por armas de fuego representa una entidad clínica compleja, asociada a riesgo importante de morbilidad y mortalidad, tanto por el compromiso estructural, como por las lesiones concomitantes en otros sistemas vitales<sup>1</sup>. Las lesiones en la región mandibular suelen ser resultado de impactos de alta energía, que provocan fracturas conminutas, pérdida ósea segmentaria, hemorragias, destrucción de tejidos blandos y riesgo de contaminación bacteriana por comunicación con el medio oral<sup>2</sup>.

Las fracturas mandibulares conminutas corresponden a lesiones en las que el hueso se fragmenta en múltiples porciones irregulares, frecuentemente asociadas a pérdida de vitalidad tisular. En este escenario, el tratamiento quirúrgico inmediato puede estar limitado por la condición general del paciente, contaminación del lecho quirúrgico y compromiso de los tejidos blandos de sostén. Estos factores aumentan el riesgo de fracaso de la fijación interna rígida (también llama-

da ORIF, por sus siglas en inglés Open Reduction and Internal Fixation). Por ello, la estabilización inicial de los segmentos óseos desempeña un papel prioritario para preservar la vitalidad tisular y/o permitir una reconstrucción funcional posterior<sup>1</sup>.

La fijación externa mandibular se ha posicionado como una técnica eficaz que permite mantener la posición de los segmentos óseos sin necesidad de abordajes amplios ni exposición de la fractura, favorece la preservación vascular y reduce del riesgo de necrosis ósea. En escenarios de trauma balístico, la fijación externa ha sido descrita tanto como tratamiento definitivo en casos seleccionados, o bien como estrategia inicial dentro de un abordaje por etapas, dependiendo del grado de conminución, el compromiso de los tejidos blandos y la condición general del paciente<sup>1-3</sup>.

Se presenta un varón de 19 años sin antecedentes médico-quirúrgicos relevantes que ingresó en Urgencias tras recibir impacto de proyectil balístico

por arma de fuego en la mandíbula. Después de la estabilización inicial y tratamiento en Unidad de Cuidados Intensivos (UCI), se realizó una interconsulta con el equipo de Cirugía Maxilofacial por heridas penetrantes y fractura mandibular. El examen físico reveló orificios compatibles con trayectoria balística y conminución ósea mandibular extensa, como también lo hizo la tomografía computarizada (Fig. 1).

Se planificó una intervención quirúrgica inicial, de manera urgente para estabilizar la fractura conminuta mediante fijación externa mandibular. En pabellón, bajo anestesia general y con intubación nasotraqueal, se colocaron tres clavos de Kirschner bicorticales en cada segmento del cuerpo mandibular utilizando un taladro percutor estéril. Posteriormente, se adaptó un tubo de silicona perforado para alojar y estabilizar los clavos, el cual se rellenoó con acrílico de autopolimerización para conferir rigidez al sistema. Se complementó la fijación con ferulización dentaria adhesiva mediante alambre en los dientes inferiores para proporcionar estabilidad secundaria. El aspecto inflamatorio de la mucosa gingival corresponde al contexto de un trauma reciente de alta energía, con edema y daño tisular agudo, que fue resuelto con el paso de los días (Fig. 2).

