

TRATAMIENTOS CON APOYO EMPÍRICO. UNA APROXIMACIÓN A SUS VENTAJAS, LIMITACIONES Y PROPUESTAS DE MEJORA

EMPIRICALLY SUPPORTED TREATMENTS. AN APPROACH TO YOUR BENEFITS,
LIMITATIONS AND SUGGESTIONS FOR IMPROVEMENT

GERARDO PRIMERO¹; JUAN ANTONIO MORIANA

(UNIVERSIDAD DE BUENOS AIRES (ARGENTINA); UNIVERSIDAD DE CÓRDOVA (ESPAÑA)
(RECIBIDO EL 2/11/2011, ACEPTADO EL 15/12/2011)

RESUMEN

El presente trabajo realiza una revisión sobre el movimiento de Tratamientos con Apoyo Empírico señalando sus principales ventajas e inconvenientes y analizando algunas propuestas para su mejora. En la primera parte del artículo, se presenta una introducción y perspectiva histórica sobre el tema. Posteriormente, desarrollamos las principales dificultades: descoordinación entre el ámbito investigador y el aplicado, su función prescriptiva, limitaciones en el uso de los diagnósticos de las principales clasificaciones internacionales, problemas de generalización y manualización de guías de tratamiento y divergencias entre los niveles de evidencia. Finalmente, se comentan las principales propuestas para superar estas limitaciones y mejorar el estudio de los tratamientos psicológicos eficaces, no solo para los trastornos mentales sino también para problemas psicológicos no clínicos.

Palabras Clave: Tratamientos con Apoyo Empírico, psicología basada en la evidencia, psicología científica, psicología clínica

ABSTRACT

In this paper we carry out a review about the Empirically Supported Treatment movement, pointing out its principal advantages and disadvantages and analyzing improvement proposals. In the first part of the article we introduce the topic, providing a historical perspective. Secondly, we focus on our main concerns: lack of coordination between research and its applications in the field, limitations when using the international classifications of mental disorders, standardization of treatment guidelines and differences between levels of evidence. Finally, the main proposals to overcome these limitations are discussed in order to improve the study of efficient psychological treatments of both mental and nonclinical psychological problems.

Keywords: Empirically Supported Treatments, Evidence-based psychology, psychological science, clinical psychology.

1 Universidad de Buenos Aires (Argentina); Universidad de Córdoba (España)

* Facultad de Psicología (Universidad de Buenos Aires - Argentina). E-mail: gerardoprimer@yahoo.com.ar

2 Departamento de Psicología (Universidad de Córdoba - España). E-mail: jamoriana@uco.es

1. Introducción y antecedentes históricos de la investigación sobre la eficacia de los tratamientos psicológicos

En la actualidad coexisten una enorme diversidad de tratamientos psicológicos (p.ej. terapia cognitiva, conductual, sistémica, gestáltica, rogeriana, interpersonal, psicoanalítica...). Los tratamientos pueden categorizarse según diversos criterios (p.ej. verbales versus no-verbales, orientados a la eliminación de conductas versus orientados a la construcción de conductas, focalizados en el cambio conductual versus focalizados en el *insight* o en la reflexión filosófica). Esta diversidad plantea una cuestión central: ¿Cómo podemos elegir adecuadamente entre todas las posibles opciones? ¿Hay algunas opciones más plausibles o eficaces para obtener nuestros objetivos? Y en caso afirmativo, ¿Cómo podemos saberlo? Esta información es relevante para cada uno de los actores (el psicólogo, la persona que requiere tratamiento, las instituciones de salud), porque una elección incorrecta puede resultar perjudicial en más de un sentido. En primer lugar, un tratamiento inadecuado puede provocar daños directos en el paciente (Lilienfeld, 2007). En segundo lugar, los tratamientos inocuos también producen perjuicios de tipo indirecto (privar a las personas del tiempo y los recursos económicos que podrían haber utilizado en tratamientos más eficaces, desprestigiar la profesión, disminuir la confianza del público). En tercer lugar, los tratamientos ineficaces ocupan recursos sociales limitados (tiempo, dinero, personas, esfuerzo), desplazando a otras opciones más eficaces.

Una característica común a los distintos tratamientos es la presencia de comportamientos (públicos y privados) realizados por el cliente y por el terapeuta, orientados a un objetivo, y guiados (al menos en parte) por un modelo teórico (el concepto de “relación terapéutica” está implícito en esta descripción, pues se refiere a un subconjunto de los comportamientos del cliente y el terapeuta). Sin embargo existen diferencias significativas en cuanto a estos aspectos: cada tratamiento categoriza a su modo los “problemas” (p.ej. síntomas, trastornos, conductas problemáticas, patrones de personalidad), las “causas” (p.ej. disfunción biológica, aprendizaje, pensamientos, mecanismos inconscientes), los “objetivos” (p.ej. reducir los síntomas, modificar las conductas, obtener *insights*), y los “medios” (p.ej. entrenamiento de habilidades, reestructuración cognitiva, asociación libre y asociación, establecimiento de una relación terapéutica). El reconocimiento de la “parcialidad” de la influencia del modelo teórico se debe a que existen comportamientos no planificados conscientemente (como consecuencia de aspectos idiosincrásicos del terapeuta y el cliente, así como del entorno de ambos), que pueden influir de forma beneficiosa o perjudicial respecto a los objetivos del tratamiento.

Históricamente, los estudios de resultados pueden agruparse en tres etapas (Goldfried y Wolfe, 1996). La primera generación de estudios, en las décadas de 1950 y 1960, centraron su atención en intentar demostrar si la psicoterapia “funcionaba”. Se hicieron revisiones de estudios de caso, y estudios poco controlados donde se comparaban sujetos tratados con diversas estrategias y sujetos

no tratados, aunque estas investigaciones adolecían de fiabilidad y de contraste de las hipótesis rivales (remisión espontánea, efecto placebo).

En la segunda generación (décadas de 1960 y 1970), la cuestión principal fue abordar qué intervenciones eran más efectivas para ciertos problemas específicos. Por aquel entonces se comparaba el grupo que recibía el tratamiento con otro grupo control (lista de espera, placebo, tratamiento rival, o tratamiento convencional). Aunque, en general, los estudios de esta época tenían importantes deficiencias metodológicas (sobre todo en las muestras de población empleadas).

Durante la tercera generación de estudios (años 80 y 90), mejoraron notablemente la calidad de estudios y se generalizó el uso de ensayos clínicos controlados (RCT, *Randomized Controlled Trials*, estudios con grupo control y asignación aleatoria de pacientes reales), meta-análisis (estudios que suman los resultados obtenidos en diversos estudios), y revisiones selectivas (explicitación de un criterio mínimo de calidad y revisión de investigaciones que cumplan el criterio, estudios bibliométricos, etc). En esta última etapa, se buscaba obtener resultados más fiables y generalizables utilizando terapeutas entrenados y manuales o guías de tratamiento, y haciendo evaluación de la adherencia, de los resultados, mejorando considerablemente la selección de pacientes reales, etc.

