

¿CÓMO SABEMOS SI NUESTROS PACIENTES MEJORAN? CRITERIOS PARA LA SIGNIFICANCIA CLÍNICA EN PSICOTERAPIA: UN DEBATE QUE SE RENUEVA

Juan Martín Gómez Penedo*
y Andrés Roussos**

Resumen

El presente artículo sintetiza un debate surgido hace más de veinticinco años, que en el último tiempo ha renovado su intensidad y relevancia: cuál es el criterio más fidedigno para establecer el cambio clínico en los estudios de eficacia y efectividad en psicoterapia. Los dos criterios tradicionales usados en los estudios de resultados – significancia estadística y significancia práctica – son aquí presentados. La significancia clínica, el más joven de los criterios, es descrito y discutido con más detalle. A lo largo del artículo se explicitan las distintas definiciones propuestas para su estudio, junto con las metodologías utilizadas para su cálculo. La aplicación de este último criterio es ilustrada a partir de un caso ficticio. El trabajo realizado muestra a la significancia clínica como un aporte fundamental tanto para la investigación en psicoterapia como para la clínica, siendo, a su vez, un recurso importante para facilitar espacios de interacción entre estas áreas.

Palabras clave: significancia clínica, cambio en psicoterapia, resultados, eficacia

Key words: clinical significance, psychotherapy change, outcome, efficacy

Introducción

El objetivo del presente estudio es describir e ilustrar los debates históricos y actuales acerca del uso de la significancia clínica como criterio para las investigaciones de eficacia y eficiencia en psicoterapia. Si bien existen otros criterios que han tenido mayor popularidad dentro de los estudios de resultados, como la significancia estadística y la significancia práctica, en las últimas décadas se observa un interés creciente de los investigadores por evaluar el impacto y valor de los dispositivos psicoterapéuticos en la vida cotidiana de los pacientes. A diferencia de la significancia estadística que estudia si existen diferencias entre las personas que llevan a cabo una psicoterapia y aquellas que no, y la significancia práctica que tiene como fin medir la magnitud de la diferencia entre estos dos grupos, la significancia clínica sugiere estudiar la eficacia de un

tratamiento indagando hasta qué punto éste permite alcanzar cambios importantes y valiosos para el paciente y su comunidad. Autores como Goldfried, Greenberg y Marmar (1990) y Kazdin (1999) consideran a la significancia clínica como uno de los avances en investigación en psicoterapia más importantes de los últimos años. Este criterio no sólo permite la evaluación de la eficacia o eficiencia de un tratamiento en investigaciones de resultados. La significancia clínica brinda la posibilidad de realizar un análisis individual de cada paciente, para observar su variabilidad intra-sujeto (Jacobson y Truax, 1991; Peterson, 2008), a diferencia de los otros criterios recién mencionados – significancia estadística y significancia práctica – que sólo facilitan un análisis de diferencias grupales. Por ello, al promover el retorno a criterios de evaluación individual (Beckstead y cols., 2003), permite a los terapeutas comprender el verdadero impacto de sus intervenciones en la práctica clínica (Ogles y Lunnen, 1996). De esta manera, la metodología de la significancia clínica puede ser aplicada por clínicos en la evaluación privada de sus pacientes, convirtiéndose en un instrumento cuyo interés excede a los investigadores. Esta cualidad del criterio de significancia clínica nos lleva a

* Juan Martín Gómez Penedo, Universidad de Belgrano

** Andrés Roussos, CONICET, Universidad de Buenos Aires, Universidad de Belgrano

E-Mail: andres.roussos@comunidad.ub.edu.ar

REVISTA ARGENTINA DE CLÍNICA PSICOLÓGICA XXI p.p. 173-190

© 2012 Fundación AIGLÉ.

considerarlo una contribución valiosa para el establecimiento de vínculos entre las áreas usualmente escindidas de la clínica y la investigación.

Contrario a lo que intuitivamente se suele suponer, este movimiento no implica un alejamiento de la estadística o de los modelos matemáticos, sino que busca complementar estos esfuerzos a la hora de establecer estimadores de las diferencias entre los valores que representan a la salud y a la enfermedad. Al mismo tiempo, la significancia clínica se sustenta en recursos matemáticos fundamentales para la medición de sus criterios. Es decir, utiliza parámetros cuantitativos para dar cuenta de los criterios cualitativos clínicos que emplea. A su vez, tampoco supone una ruptura con el concepto de significancia estadística, ya que incorpora algunos de los conceptos básicos de la misma, como es el valor p de la probabilidad. Lo que claramente diferencia a la noción de significancia clínica es la inclusión de conceptos clínicos y de nuevos parámetros estadísticos para la determinación de la eficacia o efectividad de los tratamientos, entre otras evaluaciones propias de la psicología clínica.

En el presente trabajo se describen los distintos criterios ya mencionados para la evaluación de eficacia en psicoterapia, profundizando particularmente en la significancia clínica. Se desarrollan sus características, los diferentes parámetros que se utilizan para su valoración, así como los métodos estadísticos más usados para su cálculo. Se presentan sus alcances y limitaciones. La metodología para su cálculo es ilustrada a partir de un caso ficticio (Ver Apéndice).

El camino hacia criterios de significancia clínica

La significancia estadística como punto de partida

Históricamente, tanto la eficacia como la efectividad de los tratamientos en psicoterapia se ha abordado de manera casi exclusiva mediante el uso del criterio de significancia estadística (Jacobson y Truax, 1991; Jacobson, Roberts, Berns y McGlinchey, 1999). Este parámetro busca establecer si las diferencias observadas entre dos grupos pueden ser consideradas reales o podrían, en cambio, ser atribuidas al azar (Kraemer y cols., 2003). Para ello, se estudia la probabilidad de que los datos hayan surgido, sin que existan diferencias verdaderas entre los grupos. Al leer investigaciones de eficacia es muy común encontrarse con el siguiente símbolo: $p < 0,05$. ¿Pero cuál es su significado? Implica que la probabilidad de que hayan surgido estos datos sin

que existan diferencias entre los grupos es inferior al 5%. Este porcentaje es interpretado como una probabilidad suficientemente baja como para considerar que las diferencias son reales o, en otras palabras, estadísticamente significativas (Cohen, 1994; Thompson, 2002).

Algunos autores como Durand y Barlow (2007), en una postura radical, afirman que la utilización de la estadística ha tenido un papel fundamental en la historia de la psicología, siendo uno de los elementos que permitieron el abandono de su condición de disciplina precientífica, facilitando su consolidación como ciencia. Si bien no lo explicitan, los autores aluden en esta referencia al uso de criterios de significancia estadística, entre otros, para la validación de constructos psicológicos. Aunque coincidimos con la idea de valorizar parámetros de orden cuantitativo para el establecimiento de criterios de diferenciación que usualmente se comprenden desde lo cualitativo, como en el caso de los fenómenos de orden clínico, encontramos en la visión de Durand y Barlow (2007) cierto sesgo que podría dar lugar a una mala interpretación del rol de la estadística en el ámbito de la investigación en psicología. En las últimas décadas, diversos autores identificaron algunas limitaciones de los métodos estadísticos tradicionales para la evaluación de la eficacia en psicoterapia. En contraposición con los dichos de Durand y Barlow, Cohen (1994, p. 997) afirma que la significancia estadística, “no solo falló en brindarle un soporte a los avances de la psicología como ciencia, sino que incluso impidió dicho avance”. Una de las críticas dirigidas a este criterio se centra en el hecho de que no toma en cuenta las variaciones intra-sujeto dentro de las muestras, limitándose al estudio de las diferencias grupales (Jacobson y Truax, 1991). Por otro lado, diversos autores (Jacobson y Truax, 1991; Thompson, 2002; Kraemer y cols., 2003; Peterson, 2008) afirman que uno de los déficits más destacados de la significancia estadística es que no aporta evidencia acerca de la importancia de los resultados obtenidos ni para los investigadores del área de estudio ni para la sociedad en general.

La significancia estadística permite determinar si existe una diferencia significativa entre dos grupos, pero no da cuenta de la magnitud de dicha diferencia ni del valor de los resultados obtenidos para pacientes y clínicos (Kraemer y cols., 2003). Los resultados obtenidos en los estudios que se basan en la significancia estadística, permiten sacar solo conclusiones parciales desde el punto de vista clínico; la presencia de un efecto estadísticamente significativo no garantiza el haber alcanzado efectos valiosos y significativos desde la perspectiva clínica (Jacobson y Truax, 1991). Como afirman Jacobson y colaboradores (1999, p. 306), “la mayoría de los

pacientes que consultan no buscan alcanzar simplemente un cambio estadísticamente significativo sino lograr una merma en su sufrimiento”. Esto hace que algunos autores no consideren a la significancia estadística un criterio lo suficientemente útil como para ser el único a tener en cuenta al estudiar la eficacia de los tratamientos en psicoterapia (Thompson, 2002), afirmando la necesidad de que este tipo de estudios indaguen cuáles son los beneficios de los cambios producidos por la psicoterapia, junto con su magnitud e impacto en la vida cotidiana de los pacientes (Jacobson y Truax, 1991).

