

# Confiabilidad de una escala de valoración de resúmenes para congresos

Carlos Manterola D.<sup>1,2,3</sup>, Tamara Otzen H.<sup>1,2</sup>

<sup>1</sup>Programa de Doctorado en Ciencias Médicas, Universidad de La Frontera. Temuco, Chile.

<sup>2</sup>Centro de Excelencia en estudios Morfológicos y Quirúrgicos (CEMyQ), Universidad de La Frontera. Temuco, Chile.

<sup>3</sup>Comité Científico, Sociedad de Cirujanos de Chile. Chile.

Recibido el 2021-11-08 y aceptado para publicación el 2022-03-03

**Correspondencia a:**

Dra. Tamara Otzen H.  
tamara.otzen@ufrontera.cl

## Reliability of a scale for evaluation of abstracts for conferences

**Background:** In 2013, we developed a scale to evaluate the abstracts of the congresses of the Society of Surgeons of Chile (SOCICH). **Objective:** To determine internal consistency and interobserver reliability of a scale to evaluate conference abstracts. **Material and Methods:** Reliability study. Twelve surgeons were trained virtually for 8 hours, to apply the scale. Once the training was finished, they were sent a questionnaire to evaluate the contents of the training, and several summaries-cases to be evaluated with the aforementioned scale. Descriptive statistics were applied, then the degree of agreement between observers was estimated for each item of the scale. Subsequently, intraclass correlation coefficient (ICC) was evaluated, using a mixed two-factor model where the effects of the evaluators are random and the items are fixed, using a definition of absolute agreement. In addition, the internal consistency of the items was evaluated using Cronbach's alpha, considering 95% confidence intervals (95% CI). **Results:** After analyzing the measurements of the 9 items by the 12 observers, it was verified that the ICC was 0.871; with a 95% CI of 0.700; 0.965. The internal consistency value was 0.7 considering the 9 items, it is not recommended to delete any item. **Conclusions:** The scale has good internal consistency and interobserver reliability. Therefore, it can be considered as reliable instrument to be used in the evaluation of abstracts for congresses. **Keywords:** "Congresses as Topic"[Mesh]; "Observer Variation"[Mesh]; "Reproducibility of Results"[Mesh]; intraclass correlation coefficient; internal consistency; Cronbach's alpha.

## Resumen

**Introducción:** En 2013, desarrollamos una escala, para evaluar resúmenes de congresos de la Sociedad de Cirujanos de Chile (SOCICH). **Objetivo:** Determinar consistencia interna y confiabilidad interobservador de una escala para evaluar resúmenes de congresos. **Material y Método:** Estudio de confiabilidad. Doce cirujanos fueron capacitados de forma virtual durante 8 horas, para aplicar la escala. Una vez finalizado el entrenamiento, se les envió un cuestionario para evaluar contenidos de la capacitación, y varios resúmenes-casos para ser evaluados con la escala antes señalada. Se aplicó estadística descriptiva, luego se estimó el grado de acuerdo entre observadores para cada ítem de la escala. Posteriormente, se evaluó el coeficiente de correlación (CCI), utilizando un modelo de dos factores mixtos en el que los efectos de los evaluadores son aleatorios y los ítems fijos; utilizando una definición de acuerdo absoluto. Además, se evaluó la consistencia interna de los ítems utilizando alfa de Cronbach, considerando intervalos de confianza del 95% (IC 95%). **Resultados:** Luego de analizar las mediciones de los 9 ítems por los 12 observadores, se verificó que el CCI fue de 0,871; con un IC 95% de 0,700; 0,965. El valor de la consistencia interna fue de 0,7 considerando los 9 ítems, no se recomienda eliminar ningún ítem. **Conclusión:** La escala tiene buena confiabilidad interobservador y los ítems son consistentes entre sí; por lo que puede ser considerada como un instrumento confiable para la valoración de resúmenes de congresos.

**Palabras clave:** congresos médicos; confiabilidad; reproducibilidad; confiabilidad interobservador; coeficiente de correlación intraclass; consistencia interna; alfa de Cronbach.

## Introducción

Habitualmente, quienes investigan, se reúnen en congresos y conferencias para presentar ante sus pares los resultados de sus investigaciones,

compartir esta información, y escuchar la retroalimentación por parte de ellos. Sin embargo, existe evidencia respecto de: que la mayoría de las sociedades médicas utilizan métodos subjetivos para seleccionar los resúmenes que se presenta-

rán<sup>1</sup>; y de que los resúmenes seleccionados para presentación oral son superiores a los presentados como póster<sup>2</sup>.