El antecedente más significativo de los TAEs proviene de la medicina, disciplina desde la cual se heredó la metodología de ensayos clínicos de validación de fármacos (que se intentó adaptar, con importantes diferencias, a la investigación de tratamientos psicológicos), así como la propuesta de sistematizar y hacer accesible la evidencia (Sackett, Richardson, Rosenberg y Haynes, 1997). En 1993, la *American Psychological Association (APA)* creó la Comisión sobre la Promoción y Difusión de Procedimientos Psicológicos (*Task Force on Promotion and Dissemination of Psychological Procedures*), con el fin de evaluar y determinar la eficacia de los tratamientos psicológicos para diferentes tipos de trastornos psicológicos. Entre los factores sociales que llevaron a tomar esta decisión se encontraban la presión de los seguros privados y sistemas públicos de salud para evaluar el coste-beneficio, las críticas mediáticas a las psicoterapias, el avance de los tratamientos farmacológicos en competencia con los psicológicos, el incremento de los juicios por mala praxis a psicólogos y, sobre todo, conocer cuáles eran los mejores tratamientos para cada trastorno.

2. ¿Qué son los Tratamientos con Apoyo Empírico (TAEs)?

La Comisión de la Sección 12 de la APA (2006) propuso criterios explícitos para identificar el grado de eficacia demostrada por un tratamiento, utilizando 3 categorías (bien establecido, probablemente eficaz, y experimental). En base a esos criterios de categorización, realizó revisiones selectivas de los estudios publicados (utilizando criterios explícitos de inclusión y exclusión de estudios). Finalmente, a partir de esos criterios y de las revisiones de estudios, confeccionó un conjunto de listados de tratamientos bien establecidos y probablemente eficaces.

Estos tres componentes (categorías de eficacia, revisiones selectivas y listados de tratamientos) fueron adoptados, con ciertas modificaciones, por otras instituciones, como el *National Institute for Health and Clinical Excellence* (NICE), y por otros países (p.ej. Reino Unido, Canadá, Alemania, España). En esta línea podemos denominar TAEs a aquellos tratamientos que cumplan los criterios de tratamiento “bien establecido” o “probablemente eficaz” en una revisión selectiva.

Los listados de la APA Task Force desarrollados en 1995, 1996, 1998 y 2001, y de otras instituciones que informaban sobre el nivel de evidencia de los tratamientos psicológicos, obtuvieron en general una buena aceptación entre profesionales e investigadores. Parece haber consenso respecto a la importancia de la identificación de las intervenciones más efectivas dirigidas a mejorar la práctica clínica. Aunque esto no significa que este modelo no haya recibido críticas. Es evidente que las revisiones selectivas y los listados de tratamientos constituyen un primer paso hacia un objetivo más ambicioso: una psicología aplicada basada en la evidencia. Pero todavía queda mucho por realizar para llegar a ese objetivo, y por ello resulta indispensable analizar en detalle las críticas, conservar los aciertos, y corregir las deficiencias. A continuación, hacemos un recorrido por algunas de estas cuestiones, reflexionando sobre las posibles propuestas de mejora.

3. Críticas y propuestas en relación al listado de TAEs y la metodología empleada

Posiblemente, una de las mayores contribuciones de los listados de TAEs ha implicado la creación de instituciones que actúan como mediadoras entre la investigación y la práctica clínica, y el establecimiento de criterios explícitos para juzgar la calidad de la evidencia de las diversas intervenciones (en principio, en el ámbito de la psicología clínica, aunque ya existen propuestas para extenderlo a otras áreas de la psicología, p.ej. Moran y Malott, 2004). Esta mediación implica, básicamente, dos etapas: por un lado, la evaluación de la evidencia (mediante revisiones selectivas guiadas por criterios), y por otro, el traslado de la información (mediante publicaciones, libros, manuales, cursos de entrenamiento, etc.) a los distintos actores involucrados (psicólogos, pacientes, instituciones de salud, público en general).

Sin embargo, hay que reconocer que ambas etapas pueden ser mejoradas. Las críticas a los listados de TAEs podrían clasificarse, en principio, según su focalización en las deficiencias actuales de la primera o la segunda etapa. Respecto a la primera (evaluación de la evidencia), se ha cuestionado: (1) la evaluación de la evidencia es incorrecta (o al menos, diferente entre instituciones y profesionales que la evalúan), (2) las categorías de “problemas” y “objetivos” son inadecuadas, (3) faltan datos para optar entre individualización y estandarización del tratamiento, (4) faltan datos sobre la generalización de los tratamientos, (5) faltan revisiones de otras áreas y metodologías de investigación. Respecto a la segunda etapa, se ha argumentado que (6) los listados de tratamientos resultan insuficientes para el traslado de la información a los actores involucrados. Esta clasificación nos permitirá identificar y analizar algunas de las principales propuestas para solucionar cada deficiencia.

3.1. Críticas basadas en que la evaluación de la evidencia es incorrecta o poco fiable

Una de las críticas más radicales a la revisión de TAEs es el “veredicto del pájaro Dodo” (Wampold, 1997), que defiende la equivalencia de psicoterapias (“todas ganan y todas merecen premios”). Según los defensores del veredicto del Dodo, no hay diferencias significativas de eficacia entre los tratamientos, y los beneficios terapéuticos se explican por factores comunes a los distintos tratamientos. Esta hipótesis resulta problemática, y tenemos razones para dudar de su plausibilidad (Crits-Christoph, 1996; Beutler, 2002a; Cautilli, 2006; Chambless, 2002; Lilienfeld, 2007). ¿Cómo puede ser cierto que “todos los tratamientos ganan o sirven”, si la mayoría de los tratamientos no han sido investigados en condiciones controladas? Para corregir el veredicto en pocas palabras, habría que decir que “algunos ganan, otros pierden, algunos empatan, y la mayoría no participan de la carrera”.

Los que “ganan” son principalmente los tratamientos cognitivo-conductuales y analítico-conductuales, pero también algunos tratamientos específicos provenientes de otras tradiciones teóricas (Chambless y Ollendick, 2001). La mayoría de los estudios de meta-análisis son comparaciones entre diversas formas de terapia cognitiva y conductual para trastornos de ansiedad. La conclusión de los meta-análisis no es sorprendente: algunos tratamientos similares, para los mismos trastornos, obtienen resultados similares. Pero estos hallazgos no permiten inferir que “todos los tratamientos funcionan para todos los trastornos”. De hecho, hay evidencias para sostener lo contrario: algunos tratamientos son mejores que otros para problemas y poblaciones particulares (Chambless y Ollendick, 2001).

Los que “pierden” son los tratamientos de los cuales existen evidencias de efectos perjudiciales, por ejemplo, la recuperación de recuerdos reprimidos y el *debriefing* (Lilienfeld, 2007). Aunque son pocos los estudios que reportan de efectos negativos (probablemente no lleguen a publicarse). También existen evidencias de diferencias en la relación coste/beneficio y en la probabilidad de recaídas, dependiendo del tratamiento y el trastorno.