Por lo cual, repensando la frase de Durand y Barlow (2007), podríamos sugerir que sólo a través de la interacción entre los criterios cualitativos clínicos y los modelos estadístico-matemático es posible generar un modelo genuinamente científico para los estudios en psicología clínica, sin que la inclusión de un único factor sea el responsable de delimitar la diferencia entre lo científico y lo no científico.

El desarrollo de nuevos criterios para los estudios de eficacia

Significancia práctica

Un intento de resolución del debate recién planteado han sido los estudios estadísticos basados en el Tamaño del Efecto, ya que los mismos no se limitan a establecer la presencia o ausencia de un efecto, sino que a su vez buscan explicitar su magnitud (Jacobson y Truax, 1991; Kraemer y cols., 2003). El Tamaño del Efecto es el nombre que se le asigna a un conjunto de índices que tienen por cualidad medir la magnitud de las diferencias entre dos grupos, respecto a una variable dada. Estos índices representan la base de los estudios meta-analíticos (Roussos, 2007).

Comencemos a jugar con algunas fórmulas matemáticas para comprender lo que se está descri-

biendo. A pesar de que existen diversas formas de estudiar el tamaño del efecto, la estimación d , desarrollada por Cohen (1988), es la más utilizada en los meta-análisis (Lane, 2007). Su fórmula básica es:

$$d = \frac{(M_1 - M_2)}{D}$$

Donde M_1 es la media del grupo 1, M_2 la media del grupo 2 y D el desvío estándar de ambos grupos. Para una ejemplificación de la forma de calcular el tamaño del efecto ver el Apéndice.

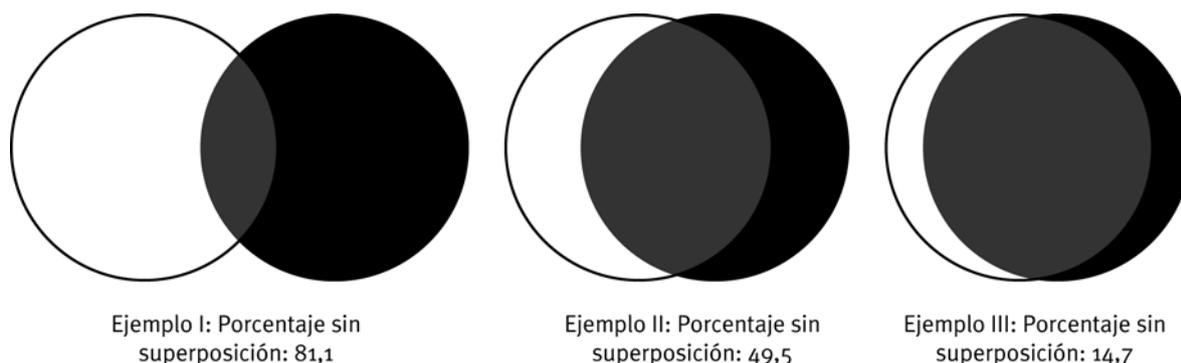
Para la interpretación de los resultados de esta fórmula, Cohen (1988) sostiene que un d igual o menor a 0,2 es considerado “pequeño”, un d igual a 0,5 es conceptualizado como “moderado”, mientras que un d igual a 0,8 implica un tamaño del efecto “grande”.

Los desarrollos en relación al tamaño del efecto posibilitan el pasaje hacia criterios de significancia práctica, mostrando la importancia de los resultados encontrados a partir de cuantificaciones que exceden la mera presencia o ausencia de una diferencia estadísticamente significativa (Peterson, 2008). Por lo tanto, en vez de limitarse a establecer si dos grupos son diferentes, buscan determinar el grado de aquella diferencia, observando los niveles de superposición de ambos grupos (Kraemer y cols., 2003) (Ver Figura 1).

La importancia práctica del tamaño del efecto ha llevado a Wilkinson y a la American Psychological Association Task Force on Statistical Inference (1999) a prescribir su presencia como parte de los resultados primarios de toda investigación en psicología en la que se comparen dos o más grupos. Es decir, en nuestro artículo, los investigadores en psicología deberíamos, según esta sugerencia, informar el tamaño del efecto.

Sin embargo, pese a que los trabajos sobre tamaño del efecto han sido un importante paso en

Figura 1. Tres ejemplos de distintos grados de superposición entre dos grupos diferentes (Roussos, 2007, p. 164)



esta área, no se puede establecer una relación directa e inmediata entre el tamaño del efecto y el valor clínico de los resultados (Jacobson y Truax, 1991). Aunque, frente a idénticos dispositivos, las investigaciones con un mayor tamaño de efecto tienen más probabilidad de alcanzar resultados clínicamente de mayor relevancia que aquellas con un menor tamaño de efecto, investigaciones en las que se manifiesta un tamaño de efecto importante pueden contar con resultados no significativos desde una valoración clínica (Jacobson y cols., 1999). En algunos casos, estudios con muy poco nivel de superposición entre los dos grupos comparados pueden dar por resultado diferencias destacadas desde el punto de vista cuantitativo pero con escaso valor para los sujetos que consultan.

Significancia clínica

En este marco, desde el área de la psicoterapia, extendiéndose posteriormente a la medicina y a las ciencias de la educación, surge un tercer criterio para evaluar la eficacia de tratamientos, denominado significancia clínica (Jacobson, Follette y Revenstorf, 1984; Jacobson y Truax, 1991), cuyo objetivo fundamental es determinar en qué medida los procesos terapéuticos permiten alcanzar resultados valiosos y significativos para los pacientes y terapeutas. Por lo tanto, la significación clínica podría definirse como la capacidad de un tratamiento de alcanzar los parámetros de eficacia establecidos por los pacientes y los psicoterapeutas (Jacobson y Truax, 1991; Kraemer y cols., 2003). Es una medida que para establecer la calidad de un abordaje observa la capacidad de este para producir modificaciones reales y tangibles en la vida de los pacientes, logrando los resultados que esperan al llevar a cabo una consulta psicológica (Kazdin, 1999; Durand y Barlow, 2007).

El estudio de la significación clínica es considerado uno de los avances metodológicos más importantes de la investigación en psicoterapia en las últimas décadas (Goldfried, Greenberg y Marmar, 1990; Kazdin, 1999). Como ya se adelantó en la introducción, este criterio permite realizar un análisis individual de la evolución de los paciente (Jacobson y Truax, 1991; Peterson, 2008), a diferencia de los dos criterios descriptos previamente. Esto lo torna un importante recurso para que los clínicos puedan evaluar el impacto de sus intervenciones y el devenir de sus pacientes a largo del tratamiento. En otras palabras, la significancia clínica le permite saber a los terapeutas si el tratamiento realizado produjo en el paciente cambios valiosos, utilizando un método más confiable que la mera valoración cualitativa del clínico.

Vale destacar que a pesar de que los propulsores de la significancia clínica mantienen una postura crítica respecto de los alcances de los anteriores criterios mencionados, los resultados de los estudios de eficacia y eficiencia que utilizan la significancia clínica no se contradicen con los resultados que se desprenden de los otros criterios, sino que tienden a agregar mayor información a estos (Jacobson y Truax, 1991). La significancia clínica no tiene como finalidad erradicar la significancia estadística ni el tamaño del efecto. Además de buscar convivir con ellos, se apoya en sus metodologías: utiliza los modelos matemáticos propios de los criterios tradicionales y los conjuga con valoraciones cualitativas de los resultados clínicos, intentando aportar nuevos datos para una comprensión de mayor complejidad del fenómeno de la eficacia en psicoterapia.

Pese a que la significancia clínica tuvo una importante proliferación desde su surgimiento, algunos autores consideran que han sido escasos los intentos de validar empíricamente este constructo (Atkins, Bedics, McGlinchey y Beauchaine, 2005; Bauer, Lambert y Nielsen, 2004; Beckstead y cols., 2003; Lambert y Ogles, 2009). Sin embargo, presentaremos en este trabajo, algunos estudios en esta dirección. Tal es el caso de Ankuta y Abeles (1993) quienes encuentran una estrecha correlación entre los cambios clínicamente significativos y el nivel de satisfacción que los pacientes tienen acerca de la psicoterapia al finalizar el proceso. En dicho estudio, aquellos que a partir de la terapia lograron alcanzar cambios clínicamente significativos, mostraron mayores niveles de satisfacción y beneficios percibidos de la psicoterapia, que aquellos que lograron cambios moderados o nulos. Por su parte, Lunnen y Ogles (1998) evaluaron la validez de la significancia clínica tomando en cuenta distintas medidas de resultados (percepción de cambio, alianza terapéutica, satisfacción con el tratamiento) desde múltiples perspectivas (visión del paciente, del terapeuta y de otros significativos). Los investigadores observaron que aquellos pacientes con cambios clínicamente significativos tenían diferencias marcadas en comparación con aquellos pacientes que no cambiaron o que empeoraron, tanto en los niveles de percepción de cambio como en la alianza, desde la perspectiva del paciente y del terapeuta, lo que representa una evidencia en favor de validez de este constructo. Sin embargo, entre los pacientes con cambios clínicamente significativos y aquellos que no cumplieron este criterio, no se observaron diferencias en cuanto a la satisfacción del tratamiento, resultado también observado por Beckstead y colaboradores (2003), ni a la percepción de cambio desde la perspectiva de otros significativos para el paciente (Lunnen y Ogles, 1998). En consonancia con este trabajo, Be-

retta y colaboradores (2005) encuentran una fuerte relación entre cambio clínicamente significativo, nivel de cambio percibido y satisfacción por el tratamiento. Otros autores, a su vez, han encontrado relaciones significativas entre significancia clínica y modificaciones en el funcionamiento global (Karpenko, Owens, Evangelista y Dodds, 2009).

