

Resultados del proceso de evaluación de resúmenes para el XCVII Congreso Chileno e Internacional de Cirugía. Estudio bibliométrico

Carlos Manterola¹, Jorge Piérart¹, Rodrigo Torres-Quevedo¹, Sergio Sotelo¹, Luis E. Burgos-de Cea¹

Results of abstract evaluation process of the XCVII Congress of the Chilean Society of Surgery

Objective: To describe the characteristics of abstracts submitted to the XCVII Chilean and International Congress of Surgery in the categories of e-poster, video, and free oral communications (FOC). **Materials and Methods:** Bibliometric study. All submitted abstracts were included. Variables analyzed were presentation format, origin, disciplinary area, study design, and level of evidence provided by FOC. A validated and reliable scale was used for FOC, and two ad hoc checklists were applied for e-posters and videos. Each abstract was independently evaluated by two blind reviewers. In case of discrepancies, arbitration was undertaken by the Scientific Committee. Descriptive and analytical statistics were applied (Student's t test for mean comparisons, χ^2 test for categorical variables), as well as the intraclass correlation coefficient to assess inter-rater agreement. **Results:** A total of 863 abstracts were analyzed. Most corresponded to e-posters (412; 47.7%), followed by FOC (390; 45.2%) and videos (61; 7.1%). Overall, 538 abstracts (62.3%) were accepted: 74.5% of e-posters, 72.1% of videos, and 47.9% of FOC. The sections with the highest number of accepted abstracts were coloproctology (19.5%) and general surgery (16.2%). Among accepted FOC, 80.2% corresponded to case series. For treatment-related scenarios, 91.4% of studies were classified as level 4 evidence. No statistically significant differences were observed in the scores assigned by different reviewers across any of the presentation formats. **Conclusion:** This study characterized the abstracts submitted to the XCVII Chilean and International Congress of Surgery. No significant differences were found between evaluations performed by different reviewers.

Key words: conferences; oral presentations; conference papers; methodology; bibliometrics; bibliographies as topic; bibliometric analysis.

Resumen

Objetivo: Describir las características de los resúmenes enviados al XCVII Congreso Chileno e Internacional de Cirugía en sus modalidades de e-póster, videos y trabajos libres orales (TLO). **Material y Método:** Estudio bibliométrico. Se incluyeron todos los resúmenes enviados. Se analizaron modalidad de presentación, origen, área disciplinar, diseño del estudio y nivel de evidencia aportado por los TLO. Se empleó una escala válida y confiable para los TLO, y dos pautas *ad hoc* para e-póster y videos. Cada resumen fue evaluado por dos observadores enmascarados. En casos de discrepancia se recurrió al arbitraje del Comité Científico. Se aplicaron estadísticas descriptivas y analíticas (t de Student para comparación de medias y χ^2 para variables categóricas) y coeficiente de correlación intraclass para evaluar concordancia entre evaluadores. **Resultados:** Se analizaron 863 resúmenes. La mayoría correspondió a e-póster (412; 47,7%), seguidos de TLO (390; 45,2%) y videos (61; 7,1%). Se aceptaron 538 resúmenes (62,3%): 74,5% de los e-póster, 72,1% de los videos y 47,9% de los TLO. Las seccionales con mayor número de resúmenes aceptados fueron coloproctología (19,5%) y cirugía general (16,2%). El 80,2% de los TLO aceptados correspondieron a series de casos. El 91,4% de los estudios son de nivel de evidencia tipo 4 para escenarios de tratamiento. No se encontraron diferencias estadísticamente significativas en los puntajes otorgados por distintos revisores, en ninguna de las modalidades evaluadas. **Conclusión:** Se caracterizaron los resúmenes enviados y aceptados al XCVII Congreso Chileno e Internacional de Cirugía. No se observaron diferencias en las evaluaciones realizadas por distintos revisores.

Palabras clave: congresos; presentaciones orales; ponencias en congresos; metodología; bibliometría; análisis bibliométrico.

¹Comité Científico Sociedad Chilena de Cirugía, Chile.

Recibido el 2025-09-11 y aceptado para publicación el 2025-10-17

Correspondencia a:
Dr. Carlos Manterola
carlos.manterola@ufrontera.cl

E-ISSN 2452-4549



Introducción

La presentación de resultados en congresos médicos es una práctica esencial del quehacer científico, pues constituye una fuente relevante sobre el avance disciplinar. Dada la amplia concurrencia a estos eventos, es crucial asegurar la calidad de la información expuesta. Sin embargo, los estudios presentados suelen carecer del rigor de revisión propio de las revistas arbitradas, y pocas sociedades científicas realizan evaluaciones sistemáticas de la calidad de los trabajos incluidos en sus reuniones¹.

Evaluar la calidad de los resúmenes permite no solo organizar y clasificar las propuestas de un congreso, sino también estimar la probabilidad de su posterior publicación en revistas con revisión por pares. De hecho, la tasa de publicación de los resúmenes presentados refleja, en buena medida, la calidad del comité de selección².

La evaluación de los resúmenes facilita su publicación en las actas del congreso, aumentando su visibilidad³. Aunque la extensión limita la descripción metodológica y de resultados, se ha demostrado que la calidad del resumen se correlaciona con la calidad científica del estudio subyacente⁴.

Diversos criterios de presentación de informes han sido utilizados para aumentar la calidad e impacto de los resúmenes de congresos⁵, especialmente para garantizar estandarización en las presentaciones. Así, se han desarrollado estudios sobre implementación de estos sistemas^{6,7}, aunque persiste la búsqueda de la herramienta de evaluación de resúmenes más adecuada⁵. Entre las opciones disponibles destacan STROBE (0-11 puntos), Timmer (0-15 puntos), NES (0-13 puntos)^{5,7,8} y nuestra escala (0-100 puntos), que cuenta con estudios de validez y confiabilidad⁹⁻¹¹. Pero independiente de la herramienta utilizada, la selección de resúmenes debe realizarse mediante una revisión por pares, enmascarada, utilizando instrumentos válidos y confiables. Sin embargo, la evidencia muestra que la preocupación por la calidad metodológica (CM) en los congresos sigue siendo insuficiente (por ejemplo en la Asociación Europea de Urología, no mejoró en una década¹²).

