



FACULTAD DE CIENCIAS DE LA SALUD
CARRERA PROFESIONAL DE ESTOMATOLOGÍA

**“APLICACIÓN Y EFICIENCIA DEL USO DE FÓRMULAS DE
TAMAÑO DE MUESTRA EN ENSAYOS CLÍNICOS
ALEATORIZADOS PUBLICADOS EN REVISTAS DE ALTO
FACTOR DE IMPACTO EN ORTODONCIA”**

Tesis para optar el grado académico de:

MAESTRO EN ORTODONCIA Y ORTOPEDIA MAXILAR

Presentado por:

Marialicia Calderón Augusto (0000-0002-3339-4398)

Asesor:

Luis Ernesto Arriola Guillen (0000-0003-0010-5948)

Lima – Perú

2022

ACTA DE SUSTENTACIÓN DE TESIS

Lima, 06 de junio de 2022.

Los integrantes del Jurado de tesis:

Presidente: Dra. Esp. Patricia Aurora Vidal Manyari.
Miembro: Mg. Esp. Gustavo Adolfo Watanabe Kanno.
Miembro: Mg. Esp. Gissella Roxana Gutiérrez Tapia.

Se reúnen para evaluar la tesis titulada: **APLICACIÓN Y EFICIENCIA DEL USO DE FÓRMULAS DE TAMAÑO DE MUESTRA EN ENSAYOS CLÍNICOS ALEATORIZADOS PUBLICADOS EN REVISTAS DE ALTO FACTOR DE IMPACTO EN ORTODONCIA.**

Presentada por el estudiante/bachiller:

- **MARIALICIA CALDERÓN AUGUSTO con DNI 43052906.**

Para optar el Grado Académico de Maestro en Ortodoncia y Ortopedia Maxilar.

Asesorada por: Dr. Esp. CD. Luis Ernesto Arriola Guillén.

Luego de haber evaluado el informe final de tesis y evaluado el desempeño del estudiante de la Carrera de Estomatología en la sustentación, se concluyen de:

Manera unánime ()

Por mayoría ()

Calificar a:

Tesis: MARIALICIA CALDERÓN AUGUSTO		Nota (en letras): Dieciocho	
Aprobado (<input type="checkbox"/>)	Aprobado - Muy buena (<input checked="" type="checkbox"/>)	Aprobado - Sobresaliente (<input type="checkbox"/>)	Desaprobado (<input type="checkbox"/>)


Los miembros del jurado firman en señal de conformidad.



Dra. Esp. Patricia Aurora Vidal Manyari
Presidente del Jurado
DNI: 08764853



Mg. Esp. Gissella Roxana Gutiérrez Tapia
Miembro 2
DNI. 06806243



Mg. Esp. Gustavo Adolfo Watanabe Kanno
Miembro 1
DNI. 40230501



Dr. Esp. CD Luis Ernesto Arriola Guillén
Asesor
DNI. 40990364

INDICE

- I. Resumen
- II. Abstract
- III. Introducción
- IV. Materiales y Métodos
- V. Resultados
- VI. Discusión
- VII. Conclusión
- VIII. Referencias Bibliográficas
- IX. Anexos

Formato de Tesis Artículo Original: Publicado en la revista Turk Journal of Orthodontics

I.- RESUMEN

Introducción: El propósito de este estudio fue determinar la frecuencia de aplicación y la pertinencia en el uso de fórmulas de tamaño de muestra utilizadas en ensayos clínicos (EC) publicados durante los años 2015-2019 en las revistas de ortodoncia de más alto factor de impacto.

Métodos: Este estudio transversal evaluó 142 EC publicados durante los años 2015-2019 en las cuatro revistas de mayor factor de impacto de la ortodoncia según el ranking SCIMAGO 2018. Se incluyeron EC aleatorizados y realizados en seres humanos. Dos ortodoncistas entrenados y calibrados evaluaron si los EC seleccionados reportaron el cálculo de muestra y si estos describieron los criterios de su cálculo adecuadamente, incluyendo el tipo de test, nivel de significancia, poder de la prueba, precisión o tamaño del efecto (diferencia clínicamente relevante) y varianza. Se determinó un cálculo adecuado cuando reportaron por lo menos 4 de 5 criterios señalados. Se utilizaron las pruebas de chi cuadrado y regresión logística binaria.