Los que “empatan” son los casos de “equivalencia local” entre tratamientos. Hay evidencias de que algunos tratamientos tienen una eficacia similar para ciertos trastornos, pero este fenómeno de equivalencia local no implica una equivalencia generalizada (“todos empatan”). Por otro lado, no es suficiente con decir que dos tratamientos son “equivalentes”, ya que pueden ser ambos ineficaces, y por lo tanto “equivalentes” en su ineficacia.

Algunos de los tratamientos de amplio uso (p.ej. psicoanalíticos y humanistas) “no participan de la carrera”, ya que no han sido estudiados de forma rigurosa (porque son prolongados, difícilmente manualizables, y con objetivos no asimilables a la reducción de síntomas). Pero estas dificultades no les eximen de la necesidad de evaluar los resultados y generar guías de práctica clínica.

En síntesis, no es cierto que “todos ganan”: algunos ganan, otros pierden, otros empatan, y la mayoría no participan de la carrera. Este veredicto corregido apoya la necesidad de identificar el grado de evidencia de cada tratamiento, y de facilitar a todos los actores involucrados el acceso a la información relevante para tomar sus decisiones. ¿Cómo se llegó entonces a defender el “Veredicto del Dodo”? Posiblemente, en base a un conjunto de errores y sesgos en el muestreo y el análisis de datos (Crits-Christoph, 1996; Beutler, 2002b; Cautilli, 2006; Chambless, 2002; Ehlers y otros, 2010). Por otro lado, existen razones políticas que explicarían la persistencia del veredicto del Dodo, ya que el eslogan “todos ganan” permite mantener una actitud permisiva en la cual los psicólogos pueden ofrecer cualquier tratamiento, sin restricciones ni penalizaciones a la difusión de tratamientos pseudocientíficos, que aún no han sido sometidos a contraste, y que podrían tener efectos perjudiciales de forma directa (produciendo daño activamente) o indirecta (privando a las personas del tiempo y los recursos económicos que podrían dedicar a tratamientos más eficaces). Además la evaluación de la evidencia sigue a menudo diferentes criterios y grados de valoración (Chambless, et al, 1996, 1998; Chorpita, et al , 2011; Moriana y Martínez, 2011a), lo que induce a pensar que la fiabilidad entre listados en cuanto a su construcción y análisis es significativamente distinto.

3.2. Críticas basadas en que las categorías de “problemas” y “objetivos” son inadecuadas

En general, las evaluaciones de TAEs utilizaron las categorías de las clasificaciones internacionales de enfermedades (CIE-10; DSM-IV-TR). Pero estas categorías, en muchos casos, tienen escasa validez predictiva para seleccionar un tratamiento eficaz. Están basadas en las apariencias (topografía), en lugar de basarse en los aspectos funcionales que podrían ser más relevantes para la elección del tratamiento. La estrategia del diagnóstico sindrómico, que ha sido útil en muchas áreas de la medicina, no ha sido tan útil en psicología, donde conductas con la misma apariencia pueden tener distintas causas (p.ej. consumo de droga reforzado positivamente por los efectos de la droga, o reforzado negativamente por el escape de situaciones aversivas) y donde el mismo proceso causal puede tener distintos resultados (p. ej. diversos trastornos de ansiedad tienen la misma función de evitación sostenida por reforzamiento negativo). Algunas enfermedades físicas también tienen estos rasgos (p. ej., el cáncer) y los investigadores médicos abandonaron las clasificaciones sindrómicas para tales casos (Hayes y Follette, 1992). Las categorías topográficas conllevan algunos problemas en la generalización de los tratamientos. Beutler menciona que, después de una década, se identificaron 150 TAEs, que cubren sólo 51 de los 366 diagnósticos del DSM (Beutler, 2002b). Entrenarse en tantos TAEs es imposible para un psicólogo, y aún haciéndolo, faltarían guías para tratar la mayoría de los problemas clínicos. En los casos de comorbilidad no hay criterios para decidir un tratamiento, y la opción de utilizar secuencialmente varios manuales no parece ser una estrategia prometedora.

Otra crítica importante señala que la mayoría de los psicólogos trabajan con personas “normales” que presentan problemas que no deberían considerarse “trastornos

mentales” (Moriana y Martínez, 2011a). Quizás deberíamos proponer en estos casos otros términos (problemas de la vida, problemas psicosociales) que facilitarían la normalización de eventos. No hay estudios sobre por qué los pacientes buscan psicoterapia: por quejas del Eje I, clínicas o subclínicas, o por cuestiones personales (relaciones problemáticas, dificultades laborales), o si se presentan con un panorama difuso que requiere clarificación de objetivos (Westen, Novotny y Thompson-Brenner, 2004a) o de conductas problema. Estudios en contextos aplicados muestran que entre un tercio y la mitad de los pacientes no pueden diagnosticarse con el DSM porque sus problemas no encajan o porque tienen niveles sub-umbral para las categorías (Stirman, DeRubeis, Crits-Christoph y Brody, 2003; Westen, Novotny y Thompson-Brenner, 2004a). Raave (2003, pp. 128) señala que, debido a que los problemas de la vida no suelen ser costeados por las compañías de seguros médicos, “hay un esfuerzo conjunto en el campo de la salud mental para definir los asuntos existenciales, angustias e incluso dilemas morales en términos de enfermedad médica y para establecer los problemas de los pacientes como poseyendo causas genéticas o debidos a un desequilibrio químico en el cerebro. Los profesionales de salud mental están aprobando la redefinición de los síntomas de los trastornos mentales encontrados en el DSM en términos de enfermedad, no simplemente para darle un criterio más científico, ni para justificar el progresivo incremento de drogas como tratamiento, sino para obligar a las compañías privadas de seguros médicos a reembolsarles sus servicios”.

El concepto de “trastorno” suele tener una connotación de “disfunción interna”, que carece de evidencias en la mayoría de los casos, y que favorece al tratamiento farmacológico por encima de las categorías y tratamientos focalizados en la interacción psicosocial. La restricción a las categorías del DSM lleva a que no se financien los estudios sobre problemas y objetivos psicosociales que no puedan ubicarse como “trastornos” (p.ej. habilidades sociales). Los objetivos y resultados tienden a restringirse a la medición de síntomas, omitiendo otros cambios relevantes (experiencias subjetivas, cambios en la calidad de vida). Y el diagnóstico basado en categorías sindrómicas, que se suponen síntomas externos de una disfunción interna, es herencia de la psiquiatría, y conserva los supuestos del modelo médico. La idea de “enfermedad mental” es cuestionable por aplicar metafóricamente un término médico (la enfermedad) a un contexto psicológico, en el cual el concepto resulta inapropiado. Como señala Freixa i Baqué (2003), cuando se infiere un “trastorno” o una “psicopatología” (por ejemplo, “fobia”) a partir de los síntomas, la única prueba de su existencia es la presencia misma de esos síntomas, mientras que cuando un médico infiere la tuberculosis a partir de los síntomas, existe otra prueba independiente de la existencia de la tuberculosis: el bacilo de Koch, que se considera la causa del conjunto de síntomas que resumimos con la etiqueta de “tuberculosis”. La diferencia es que en la fobia hay sólo dos términos (síntomas y etiqueta), mientras que en la tuberculosis hay un tercer término (el bacilo de Koch). La tuberculosis y la fobia son sólo etiquetas para resumir síntomas o conductas, pero de ningún modo pueden ser consideradas la causa de dichos síntomas. Sin embargo, a veces este razonamiento aparece implícito en el diagnóstico de trastornos: alguien realiza tales conductas porque “tiene” fobia o

porque “es” fóbico. El bacilo de Koch no es simple inferencia de los síntomas: posee una existencia independiente de los síntomas que produce. Esto no ocurre con la mayoría de los trastornos propuestos por el DSM, que no pueden identificarse en ausencia de los criterios sintomáticos y conductuales, y por lo tanto, no tienen otra función que ser etiquetas descriptivas, a veces demasiado inexactas o imprecisas.

