

“Hay vida más allá del valor p. Significancia clínica en kinesiología”

“There is life beyond the p value. Clinical significance in kinesiology”

Kinesiólogo Jorge Fuentes Contreras, PhD

Profesor adjunto Departamento de Kinesiología Universidad Católica del Maule

Título Abreviado: Más allá del valor p

Información del Artículo

Recepción: 25 de Junio de 2019

Aceptación: 28 de Julio de 2019

¿Cuándo fue la última vez que usted escuchó a un kinesiólogo, docente o investigador decir en una presentación, o leer en un artículo científico que no hubo diferencias entre los grupos estudiados, debido a que la diferencia no alcanzó a ser *estadísticamente significativa*? Probablemente, no deba hacer mucho esfuerzo para recordar y responder esta pregunta. Es altamente posible que haya sido en la última presentación o artículo que acaba de revisar.

Esto se explica porque el test de significancia estadística ha dominado la forma que los investigadores reportan los resultados de sus estudios y más importante aún, evalúan su relevancia.

La práctica del test de hipótesis ha sido históricamente utilizada para la interpretación de los resultados de estudios de distintas disciplinas, incluida la kinesiología. La determinación de la existencia o no de diferencias estadísticamente significativas es centrada en aceptar o rechazar una hipótesis “nula” o “alternativa”. La hipótesis nula (H_0) establece la no diferencia entre grupos (diferencia en las medias entre grupos = 0) o, que por ejemplo en estudios de intervención, la variable independiente no tiene efecto sobre la variable dependiente¹. Por otro lado, la hipótesis alternativa (H_1) indica la presencia de diferencias entre grupos o la existencia de un efecto².

El test de significancia estadística incluye el cálculo de un cierto nivel de probabilidad (valor p o α) que el investigador selecciona para asegurar que uno no rechaza (incorrectamente) la H_0 , debido al azar, cuando esta es verdadera (error tipo 1).

El nivel generalmente aceptado, aunque arbitrario, para el valor P es 0.05 (aunque también puede ser 0.01), lo que sugiere que existe un 95% de probabilidades que el investigador correctamente acepta la H_0 cuando no existen diferencias entre los grupos³. Así, cuando el valor p es menor al umbral seleccionado, es entonces cuando se indica que las diferencias son estadísticamente significativas, representado esto por $p < 0.05$. En este caso se rechaza la hipótesis nula aceptándose la hipótesis alternativa como probablemente cierta. En otras palabras, se podría sostener que es muy poco probable que el azar sea el responsable de las diferencias o asociaciones encontradas entre los grupos estudiados.

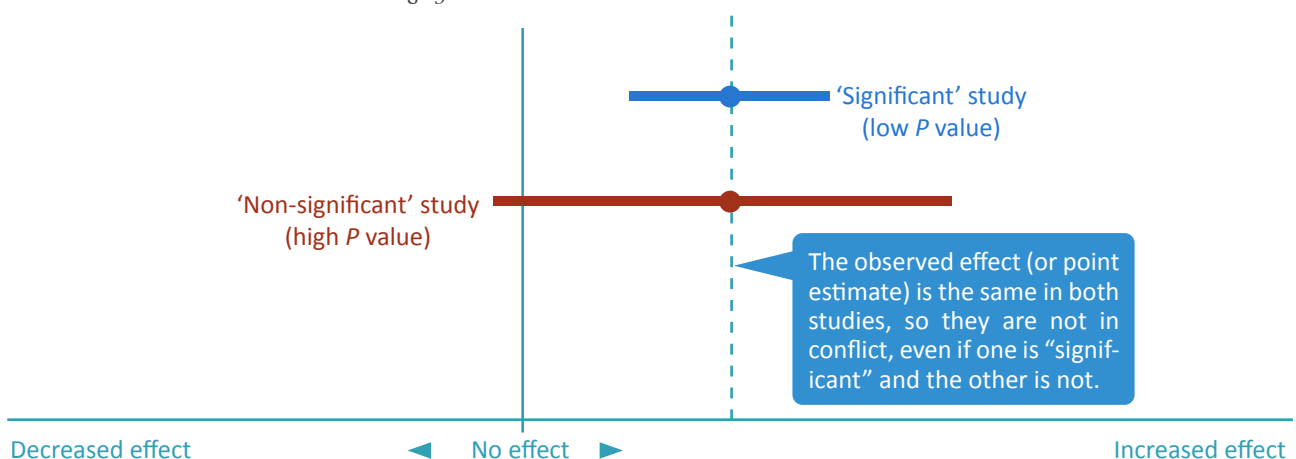
Sin embargo, el valor p falla en un elemento básico. Esto es, el tipo de información que entrega. El valor p ha sido utilizado para evaluar si los resultados son explicables por el azar o no. En base a esto, el valor p sólo provee a una respuesta dicotómica (significativo vs. no significativo)⁴, sin proporcionar información acerca de la magnitud o importancia de las observaciones medidas, ni tampoco su aplicabilidad clínica de los resultados