Aunque la calidad de un resumen refleja la del estudio original, el uso de métodos estandarizados y listas de verificación aumenta la confiabilidad entre evaluadores y permite elaborar resúmenes más completos e informativos^{9,13}. Estas listas existen para todo tipo de diseños (para algunos se dispone de múltiples versiones^{6,13}). Sin embargo, las herramientas específicas para evaluación de resúmenes de congresos son escasas^{11,14,15}. Entre estas destacan:

una para evaluar pósteres, compuesta por 30 ítems distribuidos en tres dimensiones (apariencia general, contenido y presentación), con puntuaciones entre 30 y 150 puntos¹⁴. Otra, basada en el tipo de estudio (investigación humana, ciencias básicas, otros), con puntaje máximo de 43 puntos y estudios de validez de constructo y confiabilidad interobservador⁶. Otra, propuesta por un grupo de cirujanos plásticos, con puntuaciones entre -6 y +6 puntos, que cuenta con estudios de consistencia interna, validez y confiabilidad interobservador^{15,16}; y la herramienta desarrollada para evaluar y seleccionar resúmenes del congreso de la Sociedad Chilena de Cirugía (SoChiCir), la que se aplicó por primera vez en 2014. Esta escala cuenta con estudios de consistencia interna, validez y confiabilidad interobservador, y está compuesta por 10 ítems distribuidos en seis dominios⁹⁻¹¹. Su aplicación continua, ha permitido un proceso de capacitación permanente, asegurando mediciones reproducibles entre distintos evaluadores. Además, se organizan seminarios de capacitación una o dos veces al año, que permiten pilotajes, ajustes en la redacción de ítems y dominios, y avances en la validación del instrumento¹⁰. Con ello, el punto de corte que define la aceptabilidad de un resumen aumentó desde 50 puntos (50% de cumplimiento) a 60 puntos (60% de cumplimiento, equivalente a nota 4,0 en una escala de 1,0 a 7,0)¹¹.

El objetivo de este estudio fue determinar las características de los resúmenes enviados al XCVII Congreso Chileno e Internacional de Cirugía, en sus modalidades e-póster, videos y trabajos libres orales (TLO).

Material y Método

Diseño

Estudio bibliométrico tipo meso-análisis de línea disciplinar¹⁷.

Población

Se incluyó la totalidad de los resúmenes enviados al XCVII Congreso Chileno e Internacional de Cirugía, 2025; en sus diferentes modalidades: e-póster, videos y TLO. No hubo exclusión.

Variables estudiadas

Se evaluaron las siguientes variables: modalidad de presentación (TLO, e-póster o vídeo); origen del trabajo; pertenencia a un centro universitario (sí o no); y área disciplinar (cirugía bariátrica, de cabeza y cuello, mama, general, plástica y reparadora, coloproctología; cirugía endoscópica, esófago-gástrica,

hepatobiliar; cirugía de hernias, de tórax, de trauma, vascular; oncología, ERAS (recuperación acelerada post-cirugía), simulación y educación. Asimismo, se registró el diseño del estudio (series de casos, estudios de corte transversal, de casos y controles, de cohortes, ensayos clínicos, y revisiones sistemáticas). Además, se evaluó nivel de evidencia aportado por los TLO para escenarios de tratamiento o procedimientos terapéuticos, así como para escenarios de pronóstico o historia natural, siguiendo la clasificación del Centro de Medicina Basada en Evidencia de Oxford^{18,19}. Finalmente, se determinaron las causas de no aceptación por modalidad de presentación.

Instrumentos de medición

Se utilizó una escala válida y confiable para evaluar resúmenes de TLO⁹⁻¹¹; y dos pautas *ad-hoc* para valoración de e-póster y de videos respectivamente, ambas con puntajes de 0 a 50 puntos, y punto de corte de aceptación de 30 puntos.

Proceso de medición

Participaron 85 revisores, algunos evaluando en una sola seccional y otros en varias. Cada resumen de TLO, e-póster y video fue analizado por dos observadores independientes y enmascarados, y los datos se registraron en hojas de Excel. Cada evaluador valoró una mediana de 20 resúmenes (rango: 1-95 evaluaciones). En casos de discrepancias respecto a la dicotomía “aceptado vs no aceptado”, se recurrió al arbitraje enmascarado del Comité Científico de la SoChiCir para dirimir diferencias.

Estadísticas

Se aplicó estadística descriptiva (proporciones, promedios y desviaciones estándar), analítica (t-test para comparación de promedios; Chi² o prueba

exacta de Fisher para variables dicotómicas y categóricas); y coeficiente de correlación intraclass para comparación de mediciones entre evaluadores.

Éticos

El estudio se realizó con resúmenes enmascarados respecto de sus autores, garantizando confidencialidad y anonimato.

Resultados

Se analizaron 863 resúmenes presentados en el XCVII Congreso Chileno e Internacional de Cirugía. La mayoría se presentó en modalidad e-póster (n = 412; 47,7%), seguida de TLO (390 resúmenes; 45,2%), y videos (n = 61; 7,1%). Fueron aceptados 538 resúmenes (62,3%): 307 e-póster (74,5%), 44 videos (72,1%) y 187 TLO (47,9%) (Figura 1).

Las seccionales con mayor número de resúmenes aceptados fueron coloproctología (19,5%), cirugía general (16,2%) y cirugía de tórax (12,7%). Cirugía general aportó la mayor proporción de pósteres (24,4%), y coloproctología la mayor proporción de TLO y videos (26,7% y 31,8%, respectivamente) (Tabla 1).