Una posible solución a estas críticas podría consistir en el desarrollo de nuevos sistemas de evaluación (quizás complementarios) que posean mayor utilidad clínica que las actuales clasificaciones (fomentar una “psicopatología o evaluación basada en la evidencia”), mediante la incorporación de categorías funcionales (Follette y Houts, 1996), y de sistemas de evaluación dimensionales. Wulfert, Greenway y Dougher (1996) proponen usar las categorías para mantener la comunicación con otros profesionales que utilizan este sistema diagnóstico, pero complementándolo con un análisis funcional que, a diferencia del diagnóstico sindrómico, sirva como guía para la elección del tratamiento. El DSM puede ser útil para agrupar los problemas según descripciones (a nivel descriptivo) para que los investigadores puedan acotar un área de estudio, pero es necesario reconocer sus deficiencias (no incluir todas las problemáticas tratadas por los psicólogos, sus categorías son heterogéneas en cuanto a la precisión, validez y confiabilidad, etc.).

También es posible clasificar los diversos objetivos que exceden la mera reducción de síntomas (crecimiento personal, ampliación del conocimiento de sí mismo y del mundo, construcción de nuevas formas de entender ciertas cuestiones, búsqueda y exploración de soluciones a problemas personales, reflexionar sobre la propia vida...). Ciertos problemas y objetivos pueden tener una solución objetivamente mejor que otras, mientras que en otros casos, podrían tener diversas soluciones posibles, sin criterios universales para imponer una de ellas como la “mejor”. Si el objetivo es reducir los episodios de ataques de pánico, que la reducción sea lo más significativa y duradera, y que el tratamiento sea lo más breve y eficaz, supone que haya un criterio objetivo y empírico para hacer un “ranking” de las opciones. Pero si el objetivo es que una persona reflexione sobre su proyecto de vida y su forma de relacionarse con los demás, y que esa reflexión le sirva para buscar las opciones con las que se sienta mejor, entonces hay más de una forma posible de cumplir el objetivo, y no tiene sentido fiscalizar formas “mejores” en función de una estadística acerca de qué consideraron “mejor” otras personas (en este caso quizás sea el paciente mismo quien tenga derecho a juzgar según sus propios valores o intereses). Los problemas de vida no son “enfermedades” o “trastornos” para los cuales se precisa un “tratamiento”. Un paciente podría preferir salvar su matrimonio o divorciarse, fortalecer su heterosexualidad o aceptar su bisexualidad, privilegiar objetivos financieros o artísticos. ¿Cómo se evaluarían tales cambios?

Fuera del problema de la pluralidad de valores e intereses, hay otro aspecto importante en la evaluación de resultados: la necesidad de utilizar múltiples mediciones de los resultados y según múltiples criterios. En síntesis, puede recomendarse que la próxima generación de estudios aborde el tema de los

objetivos en un análisis de múltiples niveles: evaluar cambios conductuales y experienciales, reconocer grados en los efectos (un tratamiento puede no sólo retornar al funcionamiento normal, sino exceder los niveles previos), y complementar diversas formas de medición (no solo de síntomas sino también de calidad de vida, incremento de habilidades, etc.).

3.3. Críticas basadas en que faltan datos para optar entre individualización y estandarización del tratamiento

El tratamiento manualizado se criticó por enfatizar la técnica a expensas del análisis y tratamientos individualizados, ya que sin una comprensión de los mecanismos de cambio, los clínicos se ven obligados a aplicar los tratamientos de forma rígida. Los manuales tienen distintas funciones: como guías para la práctica y el entrenamiento y formación, como operacionalización para la investigación en eficacia, y como criterios que determinan la financiación por parte de los sistemas públicos de salud. La función principal de la manualización es estandarizar las intervenciones y controlar variables extrañas, minimizando la variabilidad entre condiciones experimentales. Previo a la manualización, las diferencias entre un terapeuta y otro pueden ser tan grandes que no habría razón para considerar que aplican la misma técnica. Una ventaja secundaria de la manualización es su utilidad para el entrenamiento de profesionales facilitada por la clara definición de la conducta del terapeuta. La manualización es útil para la validez interna pues permite operacionalizar los tratamientos. Según los estudios existentes, la adherencia al manual se asocia en distintas formas con el resultado, a veces positivamente, otras negativamente, y en otros casos no se encuentra correlación (Castonguay, Goldfried, Wisner, Raue y Hayes, 1996; Feeley, DeRubeis y Gelfand, 1999; Henry, Strupp, Butler, Schacht y Binder, 1993).

Una solución para este tipo de críticas es investigar los mecanismos de cambio en lugar de “tipos” o “marcas” de terapia (Rosen y Davison, 2003; Fronjan, 2011). Plaud, Eifert y Wolpe (1998) señalan que la individualización del tratamiento y el uso de manuales no son necesariamente incompatibles. Eifert, Schulte, Zvolensky, Lejuez y Lau (1997) propusieron usar los manuales de una forma flexible y guiada por la teoría y por reglas de decisión empíricamente validadas. De hecho, un creciente número de manuales proponen la implementación flexible, y ya hay, en la actualidad, muchos aspectos individualizados en los manuales (por ejemplo, la construcción de las jerarquías de ansiedad). En el estado actual del conocimiento, ni una aplicación rígida de manuales ni un rechazo total a su manualización parecen deseables. Una posibilidad sería usar manuales estandarizados para casos simples de problemas para los cuales se conoce un TAE, y tratamientos más individualizados para casos de problemas múltiples y problemas no descritos por categorías del DSM. Otra posibilidad sería flexibilizar los manuales, subdividirlos en módulos que se aplicarían según la evaluación del caso, y continuar investigando modelos específicos para cada clase de conducta problema, que puedan guiar las intervenciones de forma flexible y focalizada en el caso particular. Quizás el mayor desafío actual consista en modificar los tratamientos manualizados para individualizar las intervenciones, basándose en principios más que en paquetes.