del estudio. Además, los valores p son frecuentemente mal interpretados o conllevan a interpretaciones engañosas. Por ejemplo, una diferencia estadística puede estar presente cuando la variabilidad inter-sujeto es baja junto con un tamaño de muestra elevada, aunque las diferencias entre los grupos sean pequeñas⁵. Por ejemplo, en el 2010, un estudio prospectivo (13 años de seguimiento) incluyendo más de 30.000 mujeres (edad promedio 54.2 años) concluyó que estas necesitan 60 minutos diarios de ejercicio moderado para prevenir el alza de peso⁶. Los autores compararon tres grupos de intervención: ejercicio con bajo nivel metabólico equivalente (< 7.5 MET horas/semana), medio (7.5 a 21 MET horas/semana) y alto (≥ 21 MET hora/semana). Los resultados indicaron que las mujeres en el grupo de ejercicio de alto nivel ganaron menos peso que las del grupo de ejercicio medio ($p=0.003$) y las del grupo de ejercicio bajo ($p= 0.002$). Algo no destacado en el artículo fue el hecho que el grupo de ejercicio alto ganó un promedio de sólo 0.12 kg menos que el grupo de ejercicio bajo y sólo 0.11 kg menos que el grupo de ejercicio medio. La explicación para que estas pobres diferencias entre grupos sean estadísticamente significativas descansa en el gran tamaño de la muestra. Estos modestos resultados cuestionan el real impacto o relevancia de la intervención.

Igualmente, se asume comúnmente que un valor que no alcanza significancia estadística significa ausencia de efecto (figura 1). Por ejemplo, un análisis de cientos de artículos en 5 revistas determinó que casi el 50% de los resultados que fueron estadísticamente no significativos fueron a su vez erróneamente etiquetados como resultados “sin diferencias” o “sin efecto.”⁷.

Es relevante entender que en ningún caso un cambio estadísticamente significativo no asegura que estos sean relevantes del punto de vista clínico. Para ilustrar este punto se presentarán algunos ejemplos más próximos a nuestra disciplina. Robertson y cols 2005⁸, condujeron un estudio de intervención cuyo propósito fue evaluar el efecto del calor terapéutico sobre la movilidad de tobillo en voluntarios sanos. Utilizando un diseño cruzado con medidas repetidas, esta investigación comparó la aplicación de calor profundo, calor superficial y ausencia de calor (control) en 22 sujetos, siendo el rango de movilidad la medida de resultado de interés. Los resultados indicaron que el aumento en dorsiflexión de tobillo en todos los sujetos (hombres y mujeres) recibiendo calor profundo fue de 1.8° , este aumento alcanzó 0.7° para el calor superficial, no generando efectos (-0.1°) para el grupo sin intervención. Estos cambios equivalen a un incremento del 5.2%, para el calor profundo, 2.0% para el calor superficial y -0.4% para el grupo control respectivamente. De manera interesante, el test de hipótesis demostró diferencias estadísticamente significativas entre el calor profundo y superficial ($p=0.015$) y entre el calor profundo y la ausencia de calor ($p=0.001$). Una conclusión simple sería que el calor profundo debiera ser la elección por sobre el calor superficial cuando se trata de mejorar la extensibilidad de la articulación de tobillo. Sin embargo, es clínicamente relevante un cambio de 1.8° ? el sentido común nos indica que no, es más, este cambio podría ser incluso un error de técnica. Un ejemplo similar corroborando el punto anterior proviene de uno de nuestros estudios que exploró las diferencias encontradas en la postura craneocervical entre sujetos sanos y con desórdenes temporomandibulares⁹. Los resultados de ese

