

La aplicación de modelos lineales jerárquicos para el estudio de la eficacia en psicoterapia.

Gómez Penedo, Juan Martín^{a, d}, Muiños, Roberto^{a, b}, Hirsch, Pablo^c, y Roussos, Andrés^{a, d}

Artículo Metodológico

Resumen

Abstract

Tabla de Contenido

Los modelos lineales jerárquicos (HLM) representan una estrategia estadística fundamental para la investigación en psicoterapia, ya que permiten superar la dependencia de las observaciones que habitualmente se presenta en sus datos. Estos métodos son útiles para estimar el cambio, desagregar fuentes de variación y analizar efectos de predictores de distintos niveles de jerarquía. Debido a que la aplicación de estos métodos requiere de un alto grado de conocimiento técnico, aún inaccesible para muchos investigadores, el objetivo de este trabajo es presentar una guía para entender, aplicar y reportar los HLM para estudiar los efectos de la psicoterapia. Para ilustrar cómo aplicar y reportar los HLM hemos utilizado una base de datos clínica real. Diseminar estos métodos en Latinoamérica puede representar una contribución tanto para la investigación como para la práctica, mejorando la solidez de los estudios clínicos y desarrollando un conocimiento robusto para optimizar los procesos y resultados en psicoterapia.

Using Hierarchical Linear Models to study psychotherapy efficacy. Hierarchical Linear Models (HLM) represents a valuable statistical tool for psychotherapy research, given that they allow dealing with the usual dependency presented in its data. These methods are useful to estimate change, disaggregate sources of variations, and analyze the effect of different level predictors. Considering that, these analyses required a highly sophisticated technical knowledge that might remain inaccessible for many researchers, the aim of this paper is to present a guide on how to understand, apply, and report HLM for psychotherapy effects research. To illustrate how to apply HLM, we have drawn on a naturalistic clinical dataset. Disseminating these methods in the Latin-America might represent a meaningful contribution both for research and practice, improving the soundness of clinical studies and helping to develop a more robust knowledge that might leads to greater process and outcome in psychotherapy.

Introducción	25
Aplicación de los modelos	30
Síntesis	35
Agradecimiento	35
Referencias	35

Palabras clave:

Modelos lineales jerárquicos, modelos de curva de crecimiento, modelos multinivel, psicoterapia.

Keywords:

Hierarchical linear models, growth curve models, multilevel models, psychotherapy.

Recibido el 25 de junio de 2018; Aceptado el 31 de diciembre de 2018

Editaron este artículo: Ángel Elgier, Paula Abate, Nadia Justel, María Victoria Ortiz y Fernando Luna

Introducción

La investigación en psicología clínica, desde sus comienzos, ha estado orientada a analizar, predecir y explicar el efecto de los tratamientos, mediante el estudio del cambio de los pacientes en el transcurso de los mismos (Gómez Penedo, Juan, Manubens, & Roussos, 2018; Ogles, 2013). Para estudiar los efectos de la terapia las investigaciones basadas en modelos de análisis

cuantitativos han utilizado históricamente métodos clásicos de estadística paramétrica, como es el caso de las correlaciones de Pearson, las regresiones de mínimos cuadrados ordinarios o logísticas, pruebas *t* de Student y Análisis de la varianza (ANOVA) (Atkins, 2005). A pesar de la difusión de estos procedimientos, en las últimas décadas diversos autores han señalado que la

^aUniversidad de Buenos Aires, Buenos Aires, Argentina

^bUniversidad Tres de Febrero, Buenos Aires, Argentina

^cCentro Privado de Psicoterapias, Buenos Aires, Argentina

^dConsejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina

*Enviar correspondencia a: Gómez Penedo, J. M. E-mail: jmgomezpenedo@gmail.com

aplicación de estos métodos implica una serie de problemáticas para la investigación en psicoterapia, ya que la estructura de los datos que habitualmente se utilizan en ese campo, viola un supuesto fundamental de estos análisis: el supuesto de independencia de las observaciones (Baldwin, Imel, Braithwaite, & Atkins, 2014). Para aplicar estas estrategias analíticas, es necesario que cada observación sea independiente de las otras, lo que implica que el comportamiento de una variable, en una determinada observación, no permita predecir su comportamiento en otras observaciones (Finch & Bolin, 2017).