Resultados: 120 publicaciones (84.5%) reportaron el cálculo de muestra en su redacción, pero solo 61 (50.8%) reportaron adecuadamente el cálculo (por lo menos 4 de 5 parámetros). Fallaron en reportar el tipo de test 101 (84.2%), el nivel de confianza 10(8.3%), el poder de prueba 3 (2.5%), no precisaron el detalle de la obtención del tamaño de la muestra 38 (31.7%). Sólo 9 (7.5%) cumplió con describir todos los 5 parámetros evaluados. Finalmente, de los trabajos que habían usado un cálculo de muestra solo 96 (80%) superaba el mínimo tamaño de muestra requerido. Por lo tanto 24 (20%) no cumplió lo mínimo requerido. Cuando se evaluó si el año de publicación estuvo relacionado a la pertinencia del reporte se encontró que por cada año que pasó, hubo 1.301 veces más probabilidad de reportar inadecuadamente el cálculo de muestra ($P=0.048$).

Conclusiones: Los EC de las revistas evaluadas frecuentemente no reportan con claridad los parámetros utilizados en el cálculo de muestra, cuestionando su validez externa. La situación está progresivamente empeorando. Futuras investigaciones deben reportar la justificación del tamaño de muestra de manera clara y transparente.

Keywords: Ensayos clínicos aleatorizados, ensayos clínicos

II.- ABSTRACT

Introduction: The purpose of this was to determine the frequency of application and the relevance in the use of sample size formulas used in clinical trials (CT) published in the years 2015-2019 in four leading orthodontic journals

Methods: This cross-sectional study evaluated 142 CT published between 2015 to 2019 in the four journals with the greatest impact factor for orthodontics according to SCIMAGO 2018 ranking. Randomized CT conducted on humans were included. Two trained and calibrated orthodontists assessed if the included CT reported their sample calculations and if they adequately described the criteria for the calculations, including the type of tests, significance levels, test powers, precision or effect sizes (clinically relevant difference) and variability. A sample size calculation was considered adequate when the five criteria were reported. Chi-square tests were used to assess differences.

Results: 120 publications (84.5%) reported the sample size calculation but only 61 (50.8%) adequately reported the calculation (at least 4 of 5 parameters). They failed to report type of test 101 (84.2%), confidence level 10 (8.3%), test power 3 (2.5%), they did not specify the detail of obtaining the sample size 38 (31.7%). Only 9 (7.5%) complied with describing all 5 parameters evaluated. Finally, of the studies that had used a sample calculation, only 96 (80%) exceeded the minimum required sample size. Therefore, 24 (20%) did not meet the minimum requirement. When evaluating whether the year of publication was related to the relevance of the report, it was found that for each year that passed, there was 1.301 times more probability of inappropriately reporting the sample calculation ($P=0.048$).

Conclusions: Published CT in four leading orthodontic journals frequently do not clearly report the parameters used for their sample calculations, questioning their external validity. The situation is progressively getting worse. Future research should report the justification for the sample size in a clear and transparent manner.

Keyword: Randomized clinical trials, clinical trials

III.- INTRODUCCIÓN

Los ensayos clínicos (EC) son estudios de valor en la literatura científica ortodóncica, debido a que permiten responder a preguntas de investigación clínica mediante cuasi-experimentación científica y facilitan a los ortodoncistas tomar posturas sobre las alternativas de tratamiento frente a un determinado diagnóstico.¹⁻⁵ Los EC correctamente ejecutados proporcionan la mejor evidencia sobre el efecto de las intervenciones sanitarias.^{1,6} Así mismo, la capacidad de extrapolar los resultados de un EC va a depender del control de sesgos que una investigación ha tenido durante su ejecución.^{7,8} Además, todo investigador que diseña un EC busca que sus resultados sean extrapolables a la práctica clínica por los ortodoncistas;^{3,4} siendo así, además de la randomización una de las condiciones que requiere un estudio para ser extrapolado, que es la capacidad que tiene un ensayo para generalizar sus resultados a individuos con semejantes criterios de selección, es el tamaño de muestra que han empleado en su investigación.^{1,7,9-11}