Figura 1: Representación gráfica de potenciales errores de interpretación del nivel de significancia estadística y la consiguiente determinación de conclusiones erradas acerca de la presencia de efectos genuinos de una intervención. Fuente: Amrhein V, Greenland S, Mc Shane B. Retire statistical significance. Comment. Nature 2019, 567: 305-307.



estudio determinaron que las diferencias entre grupos alcanzaron significancia estadística ($p < 0.05$). En otras palabras, se entiende con un 95% o más de certeza que los grupos fueron distintos, y esta diferencia fue muy grande para ser explicada por el azar. Es importante consignar, sin embargo, que estas mentadas diferencias entre ambos grupos fueron de sólo 2.6° . Sería muy poco probable que un kinesiólogo, en una evaluación de rutina podría ser capaz de detectar tan modesto cambio en la postura en sus pacientes, incluso usando lo más sofisticados instrumentos de medición, como un inclinómetro digital. Es altamente cuestionable, además, que un cambio tan discreto en la postura de un sujeto, a recordar 2.6° , pueda verse reflejado en un impacto en la función y/o sintomatología de ese paciente.

Por otro lado, resultados que no son estadísticamente significativos pueden tener relevancia clínica. Por ejemplo, un estudio evaluó la actividad electromiográfica (EGM) en sujetos con trastornos temporomandibulares. Los autores concluyeron que estos pacientes no presentaron diferencias estadísticamente significativas en la actividad de los músculos superficiales (esternocleidomastoideo y escaleno anterior) al ser comparados con sujetos sanos ($p = 0.07$) aunque cambios reflejando una diferencia clínicamente importante fueron documentados¹⁰.

Otra combinación, no muy comúnmente encontrada es la presencia de diferencias estadísticamente significativas acompañadas de cambios clínicamente relevantes. Un ejemplo de esto se da en otro de nuestros estudios¹¹. En este estudio clínico aleatorizado, 119 participantes con dolor lumbar crónico fueron tratados con corrientes interferenciales con distintos niveles de alianza terapéutica. Los cambios en intensidad de dolor y umbral de dolor en los grupos de alianza terapéutica favorecida fueron estadísticamente distintos ($p < 0.01$) a los grupos que recibieron la corriente interferencial en un contexto de alianza terapéutica limitada. Más relevante, cambios clínicamente relevantes fueron alcanzados entre estos dos grupos con una magnitud de efecto moderada para la intensidad del dolor y la sensibilidad muscular.

Es importante comprender entonces que la significancia estadística no implica que los resultados sean importantes desde una perspectiva clínica. Por ejemplo, y como se ha demostrado en los casos previos, es posible que un estudio demuestre significancia estadística sin ser los cambios clínicamente relevantes, o presentar

importancia clínica sin significancia estadística. También puede existir el caso donde se demuestran ambas, significancia clínica y relevancia estadística o no presentar ninguno de estos atributos.

El método del test de significancia estadística ha estado últimamente cuestionado, particularmente en las investigaciones originadas para dar respuestas a problemas clínicos. Por ejemplo, en el año 2016, la asociación americana de estadísticas generalizó un anuncio acerca del mal uso de la significancia estadística y el valor p , y muy recientemente, autores incluso establecen que el concepto de significancia estadística debería abandonarse^{7,12}. Mucha de la tendencia actual sugiere detener el uso del valor p en la forma dicotómica convencional para decidir si un resultado apoya o refuta una hipótesis científica. En los casos de aplicación práctica de los resultados de un estudio, las decisiones deben considerar costos, beneficios y por supuesto, la magnitud de los efectos alcanzados¹³.