Los datos que se utilizan en la investigación en psicología clínica se encuentran habitualmente *anidados*. Esto significa que están agrupados en estructuras jerárquicas que presentan diferentes niveles, lo que haría esperable la presencia de correlaciones entre las observaciones pertenecientes a un mismo nivel de agrupamiento (Baldwin et al., 2014).

Para ilustrar el fenómeno del anidamiento de datos que habitualmente ocurre en psicología clínica, imaginemos un estudio orientado a analizar los niveles de cambio en una muestra de 100 pacientes que estuvieron expuestos a 20 sesiones de un tratamiento y que completaron una medida de severidad clínica, al principio de la terapia (t1), a mitad del tratamiento (t2) y al finalizar las 20 sesiones (t3). En este caso, es esperable que los puntajes de un paciente en el momento t1, correlacionen con sus propios puntajes en el momento t2 y t3; de esta manera, diremos que las medidas repetidas del instrumento de severidad se encuentran anidadas *dentro* de los pacientes (Raudenbush & Bryk, 2002). De igual manera, los pacientes que conforman esa muestra hipotética, pudieron ser tratados por un número menor de terapeutas (por ejemplo, diez terapeutas que hayan tratado cada uno a diez pacientes de la muestra total). Como los pacientes se encuentran agrupados (es decir, anidados) dentro de un conjunto de terapeutas, también es esperable correlaciones en las observaciones presentes a este nivel (Baldwin et al., 2014). Por ejemplo, que la evolución en los niveles de severidad en los pacientes de un determinado terapeuta tiendan a ser similares entre sí y distinta a los de otros terapeutas.

El uso de métodos estadísticos clásicos frente a una violación del supuesto de independencia de las observaciones, puede llevar a sesgos

importantes en el análisis, como es el caso de una inflación de la probabilidad de cometer un error de Tipo I (Hox & Maas, 2005).

Los modelos lineales jerárquicos (HLM por su sigla en inglés, Raudenbush & Bryk, 2002) surgieron en el contexto de la psicología educacional como una estrategia alternativa para estimar parámetros de regresión en base a datos que se encuentran anidados y organizados en una estructura jerárquica, lidiando con la problemática de la dependencia en las observaciones. Estos métodos, además de superar la limitación de la dependencia de las observaciones, presentan implicancias adicionales para la investigación en psicología clínica, ya que permiten (i) una estimación robusta de las tasas de cambio de los pacientes (los efectos de los tratamientos), (ii) desagregar fuentes de variación en diferentes niveles de la estructura (efectos intra-paciente o dentro de los pacientes, efectos entre pacientes y efectos entre terapeutas), (iii) incluir predictores en cada uno de los niveles de la organización de los datos (variables que cuentan con medidas repetidas, características de los pacientes, características de los terapeutas, etc.) (Raudenbush & Bryk, 2002).

Si bien a nivel internacional los modelos HLM han tenido diversas aplicaciones recientes en el contexto de la investigación en psicoterapia (Ver por ejemplo, Constantino, Westra, Antony, & Coyne, 2017; Zilcha-Mano & Errázuriz, 2015), considerándose en la actualidad los análisis estadísticos de punta en el campo de los estudios en psicología clínica, su aplicación en la región latinoamericana ha sido mucho más limitada. Es posible que esto se deba al hecho de que realizar estos análisis requiere de un conocimiento estadístico y técnico muy sofisticado y específico, que no se encuentra ampliamente difundido en la región mediante cursos de formación o publicaciones que expliquen estos modelos en castellano.

En este contexto, el objetivo de este trabajo es realizar una presentación comprensiva de los modelos HLM para realizar investigación en psicología clínica, facilitando su aplicación en estudios en esta área de conocimiento. Para ilustrar la aplicación de estos modelos y explicar la forma de interpretar sus resultados, se utilizará una muestra clínica proveniente de un centro de psicoterapia, que será explicada en detalle más adelante. Buscamos que este trabajo ayude a

incrementar el conocimiento y disseminación de los modelos HLM entre investigadores de la región latinoamericana, proveyendo un recurso técnico valioso para analizar de manera robusta los datos de investigaciones en psicoterapia, pudiendo establecer sus efectos e identificar predictores asociados.