Los manuales pueden ser vistos como una fuente de restricción indeseable, pero también como una fuente de ayuda, que permite especificar, transmitir y acumular el conocimiento sobre un área. La flexibilidad no debería implicar una pérdida de eficacia, sino todo lo contrario: una mayor sensibilidad a los factores individuales de los tratamientos, y una mejor comprensión de los principios de cambio (en vez de focalizarse únicamente en técnicas y procedimientos). Por estas razones, es deseable que los manuales o guías sean cada vez más precisas, y a la vez más flexibles y “amigables” para el terapeuta. La búsqueda del punto óptimo de rigidez y flexibilidad en los manuales deberá evaluarse según la relación coste-beneficio en cada caso, en términos de fiabilidad científica y generalización de resultados en los contextos aplicados (Moriana y Martínez, 2011a).

3.4. Críticas basadas en que faltan datos sobre la generalización de los tratamientos

Existen dudas sobre la generalización de algunos TAEs a los contextos naturales, debido a diversas cuestiones (la artificialidad de los RCTs, las exclusiones por comorbilidad, el hecho de que hay poblaciones, problemas y trastornos aún no estudiados, etc.). La mayoría de los TAEs son paquetes heterogéneos en los que no se sabe qué funciona y por qué. En muchos casos, no hay explicaciones del mecanismo causal (cómo y por qué las intervenciones producen cambio), y no se sabe a qué componentes de un paquete (p.ej. reestructuración cognitiva, habituación, reducción del estrés, cambio en la esperanza) se atribuye el efecto. Pero es importante recordar que los problemas de generalización también afectan a la experiencia clínica: ¿cómo generalizar a un nuevo paciente la experiencia obtenida con pacientes previos? El argumento de la generalización es un arma de doble filo: si el paciente es único, no está claro cómo el terapeuta podría generalizar la experiencia clínica que obtuvo con sus clientes anteriores.

A menudo se han sometido a contrastación paquetes amplios de tratamiento, sin identificar sus mecanismos causales. La configuración del paquete es previa al testeo de hipótesis, se realiza en base a escasos datos, por lo cual es casi una cuestión de fe aceptar que el paquete elegido es superior a la infinidad de variantes posibles (O'Donohue y Yeater, 2003), incluido el propio paso del tiempo. Esto lleva a formular dudas acerca de cuáles son los mecanismos causales del tratamiento. La polémica respecto al *Eye Movement Desensitization and Reprocessing* EMDR se debe a que el movimiento ocular parece ser inocuo, y el paquete completo parece ser menos efectivo que la exposición sin agregados (Davidson y Parker, 2001). Del mismo modo, podríamos preguntarnos si otros tratamientos “funcionan” por lo que dicen sus teorías (Westen, Novotny y Thompson-Brenner, 2004b).

Otro aspecto cuestionable es la desvinculación entre el contexto aplicado, la investigación de la eficacia y la investigación básica, ya que son las teorías sobre mecanismos causales las que han permitido generar nuevos tratamientos eficaces, predecir la posibilidad de que una intervención funcione, y proponer cambios en los tratamientos. El desencuentro entre el mundo aplicado y el investigador es mayor de lo deseable (Kazdin, 2008).

La solución a estas críticas no implica rechazar las revisiones selectivas, sino mejorarlas. La próxima generación de estudios podrían analizar procedimientos (técnicas, estrategias) y teorías del cambio (principios y mecanismos), que constituyan unidades de análisis más útiles que paquetes prescriptivos para saber qué es útil, perjudicial o inocuo en cada guía de tratamiento, para realizar cambios que aumentarán la eficacia (Westen, Novotny y Thompson-Brenner, 2004b). Una clasificación basada en mecanismos causales permitiría reducir el número de componentes de los paquetes. El nivel de análisis de los paquetes es demasiado abstracto e inespecífico para ser útil como variable causal, pero resulta posible investigar otras unidades de análisis en los tratamientos: las intervenciones particulares, diseños experimentales de caso único, las variables más relevantes del paquete, los procesos y mecanismos que explican su eficacia.

3.5. Críticas basadas en que faltan revisiones de otras áreas y metodologías de investigación

Existen propuestas de que se realicen revisiones de otras cuestiones (además de la eficacia de los tratamientos) que también resultan relevantes para los psicólogos clínicos. Ya existen revisiones de algunas de cuestiones tales como los efectos de la relación terapéutica (Norcross, 2002), y los principios de cambio con apoyo empírico (Castonguay y Beutler, 2005). No existe una oposición entre relación terapéutica y técnicas, ya que la relación terapéutica está compuesta por conductas recíprocas entre el paciente y el terapeuta, y por tanto, pueden aplicarse en ellas las técnicas de análisis y modificación de conducta (Kohlenberg y Tsai, 1991).

También sería importante obtener una lista de “tratamientos empíricamente invalidados” (Beutler, 1998; Lilienfeld, 2007) y de riesgos secundarios de los tratamientos eficaces, si existieran. Aunque tal lista sería aún más compleja que la de TAEs, es parte de la responsabilidad de la disciplina informar a los actores involucrados sobre estas cuestiones.

Varias propuestas apuntan a incorporar información obtenida por metodologías diferentes a la comparación de grupos. Si bien la metodología de comparación de grupos tiene muchos aspectos valiosos y deseables, también posee sus limitaciones. El trabajo con promedios de grandes grupos es poco adecuado para registrar con suficiente detalle las variaciones conductuales. Cuando se informa de un efecto promedio, suele olvidarse que muchos (a veces la mayoría) muestran el efecto opuesto. El diseño de comparación de grupos supone que las diferencias se equilibran por la distribución aleatoria y que el promedio elimina la variación inherente al fenómeno estudiado. Pero la variabilidad es un objeto válido de estudio: lo que se elimina es precisamente lo más importante para ser investigado. La investigación aplicada se focaliza en el caso individual (cuál es el mejor tratamiento para este paciente), y los promedios (además de muestras pequeñas) no son suficientemente informativos, aunque sugieren posibles relaciones causales. En la investigación de TAEs aún queda por investigar en detalle quién se beneficia y quién no en cada tratamiento.

Aún ampliando el uso de pruebas clínicas para hacer análisis de componentes de paquetes de tratamiento (Jacobson y Christensen, 1996), es necesario utilizar otros diseños: para abordar relaciones entre intervenciones y respuestas (investigación

de proceso), para desarrollar nuevos y mejores tratamientos (paradigmas para el desarrollo de tratamiento), o para investigar casos particulares (área en la que los clínicos podrían jugar un rol central en el diseño y la realización de estudios experimentales de caso único y estudios de caso de tipo cualitativo). Varias propuestas actuales sugieren priorizar el descubrimiento de principios básicos de cambio y continuar la investigación de poblaciones y tratamientos que aún están escasamente estudiados (niños, pacientes con trastornos de personalidad, comorbilidad, terapias familiares, maritales y/o grupales, casos inferiores y superiores al umbral impuesto por las categorías del DSM). Una cuestión central es investigar no sólo qué tratamientos funcionan, sino cómo funcionan los tratamientos: comparar componentes de los paquetes (en forma individual o variando las combinaciones), focalizar no tanto en tratamientos para trastornos, sino en procesos terapéuticos para problemas funcionalmente definidos. Para ello se pueden realizar investigaciones de proceso, de “mini-resultados” dentro de cada sesión, por ejemplo utilizando análisis de secuencias conductuales (una estrategia que ya se ha aplicado exitosamente al estudio de las interacciones de pareja o de maestros-alumnos), y también podría aplicarse la metodología de comparación de grupos para evaluar intervenciones o principios específicos, abordajes generales al tratamiento y moderadores del resultado, en lugar de paquetes completos.