Comúnmente, los resultados se presentan en forma de resúmenes, documentos que deben proporcionar una síntesis eficaz de lo que se realizó, cómo se realizó y que significa lo que se obtuvo de lo realizado<sup>3</sup>. Es por ello, que el rol de un resumen no debe ser subestimado.

Generalmente, los resúmenes para un congreso son sometidos a revisión por pares, quienes aplicando diversos tipos de criterios, deciden la aceptación o rechazo de dicho resumen, o eventualmente si un resumen será presentado en forma oral, o como póster.

En las últimas décadas se han realizado numerosos esfuerzos para mejorar la calidad de los resúmenes, aplicando criterios de medicina basada en la evidencia<sup>4</sup>, diversas listas de chequeo o de verificación<sup>5</sup>, entre las que destacan iniciativas como CONSORT y STROBE<sup>6,7</sup>. Sin embargo, hacen falta herramientas aplicables de forma específica a resúmenes que se presentarán a congresos.

Por ello, en 2000, el Comité Científico de la Sociedad de Cirujanos de Chile (SOCICH) diseñó un instrumento para aplicar a los resúmenes que se presentarían en los congresos societarios; el que consideraba parámetros no discriminantes, utilizaba conceptos de difícil evaluación, y sobrevaloraba ítems, relacionados más con «la forma» que con el fondo. Esto motivó la necesidad de diseñar y construir un nuevo instrumento, válido y confiable, decisión que adoptó el Comité Científico de la SOCICH en 2013 y fue publicado en 2016. Este instrumento, está compuesto por 9 ítems agrupados en 5 dominios, con validez de fachada y contenido<sup>8</sup>; y es el que se ha venido utilizando desde entonces para la valoración de los resúmenes de los congresos de la SOCICH, y que se encuentra aún en proceso de validación y de estudios de confiabilidad.

Un instrumento de medición se considera confiable o reproducible, cuando las mediciones realizadas con él generan los mismos resultados en diferentes momentos, escenarios y poblaciones, si se aplica en las mismas condiciones; de tal forma que a mayor precisión de una medición, mayor poder estadístico en la muestra en estudio. Esta, se afecta por errores aleatorios o por azar, tanto por variabilidad del observador, del sujeto observado o del o los instrumentos utilizados para medir<sup>9</sup>.

El objetivo de este estudio fue determinar consistencia interna y confiabilidad interobservador de la escala antes señalada.

## Material y Método

El informe de este estudio se realizó siguiendo la guía GRRAS, para el reporte de estudios de confiabilidad<sup>10</sup>.

### Diseño

Estudio de confiabilidad.

### Definiciones

Se definió como confiabilidad inter observador, al grado de acuerdo obtenido entre los investigadores independientes al aplicar la escala en el mismo escenario (n = 12) en un mismo período de tiempo<sup>9</sup>. La consistencia interna se refiere al grado en que los ítems de una escala se correlacionan entre ellos, y a la magnitud en que miden el mismo constructo<sup>11</sup>.

### Instrumento

La escala, compuesta por 9 ítems distribuidos en 5 dominios, tiene como punto de corte definido para aceptar un resumen, los 55 puntos (mínimo de 0 y máximo de 100 puntos); y considera los siguientes dominios: introducción (argumentación y objetivo); material y método (diseño, muestra utilizada para la conducción del estudio, variables en estudio, análisis estadístico); resultados; conclusión; e importancia, originalidad e interés del estudio<sup>8</sup>.

### Tamaño de la muestra

Se utilizaron los criterios de factibilidad propuestos por Streiner<sup>12,13</sup>.

### Evaluadores

Participaron 12 médicos especialistas en cirugía y subespecialistas.

### Proceso de medición

Los cirujanos fueron capacitados de forma virtual durante 8 horas. Una vez finalizado el entrenamiento, a todos ellos se les envió de forma simultánea e independiente, un cuestionario para evaluar contenidos de la capacitación (12 preguntas de selección múltiple, para evaluar conceptos teóricos); y un caso-problema (resumen de un congreso de cirugía anterior) para que fuera evaluado de forma independiente, aplicando la escala; mediante lo cual, se evaluó confiabilidad interobservador.