Definiciones operacionales de la significancia clínica

Para poder evaluar la significancia clínica de un cambio en psicoterapia, en primera instancia hay que definir qué debería ser considerado un cambio significativo desde el punto de vista clínico. En otras palabras, qué cambios consideran importante conseguir a partir del proceso los clínicos y los pacientes. Desde el primer artículo de Jacobson, Follette y Revenstorf (1984), en el que se introdujo este concepto, hasta la actualidad, se han enumerado diferentes definiciones posibles.

Jacobson y Truax (1991) asumen como parámetro de eficacia el cambio del paciente hacia un funcionamiento normal. A partir de esta teorización se piensa a los pacientes que demandan terapia como parte de una población disfuncional en contraposición a la población funcional (aquellos que carecen del cuadro que padece el paciente). Un tratamiento eficaz o eficiente sería aquel que permite modificar el funcionamiento de un paciente dado, haciendo que éste pase a estar más cerca de la media de desempeño de la población funcional que de la media de la población patológica (para un ejemplo ver la Figura 2 en el Apéndice). Algunos años más tarde los mismos autores que crearon esta definición (Follette y Callaghan, 1996; Jacobson y cols., 1999) afirmaron que dicho parámetro, para algunos trastornos en los que el objetivo de alcanzar un funcionamiento normal es inviable, puede ser un poco restrictivo y excesivamente conservador, debiendo ser modificado y adecuado a la situación particular estudiada. A su vez, otros autores (Wampold y Jenson, 1986; Tingey y cols., 1996; Martinovich, Saunders y Howard, 1996; Hansen y Lambert, 1996; Kazdin, 1999) critican esta manera bimodal de pensar a la población (separando de manera tajante a una población funcional de una población disfuncional), sosteniendo que la distribución de la población se caracteriza por formar parte de un continuo que va desde el funcionamiento patológico hasta el no patológico. Es decir, consideran que no existe un punto de demarcación categórico que divida las aguas entre la salud y la patología, concibiendo estos términos como dos polos entre los cuales existen diferentes graduaciones. Para estos autores la diferencia entre la salud y la enfermedad es de naturaleza cuantitativa y no cua-

litativa. Por lo tanto, los individuos, según este paradigma, no deben ser catalogados como personas con o sin patologías, sino que deben ubicarse en este continuo según su grado de patología o salud. En conformidad con este punto, Tingey y colaboradores (1996) sostienen que se pueda encontrar una población integrada por diversos grupos claramente identificables con distintos grados de severidad sintomatológica e impacto social de sus síntomas.

A pesar de estas críticas a la definición propuesta por Jacobson y Truax (1991), la misma ha sido la más utilizada históricamente para el estudio de la significancia clínica (Sandell y cols., 1994; Kazdin, 1999; Beckstead y cols., 2003; Lambert y Ogles, 2009; Bauer, Lambert y Nielsen, 2004; Atkins y cols., 2005).

Sin embargo, debido a que esta definición, como se acaba de mostrar, no conformó a muchos investigadores del área (Kraemer y cols., 2003), se han propuesto algunas alternativas, a saber:

Cambios que reduzcan el riesgo de problemas de salud (Jacobson y Truax, 1991; Kraemer y cols., 2003).

Un nivel de cambio reconocido por personas significativas (Wolf, 1978).

Eliminación del problema que lleva a consultar (Kraemer y cols., 2003).

Mejorías en la calidad de vida (Kazdin, 1999).

Modificaciones sustanciales en las relaciones con los otros (Kazdin, 1999).

Modificaciones en los otros significativos del paciente a partir de la terapia (Kazdin, 1999).

Reducción de discapacidades funcionales (Kazdin, 1999; Karpenko, Owens, Evangelista y Dodds, 2009)

Kazdin (1999), por su lado, sostiene que incluso hasta un tratamiento cuyo efecto es la ausencia de cambios, puede generar resultados interpretables como clínicamente significativos. En trastornos caracterizados por un deterioro progresivo esperable, como pueden ser las demencias, los objetivos de la terapia no apuntan a reducir síntomas, sino a mantener o retrasar la pérdida de las distintas funciones afectadas. En estos casos, la ausencia de cambios sería el parámetro adecuado para evaluar la significancia clínica del proceso (Kazdin, 1999).

Tingey y colaboradores (1996), frente a sus críticas ya mencionadas a la definición operacional de Jacobson y Truax (1991), desarrollan una propia. En vez de dividir a la población en dos estratos, una población funcional y otra disfuncional, dividen a la misma en varios grupos en función del grado de impacto social de los síntomas y su posible intensidad. En el ejemplo que presentan, identifican cuatro poblaciones con distinto grado de impacto social de los

síntomas: 1. Un grupo caracterizado por personas que no contaban con sintomatología en base a una serie de estudios previos; 2. Un grupo de adultos con un nivel sintomático medio; 3. Paciente externados con tratamiento psicoterapéutico; 4. Pacientes internados con tratamiento psicoterapéutico. Una vez identificados los grupos, no es requisito que el paciente alcance el puntaje de la población "más funcional" para que el cambio sea considerado clínicamente significativo. Al pasar al funcionamiento de un grupo de menor severidad de síntomas e impacto social de los mismos, en relación a la población en la que estaba ubicado el paciente previo al tratamiento, se considera que el paciente alcanzó un cambio clínicamente significativo.

Para utilizar cada una de estas definiciones alternativas en el estudio de la significancia clínica, son necesarias algunas adecuaciones metodológicas en consonancia con las características específicas de cada definición. Este esfuerzo extra que demandan, no le quita valor a su potencialidad para enriquecer y refinar el constructo significancia clínica.

Métodos para el cálculo de la significancia clínica

En el siguiente apartado se presenta el método original de Jacobson y Truax (1991) para el cálculo de la significancia clínica y algunas alternativas a éste surgidas con posterioridad. La acentuación en aquel método se debe a que ha sido el más usado al operacionalizar la significancia clínica desde su surgimiento (Beckstead y cols., 2003; Bauer, Lambert y Nielsen, 2004; Atkins y cols., 2005). Según Ogles, Lunnen y Bonesteel (2001), el 53% de las investigaciones publicadas en el *Journal of Consulting and Clinical Psychology* que trabajaron con el concepto de significancia clínica, utilizaron el método de Jacobson y Truax (1991).

Todos los maneras aquí descritas de calcular la significancia clínica utilizan como definición operacional la establecida por Jacobson y Truax, con excepción del método desarrollado por Tingey y colaboradores (1996), cuya definición ya fue expuesta.

El Método tradicional de Jacobson y Truax [Método JT]

El método JT (Jacobson y Truax, 1991), se sostiene en dos constructos básicos:

1. El establecimiento de un Punto de Corte (PC), que permite deducir si el funcionamiento del pa-

ciente luego de la terapia pudo ubicarse más cerca de la media del grupo funcional que de la media del grupo disfuncional. El método indica que superado ese punto de corte el funcionamiento del paciente se acerca más al de la población funcional que al de la disfuncional.

2. El cálculo de un Índice de Confiabilidad de Cambio (ICC*) para determinar si el cambio producido tiene la suficiente fuerza estadística como para ser considerado real ($p < 0,05$).

Por lo tanto, según el método JT (Jacobson y Truax, 1991; Jacobson y cols., 1999) el paciente que logra cambios clínicamente significativos es aquel que supera el Punto de Corte calculado y que alcanza un Índice de Confiabilidad de Cambio suficiente para que dichas modificaciones sean consideradas reales.

Punto de corte. Para calcular este puntaje bisagra, que permite diferenciar a aquellos pacientes que alcanzaron un cambio clínicamente significativo de aquellos que no, Jacobson y Truax (1991) sugieren tres estrategias diferentes (denominadas estrategias a, b y c) en función de los datos con los que se cuenta y de las características diferenciales de las distribuciones de las poblaciones.

a) Esta estrategia se utiliza cuando solo se tienen datos de la población disfuncional y no se conoce la información necesaria de la población funcional o también en el caso de que existe un alto grado de superposición entre la distribución normal de ambas poblaciones. Esta estrategia da como resultado el punto de corte a partir de la sumatoria de la media de la población disfuncional (M_d) y de dos veces el desvío estándar de dicha población disfuncional (D_d). Siendo la fórmula: $a = M_d + 2D_d$.