Un problema usual es que se adopta un modelo unidireccional según el cual los investigadores producen conocimiento y los clínicos lo aplican. Este modelo aumenta la brecha investigación-práctica, dado que ignora los aportes de la clínica, y no apuesta por la coordinación y *feedback* entre ambas (Mustaca, 2004; Kazdin, 2008). En la mayoría de las ocasiones, la experiencia podría acumularse y utilizarse para generar y comprobar hipótesis si existiera más trabajo conjunto entre los investigadores y los clínicos. También habría notables mejoras si detrás de cada servicio asistencial hubiera investigadores que facilitaran el análisis y evaluación de resultados, independientemente de que fuera un estudio de caso usual (descripción del terapeuta, sin control ni evaluación sistemática), un estudio de caso cuasi-experimental (con evaluación sistemática pero sin control), un diseño experimental de caso único (con evaluación y control) o un diseño de comparación de grupos de pacientes con los mismos diagnósticos y/o características.

El diseño experimental de caso único no implica que sólo haya un participante, sino que todas las variables dependientes se registran repetidamente en el curso del experimento, y esos datos no se combinan con los de otros participantes para producir promedios de grupo (Barlow y Hersen, 1988). En estos diseños, los participantes son sus propios controles, y las comparaciones se realizan a través de condiciones experimentales. Las diferencias individuales no contribuyen a la variación, porque no se realizan comparaciones a través de los participantes. Para evaluar la efectividad de una intervención en un individuo, se compara la conducta en fases pre y post-intervención, y no cuánto se desvía la conducta del participante respecto al promedio grupal. En psicología, educación y otras áreas aplicadas, el objetivo principal es el cambio de conductas de un individuo, y no el cambio de promedios grupales, y esto implica reconocer las ventajas de los diseños de caso

único (Barlow y Hersen, 1988), a veces no suficientemente reconocida. Esto no implica una oposición entre estas metodologías, sino aspectos complementarios en función de características como la disponibilidad de sujetos, costes y beneficios, y de las preguntas que intenta responder el investigador en cada estudio, etc. Si bien se están obteniendo datos sobre estas cuestiones, todavía no está claro de qué modo los psicólogos podrían trasladar a la práctica clínica esta acumulación de información. Esto nos lleva al último grupo de críticas, referidas a las dificultades del traslado.

3.6. Críticas basadas en que los listados de tratamientos resultan insuficientes para el traslado de la información a los actores involucrados

Como ya comentamos anteriormente, el rol mediador de los listados implica básicamente dos etapas: la revisión selectiva de la evidencia mediante criterios explícitos, y el traslado de las conclusiones a la práctica clínica. Los listados se pensaron para cumplir una función descriptiva, dirigida a intentar detectar qué tratamientos cumplen un mínimo nivel de evidencia. La *APA Task Force* se limitó a divulgar las conclusiones y a recomendar la inclusión de los TAEs en la formación de los psicólogos clínicos pero, en ocasiones, las instituciones de servicios de salud plantearon la aplicación de TAEs como condición para la financiación de los tratamientos, lo cual generó la oposición de muchos psicólogos clínicos.

Para cumplir una función prescriptiva, es necesario responder diversas preguntas sobre cómo un psicólogo debería realizar la evaluación de cada paciente y cómo debería seleccionar (o diseñar) el tratamiento más adecuado en función de la información particular (la evaluación de ese paciente) y general (las revisiones de estudios publicados sobre casos similares), y capacitar a los psicólogos para que puedan aplicar los tratamientos en forma eficaz (Moriana y Martínez, 2011b). Ante la ausencia de un análisis adecuado de estas cuestiones, muchos psicólogos consideran que un listado de TAEs es insuficiente para abordar los problemas reales de la práctica clínica. Es evidente que, en muchas ocasiones, para entrenar a un terapeuta no basta solo con un listado de TAEs, un manual y un *workshop*. Un manual es condición necesaria pero no suficiente: el tratamiento eficaz requiere de un entrenamiento de los terapeutas igualmente eficaz (incluso en ocasiones la supervisión de casos).

La solución de los problemas del traslado de los TAEs a la práctica clínica requiere implementar tanto guías como cursos de entrenamiento, y comprobar empíricamente la efectividad educativa de ambos recursos. Herbert (2003) propone que el objetivo de la PBE se logra mejor con el desarrollo y difusión de guías, y que las listas de TAEs son innecesarias para ese proceso. Una guía podría incluir no sólo información sobre eficacia de tratamientos para trastornos, sino sobre aspectos axiológicos y praxiológicos, estrategias de evaluación, categorías diagnósticas con utilidad clínica, tratamientos perjudiciales (Lilienfeld, 2007), cuestiones sobre la relación terapéutica (Norcross, 2002), principios de cambio (Castonguay y Beutler, 2005), mecanismos causales, moderadores, micro-intervenciones, dificultades usuales, aspectos etiológicos, etc.

Las guías permitirían utilizar criterios de evidencia más dinámicos y complejos, en los cuales el umbral de “mejor evidencia” dependa del estado de investigación de cada área. Los RCT se basan en tratamientos muy estructurados, poblaciones cuidadosamente seleccionadas y aleatorizadas, y resultados medidos con exactitud. Éstas son ventajas para la validez interna, pero implican un riesgo para la validez externa y dejan sin evaluar mucha información relevante. Por ello, deben complementarse de forma planificada con otras metodologías. Resulta difícil consensuar una línea base universal para definir la eficacia, por esta razón, se puede utilizar la “mejor evidencia” para cada problema según diferentes criterios y distintas metodologías. La pregunta “¿qué tratamiento funciona?” se contestaría fundamentalmente con RCTs, pero hay otras preguntas relevantes. Para la pregunta “¿qué puedo hacer para ayudar a este paciente, con este problema, en este momento?” quizás resultara más adecuado contestarlas con diseños de caso único. Para la pregunta “¿cómo se relaciona el vínculo terapéutico y el resultado?” se pueden usar estudios correlacionales. De este modo, se pueden responder las distintas cuestiones relevantes para la clínica: ¿cómo definir el problema (investigación en psicopatología, estudios de validez predictiva)?, ¿qué tratamiento es eficaz/efectivo/eficiente (estudios de resultados)?, ¿qué componentes tienen efectos (estudios de componentes)?, ¿cómo se jerarquizan los tratamientos rivales (estudios comparativos)?, ¿qué tratamientos son perjudiciales o ineficaces (meta-análisis)?, ¿qué mecanismos básicos explican los efectos (investigación básica)?. Algunas guías elaboradas por el NICE están teniendo en cuenta estos aspectos informando sobre el grado de evidencia y el tipo de estudios que los sustentan.