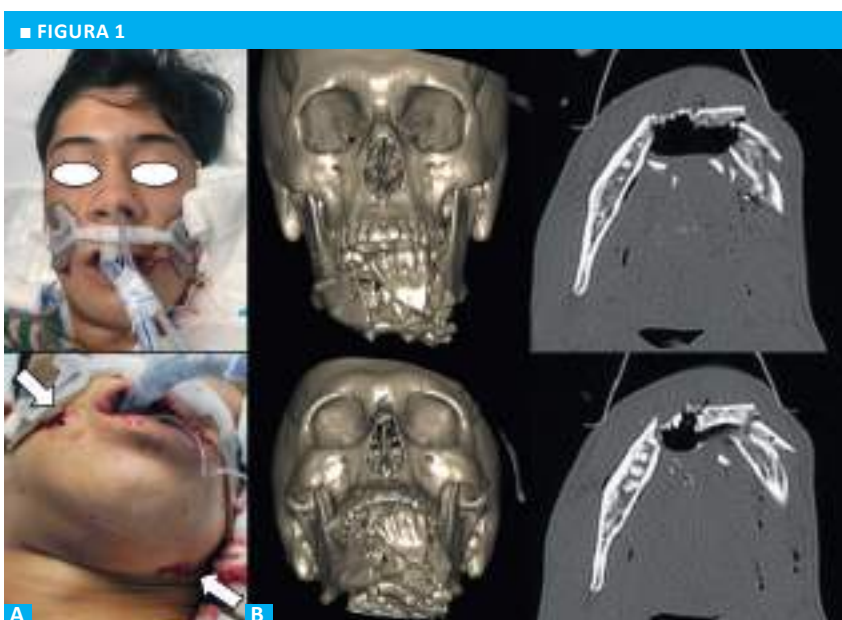
La técnica de fijación externa permitió una estabilización eficaz sin requerir exposición abierta ni despegamiento extensivo, minimizar el daño a tejidos blandos y preservar la vascularización periosteal. Durante el seguimiento de 4 meses, el paciente presentó resolución del dolor, sin complicaciones infecciosas ni fallo del dispositivo. Se mantuvo control, con énfasis en la movilidad mandibular e higiene prolija. Desde el punto de vista funcional, el paciente presentó una evolución satisfactoria, con recuperación progresiva de la apertura oral y ausencia de signos clínicos de infección o movilidad mandibular patológica durante los contro-

les. La fijación externa fue retirada a los 4 meses postoperatorios, una vez evidenciada la estabilidad clínica de los fragmentos mandibulares, sin movilidad patológica ni signos de infección activa. Se mantuvieron controles durante los 30 días posteriores (Fig. 2).

Las fracturas mandibulares conminutas por trauma balístico se asocian a fragmentación extensa, pérdida ósea y daño de tejidos blandos, lo que limita las opciones reconstructivas en la fase aguda y aumenta el riesgo de complicaciones posoperatorias<sup>4</sup>. La elección del método de estabilización debe priorizar la preservación tisular y el control de la infección<sup>4,5</sup>. Mardones y cols.<sup>2</sup> informaron resultados favorables utilizando fijación externa en fracturas mandibulares balísticas como tratamiento único. Secchi y cols.<sup>4</sup> lo usaron como tratamiento puente para una ORIF combinada con injerto de cresta ilíaca con resultados también favorables y osificación a los 3 meses.

En el presente caso, la fijación externa permitió una evolución favorable en un paciente politraumatizado que no se encontraba en condiciones de recibir una reconstrucción definitiva inmediata. El objetivo se centró en permitir una consolidación ósea y preservar la apertura bucal, para un mejor tratamiento de las demás lesiones por parte de los demás equipos (cirugía de tórax, cirugía digestiva, traumatología y coloproctología, debido a lesiones asociadas en distintos segmentos corporales).

La fijación externa mandibular ha resurgido como una herramienta valiosa como método de estabilización inicial de fracturas conminutas. Ofrece un método económico, adaptable y biomecánicamente estable. Esto resulta especialmente útil en centros con recursos limitados o en pacientes con condiciones sistémicas que contraindican cirugías extensas<sup>1,2,4</sup>. No obstante, esta técnica presenta limitaciones. Ellis y cols<sup>1</sup>. in-



A: Preoperatorio. Paciente en UCI con intubación oral. Orificios de entrada y salida de proyectil balístico (flechas) y B: Tomografía axial computarizada (reconstrucción 3D y cortes axiales).

■ FIGURA 2



Arriba: Posoperatorio inmediato de fijación externa con tutores, férula dental inferior y reconstrucción 3D. Abajo: Cinco meses de posoperatorio: fijación externa retirada, apertura y dinámica mandibular, reconstrucción 3D y cortes axiales de tomografía.

forman que la fijación externa con clavos presentó una tasa de complicaciones del 35,2%, en comparación con una fijación interna estable (10,3%). La fijación externa puede asociarse a complicaciones locales, como infecciones en los sitios de inserción de las agujas, celulitis, formación de fístulas, maloclusión y daño nervioso. El uso del acrílico requiere precauciones para evitar quemaduras cutáneas por la reacción exotérmica durante la polimerización. Además, la visibilidad y el tamaño del dispositivo pueden afectar la adherencia del paciente al tratamiento, impactando su bienestar psicológico y social<sup>2,4</sup>.