Si a partir del instrumento utilizado las mejorías se ven representadas por reducciones de los puntajes y no por incrementos (Por ejemplo una escala sintomática donde una disminución en los síntomas es interpretada como una mejoría), se debe cambiar la valencia del desvío estándar pasando a ser un valor negativo: $a = M_d + (-2D_d)$.

b) Esta estrategia, en cambio, se calcula por medio de los datos de la población funcional, restando a la media de dicha población (M_f) dos veces el desvío estándar de esta (D_f). La presente estrategia se aplica en casos en los que la distribución normal de ambas poblaciones tiene escasa o nula superposición. Su fórmula es: $b = M_f - 2D_f$.

Al igual que en la estrategia (a), frente a una escala donde las mejorías están dadas por disminuciones en el puntaje, la valencia del desvío están-

* Cada vez que se utilicen las siglas ICC en este artículo será en referencia al Índice de Confiabilidad de Cambio. No confundir con el método IntraClass Correlation conocido usualmente a partir de la misma denominación.

dar debe ser modificada por una valencia negativa. Siendo la fórmula: $b = M_f - (-2D_f)$.

c) Según los autores, es la estrategia menos arbitraria de las tres, ya que surge mediante datos de ambas poblaciones, y no se deriva de las cualidades de una sola como pasa en el caso de las estrategias (a) y (b). Cuando las varianzas de los dos grupos son iguales, la estrategia (c) se calcula sumando la media de la población funcional (M_f) a la media de la población disfuncional (M_d) y dividiendo su resultado por dos. Por ende su fórmula es: $c^i = (M_f + M_d) / 2$.

En el caso de que las varianzas de ambos grupos sean diferentes la estrategia (c) se calcula de la siguiente forma:

$$c^{ii} = \frac{(D_{-d} M_f + D_f M_d)}{(D_{-d} + D_f)}$$

Siendo D_d el desvío estándar de la población disfuncional y D_f el desvío estándar de la población funcional.

La estrategia (c) se utiliza cuando existen datos de ambas poblaciones, habiendo cierta superposición (no excesiva) en la distribución de las mismas. Recapitulando lo dicho en los párrafos anteriores, frente a escasa o nula superposición se recomienda la estrategia (b) y ante una excesiva superposición o ante el desconocimiento de los datos de la población normal la estrategia (a) es la indicada.

Índice de Confiabilidad de Cambio. Cuando existe superposición entre la distribución de ambas poblaciones, algunos casos pueden superar el punto de corte establecido sin alcanzar indicadores de un cambio confiable, desde el punto de vista estadístico. Para superar este inconveniente Jacobson, Follette y Revenstorf (1984) desarrollan un Índice de Confiabilidad de Cambio, para evaluar el nivel de cambio entre el pre-test y el post-test y su grado de confiabilidad.

A continuación se presenta la última versión de la fórmula, modificada por Christensen y Mendoza (1986) y avalada por Jacobson y Truax (1991):

$$\frac{X_{post} - X_{pre}}{S_{diff}}$$

Siendo X_{pre} el puntaje del paciente previo al tratamiento, X_{post} el puntaje del paciente posterior al tratamiento y S_{diff} el error estándar de la diferencia entre X_{post} y X_{pre} . Para realizar el cálculo de S_{diff} es necesario calcular primero el error estándar del instrumento utilizado (S_E).

La fórmula para el S_{diff} es la siguiente:

$$S_{diff} = \sqrt{2 S_E^2}$$

$$S_E = D_{pre} \sqrt{1 - r_{xx}}$$

Siendo D_{pre} el desvío estándar de la muestra previo al tratamiento y r_{xx} la confiabilidad test-retest del instrumento de medición utilizado.

Por lo tanto:

$$S_{diff} = \sqrt{2 [D_{pre} \sqrt{1 - r_{xx}}]^2}$$

Un ICC superior o igual a 1,96 o inferior o igual a -1,96, es suficiente para considerar que el cambio es real, siendo en extremo improbable que se deba al azar ($p < 0,05$). En caso de que el instrumento de medición usado sea una escala donde las mejorías se resumen en una reducción en el puntaje, las mejorías confiables serían aquellas con un ICC inferior o igual a -1,96 y los deterioros confiables los que tengan un ICC igual o superior a 1,96. Por el contrario, aquellos instrumentos en los que las mejorías están representadas por incrementos en los puntajes, los cambios positivos confiables serían los que cuenten con un ICC superior o igual a 1,96, mientras que los deterioros confiables aquellos inferiores o iguales a -1,96.

Como ya se anticipó, desde este método se sostiene que un tratamiento eficaz o eficiente (que deriva en cambios clínicamente significativos) es aquel que permite al paciente superar el PC establecido, alcanzando, a su vez, un ICC igual o superior a 1,96 o igual o inferior a -1,96, según corresponda.

Teniendo en cuenta el PC y el ICC, el método sugiere distribuir a los participantes en cuatro categorías en función de sus niveles de cambio (Jacobson y Truax, 1991; Jacobson y cols., 1999): pacientes recuperados (aquellos que alcanzaron el PC y el ICC necesario) pacientes que mejoraron (aquellos que no alcanzaron el PC pero si el ICC), pacientes que no mejoran (aquellos que no alcanzaron ninguno de los dos requisitos) y pacientes que empeoran (aquellos que alcanzaron un ICC confiable pero en dirección opuesta a la de la mejoría).

Luego del trabajo de Jacobson y Truax (1991), surgieron algunas críticas a su método, desarrollándose modelos alternativos para el cálculo de la significancia clínica.

Métodos alternativos al de Jacobson y Truax

Método GNL

Hsu (1989) fue el primero en introducir un método alternativo al de Jacobson y Truax, sosteniendo que la fórmula para el ICC desarrollada por estos no tiene en cuenta la regresión a la media al indagar las diferencias entre pre-test y post-test. Para resolver

esta falencia propone un nuevo método de cálculo: tanto al pre-test como al post-test se les resta la media hipotética de la población, y su resultado se divide por su desvío estándar hipotético. El método GNL, desarrollado por Hsu, propone observar si los cambios de los pacientes son o no superiores al valor predicho por una regresión lineal para el post-test, en función de los resultados del pre-test. En tal caso, si una mejoraría es más baja que el puntaje que la regresión a la media predice, no se considera al paciente con un cambio clínicamente significativo. El hecho de que en escasas ocasiones se conozcan la media y el desvío estándar de la población, es una de las principales limitaciones de este método (Atkins y cols., 2005). Si se desconocen los datos de la población, Atkins y colaboradores (2005) proponen utilizar en su lugar la media y el desvío estándar del pre-test de la población disfuncional.

A continuación se presenta la fórmula del método GNL (Hsu, 1989):

$$ICC = \frac{[X_{post} - M_{pob}] - r_{xx}[X_{pre} - M_{pob}]}{D_{pob} \sqrt{(1-r_{xx}^2)}}$$

Siendo X_{pre} el puntaje del paciente previo al tratamiento, X_{post} el puntaje posterior al mismo, M_{pob} la media hipotética de la población, D_{pob} el desvío estándar hipotético de la población y r_{xx} la confiabilidad test-retest del instrumento de medición utilizado.

Según la fórmula del Método GNL, un ICC igual o superior a 1,96 representa una mejoría confiable desde el punto de vista estadístico. En caso de que las mejorías, de acuerdo al instrumento usado, estén representadas por una reducción en los puntajes, un cambio positivo confiable sería aquel con un ICC inferior o igual a -1,96.

Método EN

Speer (1992) comparte la crítica de Hsu (1989) al método de JT, afirmando que este último no tiene en cuenta la regresión a la media. La propuesta alternativa de Speer, el método EN, está compuesta por dos fórmulas que permiten el cálculo de un intervalo de confianza. A continuación se presentan ambas fórmulas:

$$ICC = [r_{xx}(X_{pre} - M_{pre}) + M_{pre}] \pm 2 D_{pre} \sqrt{(1-r_{xx}^2)}$$

Siendo M_{pre} la media de la muestra previa al abordaje.

La fórmula de Speer otorga como resultados dos puntajes que representan los límites del ICC, un límite inferior y otro superior. Si el puntaje post-trata-

miento del paciente supera el límite superior implica un deterioro confiable, si el puntaje en cambio es menor al límite inferior el mismo significa una mejoría confiable. Puntajes que se encuentran entre el límite inferior y el superior, representan un cambio que no es lo suficientemente confiable.

Método HLM

Speer y Greenbaum (1995) usan un modelo de curva de crecimiento para calcular la significancia clínica. Este método requiere de al menos tres puntajes de cada paciente. Para algunos autores esto representa una ventaja, ya que se incorporan puntajes intermedios entre el pre-test y el post-test, pudiendo tener en cuenta los cambios producidos entre aquellos (Bauer, Lambert y Nielsen, 2004). La utilización de las estimaciones empíricas de Bayes permiten combinar en el análisis datos acerca de la evolución tanto del individuo como del grupo, otorgando mayor flexibilidad e información más valiosa (Bauer, Lambert y Nielsen, 2004). La fórmula es (Speer y Greenbaum, 1995):

$$\frac{B}{\sqrt{V}}$$

Siendo B la estimación empírica de Bayes de la pendiente lineal y V el desvío estándar de la estimación empírica de Bayes.