El pluralismo metodológico no implica considerar que toda evidencia es equivalente, sino que pueden clasificarse jerarquías de evidencia tal y como realmente ocurre en los listados APA y del NICE, aunque no sean totalmente coincidentes (Moriana y Martínez, 2011a). Las guías no se limitan a un estándar único para establecer el apoyo empírico. Toda metodología relevante puede usarse, teniendo en cuenta sus alcances y límites.

Las guías no requieren asumir una restricción a las categorías del DSM, ni al objetivo de eliminar los síntomas. Se pueden explorar otras categorías más útiles (grados de malestar, calidad de vida, adquisición de habilidades), basadas en los mecanismos y en la utilidad para predecir la eficacia diferencial de los tratamientos, y ampliar el análisis de otros objetivos (utilizando mediciones múltiples de cambios conductuales y experienciales, incluyendo auto-informes y estudios cualitativos). También puede incluirse el estudio de los valores (axiología). En este sentido, Barlow distingue entre tratamientos psicológicos (abordaje de “trastornos”, cuyo objetivo es la “eliminación de síntomas”) y psicoterapias (abordaje de “problemas de vida”, cuyo objetivo es el “crecimiento personal”). También podría incluirse en el debate la excesiva medicalización de los problemas en ambas perspectivas (tratamientos combinados). La reducción de síntomas suele ser más fácil de medir, pero hay otros resultados valiosos que pueden requerir formas de investigación de la experiencia subjetiva (métodos cualitativos...). A esta propuesta de Barlow, quizás fuera interesante añadir el tratamiento de problemas “puntuales”, que no

son trastornos, pero que tampoco requieren un tratamiento psicoterapéutico a largo plazo, ya que esta es una de las mayores demandas que tienen actualmente los psicólogos que trabajan en contextos aplicados y que asumen con la práctica del consejo psicológico breve, *counseling* y distintas formas de asesoramiento u orientaciones.

La información de las guías de tratamiento se debería trasladar a un formato útil para el clínico (por ejemplo, algoritmos o árboles de decisión para la elección de intervenciones). Las guías deberían testarse en relación a su aceptabilidad, practicidad, y uso. Podrían crearse instituciones mediadoras, conformadas por investigadores y psicólogos clínicos, que se encarguen de actualizar y corregir en forma constante tanto las guías como los cursos de entrenamiento. Propuestas muy interesantes en esta línea pueden observarse en el NICE, creado como institución independiente del sistema sanitario público del Reino Unido, y encargado de diseñar y evaluar guías de intervención en el ámbito sanitario. Estas instituciones deberían velar por la objetividad, por intentar que todos los colectivos estuvieran representados en el proceso de diseño y evaluación de guías y evidencias (usuarios, profesionales, investigadores, gestores...).

De esta forma, las guías podrían evitar estar sesgadas por intereses políticos (por ejemplo, las guías actuales de la Asociación Psiquiátrica Americana están sesgadas a favor de la farmacoterapia, incluso en casos en los que la psicoterapia es preferible a largo plazo), y por desacuerdos entre psicoterapeutas, pero la evidencia seguirá siendo necesaria para evaluar los tratamientos. Se puede controlar que las guías se basen en la evidencia usando revisores externos sin conflictos de intereses de distintas disciplinas (psicología y psiquiatría), y que representen a distintas teorías.

Otra propuesta consiste en diseñar manuales que permitan la individualización del tratamiento. Pueden incluir distintos módulos o técnicas que se aplican en función de las características de cada paciente. Sería posible manualizar la evaluación conductual y generar categorías funcionales que orienten la elección del tratamiento. Esta línea de investigación podría basarse en propuestas ya existentes como el análisis funcional (detección de variables que influyen en la conducta de un individuo).

4. Conclusiones

Desde la Conferencia Boulder y desde múltiples foros y autores se ha luchado por lograr una sinergia entre investigación científica y práctica clínica, pero la brecha permanece (Kazdin, 2008). Hay una carencia generalizada de formación y actualización en investigación relevante para los clínicos y una falta de realidad asistencial y práctica clínica de los investigadores. Además, la mayoría de las psicoterapias no surgieron en un contexto de investigación controlada, ni fueron testadas en tales condiciones. El movimiento de TAEs se enfrentó al problema de cómo debería ser la conexión entre investigación y práctica, generando un modelo posible. A pesar de sus limitaciones, los TAEs han significado un paso adelante respecto al aislamiento de escuelas, a la ausencia de investigación, y a una autocomplacencia o negativa de diversas tradiciones teóricas para comprobar

sus propios supuestos. El movimiento TAE colocó la evaluación empírica en un primer plano, y fomentó que los proponentes de abordajes poco investigados comenzaran a realizar investigación más rigurosa.

A pesar de sus virtudes, el análisis de los problemas de los listados de tratamientos eficaces nos permite avanzar en el conocimiento explorando posibles soluciones. De este modo, hemos analizado algunas de las críticas principales, y algunas propuestas formuladas para su resolución: el énfasis en la conexión con la ciencia básica y la explicación teórica de los mecanismos y principios de cambio, la selección de objetivos basada en los valores de los individuos y grupos (además de la mera evaluación de síntomas), la selección de las intervenciones basadas en categorías funcionales y en el conocimiento de principios de cambio (como alternativa al diagnóstico sindrómico), la comprobación del tratamiento mediante diseños de caso único (modelo clínico-experimental), la combinación de metodologías en función de los objetivos de investigación, el diseño de guías de tratamiento manualizadas pero flexibles y adaptadas a cada caso, el estudio del entrenamiento y la supervisión de psicólogos, etc.

En este sentido, llegamos a la conclusión de que el listado de TAEs actual es insuficiente para cumplir una función prescriptiva en el entrenamiento profesional y en la elección del tratamiento. Sin embargo, el debate sobre TAEs en un contexto de pensamiento crítico nos llevaría a valorarlo de forma positiva por presentar, en la actualidad, más aspectos positivos que negativos, y ser, hoy por hoy, la mejor propuesta metodológica de la psicología basada en la evidencia para el diseño, evaluación y elección de tratamientos psicológicos eficaces.

Los cambios necesarios se dirigen a identificar principios y mecanismos de cambio, abandonar el modelo médico y buscar modelos basados en principios psicológicos, comparar costes-efectividad y significación del tratamiento, investigar modelos efectivos de entrenamiento y supervisión para mantener la efectividad, y proteger (o al menos informar) de las intervenciones potencialmente perjudiciales. También sería conveniente seguir refinando las metodologías aplicadas y consensuar los niveles o grados de evidencia. Abrir el debate de estos temas implica revisar y cuestionar los supuestos propios y ajenos, y desarrollar los principios de una disciplina psicológica científica.

REFERENCIAS BIBLIOGRÁFICAS

- American Psychological Association APA (2006). Report of the 2005 Presidential Task Force on Evidence-Based. *American Psychologist*, 61, 271-285.
- Barlow, D. H., y Hersen, M. (1988). *Diseños experimentales de caso único*. Barcelona: Martínez Roca.
- Beutler, L. E. (1998). Identifying empirically supported treatments: What if we didn't? *Journal of Consulting and Clinical Psychology*, 66, 113-120.
- Beutler, L. E. (2002a). The Dodo Bird is extinct. *Clinical Psychology: Science and Practice*, 9, 30-34.
- Beutler, L. E. (2002b). It isn't the size, but the fit. *Clinical Psychology : Science and*

Practice 9, 434-438.