La ORIF consiste en la exposición de la fractura y sistemas de osteosíntesis internos, es el procedimiento de referencia (estándar de oro) para fractu-

ras mandibulares. En comparación, la fijación externa ofrece ventajas en escenarios seleccionados. Primero, preservar la irrigación perióstica y la integridad de los tejidos blandos, fundamentales para la vitalidad ósea y cicatrización. Segundo, reducir el tiempo quirúrgico y la morbilidad asociada a las incisiones amplias. Tercero, permitir el tratamiento inicial de pacientes politraumatizados o inestables que no toleran procedimientos prolongados. Finalmente, facilitar la estabilización temporal o definitiva, dependiendo de la evolución clínica y la planificación<sup>2,4,5</sup>.

En el caso presentado, la ORIF no fue considerada como primera opción, dada la condición crítica del paciente y la necesidad de tratamiento quirúrgico rápido y seguro. En contexto de un politraumatismo grave

con conminución extensa y tejidos comprometidos, la prioridad fue lograr una estabilización que permitiera controlar la fractura y preservar la vitalidad ósea. La fijación externa ha sido descrita como procedimiento de emergencia, que ofrece una solución inmediata, menos invasiva y con resultados funcionales satisfactorios, lo que justificó su elección inicial<sup>3,4</sup>.

La fijación externa mandibular mediante clavos de Kirschner, tubo de silicona y acrílico autopolimerizable debe considerarse como una alternativa de bajo

costo y adaptable para el manejo de fracturas mandibulares conminutas por arma de fuego, como tratamiento definitivo en este caso y como posible puente terapéutico hacia una reconstrucción futura. Su uso se destaca en escenarios con recursos limitados o cuando las condiciones clínicas del paciente contraindican la ORIF. Este caso refuerza el valor de individualizar los tratamientos y considerar soluciones alternativas que garanticen la estabilización ósea, el control del dolor y la preservación funcional en contextos de alta complejidad.

## ■ ENGLISH VERSION

Maxillofacial gunshot injuries are complex clinical conditions associated with a significant risk of morbidity and mortality, due both to structural damage and to concomitant injuries to other vital systems<sup>1</sup>. Mandibular injuries are usually the result of high-energy impacts, which cause comminuted fractures, segmental bone loss, hemorrhage, soft tissue destruction, and a risk of bacterial infection from oral cavity communication<sup>2</sup>.

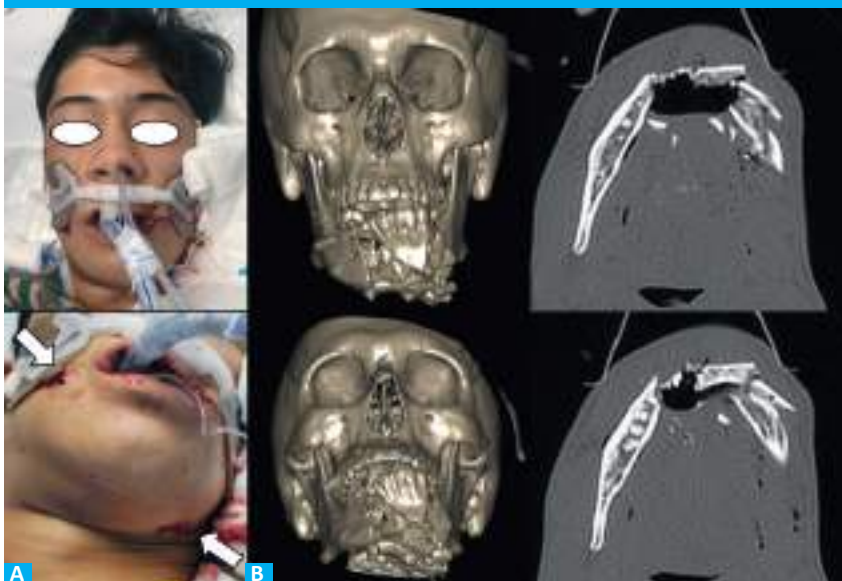
Comminuted mandibular fractures are injuries in which the bone breaks into multiple irregular fragments, often accompanied by tissue necrosis. In this scenario, the patient's general status, surgical site infection, and involvement of the supporting soft tissues may limit the need for immediate surgical treatment. These factors increase the risk of failure of open reduction and internal fixation (ORIF). For this reason, the initial stabilization of bone segments is critical for preserving tissue viability and enabling subsequent functional reconstruction<sup>1</sup>.