Una mejor alternativa sugerida, dentro del ámbito estadístico, y considerando la escasa información entregada por el valor p , es reportar además de éste último, los intervalos de confianza (IC). Los IC ofrecen una información más detallada que la contenida en los valores p . La información del IC no tiene naturaleza dicotómica. Los IC tampoco constituyen un estimado de una muestra (como lo hace el valor p), estos representan a una población mayor.

De manera simplificada, los IC describen la variabilidad entre la medida obtenida en un estudio y la medida real de la población (el valor real). Estos entregan, a través de un rango que considera un límite inferior y límite superior, el verdadero valor en la población de la variable de interés que es estimada del promedio o correlación de una muestra¹⁴. Normalmente, los CI se reportan en porcentajes, esto es 90%, 95% y 99%. Siendo el IC 95% el más común en ser presentado. Si en un estudio, usted aplica una intervención en 100 pacientes y obtiene que esta produce un aumento en el umbral de dolor promedio de 2.0 Kg (IC 95% 1.6 – 2.5), el IC 95% en palabras, estaría indicando que existe un 95% de probabilidades que, si se llevara a cabo este estudio múltiples veces en la población, el verdadero aumento del umbral de dolor estaría contenido en un IC que tendría un límite superior de 2.5 kg y un límite inferior de 1.6 kg. El promedio en umbral de dolor estimado de la muestra (2.0 kg) estaría contenido en la mitad de ese intervalo de confianza.

Desde una perspectiva práctica, los IC representan una forma de estimar la magnitud de cambio o diferencia antes-después en un grupo o entre los grupos. Así, el reporte de los IC permitiría al clínico determinar si el rango de una diferencia entre tratamientos es lo suficientemente grande como para justificar su uso. Así, los IC constituyen una ayuda en la valoración de la significancia práctica de los resultados¹⁴.

Cuando este es pequeño, frecuentemente en el caso de muestras grandes, representa un estimado más preciso con menos incertidumbre¹³. De igual forma, cuando el IC incluye al estimado nulo (valor 0 en una diferencia o 1 en una asociación) hay suficiente confianza para determinar que es estadísticamente distinto.

A pesar del aporte entregado por el reporte de los IC en la interpretación de los resultados, este es aún insuficiente para un adecuado cálculo de la magnitud de cambio en una variable y si este impacta clínicamente o hace una real diferencia en el paciente. Surge entonces la necesidad de explorar más en detalle el concepto de significancia clínica y sus implicancias para la práctica kinesiológica. La literatura presenta diversos modelos para evaluar la significancia clínica de los resultados de un estudio, métodos que serán descritos en la siguiente sección de este trabajo.

Migrando a la significancia clínica

La investigación clínica en kinesiología es central en la generación de nuevo conocimiento disciplinar y para la

validación de nuestras terapias. Sin embargo, la interpretación de los resultados de una investigación se hace muchas veces difícil. Esto impone un desafío, dado que la investigación clínica se hace útil, sólo si esta es adecuadamente interpretada.

Como se ha intencionado previamente, desde una perspectiva clínica, la presencia o no de diferencias estadísticas son de un valor limitado. El método del test de significancia estadística por sí solo no debiera ser la influencia central en la interpretación de los resultados de un estudio dirigido a evaluar el impacto de intervenciones terapéuticas³. Los clínicos debieran estar interesados en si el cambio obtenido en una variable, como resultado de una intervención, es capaz de producir una diferencia real la función de un paciente más que si las diferencias entre los grupos son o no debido al azar. Una forma inicial para mejorar la interpretación de los resultados, como se anunció previamente en este escrito, es el reporte y análisis de los intervalos de confianza, sin embargo, esta aproximación posee limitaciones al momento de ilustrar y dimensionar la significancia clínica de los resultados.

Concepto significancia clínica

En 1984 Jacobsen y cols¹⁵, fueron los primeros en proponer la significancia clínica, opuesto a la significancia estadística, como una manera de determinar el valor práctico de un tratamiento. La significancia o relevancia clínica debería reflejar la cuantía del cambio, cuanto tiempo este efecto es mantenido y si este cambio produce una verdadera diferencia en la vida del paciente¹⁶.