Las seccionales con mayor número de TLO enviados fueron coloproctología (76; 19,4%) y hepatobiliar (50; 12,8%), mientras que las seccionales con mayor proporción de aceptación de TLO fueron simulación-educación y cirugía de mamas (80,0% y 76,9%, respectivamente). Las seccionales con mayor número de e-póster enviados fueron cirugía general (102; 24,7%), coloproctología (51; 12,3%), y cirugía de tórax (50; 12,1%); aunque las que alcanzaron mayor proporción de aceptación fueron cirugía de trauma y vascular (93,8% y 92,0%, respectivamen-

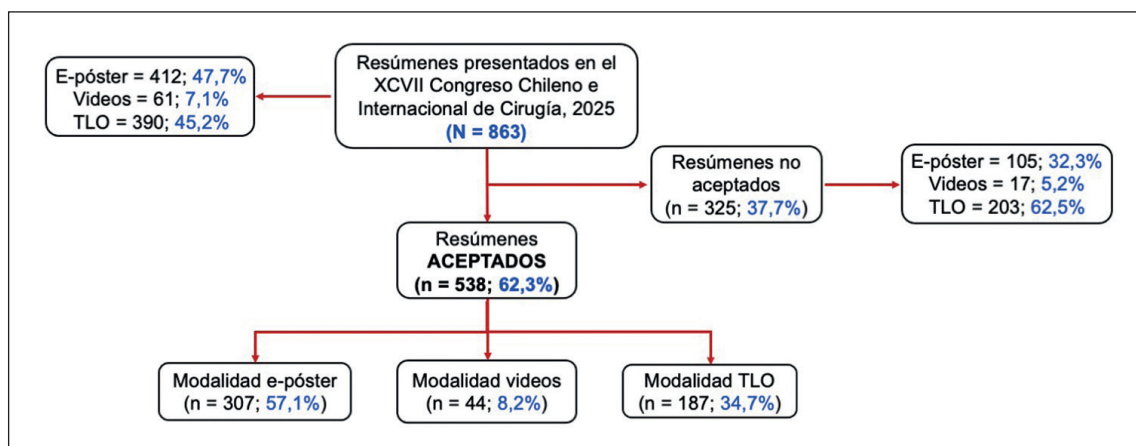


Figura 1. Diagrama de flujo de los resúmenes participantes.

te). Las seccionales con mayor número de videos enviados fueron coloproctología (19; 31,1%) y cirugía de tórax (9; 14,8%), mientras que las con mayor proporción de aceptación fueron cirugía de cabeza y cuello, hernias, oncología y vascular, todas con 100% (Tabla 2).

Tabla 1. Distribución de los trabajos aceptados por seccional y modalidad de presentación (N = 538)

Seccional	E-Póster N° (%)	TL Orales N° (%)	Videos N° (%)	Total N° (%)
Cirugía Bariátrica	2 (0,7)	1 (0,5)	3 (6,9)	6 (1,1)
Cirugía de cabeza y cuello	12 (3,9)	5 (2,7)	2 (4,5)	19 (3,5)
Cirugía de mama	11 (3,6)	10 (5,3)	---	21 (3,9)
Cirugía general	75 (24,4)	9 (4,8)	3 (6,9)	87 (16,2)
Cirugía plástica y reparadora	17 (5,6)	11 (5,9)	---	28 (5,2)
Coloproctología	41 (13,4)	50 (26,7)	14 (31,8)	105 (19,5)
Endoscopia	7 (2,3)	5 (2,7)	---	12 (2,2)
Esofagogástrico	26 (8,5)	20 (10,7)	3 (6,8)	49 (9,1)
Hepatobiliar	15 (4,9)	21 (11,2)	4 (9,1)	40 (7,4)
Hernias	14 (4,2)	6 (3,2)	2 (4,5)	22 (4,1)
Oncología	4 (1,3)	2 (1,1)	2 (4,5)	8 (1,5)
Simulación-Educación	3 (1,0)	12 (6,4)	1 (2,3)	16 (3,0)
Tórax	42 (13,7)	19 (10,2)	7 (15,9)	68 (12,7)
Trauma	15 (4,9)	5 (2,7)	2 (4,5)	22 (4,1)
Vascular	23 (7,6)	11 (5,9)	1 (2,3)	35 (6,5)
Total n (%)	307 (100)	187 (100)	44 (100)	538 (100)

Tabla 2. Distribución de TLO, e-pósteres y videos según seccional, y relación entre los enviados (N = 451) y los aceptados (n = 231)

Seccional	TLO presentados (n = 390) n° (%)	TLO aceptados (% de los presentados)	E-póster presentados (n = 390) n° (%)	E-póster aceptados (% de los presentados)	Videos presentados (n = 61) n° (%)	Videos aceptados (% de los presentados)
Bariátrica	13 (3,3)	1 (7,7)	6 (1,5)	2 (33,3)	7 (11,5)	3 (42,9)
Cabeza y cuello	24 (6,2)	5 (20,8)	26 (6,3)	12 (46,2)	2 (3,3)	2 (100)
Mama	13 (3,3)	10 (76,9)	14 (3,4)	11 (78,6)	0 (0,0)	0 (0,0)
General	35 (9,0)	9 (25,7)	102 (24,7)	75 (73,5)	4 (6,6)	3 (75,0)
Plástica y reparadora	27 (6,9)	11 (40,7)	20 (4,9)	17 (85,0)	0 (0,0)	0 (0,0)
Coloproctología	76 (19,4)	50 (65,8)	51 (12,3)	41 (80,4)	19 (31,1)	14 (73,7)
Endoscopia	8 (2,1)	5 (62,5)	11 (2,7)	7 (63,6)	0 (0,0)	0 (0,0)
Esófago-gástrica	34 (8,7)	20 (58,8)	36 (8,7)	3 (8,3)	5 (8,2)	3 (60,0)
Hepatobiliar	50 (12,8)	21 (42,0)	27 (6,6)	15 (55,6)	6 (9,8)	4 (66,7)
Hernias	14 (3,6)	6 (42,9)	16 (3,9)	14 (87,5)	2 (3,3)	2 (100)
Oncología	8 (2,1)	2 (25,0)	5 (1,2)	4 (80,0)	2 (3,3)	2 (100)
ERAS	1 (0,3)	0 (0,0)	2 (0,5)	0 (0,0)	0 (0,0)	0 (0,0)
Simulación-Educación	15 (3,8)	12 (80,0)	5 (1,2)	3 (60,0)	1 (1,6)	1 (100)
Tórax	37 (9,5)	19 (51,4)	50 (12,1)	42 (84,0)	9 (14,8)	7 (77,8)
Trauma	10 (2,6%)	5 (50,0)	16 (3,9)	15 (93,8)	3 (4,9)	2 (66,7)
Vascular	25 (6,4)	11 (44,0)	25 (6,1)	23 (92,0)	1 (1,6)	1 (100)
Total	390 (100)	187 (47,9)	412 (100)	284 (68,9)	61 (100)	44 (72,1)

Los centros que enviaron mayor número de TLO fueron Pontificia Universidad Católica de Chile (37; 9,5%), Hospital Clínico de la Universidad de Chile (30; 7,7%), Hospital Clínico San Borja Arriarán (28; 7,2%), Hospital de Puerto Montt (20; 5,1%) y Universidad de La Frontera (TLO; 4,9%). Aquellos con mayor proporción de aceptación fueron Pontificia Universidad Católica de Chile (72,9%), Hospital Clínico San Borja Arriarán (71,4%) y Hospital Clínico de la Universidad de Chile (66,7%) (Tabla 3). De los resúmenes aceptados, 24/44 videos (54,5%), 148/187 TLO (79,1%) y 182/307 e-poster (59,3%) tenían pertenencia a algún centro universitario.