Los tamaños de muestra de las investigaciones deberían ser representativos de sus poblaciones de estudio y en los estudios clínicos se busca demostrar o no un efecto clínicamente relevante a través del tamaño de muestra.^{6,10,12} La representatividad de una muestra esta dada por un adecuado tamaño de muestra y por el tipo de muestreo utilizado que debería ser probabilístico.^{13,14} Por lo tanto, el uso de una fórmula de tamaño muestral sólo es el inicio de un buen comienzo en el control de sesgos,¹⁵⁻¹⁸ ya que un estudio puede reportar el uso de una fórmula muestral pero podría cometer errores en los datos consignados en el cálculo de muestra.¹² Siendo así, es imprescindible que en la sección de materiales y métodos de los artículos se redacten los criterios consignados en el cálculo de muestra para comprender exactamente que un cálculo estuvo bien estimado.^{9,13,15,16} Con frecuencia, los errores en la consignación de datos de una fórmula muestral probablemente no van a estar en la designación del nivel de confianza, ni en el uso del poder de la prueba.^{19,20} Los problemas en el caso de una variable resultado cuantitativa empezarán en la determinación del tamaño del efecto (precisión), que es la mínima diferencia que se desea detectar entre los grupos a comparar o desde un punto de vista clínico la diferencia minimamente significativa para decidir que un tratamiento es mejor que el otro, o podrá estar en el reporte de la desviación estandar o varianza que presente el grupo control, por lo tanto una especificación de donde se obtuvieron estos datos

debería estar reportada en esta sección.¹¹ En una variable resultado cualitativa los errores que se podrían cometer puntualmente estarían relacionados a un mal reporte de las proporciones estimadas en los grupos comparados, por lo tanto la fundamentación del origen de estos datos consignados es importante para una buena estimación de muestra.^{8,21} La odontología basada en la evidencia científica reporta a los ensayos clínicos aleatorizados como los diseños que se encuentran en la punta de la pirámide de los diseños de estudios y que son la principal fuente de aporte a las revisiones sistemáticas, por lo tanto deberían tener un buen reporte de sus cálculos de muestra para permitir una mayor validez externa de sus resultados.^{5,6,15,16,18,22} Por esto, el propósito de éste estudio fue determinar la frecuencia de aplicación y la pertinencia en el uso de datos utilizados en el cálculo de fórmulas de tamaño de muestra utilizadas en los ensayos clínicos publicados durante los años 2015-2019 en las cuatro revistas de mayor factor de impacto de la ortodoncia según el ranking SCIMAGO 2018.

IV.- MATERIALES Y MÉTODOS

Este estudio descriptivo transversal fue inscrito en el comité de investigación de la Universidad Científica del Sur (Lima- Perú) con un número de protocolo 669-2019-POS8.

Se evaluaron todos los artículos correspondientes a ensayos clínicos publicados desde enero del 2015 hasta diciembre del 2019, pertenecientes a las cuatro revistas de mayor factor de impacto del área de la ortodoncia según el ranking de revistas SCIMAGO 2018 <https://www.scimagojr.com/journalrank.php?area=3500&category=3505> que incluyó las siguientes revistas: “Angle Orthodontist”, “American Journal of Orthodontics and Dentofacial Orthodontics”, “European Journal of Orthodontics”, and “Korean Journal of Orthodontics”.

Los criterios de inclusión involucrados fueron ensayos clínicos aleatorizados definidos por sus autores en el título de sus publicaciones o en la sección de metodología y que fueron realizados en seres humanos. Se excluyeron ensayos en animales, cuasi experimentos y estudios observacionales y estudios piloto (definidos así por sus propios autores en el título).

Dos ortodoncistas (MCA y LEAG) capacitados y entrenados en la parte estadística del cálculo de tamaño muestral; fueron calibrados en una prueba piloto previa realizada con 30 evaluaciones por los dos evaluadores obteniendo una concordancia intra e interobservador de Kappa de 0.93 en ambas mediciones en todas las variables evaluadas. En seguida, los evaluadores continuaron

con la búsqueda de los artículos en las revistas seleccionadas, y ubicaron los EC que cumplieron los criterios de selección establecidos, clasificándolos por revista y año. Una vez identificados, se verificó en cada uno de ellos, si reportaron el cálculo de muestra por medio del empleo de una fórmula de tamaño muestral (fórmula para comparar dos promedios o fórmula para comparar dos proporciones). Inmediatamente, se verificó en los artículos que describieron sus cálculos de muestra, si reportaban los respectivos parámetros que incluyeron el tipo de test (una cola o 2 colas), nivel de significancia, poder de la prueba, precisión (diferencia que se deseó detectar, referenciando de donde obtuvieron este dato) y la desviación estándar o varianza del grupo control. Sólo se utilizó la información reportada en la publicación. En seguida, se consideró adecuado un reporte cuando por lo menos reportaron 4 de los 5 parámetros descritos pudiendo dejar de reportar el tipo de test empleado, de lo contrario, fue considerado una descripción inadecuada. Todos los artículos fueron evaluados dos veces por los dos evaluadores con un espacio entre ellas de un mes, y en casos de discrepancias en la definición de algún criterio, el consenso de los dos evaluadores definió el cumplimiento o no de los mismos.