Tabla 1: Comparación de los tres grupos de estudio (calor profundo, calor superficial y ausencia de calor) en los cambios de dorsiflexión de tobillo. El análisis es presentado por género. Fuente: Robertson VJ, Alex R, Ward AR, Jung P. The Effect of Heat on Tissue Extensibility: A Comparison of Deep and Superficial Heating. Arch Phys Med Rehabil Vol 86, April 2005, 819-24.

Table 1: Ankle Dorsiflexion ROM in Men and Women.

Intervention (heating)	Preintervention Measure	P for 2 Preintervention Measure	Postintervention Measure	P for 2 Preintervention Measure	Difference Post to Pre
Deep					
Men	36.7 ± 5.8	0.66	38.5 ± 6.8	0.12	1.8 ± 2.2
Women	33.3 ± 4.2	0.44	35.1 ± 4.6	0.06	1.8 ± 1.7
Superficial					
Men	36.9 ± 5.8	0.67	37.5 ± 5.9	0.62	0.6 ± 1.4
Women	33.9 ± 4.4	0.84	34.7 ± 4.8	0.58	0.8 ± 1.5
No heating					
Men	38.0 ± 6.3	0.36	37.9 ± 6.3	0.07	-0.1 ± 0.7
Women	34.4 ± 3.3	0.52	34.2 ± 3.6	0.82	-0.2 ± 1.2

NOTE: Values are mean degrees ± SD. Figures are rounded to 1 or 2 decimal place(s).

Una intervención clínicamente relevante es aquella en que sus efectos son suficientemente amplios para que su aplicación supere los riesgos y costos asociados¹⁷.

Métodos de cálculo de la significancia clínica

En la actualidad existen diversos métodos para el cálculo y la determinación de la significancia clínica. Los más comunes incluyen los métodos “basados en la distribución” y los métodos “basados en el anclaje”. Entre los primeros se incluyen el cálculo de la magnitud o tamaño del efecto, la diferencia mínima clínicamente importante o cambio mínimo detectable (MCID/MDD), y el error estándar de medida (SEM). En los métodos basados en el anclaje se considera la perspectiva del paciente para la definición de la relevancia clínica. Este método compara el cambio en la variable de interés con otra medida de cambio (anclaje), usualmente la escala de evaluación cambio (GRS)¹⁸. A continuación, y en consideración al propósito eminentemente informativo de este escrito, se describirán aspectos generales de algunos de estos métodos.

Tamaño del efecto: el tamaño del efecto (en adelante ES), El ES ha sido definido por Cohen 1988¹⁹ como “grado en el cual el fenómeno está presente en la población”, por lo que mientras mayor es el ES mayor es el grado en el cual el fenómeno bajo estudio se manifiesta (Cohen, 1988 p 10)¹⁹. El ES se considera uno de los indicadores más importantes de significancia clínica. Este provee información acerca de la magnitud de una relación cuantitativa entre variables, así como también el tamaño de la diferencia entre grupos. De igual forma, la magnitud del ES ha sido interpretada como un índice de relevancia clínica^{20, 21}. A mayor ES, mayor es la diferencia entre grupos y también mayor la relevancia clínica de los resultados²¹.

Comúnmente se calcula dividiendo la diferencia entre los promedios de los grupos (inter grupos) o las diferencias antes y después (intra grupo) por la desviación estándar de la línea de base, la del grupo control o por la desviación estándar agrupada (figura 2)^{19, 22} estableció una clasificación que incluye los valores de ES < 0.2, ES= 0.5 y ES ≥ 0.8 para denominar una magnitud de efecto pequeña, moderada y grande respectivamente. Una característica importante del ES consiste en la flexibilidad de esta medida. El ES puede ser aplicada a cualquier tipo de diseño de estudio, así su utilización como indicador de significancia clínica puede ser aplicable no sólo en estudios de intervenciones, sino que

en otros tipos de investigación no-experimental (ej. correlacional, transversal)²³.