Método HA

Hageman y Arrindell (1999) sostienen la necesidad de distinguir entre el cambio individual y el cambio grupal en el cálculo de la significancia clínica. Los autores no solo introducen modificaciones en la forma de determinar la confiabilidad de los cambios, como lo hacen las propuestas alternativas ya descritas, sino que también son los primeros en establecer nuevos parámetros para la determinación del Punto de Corte. Tan es así que el método HA presenta dos nuevos índices: 1. La confiabilidad de cambio individual (ICCind), en donde un resultado superior a 1,65 implica un deterioro del paciente, un puntaje entre 1,65 y -1,65 muestra un cambio no confiable y un valor inferior a -1,65 muestra un cambio positivo confiable. 2. El índice de Significancia Clínica del Cambio (SCind), que modifica el punto de corte del método JT calculado a partir de puntajes verdaderos y coeficientes de confiabilidad. Este índice, al igual que el método tradicional, permite clasificar a los pacientes en cuatro categorías: deteriorados, sin cambio, con progreso y recuperados. Su fórmula para el ICCind es (Hageman y Arrindell, 1999):

$$ICCind = \frac{(X_{post} - X_{pre})r_{dd} + (M_{post} - M_{pre})(1 - r_{dd})}{(\sqrt{r_{dd}})(\sqrt{2\frac{S_E^2}{E}})}$$

$$r_{dd} = \frac{D_{pre}^2 r_{xx(pre)} + D_{post}^2 r_{xx(post)} - 2 D_{pre} D_{post} r_{pre*post}}{D_{pre}^2 + D_{post}^2 - 2 D_{pre} D_{post} r_{pre*post}}$$

$$r_{xx(pre)} = \frac{D_{pre}^2 - S_E^2}{D_{pre}^2}$$

$$r_{xx(post)} = \frac{D_{post}^2 - S_E^2}{D_{post}^2}$$

$$S_E = D_{pre} \sqrt{1 - r_{xx}}$$

A su vez, la fórmula de los autores para el punto de corte del SCind es:

$$\frac{M_{post} + (X_{post} - M_{post}) r_{xx(post)} - TRC}{\sqrt{(r_{xx(post)})^2 S_E^2}}$$

Siendo TRC el puntaje de corte verdadero (true cutoff score).

Usando el puntaje de corte (a), el TRC se calcula de la siguiente manera:

$$TRC = M_{pre} - 2 D_{pre} \sqrt{r_{xx(pre)}}$$

Método de Blanchard y Schwarz

Otro método para el estudio de la significancia clínica es el desarrollado por Blanchard y Schwarz (1988). Los autores con tal objetivo miden la severidad y frecuencia de la sintomatología antes y después del tratamiento, considerando una reducción del 50% en dichas dimensiones de los síntomas, un cambio clínicamente significativo. Más allá de su simpleza, el método despierta dudas ya que la designación del 50% como punto de corte entre los pacientes que alcanzaron o no un cambio significativo es arbitraria y no surge de datos poblacionales previos. Esto llevaría a pensar que en algunos casos una reducción aun superior al 50% de la severidad y frecuencia de la sintomatología, puede no ser suficiente para alcanzar cambios valiosos para los pacientes, de la misma manera que una reducción inferior al 50% de la sintomatología puede implicar un cambio significativo.

Método de Tingey y Colaboradores

Como ya se expuso, Tingey y colaboradores (1996) critican la propuesta de Jacobson y Truax de dividir a la población en dos estratos (uno funcional y otro disfuncional), al considerar que la misma puede estar integrada por múltiples grupos con diferentes graduaciones de sintomatología e impacto social, ubicados en un continuo. Por ello definen distintas poblaciones en relación a la severidad de sus síntomas y a su impacto social. Una vez identificadas las poblaciones, se deben ubicar puntos de corte entre los límites de las poblaciones adyacentes en el continuo. Para comprobar la diferencia de las poblaciones se pueden realizar t-test (test estándar), determinando si las diferencias en las distribuciones de ambas son estadísticamente significativas, y d-test (tamaño de efecto) para determinar la magnitud de dicha diferencia. Calculados los puntos de corte y el índice de confiabilidad, el paciente que alcanza un cambio clínicamente significativo es aquel que logra un ICC confiable y supera el punto de corte de la población que se encuentra adyacente a la que pertenecía antes del tratamiento. La forma de calcular el PC y el ICC es la misma que la propuesta por Jacobson y Truax (1999). Por ejemplo, un paciente perteneciente a la categoría con un nivel elevado de sintomatología tendría un cambio clínicamente significativo al superar el punto de corte que existe entre dicha población y la de pacientes con un nivel medio de sintomatología, alcanzando también el ICC necesario.

Follette y Callaghan, (1996) critican las modificaciones sugeridas por Tingey y colaboradores. Sostienen que este método se aleja del concepto de significancia clínica, cuyo objetivo original era incluir la visión de los pacientes al evaluar los resultados de la terapia. Según sus propias palabras, "los clientes no están interesados en cuántos puntos en una escala cambiaron, sino en alcanzar un cambio cualitativo que ellos perciban como importante" (Follette y Callaghan, 1996, p. 138). Sin embargo, los autores afirman que es posible la identificación de un grupo intermedio entre la población funcional y la disfuncional, en casos en los que es inviable que el paciente pase de una población a la otra, debido a que históricamente no se ha reportado casuística que cumpliera dicho objetivo (tal es el ejemplo de los pacientes con autismo o esquizofrenia) No obstante lo cual, alertan que esta adecuación puede llevar a que los clínicos e investigadores se conformen con resultados limitados, y no promuevan la búsqueda de tratamientos superadores.

La aplicación de algunos de los métodos aquí descriptos (Método JT, Método GNL, Método EN y Método HA) ha sido ejemplificada en el Apéndice de

este trabajo, con el fin de facilitar su comprensión y favorecer su eventual aplicación.

Conclusiones y Discusión

Los modelos estadísticos han tenido un papel destacado en los estudios de eficacia en psicología clínica. A pesar de sus limitaciones y de los nuevos criterios surgidos para los estudios de resultados, la significancia estadística sigue siendo el parámetro más utilizado para la evaluación de eficacia en psicoterapia (Jacobson y cols., 1999). Sin embargo, a diferencia de lo que siguieron Durand y Barlow, no consideramos que la inclusión de los modelos matemáticos haya sido el elemento esencial que permitió a la psicología alcanzar el estatus de disciplina científica, sino el uso de criterios y métodos sistemáticos y los esfuerzos realizados para operacionalizar las variables incluidas en sus construcciones teóricas.

La propuesta de incluir como criterio en las investigaciones de resultados el valor y sentido de los cambios alcanzados no implica un intento de abandonar el uso de la estadística. La significancia clínica incluye en su seno elementos estadísticos fundamentales. Tal es el caso del valor p al momento de evaluar la confiabilidad de los cambios o las fórmulas usadas para determinar los puntos de corte. Pese al avance progresivo de la significancia clínica, tanto la significancia estadística como el tamaño del efecto seguirán teniendo su espacio e importancia en las investigaciones de eficacia (Ogles y Lunnen, 1996). Además, es importante tomar en consideración que los resultados de la significancia clínica no se contraponen a los de la significancia estadística, sino que tienden a agregar mayor información a esta (Jacobson y Truax, 1991). Tal condición nos lleva a considerar que el vínculo entre estos dos criterios de significancia no debería pensarse en términos de oposición sino de complementariedad. La conjunción de modelos matemáticos y criterios clínicos cualitativos al momento de evaluar los efectos de la psicoterapia puede permitir una comprensión más compleja de la valoración de la salud y la patología, y de los procesos de cambio involucrados en la práctica clínica.

Una observación que se desprende de las investigaciones que utilizan a la significancia clínica como parámetro, es que los resultados de eficacia obtenidos a partir de ella son mucho más humildes que los que podrían esperar los clínicos y los que se derivan de otros criterios (Jacobson y Truax, 1991), detectando incluso que un porcentaje de pacientes empeoran a partir de los tratamientos (Lambert y Ogles, 2009). En estudios como los de Ankuta y Abeles (1993), Lunnen y Ogles (1998), Currie, Wilson y Curran (2002), Beckstead y colaboradores (2003)

y Beretta y colaboradores (2005) el porcentaje de pacientes con cambios clínicamente significativos se ubicó entre el 7% y el 33%. En cuanto al porcentaje de pacientes que empeoran a partir de la psicoterapia, Lambert y Ogles (2009) afirman que se ubica entre el 5% y 10% (Para algunos ejemplos ver Lunnen y Ogles, 1998; Beretta y cols., 2005). A pesar de que la utilización de la significancia clínica aporta resultados más humildes y, por ello, menos esperanzadores para los que nos dedicamos al ejercicio de la psicoterapia, es necesario plantearse si estos resultados no son más realista y, por lo tanto, si no nos brindan un conocimiento más confiable y verídico de las consecuencias de nuestro accionar clínico.