Características principales de los modelos HLM

Los modelos HLM son una herramienta estadística que le permite al investigador analizar información recolectada en diferentes niveles de organización. Estos modelos están basados en ecuaciones de regresión, pero incorporan una innovación fundamental respecto de los procedimientos de regresión de mínimos cuadrados ordinarios: En vez de tener solamente un término de error en sus ecuaciones, estiman el error (denominado efectos aleatorios) en cada nivel de organización de la información, permitiendo la medida de efectos grupales (Atkins, 2005; Finch & Bolin, 2017). Por ello, estos modelos también se conocen como *modelos mixtos*, ya que estiman efectos fijos (por ejemplo, la media de un grupo en una medida) y efectos aleatorios (el grado en el cual los participantes difieren respecto de esa media). Esta característica es la que le permite a los modelos HLM lidiar con la correlación entre las observaciones (Wang, Xie, & Fisher, 2012). Además, estos modelos son conocidos como *modelos multi-nivel*, debido a que posibilitan desagregar fuentes de variación en base a los niveles jerárquicos identificados (Hofmann, 1997). De esta manera, estiman la proporción de la varianza de una variable dependiente (por ejemplo, el grado de ansiedad de un paciente) que es explicada por los diferentes niveles del modelo (por ejemplo, cambios dentro del paciente, diferencias entre los pacientes y diferencias entre los terapeutas).

Específicamente para el campo de la psicología clínica, los modelos HLM permiten estimar la tasa de cambio de los pacientes, tanto a nivel individual como grupal (Curran & Bauer, 2011). Estos modelos específicos para estimación de cambio, que se denominan modelos de curva de crecimiento (GCM, por su sigla en inglés), admiten identificar y estimar distintas trayectorias posibles de evolución (Baldwin et al., 2014). Por ejemplo, algunos pacientes pueden presentar una

reducción progresiva en sus niveles de severidad (modelo lineal de cambio) mientras que otros podrían manifestar una reducción pronunciada en las primeras sesiones que tienda a desacelerarse hacia un cambio menor con el paso del tiempo (modelo cuadrático de cambio). HLM otorga la posibilidad de estimar la trayectoria que mejor se ajusta a los datos en promedio, así como también las trayectorias individuales de cada uno de los sujetos de la muestra (Curran & Bauer, 2011).

A su vez, a partir de los modelos HLM se pueden ajustar funciones para distintos momentos de medida (modelos *piecewise*), estimando diferentes trayectorias para cada uno de los periodos estudiados (Wang et al., 2012). Esto tiene implicancias particularmente importantes para la investigación en psicoterapia; por ejemplo, cuando se evalúan las tasas de cambio de un paciente durante el tratamiento, en comparación con un período de seguimiento posterior al abordaje, se esperan tasas de cambio diferentes en ambos momentos (Ver Westra, Constantino, & Antony, 2016).

Además, los modelos HLM pueden analizar información desbalanceada (es decir, medida en distintos momentos) e incluir en sus análisis casos que cuentan con información pérdida (Skrondal & Rabe-Hesketh, 2004). Estrategias de análisis clásicas como el ANOVA multi-factorial y el ANOVA de medidas repetidas, no podrían trabajar con datos de esa naturaleza, debiendo ser excluidos los casos que cuentan con datos faltantes, con el riesgo de producir sesgos sistemáticos (Power & Freeman, 2012).

Debajo presentaremos dos modelos básicos para representar las principales fuentes de agrupación de información en los estudios en psicoterapia. La presentación de los modelos se realizará a partir de una base de datos clínica que utilizaremos a fines ilustrativos (Ver Aplicación de los modelos HLM a una base de datos clínica), en donde 1154 pacientes de un centro privado de psicoterapia completaron una medida de auto-reporte respecto de su nivel de bienestar, después de cada una de las primeras cuatro sesiones (semanales) de psicoterapia.