La mayoría de los resúmenes aceptados en modalidad TLO correspondió a series de casos o a diseños no identificables por los revisores (80,2%). Asimismo, el 91,4% y el 85,0% de los estudios presentaron nivel de evidencia tipo 4 para escenarios de tratamiento y de pronóstico respectivamente (Tabla 4).

Las causas de no aceptación más frecuentes fueron, para TLO: “reporte de caso, debió presentarse como póster” (40,2%) y “metodología pobre o incompleta” (16,1%). Para e-póster: “aporte discreto o poco relevante al conocimiento” (30,3%) y “descripción insuficiente o caracterización inadecuada de la muestra” (13,4%). Para videos: “aporte discreto o poco relevante al conocimiento” (35,3%) y “video de calidad subóptima” (29,4%) (Tabla 5).

No se evidenciaron diferencias estadísticamente significativas en los puntajes de las evaluaciones realizadas por diferentes revisores para ninguna de las modalidades (Tabla 6).

Al desagrupar los resúmenes aceptados y no aceptados, se observaron diferencias estadísticamente significativas en las evaluaciones de los e-póster, aunque la correlación indicó una relación negativa media y significativa. En los TLO aceptados, la correlación fue positiva, pequeña y significativa. Se verificó correlación positiva en la confiabilidad interobservador de los TLO aceptados y rechazados, así como en los pósteres y videos aceptados. Por otro lado, se constató correlación negativa en la confiabilidad interobservador de los pósteres y de los videos rechazados (Tabla 7).

Discusión

Los eventos científicos son espacios clave para la difusión de conocimientos y fomentan la interacción entre profesionales, académicos y estudiantes; facilitando el intercambio de resultados e impulsando la educación e investigación. Los congresos suelen

Tabla 3. Origen de los TLO presentados y su relación con centros universitarios

Centros de origen principal	Presentados n° (%)	Aceptados n° (%)
Pontificia Universidad Católica de Chile	37 (9,5)	27 (72,9)
Hospital Clínico Universidad de Chile	30 (7,7)	20 (66,7)
Hospital Clínico San Borja Arriarán	28 (7,2)	20 (71,4)
Hospital Puerto Montt*	20 (5,1)	12 (60,0)
Universidad de La Frontera	19 (4,9)	8 (42,1)
Hospital Militar de Santiago*	16 (4,1)	7 (43,8)
Universidad Austral de Chile*	14 (3,6)	9 (64,3)
Hospital Padre Hurtado*	13 (3,3)	1 (7,7)
Hospital San Juan de Dios	13 (3,3)	8 (61,5)
Clínica Universidad de los Andes	12 (3,1)	5 (41,7)
Hospital Dr. Sótero del Río*	11 (2,8)	8 (72,7)
Hospital Dr. Franco Ravera Zunino*	11 (2,8)	5 (45,5)
Hospital Federico Abete – Argentina	9 (2,3)	2 (22,2)
Hospital del Salvador	9 (2,3)	4 (44,4)
Hospital de Antofagasta*	9 (2,3)	2 (22,2)
Hospital Regional de Talca*	9 (2,3)	2 (22,2)
Universidad de Concepción*	8 (2,1)	4 (50,0)
Universidad Finis Terrae*	8 (2,1)	2 (25,0)
Hospital Dr. Víctor Ríos Ruiz*	7 (1,8)	1 (14,3)
Clínica Alemana de Santiago*	7 (1,8)	5 (71,4)
Clínica Indisa*	6 (1,5)	5 (83,3)
Clínica Dávila	5 (1,3)	2 (40,0)
Hospital Roosevelt, Guatemala	4 (1,0)	0 (0,0)
Royal North Shore Hospital, Australia	4 (1,0)	4 (100)
Instituto Nacional del Tórax*	4 (1,0)	2 (50,0)
Clínica Santa María	4 (1,0)	0 (0,0)
Otras instituciones**	30 (7,7)	12 (40,0)
Otras instituciones***	26 (6,7)	8 (30,8)
Otras instituciones****	17 (4,4)	2 (11,8)
Total	390 (100)	187 (100)

*Sólo o asociado con alguna institución universitaria u hospitalaria. **Clínica Andes Salud El Loa, Clínicas RedSalud, Hospital Barros Luco Trudeau, Hospital Base San José de Osorno, Hospital Clínico Magallanes, Hospital Dr. Luis Tisné Brousse, Hospital del Cobre, Hospital El Pino, Hospital Félix Bulnes, y Hospital Regional de Coyhaique. Todos ellos aportaron con 3 TLO cada cual. ***Complejo Asistencial Padre Las Casas, Hospital de Coquimbo, Hospital de Constitución, Hospital de la Fuerza Aérea de Chile, Hospital del Trabajador, Hospital de Urgencia Asistencia Pública, Hospital de Villarrica, Hospital DIPRECA, Hospital Dr. Gustavo Fricke, Hospital Mauricio Heyermann, Hospital Naval Almirante Nef, Hospital San Juan de Dios de La Serena, SOLCA-Guayaquil. Todos ellos aportaron con 2 TLO cada cual. ****CESFAM Flor Fernández, Clínica Cordillera, Hospital Base de Linares, Hospital Claudio Vicuña, Hospital Clínico Herminda Martín, Hospital Clínico Mutual de Seguridad, Hospital Las Higueras, Hospital San Juan de Dios de San Fernando, Hospital Santa Cruz, Hospital 21 de mayo-Taltal, Instituto Nacional del Cáncer, The Royal Marsden Hospital, Universidad de Atacama, Universidad de O’Higgins, Universidad de Santiago, USS, Universidad de Valparaíso. Todos ellos aportaron con 1 TLO cada cual.