In addition to promoting vascular preservation

and lowering the risk of bone necrosis, mandibular external fixation has established itself as an efficient method for preserving the position of the bone segments without requiring extensive approaches or exposure of the fracture. In ballistic trauma scenarios, external fixation has been described as both a definitive treatment in selected cases and an initial strategy within a staged approach. This depends on the degree of comminution, soft tissue involvement, and the patient's general status<sup>1-3</sup>.

A 19-year-old male patient with no significant medical or surgical history was admitted to the emergency department due to a gunshot injury to the mandible. Following initial stabilization and treatment in the intensive care unit (ICU), the maxillofacial surgery team was consulted due to penetrating injuries and a mandibular fracture. The physical examination revealed orifices consistent with the trajectory of the bullet and extensive mandibular bone comminution, a finding that was subsequently confirmed by a computed tomography scan (Fig. 1).

■ FIGURE 1



A: Preoperative. Patient in the ICU with orotracheal intubation. Bullet entry and exit wounds (arrows). B: Computed tomography scan (3D reconstruction and axial sections).

An initial emergency surgery was planned to stabilize the comminuted fracture using external mandibular fixation. In the operating room, under general anesthesia and with nasotracheal intubation, three bicortical Kirschner wires were placed in each segment of the maxillary body using a sterile bone drill. Subsequently, a perforated silicone tube was adapted to hold and stabilize the wires. The tube was then filled with self-curing acrylic resin, which provided the necessary rigidity to the system. Fixation was supplemented with adhesive wire splinting of the lower teeth to provide secondary stability. The inflammatory appearance of the gingival mucosa is consistent with a recent high-energy trauma, involving edema and acute tissue damage, which resolved over the course of several days. 2).

External fixation provided effective stabilization without requiring open exposure or extensive detachment, minimizing soft tissue damage and preserving periosteal blood supply. During the 4-month follow-up period, pain resolved, and there were no complications related to infections or device failure. The patient was closely monitored, with a focus on mandibular mobility and proper oral hygiene. From a functional perspective, the patient experienced satisfactory progress, with gradual recovery of mouth opening and no clinical signs of infection or abnormal mandibular mobility during follow-up visits. The external fixation was removed 4 months postoperatively, once the mandibular fragments demonstrated clinical stability, with no abnormal mobility or signs of active infection. He was monitored for the following 30 days (Fig. 2).

■ FIGURE 2



Top: Immediate postoperative period following external fixation with pins, lower dental splint, and 3D reconstruction. Bottom: Five months following surgery. Removal of external fixation, mandibular opening and motion dynamics, 3D reconstruction, and axial CT sections.

Comminuted mandibular fractures resulting from ballistic trauma are often associated with bone loss and extensive soft-tissue damage, limiting the use of reconstructive options and increasing the risk of postoperative complications<sup>4</sup>. When selecting a stabilization method, the primary considerations should be tissue preservation and infection control<sup>4,5</sup>. Mardones et al.<sup>2</sup> reported favorable outcomes using external fixation as the sole treatment for ballistic mandibular fractures. Secchi et al.<sup>4</sup> used external fixation as a bridging treatment for ORIF combined with iliac crest bone graft, achieving favorable results and ossification within three months.

In this case, external fixation allowed for a favorable outcome in a patient with multiple traumas who was not in good condition to undergo immediate definitive reconstruction. The primary objective was to facilitate bone healing and preserve mouth opening, thereby enabling more effective treatment of the associated injuries by other teams, including thoracic surgery, gastrointestinal surgery, orthopedic surgery, and colon and rectal surgery, due to the presence of concurrent injuries in various parts of the body.