A partir de esta información, primero presentaremos un modelo de dos niveles para estimar el cambio lineal de los pacientes (sesión a sesión), en donde las medidas repetidas de bienestar se anidan (o agrupan) dentro de los pacientes que las completan. Luego

presentaremos un modelo de tres niveles en donde las medidas repetidas del bienestar se agrupan *dentro de los pacientes* que las completan y los pacientes se anidan *dentro de los terapeutas* tratantes. En la [Figura 1](#) se presenta una ilustración de los distintos niveles de anidamiento en una base de datos de esta naturaleza, correspondiente a un modelo de tres niveles

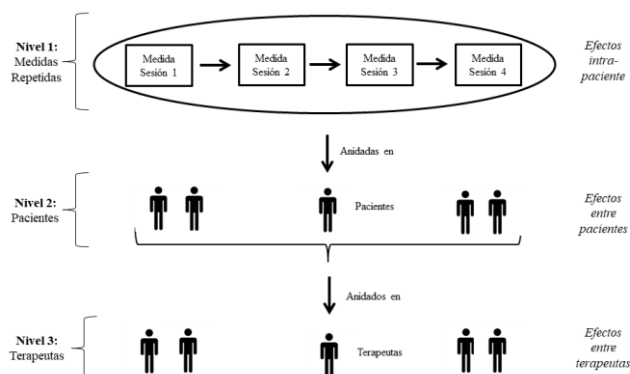


Figura 1. Estructura de anidación de datos en un diseño habitual de investigación en psicoterapia (Modelo de tres niveles).

Modelo de curva de crecimiento de dos niveles: Estimando el cambio del paciente

El modelo HLM básico para estimación del cambio en psicoterapia es un modelo de curva de crecimiento, que cuenta con dos niveles. En el primer nivel de la ecuación del modelo (efectos *dentro de los pacientes*), los puntajes de la variable dependiente que se está estudiando (la medida de resultados) en un paciente, son estimados a partir de dos elementos: (i) una constante que representa el valor de la variable cuando todos los demás predicadores están fijados en cero, y (ii) el tiempo en que se realizó cada medición, definido en semanas o en número de sesiones, que va a representar en su forma más simple a una pendiente lineal (como se ha anticipado, existen otros modelos de trayectorias como el cuadrático). Las estrategias para determinar la posición del cero en la serie temporal se denominan estrategias de centramiento de las variables, y son fundamentales para los análisis y su interpretación, ya que determinan el sentido de la *constante*. Finalmente, este modelo tiene un término de error, que permite que cada paciente

varíe respecto de los puntajes estimados para cada uno de los puntos de medición de la variable dependiente.

Luego, en el segundo nivel del modelo (que representa los efectos *entre los pacientes*), los parámetros que eran predictores en el primer nivel, se convierten ahora en variables de resultado de dos ecuaciones separadas. La *constante* en un paciente (valor de la variable dependiente en el momento = 0) se predice a partir del promedio de puntajes de la población en dicho momento, mientras que el efecto del tiempo en un paciente (la pendiente lineal) se predice mediante el promedio de cambio de la población. Adicionalmente, en el segundo nivel existen términos de error en cada una de las ecuaciones, que le permiten a cada uno de los participantes variar en sus puntajes respecto de los promedios poblacionales de la constante y la pendiente lineal.

Resumiendo, estos modelos utilizan tanto información individual como muestral (datos agregados de todos los participantes) para estimar los parámetros, permitiendo calcular tanto parámetros fijos como aleatorios. Como los modelos HLM estiman tanto coeficientes como errores estándar, para cada efecto fijo o aleatorio, estos análisis brindan pruebas de hipótesis respecto de los parámetros de nivel superior (Atkins, 2005). Además, al desagregar la varianza explicada por cada uno de los niveles del modelo, esto posibilita el cálculo de Coeficientes de Correlación Intra Clase (ICC por su sigla en inglés), para determinar el porcentaje de varianza explicada por cada uno de los niveles de agrupación, en la variable dependiente.

A partir del ejemplo introducido más arriba (medidas repetidas de un instrumento de bienestar en una población de pacientes de psicoterapia), la fórmula de un primer modelo de curva de crecimiento de dos niveles, podría ser la siguiente:

Ejemplo A. Modelo de curva de crecimiento de dos niveles.

Nivel 1

$$\text{Bienestar}_{ti} = \pi_{0i} + \pi_{1i} * (\text{Semana}4_{ti}) + e_{ti}$$

Nivel 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

Donde en el nivel 1, el bienestar del paciente *i* en el momento *t* es predicho a partir de su bienestar en la semana 4 (constante, π_{0i}) y del cambio semanal en su bienestar (pendiente lineal,

π_{1i}). Cabe recordar que el tiempo como predictor está expresado en semanas y se encuentra centrado al final de la serie (en la semana 4), lo que determina la interpretación de la constante (π_{0i}). El término aleatorio e_{ti} permite al paciente i variar respecto de su puntaje de bienestar estimado para el momento t . En el nivel 2 del modelo, la constante del paciente (π_{0i}) es predicha a partir del promedio poblacional de bienestar en la semana 4 (β_{00}), mientras que la tasa de cambio del paciente es predicha mediante el promedio poblacional de cambio semanal (β_{10}). Los términos aleatorios r_{0i} y r_{1i} le permiten al paciente i variar respecto de las medias de bienestar en la semana 4 y su cambio semanal.