Tabla 4. Diseños de estudio reportados en los TLO

Tipo de diseños	Nivel de evidencia tratamiento	Nivel de evidencia pronóstico	Presentados (n = 390) n° (%)	Aceptados (n = 187) n° (%)
Ensayo clínico	1b	---	1 (0,3)	1 (100)
Revisión sistemática	3a	2a	4 (1,0)	3 (75,0)
Estudios de casos y controles anidado	3b	2b	2 (0,5)	2 (100)
Estudios de corte transversal	3b	2c	4 (1,0)	2 (50,0)
Estudio de cohorte retrospectiva	4	2b	16 (4,1)	12 (75,0)
Estudios reportados como “comparativos” sin diseño identificado por los autores*	4	4	55 (14,1)	9 (16,4)
Poblacionales y ecológicos	2c	2c	10 (2,6)	8 (80,0)
Serie de casos	4	4	257 (65,9)	133 (51,8)
Inclasificables. Diseño no identificable por los revisores**	---	---	41 (10,5)	17 (41,4)
Total			390 (100)	187 (100)

*Se da una equivalencia de estudios de cohorte de mala calidad. **No es posible darles alguna equivalencia a los diseños tradicionales.

Tabla 5. Causas más frecuentes de no aceptación de resúmenes registradas por los revisores, según modalidad de presentación

Causas	TLO no aceptados (n = 203) n° (%)	E-póster no aceptados (n = 105) n° (%)	Videos no aceptados (n = 17) n° (%)
Reporte de caso. Debió presentarse como póster	100 (40,2)	---	---
Metodología pobre o incompleta	40 (16,1)	10 (8,4)	---
Aporte discreto o poco relevante al conocimiento actual	27 (10,8)	36 (30,3)	6 (35,3)
Deficiencias metodológicas graves	19 (7,6)	8 (6,7)	---
Resultados poco claros, incompletos o irrelevantes	13 (5,2)	---	---
Resultados discutibles	12 (4,8)	3 (2,5)	---
Descripción insuficiente o caracterización inadecuada	11 (4,4)	16 (13,4)	---
Información incompleta	---	12 (10,1)	---
Presentación poco clara o desordenada	7 (2,8)	13 (10,9)	---
Resumen mal redactado o mal estructurado	6 (2,4)	---	---
No es un trabajo de investigación (revisión bibliográfica)	6 (2,4)	---	---
Errores de concepto técnicos o anatómicos	3 (1,2)	6 (5,0)	---
Indicación injustificada de la técnica presentada	2 (0,8)	6 (5,0)	---
Sólo aumentó número de casos respecto del año anterior	2 (0,8)	---	---
Video de calidad subóptima	---	---	5 (29,4)
Debió presentarse como video o como TLO	---	2 (1,7)	---
No aplica a la disciplina (no quirúrgicos)	---	2 (1,7)	---
Otros	---	5 (4,2) *	6 (35,3) #
Total	249 **	119 **	17 **

*Problemas de redacción, resumen duplicado, estudio no realizado (es un protocolo), introducción inadecuada y objetivos vagos, deja más interrogantes que enseñanzas; uno de cada cual. #Procedimiento de alto riesgo, deja más interrogantes que enseñanzas, disección atípica, descripción insuficiente, presentación muy acotada de la técnica, video de simulación y no real; uno de cada cual. **En algunos casos hay más de un comentario. No hubo comentarios por parte de los evaluadores en 17 resúmenes de TLO, en 17 de e-póster, y en 6 videos.

recibir gran número de aportaciones, por ello, se recurre a un proceso de evaluación por pares, cuyo objetivo es asegurar que los resúmenes cumplan con las normas, tengan estructura adecuada, valor científico y CM suficiente¹¹.

Algunas sociedades científicas, como *American Academy of Dermatology*, *American Heart Association* y *Society of the Knee*, disponen de instrumentos de valoración de resúmenes²⁰; y suelen contar con mayor número de evaluadores, lo que permite que cada uno valore 4 a 8 resúmenes²¹, en contraste con la nuestra mediana de 20 por evaluador.

Se analizaron 863 resúmenes del XCVII Congreso Chileno e Internacional de Cirugía, de los cuales 390 TLO tenían potencial de publicación, muy por encima del promedio anual de 60 trabajos publicados en la revista de la sociedad en los últimos 5 años.

Aunque la evidencia sugiere que la calidad de los resúmenes puede determinar el tipo de presentación, asignándose a presentaciones orales los trabajos de mejor calidad respecto de los pósteres²²; en la So-ChiCir se estableció que el formato de presentación reflejara la intención de los autores: videos para técnicas quirúrgicas, e-póster para reportes de casos y TLO para casuísticas o estudios más complejos. Aun así, la evaluación sistemática mediante instrumentos *ad-hoc* mejora la calidad de lo aceptado, destacando los instrumentos *Timmer*, *Rowe*, *STRO-*

BE, *NES*, *CONSORT for Abstract*, *PRISMA-A* y *VAS Scoring*^{6-8,11,20,23-27}.

Los estudios observacionales constituyen la mayoría de los resúmenes orales y de póster en congresos científicos, mientras que los analíticos y experimentales son minoritarios^{7,8,28}. En nuestro caso, los observacionales de bajo nivel de evidencia representaron el 91,4%, hallazgo consistente con reportes de la Sociedad de Cirugía Vasculard de Canadá, que constató que la mayoría de sus resúmenes tenían NE 3 o 4²⁹; y de la Sociedad Brasileña de Coloproctología, que reportó 30,0% de estudios retrospectivos y 49,4% de reportes de casos (NE tipo 4²²).