Análisis Estadístico

Todos los análisis fueron realizados utilizando el software para Windows de SPSS (versión 24.0; IBM, Armonk, NY). Primero se describió el cumplimiento del uso de las fórmulas de tamaño de muestra en general y en cada una de las cuatro revistas evaluadas, después se evaluó si la descripción de estas cumplió con la descripción adecuada de los parámetros requeridos para el cálculo de muestra. Luego se evaluó la asociación entre los criterios cumplidos, el número de criterios cumplidos, la revista evaluada y el año de publicación por medio del test de Chi-cuadrado de Pearson. Finalmente, se aplicó un test de regresión logística binaria para predecir la eficiencia del reporte del cálculo de muestra (por lo menos 4 de los 5 parámetros descritos) según el tipo de revista y el año de publicación. El nivel de significancia establecido fue $P < 0.05$.

V.- RESULTADOS

Se encontraron 142 ensayos clínicos en las cuatro revistas evaluadas, de las cuáles 120 (84.5%) si habían reportado el cálculo de muestra y sólo 22 (15.5%) no reportaron cálculo de muestra en su redacción. Las cuatro revistas superaron el 80% en la descripción del cálculo muestral. Sin embargo, cuándo se evaluó si describieron de manera completa su cálculo de muestra, se observó que en general no informaron acerca del tipo de test evaluado 101 (84.2%), fallaron en reportar el nivel de confianza en 10 (8.3%), no describieron el poder de la prueba en 3 (2.5%),

no describieron la precisión 19 (15.8%) y no describieron la variabilidad 49(40.8%), en este último rubro el AJODO fue la que describió este rubro en mayor porcentaje (77.4%, $p=0.011$).

(Tabla I). Si evaluamos el cumplimiento completo de la descripción de los requisitos para poder hacer un cálculo de muestra, en general solo 9 (7.5%) cumplió con describir todos los parámetros, y 61 (50.8%) cumplió por lo menos 4 parámetros y sólo 30 (25%) cumplió tres parámetros. (Tabla II).

Cuando se evaluó lo adecuado del reporte, considerando los 5 parámetros se encontró que de los 120 artículos encontrados sólo 9 (7.5%) reportaron adecuadamente el cálculo sin encontrar asociación significativa ($p=0.346$) con alguna revista en específico, y cuando se evaluó si el año de publicación estuvo relacionado a la eficiencia del reporte no se observó asociación significativa ($p=0.083$), sin embargo se notó que el año que mejor se hicieron los reportes fue el 2015. (Tabla III)

Cuando se evaluaron algunos parámetros específicos del cálculo de muestra, se observó cuán frecuentemente describían los detalles de la obtención de la precisión del cálculo de muestra y se observó que en general solo 73 (60.8%) reportaba detalles de la precisión y estos eran obtenidos de la literatura, sólo 9 (7.5%) utilizaba una prueba piloto previa para hacer su cálculo de muestra, mientras que 38 (31.7%) no describía como obtuvieron la precisión de su cálculo. Finalmente, de los trabajos que habían usado un cálculo de muestra solo 96 (80%) superaba el mínimo tamaño de muestra requerido. Por lo tanto 24 (20%) no cumplió lo mínimo requerido. (Tabla IV).

La tabla V presenta una regresión logística binaria para predecir la eficiencia del reporte del cálculo de muestra (por lo menos 4 parámetros descritos) según el tipo de revista y el año de publicación, encontrándose que por cada año que pase hay 1.301 más riesgo de reportar inadecuadamente el cálculo de muestra $P=0.048$.

VI.- DISCUSIÓN

El cálculo de muestra en las investigaciones ha sido creado para tener un número representativo de individuos de una población de estudio.^{6,10,12} Aunque, idealmente los estudios deberían evaluar a toda la población de estudio, por razones de dinero y tiempo esta posibilidad es poco frecuente y la mayoría de investigaciones usan muestras no representativas y equivocadamente buscan extrapolar sus resultados a su población de estudio.²³ Sin embargo, para lograr una validez externa de los resultados, el estudio debe