Para evaluar consistencia interna de la escala se les enviaron distintos resúmenes a los participantes, considerando un total de 24 evaluaciones con la escala antes señalada.

Los datos de todas las mediciones fueron recogidos de forma codificada y enmascarada, en una planilla Excel.

**ARTÍCULO ORIGINAL**

**Tabla 1. Características de los observadores**

Variables	n	%
Sexo		
Masculino	8	66,7
Femenino	4	33,3
Subespecialidad		
Tórax	5	41,8
Esófago-gástrico	3	25,0
Cabeza-cuello	1	8,3
Colorectal	1	8,3
Mama	1	8,3
Vascular	1	8,3

**Tabla 2. Resultados de la evaluación de objetivos teóricos administrados durante la capacitación para la utilización de la escala**

Nº pregunta	Respuestas correctas	Respuestas incorrectas	% respuestas correctas
1	12	0	100
2	12	0	100
3	11	1	91,7
4	9	3	75,0
5	9	3	75,0
6	12	0	100
7	9	3	75,0
8	12	0	100
9	3	9	25,0
10	1	11	8,3
11	9	3	75,0
12	11	1	91,7
Total	110	34	76,4

**Tabla 3. Porcentaje de acuerdo entre observadores para cada ítem de la escala**

Dominios e ítems de la escala	% acuerdo entre observadores
Dominio 1 (introducción)	
Antecedentes	83,3
Objetivos	83,3
Dominio 2 (material y método)	
Diseño	58,3
Muestra	100,0
Variables	91,7
Análisis	66,7
Dominio 3 (resultados)	100,0
Dominio 4 (conclusiones)	91,7
Dominio 5 (importancia)	50,0

**Análisis estadístico**

Después de realizar un análisis exploratorio de los datos se aplicó estadística descriptiva. Luego se estimó el grado de acuerdo entre observadores para cada ítem de la escala. Posteriormente se evaluó el coeficiente de correlación intraclase, utilizando un modelo de dos factores mixtos en donde los efectos de los evaluadores son aleatorios y los ítems son fijos, utilizando una definición de acuerdo absoluto. Finalmente, se evaluó consistencia interna de los ítems de la escala utilizando alfa de Cronbach. Los análisis se realizaron utilizando el software IBM SPSS versión 26, considerando los intervalos de confianza del 95% (IC 95%).

**Aspectos éticos**

Se guardó la confidencialidad de los participantes, quienes autorizaron la utilización de sus datos. Respetando los principios de la declaración de Helsinki<sup>14</sup>.

**Resultados**

Las características de los cirujanos que participaron como observadores se pueden apreciar en la Tabla 1. Todos tienen una mediana de edad de 44 años (mínima de 40 y máxima de 65); y una mediana de experiencia profesional de 19 años.

En la evaluación realizada al finalizar la capacitación, se constató que la mediana de respuestas correctas fue de 70,9% (mínimo de 58,3% y máximo de 83,3%). El detalle de los resultados de esta evaluación se resume en la Tabla 2.

El puntaje total de la escala aplicado al escenario problema por los 12 observadores tuvo una mediana de 31 puntos. Por otro lado, ninguno de los evaluadores asignó más de 40 puntos, por ende, todos coincidieron en que el escenario/caso no debía ser aceptado para ser presentado en el congreso.

El porcentaje de acuerdo entre observadores para cada ítem de la escala se puede observar en la Tabla 3, destacando un excelente nivel de acuerdo en los dominios 1, 3 y 4. Heterogéneo entre los ítems que componen el dominio 2; y pobre en el dominio 5.

Después de analizar las mediciones de los 9 ítems por los 12 observadores, se verificó que el coeficiente de correlación intraclase de la escala fue 0,871; con IC 95% de 0,700; 0,965. Sin embargo, al valorar los dominios que la integran por separado, verificamos un buen desempeño de los dominios 1, 3 y 4, heterogéneo para el dominio 2, y pobre en el dominio 5.