Como se ha mencionado en el cuerpo del trabajo, luego de surgido el método de tradicional de cálculo de la significancia clínica, creado por Jacobson y Truax (1991), diferentes métodos alternativos se han desarrollado para suplir limitaciones estadísticas identificadas en el primero. Existen al menos tres posturas respecto de la diversificación de métodos:

1. La primera postura sostiene que la existencia, el uso y la proliferación de diferentes métodos para el cálculo de la significancia clínica complica los intentos de comparar resultados de distintos estudios (Lambert y Ogles, 2009). Según este paradigma, un paso fundamental para el avance del criterio de significancia clínica sería que la comunidad científica pudiera alcanzar cierto grado de consenso acerca de cuál es el método más adecuado para su cálculo, generalizando su uso, y permitiendo, así, transpolar los resultados de las diferentes investigaciones que se dedican a los estudios de eficacia a partir de la significancia clínica.

En este contexto, un conjunto de investigaciones tuvieron como objetivo comparar los diferentes métodos para favorecer la selección del más preciso. McGlinchey, Atkins y Jacobson (2002) no encontraron diferencias significativas entre los resultados de cada uno de ellos. Este dato se encuentra en consonancia con lo afirmado por Maassen (2000), quien sostiene que frente a ausencia de información (media y desvío estándar) de la población, los resultados de los diferentes métodos serían similares.

Sin embargo, a partir de otros estudios, se pudieron identificar ciertas diferencias y similitudes entre los distintos métodos. Bauer, Lambert y Nielsen (2004) y Atkins y colaboradores (2005), encontraron que los métodos con mayor nivel de concordancia son el método JT (Jacobson y Truax, 1991) y el método GNL (Hsu, 1989), llegando a porcentajes de confiabilidad en la clasificación de la significancia clínica del 96%. Es interesante observar que a pesar de que el método GNL toma en cuenta la regresión a la media en el cálculo de la significancia clínica, a diferencia del método JT, los resultados de ambos

son prácticamente iguales. Esto podría sugerir que el método JT no se ve afectado por la amenaza de la regresión a la media tal como suponía Hsu (1989). Asimismo, en ambos estudios se observó que los resultados de los métodos de JT y GNL fueron los más moderados a la hora de clasificar pacientes con cambios clínicamente significativos, siendo el método EN (Speer, 1992) el método que más pacientes identificaba con cambios clínicamente significativos y el método HA (Hageman y Arrindell, 1999) el que menos identificaba.

Estos resultados llevan a diversos investigadores a proponer el uso del método tradicional de Jacobson y Truax como método unificado para el estudio de la significancia clínica (Maassen, 2000; Bauer, Lambert y Nielsen, 2004; Atkins y cols., 2005; Lambert y Ogles, 2009). En primer lugar, porque es por amplio margen el más utilizado y difundido en las investigaciones que utilizan la significancia clínica (Bauer, Lambert y Nielsen, 2004; Atkins y cols., 2005). En segundo lugar, por la facilidad para su aplicación (Bauer, Lambert y Nielsen, 2004; Atkins y cols., 2005). Y por último, por arrojar resultados moderados en comparación con otros métodos que, con idénticos datos, clasifican menores o mayores porcentajes de pacientes con cambios significativos (Bauer, Lambert y Nielsen, 2004).

2. Una segunda postura, contrapuesta a la anterior, sostiene que sería fundamental que la definición utilizada para calcular la significancia clínica de cada paciente y su correspondiente método, se establezca en función de la búsqueda particular de cada proceso terapéutico (Kazdin, 1999), acercándose a un método más idiográfico que tenga en cuenta los objetivos específicos de cada sujeto que consulta (Lambert y Ogles, 2009). El trabajo de Asenlof, Denison y Lindberg (2005), es un ejemplo en esta dirección: Los autores desarrollan un método, basado en el JT, que utiliza parámetros individuales para evaluar la presencia o ausencia de cambios clínicamente significativos. Disentimos con Kazdin en este punto. Sin desestimar el valor de la dimensión idiográfica en la evaluación de los resultados, siguiendo su propuesta existirían tantos métodos como pacientes, por lo cual los resultados de una investigación serían imposibles de contrastar con los de otras. Por otra parte, resultaría problemático que cada clínico tenga que establecer una definición operacional particular de cambio clínicamente significativo para cada paciente, adecuando la metodología para su cálculo, ya que en muchas ocasiones los clínicos no cuentan con los conocimientos metodológicos específicos necesarios para llevar a cabo estas acciones. No obstante, es necesario remarcar que aunque la definición operacional de significancia clínica propuesta por Jacobson y Truax (alcanzar

un nivel de funcionamiento normal) busca representar lo que sería un cambio significativo para los pacientes, ésta no deja de ser una definición surgida desde el campo de la investigación. Si la verdadera intención detrás de la significancia clínica es saber cuándo un paciente alcanza cambios clínicamente significativos, son necesarias investigaciones para validar o corregir esta definición, que indaguen la clase y el nivel de cambio considerado significativo desde la perspectiva de los pacientes (Kazdin, 1999).

3. La propuesta de Jacobson y colaboradores (1999) representa una postura intermedia en este debate. Los autores afirman que en lugar de buscar indagar cuál es el mejor método para estudiar la significancia clínica, lo más adecuado es determinar qué método es el más útil para estudiar la significancia clínica en cada trastorno o problema clínico específico. Adherimos a esta postura ya que cada patología conlleva características diferentes y una evolución disímil, por lo cual lo más apropiado sería contar con una batería de estrategias a ser aplicadas en cada caso particular, según sus singularidades. Sin embargo, contar con una batería no implica una cantidad infinita de métodos sino con un conjunto limitado, que permita la validación empírica de cada uno de ellos en relación a sus correspondientes patologías de incumbencia y la posibilidad de comparar resultados.

Más allá de estos posicionamientos, diversos autores coinciden en que para poder discernir cuál de todos los métodos diseñados es el más preciso, ya sea para el cálculo de la significancia clínica a partir de un método unificado o para determinar el método más preciso para cada patología en particular, son necesarias más investigaciones empíricas (Bauer, Lambert y Nielsen, 2004; Lambert y Ogles, 2009) y, como afirman Jacobson y colaboradores (1999), menos digresiones y debates estadísticos al respecto. Hasta ahora, esta sugerencia de Jacobson y colaboradores (1999), no se ha traducido en acciones concretas. Esperamos no haber contribuido con este artículo a estos debates estériles, sino todo lo contrario: haber fomentado el incremento de estudios en busca de la validación empírica del constructo significancia clínica y sus diversos métodos. Aunque en los últimas décadas se han acumulado algunos trabajos que aportaron evidencia empírica en favor de la validez de la significancia clínica (Lambert y Ogles, 2009), mayores esfuerzos son necesarios en esta dirección para validar este criterio. Con tal fin sería fundamental que se multiplicaran las investigaciones que estudian la eficacia de los tratamientos a través de la significancia clínica y, a su vez, que proliferen aquellas que indagan la relación entre la significancia clínica y otras medidas

de mejoría de los pacientes. Asimismo, es necesario profundizar los intentos de validar las técnicas y escalas usadas para la medición del cambio en los estudios de significancia clínica (Beckstead y cols., 2003).

Otro aspecto perfectible del criterio de significancia clínica son sus definiciones operacionales. Sus cuatro categorías (recuperado, con mejorías, sin cambios y con cambios negativos) se presentan un tanto rústicas y acotadas para la descripción de un fenómeno tan complejo y con tantos matices posibles como es el de los resultados en psicoterapia. Investigaciones empíricas que busquen refinar estas categorías en busca de la contemplación de mayor cantidad de matices de los fenómenos de cambio en psicoterapia, serían un importante aporte.

Estas cuentas pendientes no le restan importancia al valor de la significancia clínica ni a su potencialidad en el campo de los estudios de eficacia. Tan es así que Lambert y Ogles (2009) recomiendan que en todos los estudios de eficacia en psicoterapia se presenten resultados de significancia clínica. Los autores consideran que la significancia clínica no solo aporta resultados más valiosos y significativos desde la perspectiva clínica, sino que también es un recurso importante para establecer puentes entre la práctica clínica y la investigación. Vemos en este artículo, cómo la significancia clínica puede ser un punto de contacto en donde el espacio de separación entre clínicos e investigadores (la usualmente denominada "brecha" entre estas dos áreas) se convierte en un espacio de unión e interacción con la potencialidad de enriquecer tanto los modelos explicativos surgidos y validados por los investigadores como las prácticas de los terapeutas. En muchas ocasiones los clínicos le asignan escaso valor a las investigaciones en psicoterapia (Morrow-Bradley y Elliot (1986), teniendo, por lo tanto, limitada influencia en su práctica. El estudio de la significancia clínica, al centrarse en las modificaciones reales que la psicoterapia produce en el día a día de los pacientes, podría ayudar a que las investigaciones de resultado tengan más influencia en la labor de los terapeutas y, por carácter transitivo, en sus pacientes. Por otra parte, la significancia clínica es un instrumento que los psicoterapeutas pueden utilizar para evaluar la evolución de sus pacientes, determinando si a partir del proceso alcanzaron o no cambios valiosos (Beckstead y cols., 2003).