External mandibular fixation has reemerged as a valuable tool for the initial stabilization of comminuted fractures. The method is cost-effective, adaptable and biomechanically stable. This is particularly useful in centers with limited resources or for patients with systemic conditions that preclude prolonged surgical procedures<sup>1,2,4</sup>. However, this technique has some limitations. Ellis et al.<sup>1</sup> reported that patients treated with external pin fixation had a 35.2% complication rate compared with stable internal fixation (10.3%). External fixation may be associated with local complications, such as pin site infections, cellulitis, fistulization, malocclusion, and nerve injury. The use of acrylic resin requires precautions to prevent skin burns caused by the exothermic reaction during polymerization. Additionally, the visible appearance and size of the device can affect patient adherence to treatment,

impacting their psychological and social well-being<sup>2,4</sup>.

ORIF is the gold standard procedure for mandibular fractures and involves exposure of the fracture site and the use of internal fixation devices. External fixation offers advantages in certain situations. First, external fixation preserves blood flow supply to the periosteum and the integrity of soft tissues, which are essential for bone vitality and healing. Second, it reduces the operative time and morbidity associated with large incisions. Third, patients with polytrauma or unstable conditions who cannot tolerate prolonged procedures can be given initial management. Finally, the procedure facilitates temporary or permanent stabilization, depending on the clinical course and treatment plan<sup>2,4,5</sup>.

In the case here described, ORIF was not considered the first option, given the patient's critical condition and the need for rapid and safe surgical treatment. In the context of severe polytrauma with extensive comminution and soft tissue involvement, the priority was to achieve stabilization that would allow for fracture control and preservation of bone viability. External fixation has been described as an emergency procedure that offers an immediate, less invasive solution with satisfactory functional outcomes. That is why we chose it as the initial treatment<sup>3,4</sup>.

External mandibular fixation using Kirschner wires, a silicone tube, and self-curing acrylic resin should be considered a cost-effective and adaptable alternative for the management of comminuted mandibular fractures caused by gunshot wounds. This method served as the definitive treatment in this case and as a potential bridging treatment to future reconstruction. It is particularly useful in settings with limited resources or when the patient's clinical condition precludes ORIF. This case underscores the importance of tailoring treatment plans and considering alternative solutions to ensure bone stabilization, pain management, and functional preservation in highly complex clinical settings.

## Referencias bibliográficas /References

1. Ellis E, Muniz O, Anand K. Treatment considerations for comminuted mandibular fractures. *J Oral Maxillofac Surg* [Internet]. 2003;61(8):861-70. DOI: 10.1016/s0278-2391(03)00249-0.
2. Mardones-M M, Torres-M C, Gunckel-M R, Bravo-A R, Mardones-M M, Torres-M C, et al. Fracturas Mandibulares Conminutadas por Impacto de Proyectoil Balístico: Resultados Clínicos de Tratamiento Con Método de Fijación Alternativo. *Int J Odontostomat* [Internet]. 2021; 15(2):363-9. <http://dx.doi.org/10.4067/S0718-381X2021000200363>.
3. Elbir B, Kolsuz N, Varol A. External mandibular fixation for gunshot fractures: report of 2 cases. *Ulusal Travma ve Acil Cerrahi Dergisi*. 2023;29(6):741–5. doi: 10.14744/tjtes.2022.77315
4. Secchi-A A, Mordoh-C S, Cristi-R E, Díaz-B E, Passalacqua-M G, Yanine-M N, et al. Algoritmo para el Manejo del Trauma Balístico Maxilofacial. Caso Clínico y Revisión de la Literatura. *Int J Odontostomat*. 2022;16(2):241-8. <http://dx.doi.org/10.4067/S0718-381X2022000200241>.
5. Rogol EV. External fixation for tissue preservation following complex mandible trauma and its sequelae: reminder of a time-honored method and case series of 23 patients. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2025;139(5):501-8. doi: 10.1016/j.oooo.2024.10.113.