tener representatividad de su muestra que está relacionada primero con el tamaño de muestra calculado mediante fórmulas de tamaño muestral y segundo con el tipo de muestreo utilizado que debería ser probabilístico, asegurando que todos los individuos de una población de estudio tengan la misma probabilidad de ser elegidos en el estudio diseñado.^{6,10-14} Siendo así, puede ser muy común que las investigaciones cometan errores durante el cálculo de tamaños de muestra por falta de entrenamiento en este tópico y por falta de difusión sobre la importancia metodológica del mismo, que consecuentemente afecta la representatividad de sus muestras y por ende disminuye el valor de sus investigaciones.^{12,17} Actualmente, los ensayos clínicos son considerados los mejores diseños de investigación porque tienen un gran aporte al área clínica y porque permiten el desarrollo de las revisiones sistemáticas, por lo tanto también deben cumplir con la obtención de muestras representativas, y mucho más hoy que la publicación de este diseño es cada vez más exigente por parte de las mejores revistas de la ortodoncia.^{24,25} Por esto, el propósito de esta investigación fue determinar la frecuencia de aplicación y la eficiencia del uso de fórmulas de tamaño de muestra utilizadas en los ensayos clínicos publicados durante los años 2015-2019 en las cuatro revistas de mayor factor de impacto de la ortodoncia según el ranking SCIMAGO 2018.

Existen pocos antecedentes en otras áreas de la odontología, que evaluaron si los cálculos de muestra de EC se desarrollaron adecuadamente.^{24,25} En el área de prostodoncia una revisión de 42 ECA publicados en un período de 10 años entre el 2008 y el 2017 en las principales revistas del área, concluyó que el 50% de las publicaciones no describieron eficientemente el cálculo de muestra.²⁵ Asimismo, en endodoncia una revisión de 50 EC publicados entre los años 2000-2001 y 2009-2010 en dos revistas de alto impacto del área, concluyó que el 60% de ellas no indicaron cómo se determinó el tamaño de muestra. Aunque esta cifra mejoró significativamente entre los años 2009 y 2010,²⁴ estos valores aún no son alentadores. En ortodoncia a conocimiento de los autores sólo se realizó un estudio el año 2011 que evaluó cuán frecuentemente los investigadores reportaron el cálculo de tamaño de muestra en estudios publicados en revistas de Brasil y EEUU, concluyendo que los investigadores y el consejo editorial de estas revistas deberían tener una mayor preocupación por los errores relacionados con el uso de un tamaño de muestra inadecuado.²⁶

En este estudio se evaluó 142 EC publicados desde el 2015 al 2019 en cuatro revistas importantes de la ortodoncia según el ranking SCIMAGO 2018 que es muy reconocido en la actividad académica, encontrando que el 84.5% de las publicaciones reportaron el uso de un cálculo de muestra, cifra que sería alentadora para la metodología y para la validez externa de

un estudio, teniendo en cuenta que cada vez se debería ser más exigente con el cumplimiento de los criterios metodológicos en las publicaciones de alto nivel. Sin embargo, cuándo se evaluó si estas publicaciones reportaron eficientemente su cálculo de muestra, entendiéndose como eficiente si es que describieron por lo menos cuatro de los cinco criterios para el cálculo de muestra considerados en toda fórmula que incluyeron el nivel de confianza, el poder de la prueba, la precisión y la variabilidad de los resultados, en general sólo el 50.8% de las publicaciones cumplieron con reportar estos parámetros eficientemente, es decir sólo esta cantidad sería considerada bien descrita y por ende bien calculada. Es importante señalar que se consideró eficiente el cálculo aun cuando no se reportó el tipo de test evaluado asumiendo que su falta de redacción pudo ser porque históricamente este criterio no ha sido tan reportado y difundido de reportar aun cuando se debiese describir.

Al comparar los porcentajes de reporte del cálculo de muestra con años anteriores específicamente entre el 2005 y el 2008, sólo 3% reportó este cálculo en revistas brasileras y 21% en revistas americanas.²⁶ Estas cifras aumentaron considerablemente (84.5%) como lo demostró este estudio 50.8%, si se toman en cuenta los artículos que lo reportaron de forma eficiente. Esta información nos lleva a la reflexión de que, pese a que las cifras de descripción del cálculo de muestra aumentaron entre los años 2015 y 2019 en estas revistas de ortodoncia evaluadas, consideradas las de mayor impacto, la descripción de su cálculo aún es incompleta evitando que otros investigadores repliquen las mismas en otras poblaciones de estudio por falta de datos, y lo peor es que existe la posibilidad de que al no redactarlas bien, sus cálculos sean deficientes afectando la representatividad de sus muestras y por lo tanto la validez externa de sus resultados. Otro aspecto a resaltar es que mediante la regresión logística binaria empleada se pudo detectar que el año de publicación tuvo influencia directa en un reporte eficiente del cálculo de muestra, encontrándose que por cada año que pase es decir un año más reciente existe 1.301 probabilidades de reportar inadecuadamente el cálculo de muestra $P=0.048$, condición que deberían tener en cuenta los editores e investigadores para no dejar de cumplir eficientemente este reporte.