Figura 2: Fórmula para el cálculo de la magnitud de efecto (ES). Portney L, Watkins M. In: Mehalik C, editor. *Power and sample size. Foundations of Clinical Research, New Jersey: Prentice Hall Health; 2000. p. 705-30.*

$$\overline{ES} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{S_{pooled}}$$

Donde:

- ES = effect size
- \overline{X}_{G1} = mean group 1
- \overline{X}_{G2} = mean group 2
- S_{pooled} = pooled standard deviation (SD)

Para calcular la desviación estándar agrupada, se utiliza la siguiente fórmula:

$$S_{pooled} = \sqrt{\frac{(S_1)^2 (n_1 - 1) + (S_2)^2 (n_2 - 1)}{n_1 + n_2 - 2}}$$

Donde:

- S_1 = SD group 1
- S_2 = SD group 2
- n_1 = sample size for group 1
- n_2 = sample size for group 2

MCID: este término se refiere al cambio más pequeño en una variable de interés que es considerado importante por parte del paciente o tratante¹⁸. Cambios en las variables que excedan ese valor pueden ser considerados clínicamente relevantes. Un método para calcular esta medida es multiplicar la magnitud de efecto (ES) de la diferencia obtenida entre los grupos considerado como importante (0.2 ó 0.5 de acuerdo con Cohen) por la desviación estándar agrupada en la línea de base entre los dos grupos (figura 3)²⁴. En términos de interpretación, por ejemplo, una diferencia media obtenida entre grupos (2.5) que sea mayor que el valor de MCID entregado por la fórmula (2.0) es indicativo que el valor es considerado clínicamente relevante^{21, 24}.

Figura 3: Fórmula para el cálculo de la MCID. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau L, Goodwin P. *Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. Journal of Clinical Epidemiology 2007, 60 (5): 448-55.*

$$MID = 0.2 \times S_{pooled\ baseline}$$

or

$$MID = 0.5 \times S_{pooled\ baseline}$$

Conclusiones

La pregunta si un paciente mejoró en o no en una forma significativa en un dominio específico (dolor, función, velocidad de la marcha) es fundamental para mejorar el proceso de toma de decisión clínica respecto de una intervención determinada. En kinesiología, algunos intentos se han hecho para incorporar el concepto de relevancia o significancia clínica^{1, 9, 10, 11, 21, 25, 26}. En la actualidad este se ha constituido en un elemento cada vez más recurrido para facilitar la transferencia del conocimiento derivado de las investigaciones a la práctica clínica. Sin embargo, se requiere un esfuerzo mayor.

Los kinesiólogos deben abandonar la práctica de sólo considerar el análisis simplista de la interpretación de la significancia estadística (valor p). De igual manera, los investigadores conduciendo estudios clínicos en kinesiología tienen la obligación de determinar y reportar la significancia clínica para una mejor interpretación de los resultados. Es recomendable, en la medida de lo posible, el uso de distintos métodos que incluyan la perspectiva del clínico (ES, MCID) e idealmente el complemento desde la óptica del paciente (GRS). Esto es crucial para optimizar el proceso de toma de decisión clínica además de simplificar y facilitar la transferencia de conocimiento en la práctica kinésica.

Referencias

1. Armijo-Olivo S, Warren S, Fuentes J, Magee DJ. Clinical relevance vs. statistical significance: using neck outcomes in patients with temporomandibular disorders as an example. *Manual Therapy* 2011a, (16): 563-72.
2. Schober P, Bossers SM, Schwarte LA. Statistical significance versus clinical importance of observed effect sizes: What do P-values and confidence intervals really represent? *Anesthesia & Analgesia* 2018, 126(3): 1068-72.
3. Page Phil. Beyond statistical significance: clinical interpretation of rehabilitation research literature. *The International Journal of Sports Physical Therapy* 2014, 9 (5):726-736.
4. Sterne JAC, Smith GD. Sifting the evidence- what's wrong with significance tests? *British Medical Journal* 2001; 322 (7280): 226-31.
5. Millis SR. Emerging standards in statistical practice:

implications for clinical trials in rehabilitation medicine. *American Journal of Physical Medicine and Rehabilitation* 2003; 82 (suppl).