Un aspecto adicional de interés surge al analizar las causas de rechazo de resúmenes. Llama la atención que el 40,2% de los TLO fue rechazado por tratarse de reportes de casos que debían ir a e-póster, constituyéndose en la causa más común de rechazo en todas las modalidades. Este hallazgo debe considerarse cuidadosamente por los autores, pues evidencia un error recurrente en el proceso de envío. Por igual, es relevante la elevada proporción de rechazos atribuibles a deficiencias metodológicas: “metodología pobre o incompleta” explicó el 23,7% de rechazos de TLO y 15,1% de los e-póster. Esto, enfatiza la necesidad urgente de que los grupos de trabajo fortalezcan su rigor metodológico antes de realizar envíos. Por otra parte, la considerable proporción de resúmenes evaluados como “aporte

Tabla 6. Puntuaciones del total de aportaciones según modalidad de presentación

Modalidad de presentación	Puntuación mín. - máx.	Puntuación revisor 1 Media ± DE	Rango para revisor 1	Puntuación revisor 2 Media ± DE	Rango para revisor 2	p
E-póster (n = 412)	0 - 50	38,8 ± 11,3	5 - 50	37,5 ± 11,7	5 - 50	0,1051
TLO (n = 390)	0 - 100	61,8 ± 19,8	5 - 100	60,0 ± 21,5	5 - 100	0,2243
Videos (n = 61)	0 - 50	36,6 ± 9,9	15 - 50	33,7 ± 9,7	10 - 50	0,1048

Tabla 7. Puntuación de las aportaciones aceptadas y rechazadas según modalidad de presentación

Modalidad de presentación	Rechazados (n = 325)				Aceptados (n = 538)			
	Evaluación 1 (n° casos) Media ± DE	Evaluación 2 (n° casos) Media ± DE	r Pearson	p	Evaluación 1 (n° casos) Media ± DE	Evaluación 2 (n° casos) Media ± DE	r Pearson	p
E-póster (n = 412)	(n=105) 29,0±11,8	(n=105) 25,2±10,5	-0,361*	0,002	(n=307) 42,1±8,9	(n=307) 41,0±9,0	0,01*	0,862
Videos (n = 61)	(n=17) 25,9±5,5	(n=17) 26,2±8,1	-0,319#	0,212	(n=44) 40,7±8,1	(n=44) 36,3±8,8	0,148*	0,338
TLO (n = 390)	(n=203) 50,3±18,0	(n=203) 45,8±17,6	0,084*	0,233	(n=187) 74,4±12,9	(n=187) 75,5±12,9	0,226#	0,002

TLO: Trabajos libres orales. *Existe correlación positiva pequeña, pero no significativa. #Existe correlación positiva pequeña y significativa. *Existe correlación negativa media y significativa. #Existe correlación negativa media pero no significativa

discreto o poco relevante al conocimiento” (10,8% en TLO, 30,3% en pósteres y 35,3% en videos), plantea la necesidad de reflexionar acerca de la pertinencia de remitir volúmenes elevados de resúmenes que no generan contribuciones significativas. También se identificaron casos de duplicación de resúmenes y de trabajos previamente presentados con modificaciones mínimas, hallazgos coincidentes con reportes previos que destacan problemas de redacción, ausencia de objetivos claros y desorden en la estructura del resumen (Tabla 5).

Por otra parte, al observar el nivel de desacuerdo entre revisores, verificamos que esto continúa siendo un problema. En esto también hay evidencia anterior que sustenta que el desacuerdo es mayor o aumenta con ciertos parámetros subjetivos de los sistemas de evaluación, como “importancia y originalidad de la propuesta”²³; lo que ha determinado que los valores del coeficiente kappa y de los coeficientes de correlación intraclase, rara vez superen el 0,40^{15,23,27,28,31,32}.

En este estudio, los coeficientes de correlación demostraron compatibilidad positiva muy pequeña, pero no significativa, entre revisores en la evaluación de e-póster y videos aceptados, y para los TLO rechazados ($r = 0,01$; $r = 0,148$; y $r = 0,083$,

respectivamente); así como una correlación positiva pequeña y significativa en la evaluación de los TLO aceptados ($r = 0,226$; $p = 0,002$). Valores que, en general, resultan aceptables y consistentes con estudios similares, los cuales sugieren que un grado de concordancia moderado parece ser el más adecuado en este tipo de mediciones, ya que un alto grado de concordancia, aunque indica confiabilidad del sistema, también podría reflejar sesgos, tanto positivos como negativos^{23,33}.

Es así como en relación con los resúmenes rechazados, particularmente en los e-póster y videos, se observó una correlación negativa. Este hallazgo podría explicarse por el uso de instrumentos de evaluación que aún no ha sido validados ni sometido a estudios de confiabilidad, lo que limita sus capacidades para discriminar³³. Asimismo, la superposición entre criterios técnicos y metodológicos parece generar conflictos en la aplicación de instrumentos, llevando a que, en ciertos casos, los evaluadores no apliquen de forma estricta el instrumento. En contraste, en los resúmenes aceptados –independientemente de la modalidad– se constató correlación positiva de magnitud pequeña a media (Figura 2). Este resultado sugiere que, una vez superado el