Por otro lado, existe la probabilidad de que las publicaciones pudieron omitir estos datos en su reporte, pero que, si hayan hecho bien los cálculos de muestra, sin embargo, esa interpretación es subjetiva porque también existe la posibilidad que los hayan hecho de forma inadecuada y por lo tanto este reporte debe mejorar en las futuras publicaciones ya que tienen una relación directa con la validez externa de los resultados.^{24,25,27,28}

Es importante notar que de los 142 estudios encontrados todos tuvieron como variable resultado una variable cuantitativa, por lo que se evaluó características específicas del cálculo de muestra para dos promedios. En general, se observó que las publicaciones no informaron acerca del tipo de test evaluado (84.2%), es decir no precisaron si el test utilizado era unilateral o bilateral, fallaron en reportar el nivel de confianza en (8.3%) siendo este el criterio más sencillo de reportar, pero no lo hicieron todas las publicaciones, no describieron el poder de la prueba en (2.5%), no describieron la precisión o el tamaño del efecto (diferencia clínicamente relevante) (15.8%), siendo este aspecto importante de reportar porque muchas investigaciones pueden errar en este criterio ya que es necesario especificarlo y cuantificarlo para que los revisores y lectores sepan si esta diferencia reportada está relacionada a la realidad clínica y sobre todo sepan de donde proviene esta cifra.²³ Muchas veces los autores en esta sección colocan valores de precisión que cuando se comparan con sus resultados no guardan relación entre ellos ya que se observan cifras distintas como si se tratase de dos estudios diferentes, o a veces estas cifras no guardan relación con la diferencia clínicamente relevante y en los estudios las presentan como si realmente lo fueran, por eso un buen reporte de esta sección se hace indispensable. Así mismo las publicaciones no describieron la variabilidad para el cálculo de muestra (40.8%), por lo tanto, no se puede saber que tanta dispersión esperaban encontrar estas publicaciones al momento de planificar sus estudios, en este último rubro el AJODO fue la revista que mejor describió este criterio (77.4%) respecto a las otras $P=0.011$. Mas allá, cuando se evaluaron algunos otros parámetros específicos del cálculo de muestra, como la precisión, se observó que el 60.8% la obtenía de la literatura científica, sólo el 7.5% la obtenía de una prueba piloto previa, mientras que el 31.7% no describía como la obtenían siendo este un parámetro importante para considerar en un cálculo de muestra. Un punto importante de los estudios evaluados es que la mayoría de las publicaciones (80%) superaban el mínimo tamaño de muestra requerido por sus cálculos de muestra, lo que es considerado como una buena característica metodológica.

Finalmente, en base a los EC evaluados del área de ortodoncia se puede determinar que, aunque actualmente estas publicaciones están reportando en mayor cantidad el uso de fórmulas de cálculo de muestra, situación que es buena para la ortodoncia basada en la evidencia científica, aún hay deficiencias cuando éstas describen cómo determinaron las mismas, llevando consecuentemente a la incertidumbre sobre la validez externa de sus resultados. Por lo tanto, estas consideraciones deben ser tomadas en cuenta en las futuras publicaciones ya que mejorarán el nivel metodológico de las mismas.

VII.- CONCLUSIONES

Los EC evaluados en cuatro revistas de ortodoncia de alto factor de impacto mayoritariamente cumplen con el uso de cálculos de muestra, pero frecuentemente no reportan eficientemente los parámetros utilizados, cuestionando su validez externa. La situación esta progresivamente empeorando. Estos resultados deben considerarse en las futuras investigaciones para mejorar su calidad.