6. Lee MI, Djoussé L, Sesso HD, Wang JE. Physical activity and weight gain prevention. *JAMA* 2010; 303: 1173-79.
7. Amrhein V, Greenland S, Mc Shane B. Retire statistical significance. Comment. *Nature* 2019, 567: 305-307.
8. Robertson VJ, Ward AR, Jung P. The Effect of Heat on Tissue Extensibility: A Comparison of Deep and Superficial Heating. *Arch Phys Med Rehabil* 2005; 86: 819-25.
9. Armijo Olivo S, Rappoport K, Fuentes J, Gadotti IC, Major PW, Warren S, Thie NMR, Magee DJ. Head and cervical posture in patients with temporomandibular disorders (TMD). *Journal of Orofacial Pain* 25(3): 199-209, 2011b.
10. Armijo-Olivo S, Silvestre R, Fuentes J, da Costa B, Gadotti I, Warren S, et al. Electromyographic activity of the cervical flexor muscles in patients with temporomandibular disorders while executing the craniocervical flexion test (CCFT); a cross-sectional study. *Phys Ther.* 2011c Aug;91(8):1184-97.
11. Fuentes J, Armijo Olivo S, Funabashi M, Miciak M, Dick B, Warren S, Rashedi S, Magee DJ, Gross D. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: a randomized controlled study. *Physical Therapy Journal* 94 (4) 477-89, 2014.
12. Wasserstein, R. L., Schirm A, Lazar NA. Am Stat. <https://doi.org/10.1080/00031305.2019.1583913> (2019).
13. Martínez-Esquerro JD, Riojas-Garza A, Rendón-Macías ME. Significancia clínica sobre significancia estadística. Cómo interpretar los intervalos de confianza a 95%. *Rev Aler Mex* 2017; 64 (4): 477-86.
14. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Australian Critical Care* 2010, 23: 93-97.
15. Jacobsen NS, Follete WC, Revenstorf D. Physiotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behavior*

Therapy 1984; 15(4): 336-52.

16. LeFort S. The statistical versus clinical significance debate. *Journal of Nursing Scholarship* 1993, 25 (1):58

17. Ferreira M, Herbert R. What does clinically important really mean? *Australian Journal of Physiotherapy* 2008; 54 (4): 229-230.

18. Jaeschke R, Singer j, Guyatt G. Measurement of health status ascertaining the minimal clinically important difference. *Control Clin Trials* 1989, 10(4): 407-415.

19. Cohen J. Statistical power analysis for the behavioral sciences. *Hillsdale, New Jersey: Academic Press, Inc;* 1988. p. 1-17 [chapter 1].

20. Kirk RE. Practical significance: a concept whose time has come. *Educational and Psychological Measurement* 1996; 56(5):746-59.

21. Musselman KE. Clinical significance testing in rehabilitation research: what, why, and how? *Physical Therapy Reviews* 2007; 12 (4):287-96.

22. Portney L, Watkins M. In: Mehalik C, editor. Power and sample size. *Foundations of Clinical Research, New Jersey: Prentice Hall Health;* 2000. p. 705-30.

23. Callahan J, Reio Jr Tg. Making subjective judgments in quantitative studies: the importance of using effect sizes and confidence intervals. *Human Resource Development Quarterly* 2006, 17 (2): 159-73.

24. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau L, Goodwin P. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *Journal of Clinical Epidemiology* 2007, 60 (5): 448-55.

25. Fuentes J, Armijo Olivo S, Magee D, Gross DP. Does Amplitude Modulated Frequency (AMF) have a Role in the Hypoalgesic Response of Interferential Current on Pressure Pain Sensitivity in Healthy Subjects? A Randomized Cross-over Study. *Physiotherapy* 96: 22-29, 2010.

26. Fuentes J, Armijo Olivo S, Gross D, Magee D. A Preliminary Investigation into the Effects of Placebo and Active Interferential Current Therapy on Pressure Pain Sensitivity: A Random Crossover Placebo Contro-

lled Study. *Physiotherapy* 97; 291-301, 2011.

Correspondencia

jorgef@ualberta.ca