El valor de la consistencia interna de los 9 ítems

Tabla 4. Alfa de Cronbach si se suprime un ítem

Ítem	Media $\pm$ DE	Alfa de Cronbach
1	4,17 $\pm$ 1,90	0,69
2	6,25 $\pm$ 2,21	0,68
3	3,00 $\pm$ 2,57	0,69
4	6,46 $\pm$ 3,12	0,72
5	1,88 $\pm$ 2,47	0,65
6	3,96 $\pm$ 4,17	0,64
7	7,08 $\pm$ 4,87	0,69
8	4,79 $\pm$ 2,75	0,68
9	3,96 $\pm$ 3,29	0,63

DE: Desviación estándar.

de la escala fue de 0,7 en las 24 mediciones realizadas. Al considerar el alfa si algún elemento se suprime, ninguno de los valores fue significativamente alto, por lo que no se recomienda eliminar ítems (Tabla 4).

## Discusión

Hay una serie de razones para presentar los resultados de una investigación en congreso. Entre ellas, destaca el hecho que esta instancia ayuda en la preparación del manuscrito que se enviará a publicación; y que los comentarios inmediatos a la presentación permitirán enriquecer el manuscrito antes de su envío<sup>15</sup>. Y esto, radica en el hecho que la única forma en que científicos se comunican entre sí, es a través de la publicación de manuscritos, que representa un verdadero arte, desarrollando habilidades de “redacción” de un artículo científico, que en definitiva nos lleva a la máxima “publicar o perecer”<sup>16</sup>. Sin embargo, la redacción de un buen resumen no es tan simple<sup>3,17</sup>.

Existen algunos instrumentos desarrollados a lo largo de la historia que han buscado este mismo objetivo, como la propuesta de Timmer et al, publicada en 2003, compuesta por 19 ítems de respuesta ordinal, que cuenta con estudios de validez y confiabilidad, pero de compleja aplicación, pues la puntuación se calcula como la relación entre puntajes obtenidos y puntuación máxima posible, la que a su vez depende del número de ítems que son aplicables según el tipo de investigación<sup>18</sup>. La otra, es la de van der Steen et al, publicada en 2003 y 2004, y que consiste en asignar entre -6 puntos (inaceptable) a 6 puntos (excelente) a cada resumen

(con base en que los resúmenes debían incluir: una Indicación/Propósito, Material y métodos, Resultados y una Conclusión (cada criterio puntúa de la siguiente forma: abordado de manera insuficiente (-1), mal pero aceptablemente (0) o suficientemente (+1)). Además, a cada resumen se le asigna una sexta puntuación por el impacto esperado en los participantes de la reunión, que va de -1 para resúmenes no valorados hasta +1 cuando se recomienda su presentación; formato que determinó variabilidad considerable entre revisores<sup>19,20</sup>.

Al planificar el estudio, y en consideración de las experiencias anteriores, constatamos que la aplicación del instrumento requería de un proceso de calibración entre los usuarios (capacitación mínima de las bases de su origen y su ulterior aplicación); de lo contrario, se corría el riesgo de tener criterios de medición e interpretación distintos y por ende resultados dispares, lo que constituiría un error sistemático en su aplicación y por tanto, un sesgo de medición y de clasificación<sup>21</sup>.

La capacitación realizada fue acompañada de una prueba de salida, que permitió constatar que el 76,4% de las preguntas fueron bien contestadas por los participantes, con la mayor parte de ellas con excelente desempeño (Tabla 2); con excepción de las preguntas 9 y 10 (opciones de dividir de forma esquemática una ponencia en un congreso, y etapas fundamentales del reporte de resultados de un resumen para un congreso respectivamente). No obstante ello, es relevante señalar que esta capacitación, a pesar del tiempo (8 horas) y formato (*on-line*), permitieron que los participantes pudieran cumplir de forma bastante adecuada con la evaluación del escenario/caso, que como se mencionó, correspondía a un resumen verdadero y anónimo, presentado en algún congreso anterior de la SOCICH, en términos de que en forma unánime fue rechazado por los evaluadores.

Pudimos verificar que la escala como constructo, presenta una buena confiabilidad interobservador (CIC de 0,871); independiente que el porcentaje de acuerdo entre observadores para algunos dominios no fuera adecuado. En el caso del dominio 2, el acuerdo fue heterogéneo, posiblemente porque está compuesto por ítems que miden elementos metodológicos diferentes (el ítem que produjo mayor conflicto fue “diseño del estudio”, en el que 5 evaluadores confundieron serie de casos con diseño inclasificable). Y en el caso del dominio mono-itémico 5, en el que el grado de acuerdo entre observadores fue pobre; esto puede deberse a la subjetividad de un elemento cuyo contenido es más opinable que medible.