VIII.- REFERENCIAS BIBLIOGRÁFICAS

1. Berger VW, Bears JD. When can a clinical trial be called 'randomized'? *Vaccine* 2003;21:468-472.
2. Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *Journal of clinical epidemiology* 2011;64:136-144.
3. Lempesi E, Koletsi D, Fleming PS, Pandis N. The reporting quality of randomized controlled trials in orthodontics. *Journal of Evidence Based Dental Practice* 2014;14:46-52.
4. Benson P. *Systematic reviews—professional masochism or important process?:* SAGE Publications Sage UK: London, England; 2008.
5. Zuccati G, Clauser C, Giorgetti R. Randomized clinical trials in orthodontics: Reality, dream, or nightmare? *American Journal of Orthodontics and Dentofacial Orthopedics* 2009;136:634-637.
6. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *Journal of clinical epidemiology* 2002;55:783-786.
7. Manríquez J, Valdivia G, Rada G, Letelier LM. Análisis crítico de ensayos clínicos randomizados publicados en revistas biomédicas chilenas. *Revista médica de Chile* 2005;133:439-446.
8. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine* 2010;8:18.
9. Lazcano-Ponce E, Salazar-Martínez E, Gutiérrez-Castrellón P, Angeles-Llerenas A, Hernández-Garduño A, Viramontes JL. Ensayos clínicos aleatorizados: variantes, métodos de aleatorización, análisis, consideraciones éticas y regulación. *Salud pública de México* 2004;46:559-584.
10. Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. *Epidemiologic reviews* 2002;24:26-38.
11. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj* 2017;358:j4008.
12. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses;: power and sample size for the interaction test. *Journal of clinical epidemiology* 2004;57:229-236.
13. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *Jama* 1994;272:122-124.
14. Moher D, Cook D, Jadad A, Tugwell P, Moher M, Jones A et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses; 1999.
15. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj* 2011;343:d5928.
16. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions.* John Wiley & Sons; 2011.
17. Saltaji H, Armijo-Olivo S, Cummings GG, Amin M, Flores-Mir C. Randomized clinical trials in dentistry: Risks of bias, risks of random errors, reporting quality, and methodologic quality over the years 1955-2013. *PLoS One* 2017;12:e0190089.

18. Bigby M, Williams H. Appraising systematic reviews and meta-analyses. *Arch Dermatol* 2003;139:795-798.
19. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F et al. Evaluating non-randomised intervention studies. *Health technology assessment (Winchester, England)* 2003;7:iii-x, 1-173.
20. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. *Medical uses of statistics* 1992:357-373.
21. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *Jama* 1996;276:637-639.
22. Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. *Journal of orthodontics* 2003;30:309-315.
23. Bondemark L, Ruf S. Randomized controlled trial: the gold standard or an unobtainable fallacy? *Eur J Orthod* 2015;37:457-461.
24. Shahravan A, Haghdoost AA, Rad M, Hashemipoor M, Sharifi M. Sample size calculation of clinical trials published in two leading endodontic journals. *Iran Endod J* 2014;9:56-60.
25. Dumbrigue HB, Dumbrigue EC, Dumbrigue DC, Chingbingyong MI. Reporting of Sample Size Parameters in Randomized Controlled Trials Published in Prosthodontic Journals. *J Prosthodont* 2019;28:159-162.
26. Normando D, Almeida MAO, Quintão CCA. Analysis of the use of sample size calculation and error of method in researches published in Brazilian and international orthodontic journals. *Dental Press J Orthod*. 2011 Nov-Dec;16(6):33.e1-9.
27. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ* 2017;357:j2490.
28. Koletsi D, Pandis N, Polychronopoulou A, Eliades T. What's in a title? An assessment of whether randomized controlled trial in a title means that it is one. *Am J Orthod Dentofacial Orthop* 2012;141:679-685.

IX.- ANEXOS

Table I. Evaluation of the parameters reported for sample size calculation.

Parameter	AO = 49 (100%)		AJODO = 35 (100%)		EJO = 52 (100%)		KJO = 6 (100%)		Total = 142 (100%)		P
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
Description of a sample calculation	40 (81.6)	9 (18.4)	31 (88.6)	4 (11.4)	44 (84.6)	8 (15.4)	5 (83.3)	1 (16.7)	120 (84.5)	22 (15.5)	0.860
Description of specific reports	AO = 40 (100%)		AJODO = 31 (100%)		EJO = 44 (100%)		KJO = 5 (100%)		Total = 120 (100%)		
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	
Type of test	3 (7.5)	37 (92.5)	7 (22.6)	24 (77.4)	8 (18.2)	36 (81.8)	1 (20.0)	4 (80.0)	19 (15.8)	101 (84.2)	0.335
Confidence level (%)	37 (92.5)	3 (7.5)	27(87.1)	4 (12.9)	41 (93.2)	3 (6.8)	5 (100.0)	0 (0.0)	110 (91.7)	10 (8.3)	0.689
Test Power	38 (95.0)	2 (5.0)	31 (100.0)	0 (0.0)	44 (100.0)	0 (0.0)	4 (80.0)	1 (20.0)	117 (97.5)	3 (2.5)	0.026*
Effect size	31 (77.5)	9 (22.5)	31 (100.0)	0 (0.0)	35 (79.5)	9 (20.5)	4 (80.0)	1 (20.0)	101 (84.2)	19 (15.8)	0.047*
Variability	18 (45.0)	22 (55.0)	24 (77.4)	7 (22.6)	28 (63.6)	16 (36.4)	1 (20.0)	4 (80.0)	71 (59.2)	49 (40.8)	0.011*

Chi-square test

* significant

AO “Angle Orthodontist”, AJODO “American Journal of Orthodontics and Dentofacial Orthopedics”, EJO “European Journal of Orthodontics”, KJO “Korean Journal of Orthodontics”

Table II. Number of parameters reported during sample size calculation in the journals evaluated.