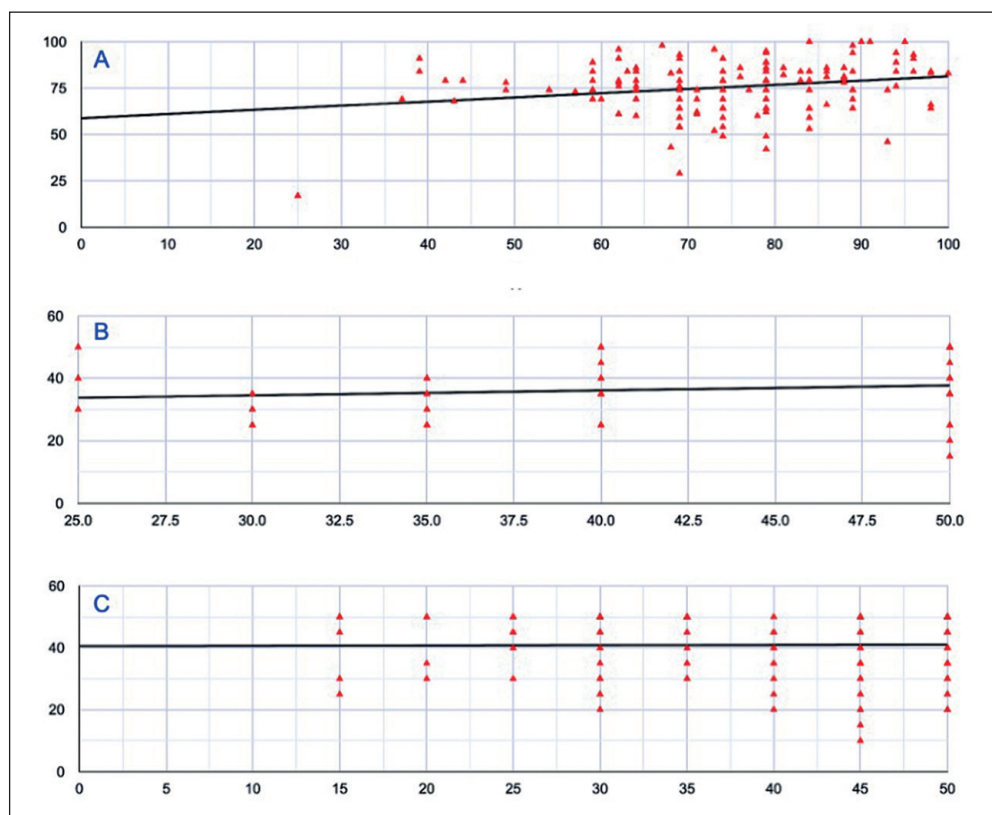


Figura 2. Gráficos de correlación de las dos mediciones entre distintos evaluadores por modalidad de presentación de los resúmenes aceptados. **A.** Evaluación de los TLO: Existe correlación positiva pequeña y significativa ($r = 0,226$), $p = 0,02$. **B.** Evaluación de los videos: Existe correlación positiva pequeña, pero no significativa ($r = 0,148$), $p = 0,338$. **C.** Evaluación de los e-póster: Existe correlación positiva pequeña, pero no significativa ($r = 0,01$), $p = 0,862$.

umbral de relevancia técnica, la valoración de los aspectos metodológicos adquiere mayor peso en la evaluación, favoreciendo una mayor confiabilidad interobservador³³.

Limitaciones del estudio

Se incluyeron todo tipo de estudios, predominando los observacionales. Los revisores no contaban con formación específica en estadística o epidemiología, lo que dificulta la interpretación de análisis de correlación entre evaluadores. Además, criterios subjetivos como la "importancia" pueden ser evaluados de forma diferente por distintos revisores, afectando la confiabilidad interobservador.

Conclusiones

El sistema actual de evaluación de resúmenes de congresos parece adecuado, aunque se recomienda organizar reuniones de evaluación con amplia participación de revisores acreditados y realizar estudios adicionales que permitan retroalimentación crítica. Se sugiere revisar periódicamente la escala utilizada para fomentar la colaboración entre evaluadores y mejorar la confiabilidad interobservador. La capacitación continua de los revisores y la retroalimentación sistemática podrían reducir los desacuerdos, y promover un enfoque de revisión por pares negativa (rechazando envíos inapropiados), podría ser beneficioso para mejorar la calidad de los resúmenes aceptados.

Contribución de los autores

- Concepción y diseño del estudio: Carlos Manterola, Jorge Piérart, Rodrigo Torres-Quevedo, Sergio Sotelo y Luis Burgos de Cea.
- Adquisición, análisis o interpretación de los datos: Carlos Manterola, Jorge Piérart, Rodrigo Torres-Quevedo, Sergio Sotelo y Luis Burgos de Cea.
- Redacción del manuscrito: Carlos Manterola, Jorge Piérart, Rodrigo Torres-Quevedo, Sergio Sotelo y Luis Burgos de Cea.
- Aprobación definitiva de la versión que se presenta: Carlos Manterola, Jorge Piérart, Rodrigo Torres-Quevedo, Sergio Sotelo y Luis Burgos de Cea.

Cumplimiento de normas éticas

Los autores declaran que no presentan ningún conflicto de interés en la realización de este estudio.

Además, informan que, al tratarse de un estudio bibliométrico, no requirió intervención directa en seres humanos, sino únicamente la información contenida en los resúmenes enviados al congreso societario de 2025.

Financiación: Ninguna.

Conflictos de interés: Ninguno.

El estudio se realizó con resúmenes enmascarados respecto de sus autores, garantizando confidencialidad y anonimato.

Bibliografía

1. Castaldi S, Giacometti M, Toigo W, Bert F, Siliquini R. Analysis of full-text publication and publishing predictors of abstracts presented at an Italian public health meeting (2005-2007). *BMC Res Notes*. 2015;8:492. <https://doi.org/10.1186/s13104-015-1463-7>.
2. Çekmecioglu BT, Kozanhan B, Eren G. Publication rate of abstracts orally presented at the Turkish Society of Anaesthesiology and Reanimation National Congresses. *Turk J Anaesthesiol Reanim*. 2019;47(2):151-7. <https://doi.org/10.5152/TJAR.2019.33603>.
3. dos Santos EF, Pereira MG. Quality of the structured abstracts presented at a congress. *Rev Assoc Med Bras*. 2007;53(4):355-9. <https://doi.org/10.1590/s0104-42302007000400023>.
4. Wiehn J, Nonte J, Prugger C. Reporting quality for abstracts of randomised trials on child and adolescent depression prevention: a meta-epidemiological study on adherence to CONSORT for abstracts. *BMJ Open*. 2022;12(8):e061873. <https://doi.org/10.1136/bmjopen-2022-061873>.
5. Hasbahçeci M, Başak F, Acar A, Şişik A. Analysis of reporting quality for oral presentations of observational studies at 19th National Surgical Congress: Proposal for a national evaluation system. *Ulus Cerrahi Derg*. 2016;32(4):267-74. <https://doi.org/10.5152/UCD.2016.3195>.
6. Timmer A, Sutherland LR, Hilsden RJ. Development and evaluation of a quality score for abstracts. *BMC Med Res Methodol*. 2003;3:2. <https://doi.org/10.1186/1471-2288-3-2>.
7. Yoon U, Knobloch K. Assessment of reporting quality of conference abstracts in sports injury prevention according to CONSORT and STROBE criteria and their subsequent publication rate as full papers. *BMC Med Res Methodol*. 2012;12:47. <https://doi.org/10.1186/1471-2288-12-47>.
8. Knobloch K, Yoon U, Rennekampff HO, Vogt PM. Quality of reporting according to the CONSORT, STROBE and Timmer instrument at the American Burn Association (ABA) annual meetings 2000 and 2008. *BMC Med Res Methodol*. 2011;11:161. <https://doi.org/10.1186/1471-2288-11-161>.
9. Manterola C, Otzen T. Diseño, construcción y validación inicial de un instrumento de medición para