**Tabla 5. Propuesta de modificaciones de la escala de medición de calidad de resúmenes para congresos**

Dominios / Ítems	Puntaje
<p>1. Introducción</p> <p>Argumentación del porque se desarrolla el estudio (0 - 5) Adecuada: 5 puntos Inadecuada: 0 punto</p> <p>Objetivo del estudio (0 - 10 puntos) Claro, preciso y conciso: 10 puntos Vago: 5 puntos Sin objetivo: 0 punto</p>	0 - 15
<p>2. Material y Método</p> <p>Diseño utilizado en el estudio (0 - 10 puntos) RS c/MA, EC c/ AA y estudios experimentales: 10 puntos RS s/MA, EC s/ AA y de cohorte prospectiva: 8 puntos Casos y controles; corte transversal; pruebas diagnósticas: 6 puntos Cohorte retrospectiva y serie de casos: 4 puntos Diseño inclasificable o inespecífico: 2 puntos No menciona diseño utilizado: 0 punto</p> <p>Características de la población estudiada: (0 - 15 puntos) <math>\Sigma</math> de los 3 ítems Se definen las características de la población (criterios de inclusión/exclusión): 5 Se realiza estima del tamaño de la muestra: 5 puntos Se utiliza alguna técnica de muestreo: 5 puntos No define criterios de selección; tamaño de muestra ni técnica muestreo: 0 punto</p> <p>Variables del estudio (0 - 15 puntos) Especifica la variable resultado y otras variables de interés: 10 puntos Sólo menciona variables a estudiar: 5 puntos No especifica variables a estudiar: 0 punto</p> <p>Análisis estadístico aplicado (0 - 10 puntos) Estadística analítica: 10 puntos Estadística descriptiva: 5 puntos No utiliza herramientas estadísticas: 0 punto</p>	0 - 55
<p>3. Resultados</p> <p>Acordes a objetivos y metodología empleada (0 -15 puntos) Completos: 15 puntos Incompletos: 5 puntos No acordes a la metodología empleada: 0 punto</p>	0 - 15
<p>4. Conclusión (sólo si corresponde)</p> <p>Congruente con los resultados y objetivos: 10 puntos Vaga, poco congruente: 5 puntos No congruente o no corresponde plantearla: 0 punto</p>	0 - 10
<p>5. Importancia, originalidad e interes del estudio</p> <p>Relevante. Aporte al conocimiento actual: 5 puntos Poco relevante. Aporte discreto al conocimiento actual: 0 puntos</p>	0 - 5
Puntaje ( $\Sigma$ dominios 1 al 5)	
Puntaje final ideal	0 - 100
Puntaje de corte actual para aceptación del resumen	55

\*En azul los modificaciones propuestas a la escala.

Por otra parte, se comprobó que las respuestas entre los ítems de la escala están correlacionadas entre sí, orientadas a la medición del mismo constructo, ya sea de forma favorable o desfavorable, con un alfa de Cronbach de 0,70; sin que sea necesario eliminar algún ítem para incrementar la consistencia interna de la escala (Tabla 4).

Respecto de limitaciones del estudio, se puede mencionar el poco tiempo de entrenamiento para los observadores, lo que podría asociarse a cierto riesgo de sesgo de medición, y por ende, influir en los resultados finales, de tal forma que para reproducir estos resultados nos parece necesario contar con mayor tiempo de entrenamiento, o una re-capacitación periódica. Por otra parte, nos parece que el dominio 5, debe ser evaluado respecto de su real utilidad como parte del instrumento, pues es el que menos concordancia tiene y resulta poco discriminante; por lo que pudiera reducirse su impacto en el puntaje total; redistribuyendo la puntuación restada, al dominio material y método, por ejemplo a “características de la población en estudio” (Tabla 5).

Por otra parte, sería fundamental replicar estos análisis con un mayor número de sujetos. Considerando que cuando se utiliza alfa de Cronbach para evaluar consistencia interna se recomienda utilizar muestras mayores a 100 sujetos o diez participantes por cada ítem<sup>22,23</sup>. Además, en caso de obtener una mayor muestra se podrían hacer análisis de regresiones que permitan determinar el peso específico de cada ítem en el constructo y las dimensiones que lo componen.