Number of parameters reported	AO	AJODO	EJO	KJO	Total = 120 (100%)	<i>P</i>
	n (%)	n (%)	n (%)	n (%)	n (%)	
5	1 (2.5)	4 (12.9)	4 (9.1)	0 (0)	9 (7.5)	
4	16 (40.0)	19 (61.3)	25 (56.8)	1 (20.0)	61 (50.8)	
3	13 (32.5)	8 (25.8)	6 (13.6)	3 (60.0)	30 (25.0)	0.063
2	9 (22.5)	0 (0.0)	9 (20.5)	1 (20.0)	19 (15.9)	
1	1 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.8)	

Chi-square test

AO “Angle Orthodontist”, AJODO “American Journal of Orthodontics and Dentofacial Orthopedics”, EJO “European Journal of Orthodontics”, KJO “Korean Journal of Orthodontics”

Table III. Efficiency of sample calculation reporting (reported 5 parameters) according to the type of journal and the year of publication.

Journal	Efficient reporting	Without efficient reporting	Total
	n (%)	n (%)	n (%)

AO	1 (2.5)	39 (97.5)	40 (100)
AJODO	4 (12.9)	27 (87.1)	31 (100)
EJO	4 (9.1)	40 (90.9)	44 (100)
KJO	0 (0.0)	5 (100.0)	5 (100)
All Journals	9 (7.5)	111 (92.5)	120 (100)
Year of publication			
2015	3 (16.7)	15 (83.3)	18 (100)
2016	1 (4.2)	23 (95.8)	24 (100)
2017	3 (18.8)	13 (81.3)	16 (100)
2018	2 (7.1)	26 (92.5)	28 (100)
2019	0 (0.0)	34 (100.0)	34 (100)

Chi-square test for the journal evaluated $P=0.346$

Chi-square test for year of publication $P=0.083$

AO “Angle Orthodontist”, AJODO “American Journal of Orthodontics and Dentofacial Orthopedics”, EJO “European Journal of Orthodontics”, KJO “Korean Journal of Orthodontics”

Table IV. Description of other specific parameters reported during sample size calculation.

Other specific parameters evaluated	AO	AJODO	EJO	KJO	Total = 120 (100%)	<i>P</i>
	n (%)	n (%)	n (%)	n (%)	n (%)	

Details of obtaining the effect size

Table V. Regresión logística binaria para predecir la eficiencia del reporte del cálculo de muestra (por lo menos 4 parámetros descritos) según el tipo de revista y el año de publicación.

Predictor variable	p	Exp(B)	95% CI				p
			Lower Limit	Upper Limit	Lower Limit	Upper Limit	
Not precise	19 (47.5)	4 (12.9)	14 (31.8)	1 (20.0)	38 (31.7)	0.084	
Previous pilot test	2 (5.0)	4 (12.9)	3 (6.8)	0 (0.0)	9 (7.5)		
Obtained from literature	19 (47.5)	23 (74.2)	27 (61.4)	4 (80.0)	73 (60.8)		
Exceeded the minimum sample size calculated	0.24						
Yes	0.1126 (65.0)	0.4626 (83.9)	39 (86.9)	5 (100.0)	20 (80.0)	0.027*	
No	0.2314 (35.0)	0.5895 (16.1)	5 (11.4)	0 (0.0)	14 (20.0)		
Chi-square test	0.395	2.726	0.27		27.473		
Year	0.048*	1.301	1.003		1.687		
Constant	0.563	0.776					

AO "Angle Orthodontist", AJODO "American Journal of Orthodontics and Dentofacial Orthopedics", EJO "European Journal of Orthodontics", KJO "Korean Journal of Orthodontics"

AO "Angle Orthodontist", AJODO "American Journal of Orthodontics and Dentofacial Orthodontics", EJO "European Journal of Orthodontics", KJO "Korean Journal of Orthodontics"