Los debates históricos y actuales al respecto, muestran a la significancia clínica como un aporte fundamental para la investigación en psicoterapia (Goldfried, Greenberg y Marmar, 1990; Kazdin, 1999) y, al mismo tiempo, como un importante desafío que requiere de mayores esfuerzos empíricos y metodológicos. En nuestras manos, las de los clí-

nicos y los investigadores en psicoterapia, está el destino de este constructo y de su potencialidad de ejercer una influencia franca en la evaluación de los resultados, y por ende, en la mejoría de los procesos terapéuticos.

BIBLIOGRAFÍA

- Ankuta, G. Y. y Abeles, N. (1993). Client satisfaction, clinical significance and meaningful change in psychotherapy. *Professional Psychology: Research and Practice*, 24(1), 70-74.
- Asenlof, P., Denison, E. y Lindberg, P. (2005). Idiographic outcome analyses of the clinical significance of two interventions for patients with musculoskeletal pain. *Behavior Research and Therapy*, 44, 947-965.
- Atkins, D. C., Bedics, J. D., McGlinchey, J. B. y Beauchaine, T. P. (2005). Assessing clinical significance: does it matter which method we use? *Journal of Consulting and Clinical Psychology*, 73(5), 982-989.
- Bauer, S., Lambert, M. J. y Nielsen, S. L. (2004). Clinical significance methods: a comparison of statistical techniques. *Journal of personality assessment*, 82(1), 60-70.
- Beckstead, D., Hatch, A., Lambert, M., Eggett, D., Goates, M. y Vermeersch, D. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *The Behavior Analyst Today*, 4, 79-90.
- Beretta, V., Roten, Y. de, Drapeau, M., Kramer, U., Favre, N. y Despland, J. (2005). Clinical significance and patients perceived change in four sessions of brief psychodynamic intervention: characteristics of early responders. *Psychology and psychotherapy: Theory, Research and Practice*, 78, 347-362.
- Blanchard, E. B. y Schwarz, S. P. (1988). Clinically significant changes in behavioral medicine. *Behavioral Assessment*, 10, 171-188.
- Christensen, L. y Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, 17, 305-308.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1994). The Earth is Round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Currie, S. R., Wilson, K. G. y Curran, D. (2002). Clinical significance and predictors of treatment response to cognitive-behavior therapy for insomnia secondary to

- chronic pain. *Journal of Behavioral Medicine*, 25(2), 135-153.
- Durand, M. y Barlow, D. (2007). *Psicopatología. Un enfoque integral de la psicología anormal* (4ta ed.). México, DF: International Thompson Editores.
- Follette, W. y Callaghan, G. (1996). The Importance of the Principle of Clinical Significance — Defining Significant to Whom and for What Purpose: A Response to Tingey, Lambert, Burlingame, and Hansen. *Psychotherapy Research*, 6(2), 133-143.
- Goldfried, M. R., Greenberg, L. S. y Marmar, C. (1990). Individual psychotherapy: process and outcome. *Annual Review of Psychology*, 41, 659-688.
- Hageman, W. J. y Arrindell, W. A. (1999). Establishing clinically significant change: Increment of precision and the distinction between individual and group level of analysis. *Behavior Research and Therapy*, 37, 1169-1193.
- Hansen, N. B. y Lambert, M. J. (1996). Clinical significance: An overview of methods. *Journal of Mental Health*, 5(1), 17-24.
- Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment*, 11, 459-467.
- Jacobson, N. S., Follette, W. C. y Revenstorf, D. (1984). Toward a standard definition of clinically significant change. *Behavior Therapy*, 17, 308-311.
- Jacobson, N. S., Roberts, L. J., Berns, S. B. y McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of consulting and clinical psychology*, 67(3), 300-307.
- Jacobson, N. S. y Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Karpenko, V., Owens, J. S., Evangelista, N. M. y Dodds, C. (2009). Clinically Significant Symptom Change in Children With Attention-Deficit / Hyperactivity Disorder: Does It Correspond With Reliable Improvement in Functioning? *Journal of Clinical Psychology*, 65(1), 76-93.
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*, 67(3), 332-339.
- Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J. y Harmon, R. J. (2003). Measures of Clinical Significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(12), 1524-1529.
- Lambert, M. J. y Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: the need for a common procedure and validity data. *Psychotherapy research*, 19(4-5), 493-501.
- Lane, D. M. (2007). HyperStat Online, Fecha de acceso 8 de febrero, 2012. Disponible en: <http://davidmlane.com/hyperstat/>
- Lunnen, K. M. y Ogles, B. M. (1998). A multiperspective, multivariable evaluation of reliable change. *Journal of Consulting and Clinical Psychology*, 66(2), 400-410.
- Maassen, G. H. (2000). Principles of Defining Reliable Change Indices. *Journal of Clinical and Experimental Neuropsychology*, 22(5), 622-632.
- Martinovich, Z., Saunders, S. y Howard, K. (1996). Some Comments on "Assessing Clinical Significance." *Psychotherapy Research*, 6(2), 124-132.
- McGlinchey, J. B., Atkins, D. C. y Jacobson, N. S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy*, 33, 529-550.
- Meyer, T.J., Miller, M.L., Metzger, R.L. y Borkovec, T.D. (1990). Development and Validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, 28, 487-495.
- Morrow-Bradley, C. y Elliott, R. (1986). Utilization of psychotherapy research by practicing psychotherapists. *American Psychologist*, 41, 188-197.
- Ogles, B. y Lunnen, K. (1996). Assessing outcome in practice. *Journal of Mental Health*, 5(1), 35-46.
- Ogles, B., Lunnen, K. y Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, 21, 421-446.
- Peterson, L. S. (2008, febrero). "Clinical" significance: "clinical" significance and "practical" significance are not the same. Trabajo presentado en The Annual Meeting of the Southwest Educational Research Association, New Orleans, Estados Unidos de América.
- Rodríguez-Biglieri, R. y Vetere, G. L. (2011). Psychometric Characteristics of the Penn State Worry Questionnaire in an Argentinean Sample: A Cross-Cultural Contribution. *The Spanish Journal of Psychology*, 14(1), 452-463.
- Roussos, A. J. (2007). Uso de métodos meta-analíticos en psicología clínica. *Revista Argentina de Clínica Psicológica*, 16, 161-169.
- Sandell, R., Alfredsson, E., Berg, M., Crafoord, K., Lagerlöf, A., Cohn, T. y Rugolska, A. (1994). Turing's game and the clinical significance of outcome with borderline patients at a day hospital. *Journal of Clinical Psychology*, 50(3), 406-414.

- Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*, 402-408.
- Speer, D. C. y Greenbaum, P. E. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology, 63*, 1044-1048.
- Thompson, B. (2002). "Statistical," "Practical," and "Clinical": How Many Kinds of Significance Do Counselors Need to Consider? *Journal of Counseling & Development, 80*, 64-71.
- Tingey, R., Lambert, M., Burlingame, G. y Hansen, N. (1996). Assessing Clinical Significance: Proposed Extensions to Method. *Psychotherapy Research, 6*(2), 109-123.
- Titov, N., Andrews, G., Robinson, E., Schwencke, G., Johnston, L., Solley, K. y Choi, I. (2009). Clinician-assisted Internet-based treatment is effective for generalized anxiety disorder: randomized controlled trial. *Australian and New Zealand Journal of Psychiatry, 43*, 905-912.
- Wampold, B. E. y Jensen, W R. (1986). Clinical significance revisited. *Behavior Therapy, 17*, 302-305.
- Wilkinson, L. y American Psychological Association Task Force on Statistical Inference Inference, T. F. on S. (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist, 54*(8), 594-604.
- Wolf, M. (1978). Social validity: the case for subjective measurement or hoe applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-214.

Abstract: The present article synthesizes a debate emerged over twenty five years ago, that has renewed it's intensity and relevance: What is the most reliable criteria for establishing clinical change for efficacy and effectiveness studies in psychotherapy? The two traditional criteria used for outcome studies – statistical significance and practical significance – are presented. Clinical significance, the youngest criteria developed, is described and discussed in detail. The different clinical significance definitions and methods for it's calculation are explained. The application of the clinical significance criteria is illustrated by a fictitious case. This study shows clinical significance as an important resource not only for psychotherapy research, but also for clinical practice, being also an important contribution to facilitate the interaction between these areas.