- la evaluación de resúmenes para presentaciones en congresos. *Rev Chil Cir.* 2016;68(4):302-9. <https://doi.org/10.1016/j.rchic.2015.11.002>.
10. Manterola C, Otzen T. Confiabilidad de una escala de valoración de resúmenes para congresos. *Rev Cir.* 2022;74(5):480-6. <https://doi.org/10.35687/s2452-454920220051518>.
 11. Manterola C, Piérart J, Torres-Quevedo R, Sotelo S, Burgos de Cea L. Escala de Medición para Evaluación de Resúmenes de Congresos. *Int J Morphol.* 2025;43(3):1082-9. <https://doi.org/10.4067/s0717-95022025000301082>.
 12. De Sio M, Yakoubi R, De Nunzio C, Damiano R, Balsamo R, Di Palma, C, et al. Reporting Quality of Abstracts Presented at the European Association of Urology Meeting: A Critical Assessment. *J Urol.* 2012;188(5):1883-6. <https://doi.org/10.1016/j.juro.2012.07.041>.
 13. Otzen, T, Manterola C, Mora M, Quiroz G, Salazar P, García N. Statements, Recommendations, Proposals, Guidelines, Checklists and Scales Available for Reporting Results in Biomedical Research and Quality of Conduct. A Systematic Review. *Int J Morphol.* 2020;38(3):774-86. <https://doi.org/10.4067/S0717-95022020000300774>.
 14. Bushy A. A rating scale to evaluate research posters. *Nurse Educ.* 1991;16(1):11-5. <https://doi.org/10.1097/00006223-199101000-00008>.
 15. van der Steen LP, Hage JJ, Kon M, Mazzola R. Reliability of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg.* 2003;111(7):2215-22. <https://doi.org/10.1097/01.PRS.0000061092.88629.82>.
 16. van der Steen LP, Hage JJ, Kon M, Monstrey SJ. Validity of a structured method of selecting abstracts for a plastic surgical scientific meeting. *Plast Reconstr Surg.* 2004;113(1):353-9. <https://doi.org/10.1097/01.PRS.0000097461.50999.D4>.
 17. Manterola C, Rivadeneira J, Salgado C. Estudios bibliométricos. Una opción para desarrollar investigación en cirugía y disciplinas afines. *Rev Cir.* 2024;76(2):147-56. <https://doi.org/10.35687/s2452-454920240021890>.
 18. Centre for Evidence-Based Medicine (CEBM). Levels of Evidence (March 2009). Disponible en: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009>. Acceso el 4 de septiembre de 2025.
 19. Manterola C, Asenjo-Lobos C, Otzen T. Jerarquización de la evidencia. Niveles de evidencia y grados de recomendación de uso actual. *Rev Chil Infectol.* 2014;31(6):705-18. <https://doi.org/10.4067/S0716-10182014000600011>.
 20. Rahbek O, Jensen S L, Lind M, Penny J O, Kalleose T, Jakobsen T, et al. Inferior reliability of VAS scoring compared with International Society of the Knee reporting system for abstract assessment. *Dan Med J.* 2017; 64(4):A5346.
 21. Kuczmarski TM, Raja AS, Pallin DJ. How do Medical Societies Select Science for Conference Presentation? How Should They? *West J Emerg Med.* 2015;164):543-50. <https://doi.org/10.5811/westjem.2015.5.25518>.
 22. Samartine Junior H, Paiva DF, Gracitelli GB, Mazzini LR, Levy NG, Aquino JLB, et al. Bibliometric analysis and conversion rate of abstracts presented at the Brazilian Congress of Coloproctology into publication of full articles. *Rev Col Bras Cir.* 2023;50:e20233560. <https://doi.org/10.1590/0100-6991e-20233560-en>.
 23. Rowe BH, Strome TL, Spooner C, Blitz S, Grafstein E, Worster A. Reviewer agreement trends from four years of electronic submissions of conference abstract. *BMC Med Res Methodol.* 2006;6:14. <https://doi.org/10.1186/1471-2288-6-14>.
 24. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF, CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med.* 2008;5(1):e20. <https://doi.org/10.1371/journal.pmed.0050020>.
 25. Hasbahçeci M, Başak F, Uysal Ö. Evaluation of reporting quality of the 2010 and 2012 National Surgical Congress oral presentations by CONSORT, STROBE and Timmer criteria. *Ulus Cerrahi Derg.* 2014;30(3):138-46. <https://doi.org/10.5152/UCD.2014.2722>.
 26. Maticic K, Martinic MK, Puljak L. Assessment of reporting quality of abstracts of systematic reviews with meta-analysis using PRISMA-A and discordance in assessments between raters without prior experience. *BMC Med Res Methodol.* 2019;19(1):32. <https://doi.org/10.1186/s12874-019-0675-2>.
 27. Van der Weegen W, Van Egmond JC, Geuze RE, Gosens T, Snoeker B, Poolman RW; Dutch Orthopedic Association Scientific Committee Abstract Group. A simplified, 2-question grading system for evaluating abstracts in orthopedic scientific meetings: a serial randomization study. *Acta Orthop.* 2024;95:180-5. <https://doi.org/10.2340/17453674.2024.40504>.
 28. Rothwell PM, Martyn CN. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain.* 2000;123:1964-69. <https://doi.org/10.1093/brain/123.9.1964>.
 29. Naji F, Javidan AP, Khan S, Srivatsav V, Rapanos T, Harlock J. Level of clinical evidence presented at the Society for Vascular Surgery Annual Meeting during a 5-year period (2012-2016). *J Vasc Surg.* 2018;67(3):951-9. <https://doi.org/10.1016/j.jvs.2017.10.053>.
 30. Alexandrov AV, Hennerici MG. How to prepare and deliver a scientific presentation. Teaching course presentation at the 21st European Stroke Conference, Lisboa, May 2012. *Cerebrovasc Dis.* 2013;35:202-8. <https://doi.org/10.1159/000346077>.
 31. Montgomery AA, Graham A, Evans PH, Fahey T. Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. *BMC Health Serv Res.* 2002;2:8. <https://doi.org/10.1186/1472-6963-2-8>.
 32. Deveugele M, Silverman J. Peer-review for selection of oral presentations for conferences: Are we reliable? *Patient Educ Couns.* 2017;100(11):2147-50. <https://doi.org/10.1016/j.pec.2017.06.007>.
 33. Manterola C, Grande L, Otzen T, García N, Salazar P, Quiroz G. Reliability, precision or reproducibility of the measurements. Methods of assessment, utility and applications in clinical practice. *Rev Chilena Infectol.* 2018;35(6):680-8. <https://doi.org/10.4067/S0716-10182018000600680>.