Nos parece interesante comentar también, que estos resultados permitirán continuar trabajando con la aplicabilidad de la escala, a modo de línea de investigación. Por ejemplo, sería interesante realizar la determinación de confiabilidad intra-observador (que requiere de una metodología diferente a la que

se utilizó); y ejecutar posteriormente el estudio de validez de constructo y generar opciones de puntos de corte, de modo de tal de ofrecer esta variable como un punto basado en el método y no en pareceres.

### Conclusión

Se puede concluir señalando que la escala tiene buena confiabilidad interobservador y consistencia interna, por lo que puede ser considerada como un instrumento a utilizar en la valoración de resúmenes para congresos. Se deben continuar realizando estudios tendientes a perfeccionar y refinar cada vez más este instrumento.

### Recomendación

Esta escala puede ser considerada como instrumento a utilizar en la valoración de resúmenes para congresos.

### Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que en este manuscrito no se han realizado experimentos en seres humanos ni animales.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Financiación:** Ninguna.

**Conflictos de interés:** Ninguno.

### Bibliografía

- Kuczmarski TM, Raja AS, Pallin DJ. How do Medical Societies Select Science for Conference Presentation? How Should They? *West J Emerg Med.* 2015;16:543-50. doi: 10.5811/westjem.2015.5.25518.
- Dos Santos EF, Pereira MG. Quality of the structured abstracts presented at a congress. *Rev Assoc Med Bras.* 2007;53:355-9. doi: 10.1590/s0104-42302007000400023.
- Manterola C, Pineda V, Vial M, Grande L. How should the results of a scientific study be presented? I. The oral presentation. *Cir Esp.* 2007;81:12-7. doi: 10.1016/s0009-739x(07)71250-2.
- Manterola C, Asenjo-Lobos C, Otzen T. Hierarchy of evidence: levels of evidence and grades of recommendation from current use. *Rev Chilena Infectol.* 2014;31:705-18. doi: 10.4067/S0716-10182014000600011.
- 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;38:774-86. doi: 10.4067/S0717-95022020000300774.
- 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:e20. doi: 10.1371/journal.pmed.0050020.

## ARTÍCULO ORIGINAL

7. 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. doi: 10.1186/1471-2288-11-161.
8. 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;68:302-9. doi: 10.1016/j.rchic.2015.11.002.
9. 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;35:680-8. doi: 10.4067/S0716-10182018000600680.
10. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96-106. doi: 10.1016/j.jclinepi.2010.03.002.
11. Hebson RK. Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha. *Meas Eval Couns Dev.* 2001;34:177-89. doi: 10.1080/07481756.2002.12069034.
12. Streiner DL. Measure for measure: new developments in measurement and item response theory. *Can J Psychiatry.* 2010;55:180-6. doi: 10.1177/070674371005500310.
13. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4<sup>th</sup> ed. Oxford (GB): Oxford University Press 2008.
14. Helsinki Statement. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 64th WMA General Assembly, Fortaleza, Brazil, October 2013. Disponible en: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects>.
15. Wood GJ, Morrison RS. Writing abstracts and developing posters for national meetings. *J Palliat Med.* 2011;14:353-9. doi: 10.1089/jpm.2010.0171.
16. Sengupta S, Shukla D, Ramulu P, Natarajan S, Biswas J. Publish or perish: The art of scientific writing. *Indian J Ophthalmol.* 2014;62:1089-93. doi: 10.4103/0301-4738.146760.
17. Alexandrov AV, Hennerici MG. Writing good abstracts. *Cerebrovasc Dis.* 2007;23:256-9. doi: 10.1159/000098324.
18. Timmer A, Sutherland LR, Hilsden RJ. Development and evaluation of a quality score for abstracts. *BMC Med Res Methodol.* 2003;3:2. doi: 10.1186/1471-2288-3-2.
19. 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;113:353-9. doi: 10.1097/01.PRS.0000097461.50999.D4.
20. 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;111:2215-22. doi: 10.1097/01.PRS.0000061092.88629.82.
21. Manterola C, Otzen T. Bias in clinical research. *Int J Morphol.* 2015;33:1156-64. doi: 10.4067/S0717-95022015000300056.
22. Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol.* 1993;78:98-104. doi: 10.1037/0021-9010.78.1.98.
23. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods.* 2000;5:343-55. doi: 10.1037/1082-989x.5.3.343.