Recibido: 11-4-12
Aceptado: 15-5-12

APÉNDICE

Ilustración de los métodos más utilizados para el cálculo de la significancia clínica

Para ilustrar la aplicación de las fórmulas más difundidas para el cálculo de la significancia clínica se utilizarán los datos de un estudio realizado en la Argentina por Rodríguez-Biglieri y Vetere (2011) acerca del Penn State Worry Questionnaire ([PSWQ] Meyer, Miller, Metzger y Borkovec, (1990), donde se presentan los puntajes promedio de la población general y de pacientes diagnosticados con Trastorno de Ansiedad Generalizada [TAG]. Los datos sobre el post-test del grupo experimental y grupo de control fueron tomados de un estudio de eficacia sobre de un tratamiento para pacientes con TAG, realizado por Titov y colaboradores (2009), que utiliza como instrumento de medición el PSWQ.

Datos utilizados para la ejemplificación

Media de la población disfuncional (M_d); Media del pre-test del grupo de tratamiento (M_{pre}): 70,15
Desvío estándar de la población disfuncional (D_d); Desvío estándar del pre-test del grupo de tratamiento (D_{pre}): 6,31
Media de la población funcional (M_f); Media hipotética de la población (M_{pob}): 42,97
Desvío estándar de la población funcional (D_f): 10,90
Media del post-test del grupo experimental (M_{post}): 56,75
Desvío estándar del post-test del grupo experimental (D_{post}): 10,78
Media del post-test del grupo de control ($M_{con/post}$): 66,14
Índice de Confiabilidad test-retest (r_{xx}): 0,86
Puntaje del pre-test del paciente (X_{pre}): 67
Puntaje del post-test del paciente (X_{post}): 56
Correlación entre el pre-test y el post-test ($r_{pre*post}$): 0,63

Tamaño del efecto

$$d = \frac{M_{post} - M_{(con/post)}}{D}$$

$$d = \frac{56,75 - 66,14}{10,78} = \boxed{0,87}$$

Siguiendo los criterios de Cohen (1988), el presente tamaño del efecto sería interpretado como “grande”, al ser superior a 0,80.

Método JT

2a) Estrategias para determinar el Punto de Corte:

Estrategia (a)

$$a = M_d + 2D_d$$

$$a = 70,15 + 2(-6,31) = \boxed{58}$$

Estrategia (b)

$$b = M_f - 2D_f$$

$$b = 42,97 - 2(-10,90) = \boxed{62}$$

Estrategia (c)

Asumiendo distribución normal:

$$c^i = (M_f + M_d) / 2 =$$

$$c^i = (42,97 + 70,15) / 2 = \boxed{56}$$

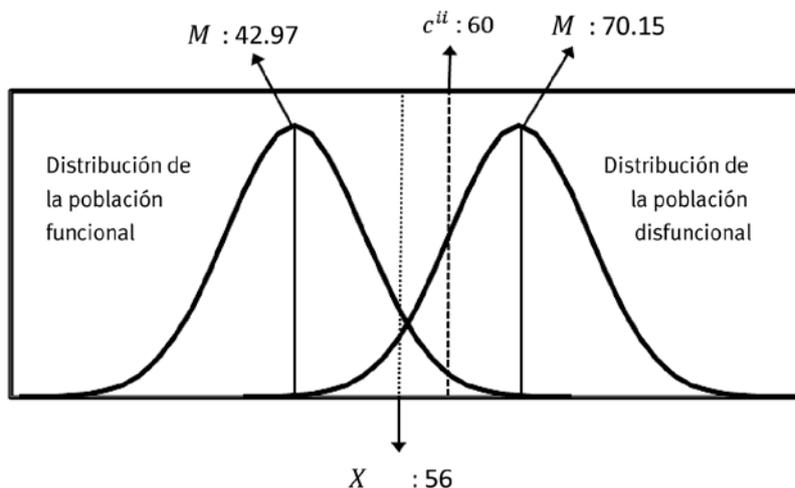
Si las varianzas de los dos grupos no son iguales la formula a aplicar es:

$$c^{ii} = \frac{(D_d M_f + D_f M_d)}{(D_d + D_f)}$$

$$c^{ii} = \frac{(6,31 \cdot 42,97 + 10,90 \cdot 70,15)}{(6,31 + 10,90)} = \boxed{60}$$

El puntaje del paciente ficticio (56) alcanza los puntos de corte de las tres estrategias (a: 58; b: 62; c: 56/60,18). Sin embargo, por el grado moderado de superposición de las distribuciones de las dos poblaciones y por el hecho de conocer los respectivos datos de ambas, la estrategia adecuada es la (c) en su adaptación frente a varianzas disimiles.

Figura 2. Presentación de la distribución normal y media de la población funcional y la población disfuncional, el punto de corte según la estrategia cii y el puntaje del paciente ficticio



2b) Índice de Confiabilidad de Cambio:

$$I_{CC} = \frac{X_{post} - X_{pre}}{S_{diff}}$$

$$S_{diff} = \sqrt{2 S_E^2}$$

$$S_E = D_{pre} \sqrt{1 - r_{xx}}$$

$$S_E = 6,31\sqrt{(1-0,86)}$$

$$S_{diff} = \sqrt{2[6,31\sqrt{1-0,86}]^2} = 3,34$$

$$ICC = \frac{(56-67)}{(3,34)} = -3,29$$

Al ser el ICC de este caso ficticio (-3,29) inferior a -1,96, se puede afirmar que el cambio es confiable ($p < 0,05$).

Al alcanzar el punto de corte, logrando un ICC adecuado, se concluye que los cambios del paciente ficticio deben ser considerados clínicamente significativos en función del método JT.

Método GNL

$$ICC = \frac{[X_{post} - M_{pob}] - r_{xx}[X_{pre} - M_{pob}]}{D_{pob} \sqrt{1-r_{xx}^2}}$$

$$ICC = \frac{[56 - 42,97] - 0,86[67 - 42,97]}{6,13 \sqrt{(1-0,86^2)}} = -2,28$$

Al alcanzar un ICC de -2,28 (inferior a -1,96), se concluye que los cambios del paciente ficticio deben ser considerados confiables según este método.

Método EN

$$ICC = [r_{xx}(X_{pre} - M_{pre}) + M_{pre}] \pm 2 D_{pre} \sqrt{(1-r_{xx})}$$

$$\text{Límite S del ICC} = [0,86(67 - 70,15) + 70,15] + 2 \cdot 6,31 \sqrt{1-0,86} = 72,11$$

$$\text{Límite I del ICC} = [0,86(67 - 70,15) + 70,15] - 2 \cdot 6,31 \sqrt{1-0,86} = 62,77$$

Siendo el puntaje del paciente posterior al tratamiento (X_{post}) de 56, se considera que el mismo representa una mejoría confiable ya que está por debajo del Límite Inferior del ICC (62,77).

Método HA

Índice de Confiabilidad de Cambio Individual:

$$ICC_{ind} = \frac{(X_{post} - X_{pre}) r_{dd} + (M_{post} - M_{pre})(1 - r_{dd})}{(\sqrt{r_{dd}})(\sqrt{2S_E^2})}$$

$$S_E = D_{pre} \sqrt{1-r_{xx}}$$

$$S_E = 6,31 \sqrt{(1-0,86)} = 2,36$$

$$r_{dd} = \frac{D_{pre}^2 r_{xx(pre)} + D_{post}^2 r_{xx(post)} - D_{pre}^2 D_{post} r_{xx(pre)} r_{xx(post)}}{D_{pre}^2 + D_{post}^2 - D_{pre}^2 D_{post} r_{xx(pre)} r_{xx(post)}}$$

$$r_{xx(pre)} = \frac{D_{pre}^2 - S_E^2}{D_{pre}^2}$$

$$r_{xx(pre)} = \frac{6,31^2 - 2,36^2}{6,31^2} = 0,86$$

$$r_{xx(post)} = \frac{D^2_{(post)} - S_E^2}{D^2_{(post)}}$$

$$r_{xx(post)} = \frac{10,78^2 - 2,36^2}{10,78^2} = 0,95$$

$$r_{dd} = \frac{D^2_{pre} r_{xx(pre)} + D^2_{post} r_{xx(post)} - 2 D_{pre} D_{post} r_{(pre*post)}}{D^2_{pre} + D^2_{post} - 2 D_{pre} D_{post} r_{xx(pre)} r_{xx(post)}} =$$

Siendo el $r_{(pre*post)}$ la correlación entre los puntajes del pre-test y el post-test.

$$r_{dd} = \frac{6,31^2 * 0,86 + 10,78^2 * 0,95 - 2 * 6,31 * 10,78 * r_{(pre*post)}}{6,31^2 + 10,78^2 - 2 * 6,31 * 10,78 * 0,86 * 0,95} = 0,84$$

$$ICCind = \frac{(56 - 67 r_{dd}) + (56,75 - 70,15)(1 - r_{dd})}{(\sqrt{r_{dd}})(\sqrt{2 * 2,36^2})} = \boxed{-1,11}$$

Al tener un ICCind que supera el punto de -1,65 se considera que el paciente no alcanzó un cambio confiable y, por ende, no es clasificado como un cambio clínicamente significativo.