- Beutler, L., Castonguay, L.G. (2005). *Empirically defined principles of therapeutic change*. Oxford University Press.
- Castonguay, L. G., Goldfried, M. R., Wisner, S. L., Raue, P. J. y Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology*, *64*, 497-504.
- Castonguay, L.G. y Beutler, L. E. (Eds) (2005). *Principles of therapeutic change that work*. New York: Oxford University Press.
- Cautilli, J. (2006). Dodo-Bird or dressing up the pig and the politics of psychotherapy: a reply to Vandenberghe and de Sousa. *International Journal of Behavioral Consultation and Therapy 2*, p. 305.
- Chambless, D. y Ollendick, T. (2001). Empirically Supported Psychological Interventions: Controversies and Evidence. *Annual Review of Psychology*, *52*, 685-716.
- Chambless, D. (2002). Beware the Dodo Bird: the dangers of overgeneralization. *Clinical Psychology: Science and Practice 9*, 13-16.
- Chambless, D., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., Crits Christoph, P....Woody, S.R.. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, *51*, 3-16.
- Chambless, D., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P....McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49*, 5-14.
- Chorpita, B.F., Daleiden, E.L., Ebesutani, C., Young, J., Becker, K.D., Nakamura, B.J.,...Starace, N. (2011). Evidence-based treatments for children and adolescents: An update review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice*, *18*, 154-172.
- Crits-Christoph, P. (1996). The dissemination of efficacious psychological treatments. *Clinical Psychology: Science and Practice*, *3*, 260-263.
- Crits-Christoph, P., Frank E., Chambless D. L., Brody C. y Karp J. F. (1995). Training in empirically- validated treatments : What are clinical Psychology students learning? *Professional Psychology: Research and Practice*, *26*, 514-22.
- Davidson, P. R. y Parker, C. H. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of Consulting and Clinical Psychology*, *69*, 305-316.
- Echeburúa, E., Salaberría, K., de Corral, P., Polo-López, R. (2010). Terapias Psicológicas Basadas en la Evidencia: Limitaciones y retos de futuro. *Revista Argentina de Clínica Psicológica*, *3*, 247-256.
- Ehlers, A., Bisson, J., Clark, D.M. y Yule, W. (2010). Do all psychological treatments really work the same in posttraumatic stress disorder? *Clinical Psychological Review*, *30*, 269-276.
- Eifert, G. H., Schulte, D., Zvolensky, M. J., Lejuez, C. W. y Lau, A. W. (1997).

- Manualizing behavior therapy: Merits and challenges. *Behavior Therapy*, 28, 499-509.
- Feeley, M., DeRubeis, R. J. y Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 67, 578-582.
- Follette, W. C. y Houts, A. C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM. *Journal of Consulting and Clinical Psychology*, 64, 1120-1132.
- Freixa I Baqué, E. (2003). ¿Qué es conducta? *International Journal of Clinical and Health Psychology*, 3, 595-613.
- Frojan, M.X. (2011). ¿Por qué funcionan los tratamientos psicológicos? *Clínica y Salud*, 22, 201-204.
- Goldfried, M. y Wolfe, B. (1996). Psychotherapy practice and research: Repairing a strained alliance. *American Psychologist*, 51, 1007-1016.
- Hayes, S.C. y Follette, W. C. (1992). Can functional analysis provide a substitute for syndromal classification? *Behavioral Assessment*, 14, 345-365.
- Henry, W.P., Strupp, H.H., Butler, S.F., Schacht, T.E. y Binder, J.L. (1993). Effects of training in time limited dynamic psychotherapy: Changes in therapist behavior, *Journal of Consulting and Clinical Psychology*, 61, 434-440.
- Herbert, J. D. (2003). The Science and Practice of Empirically Supported Treatments. *Behavior Modification* 27, 412-430.
- Jacobson, N. S. y Christensen, A. (1996). Studying the effectiveness of Psychotherapy. How well can clinical trials do the job? *American Psychologist*, 51, 1031-1039.
- Kazdin, A.E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146-159.
- Kohlenberg, R. J. y Tsai, M. (1991). *Functional Analytic Psychotherapy: A guide for creating intense and curative therapeutic relationships*. New York: Plenum.
- Lilienfeld, S. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53-70.
- Moran, D. y Malott, R. (2004). *Evidence-Based Educational Methods*. Elsevier.
- Moriana, J.A. y Martínez, V. (2011a). La psicología basada en la evidencia y el diseño y evaluación de tratamientos psicológicos eficaces. *Revista de Psicopatología y Psicología Clínica*, 16, 81-100.
- Moriana, J.A. y Martínez, V. (2011b, Noviembre 2). ¿Cómo se eligen, diseñan y evalúan los tratamientos e intervenciones psicológicas? *Infocop Online*. Recuperado desde http://www.infocop.es/view_article.asp?id=3626&cat=
- Mustaca, A. E. (2004). Tratamientos psicológicos eficaces y ciencia básica. *Revista Latinoamericana de Psicología*, 36, 11-20.
- Norcross, J.C. (2002). *Psychotherapy relationships that work*. New York: Oxford

University Press.

- O'Donohue, W. y Yater, E. A. (2003). Individuating psychotherapies. *Behavior Modification*, 27, 313-321.
- Plaud, J. J., Eifert, G. H. y Wolpe, J. (1998). The Role of Theory in Behavior Therapy: Conceptual and Practical Conclusions. En Plaud & Eifert (Eds.). *From Behavior Theory to Behavior Therapy*. Allyn & Bacon.
- Rosen, G.M. y Davison, G.C. (2003). Psychology Should List Empirically Supported Principles of Change (ESPs) and Not Credential Trademarked Therapies or Other Treatment Packages. *Behavior Modification* 27, 300-312.
- Sackett, D.L., Richardson W.S., Rosenberg, W. y Haynes, R.B. (1997). *Evidence-Based Medicine*. New York: Churchill Livingstone.
- Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P. y Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, 71, 963-972.
- Wampold, B. E. (1997). Methodological problems in identifying efficacious psychotherapies. *Psychotherapy Research*, 7, 21-43.
- Westen, D., Novotny, C. M. y Thompson-Brenner, H. (2004a). The Empirical Status of Empirically Supported Psychotherapies: Assumptions, Findings, and Reporting in Controlled Clinical Trials. *Psychological Bulletin*, 130, 631-663.
- Westen, D., Novotny, C. M. y Thompson-Brenner, H. (2004b). The Next Generation of Psychotherapy Research: Reply to Ablon and Marci (2004), Goldfried and Eubanks-Carter (2004), and Haaga (2004). *Psychological Bulletin*, 130, 677-683.
- Wulfert, E., Greenway, D. E. y Dougher, M. J. (1996). A logical functional analysis of reinforcement-based disorders: Alcoholism and pedophilia. *Journal of Consulting and Clinical Psychology*, 64, 1140-1151.