Estos modelos nos dan la posibilidad de estimar puntajes verdaderos de la constante y la pendiente lineal, en cada participante y, a su vez, en la población. Los estimadores de los puntajes verdaderos de la constante y la pendiente lineal (denominados Estimadores Empíricos de Bayes; Raudenbush & Bryk, 2002) pueden ser utilizados a su vez como datos para posteriores análisis, como puede ser en el estudio de efectos indirectos de mediación (Constantino et al., 2017; Etchebarne, O'Connell, & Roussos, 2008). Además, estos análisis nos permiten incluir predictores en cada uno de los niveles del modelo, estudiando el efecto sobre los puntajes de bienestar o su tasa de cambio, de otras variables medidas dentro de los pacientes (co-variables temporales con medidas repetidas de nivel 1) o entre pacientes (predictores de nivel 2).

En el caso de que se quisiera estudiar el efecto diferencial de un tratamiento en comparación con un grupo de control, se debería incluir la variable tratamiento dicotomizada (por ejemplo, grupo de tratamiento = 1, grupo de control = 0) como un predictor de nivel 2. Esto permitiría estudiar los efectos del tratamiento tanto sobre la constante (puntaje al final de la serie temporal, en este caso) como sobre la tasa de cambio de los pacientes durante el abordaje. De igual manera, si el diagnóstico de los pacientes no es una constante en el estudio, habiendo por ejemplo pacientes con trastornos de ansiedad y pacientes con trastornos depresivos, esa misma variable dicotomizada (ejemplo, trastornos de ansiedad = 1, trastornos depresivos = 0) podría también incluirse como un predictor de nivel 2. Al combinar los dos predictores (tratamiento y diagnóstico) el modelo nos daría la posibilidad de

estimar el efecto del tratamiento, controlando por el efecto del diagnóstico de los pacientes. Yendo un paso más allá, la inclusión de un término interactivo (tratamiento x diagnóstico) en este último modelo, configuraría un análisis de moderación, pudiendo determinar, por ejemplo, si el tratamiento produce mayores niveles de cambio en pacientes con un tipo de diagnóstico particular (Gómez Penedo, Constantino, Coyne, Westra, & Antony, 2017; Hayes, 2017).

Modelo de curva de crecimiento de tres niveles: El efecto de los terapeutas

Como se mencionó más arriba, en la investigación en psicoterapia, ya sea en estudios clínicos controlados o en trabajos en contexto natural, los pacientes suelen estar agrupados en conjuntos menores de terapeutas tratantes. Los análisis HLM permiten aplicar modelos de tres niveles, para dar cuenta de esa estructura de agrupación, en donde las medidas repetidas de la variable dependiente se encuentran agrupadas en un conjunto de pacientes y estos, a su vez, en un conjunto de terapeutas. Esta estrategia analítica, resuelve la problemática de una fuente potencial de correlación entre las observaciones, brindándonos también información respecto del efecto de los terapeutas, sobre la variable dependiente. A continuación se presenta la fórmula para el modelo de tres niveles, basado en el material clínico ilustrativo utilizado.

Ejemplo B. Modelo de curva de crecimiento de tres niveles.

Nivel 1

$$\text{Bienestar}_{tij} = \pi_{0ij} + \pi_{1ij} * (\text{Semana4}_{tij}) + e_{tij}$$

Nivel 2

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

Nivel 3

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

Aunque las ecuaciones del nivel 1 y nivel 2 son similares a las del modelo de dos niveles, la inclusión del nuevo subíndice j , modifica la interpretación de todos los parámetros del modelo.

En este modelo, en el nivel 1 el puntaje de bienestar en el momento t del paciente i tratado por el terapeuta j es estimado a partir del bienestar del paciente al finalizar la sesión 4 (constante del modelo; π_{0ij}) y el cambio lineal sesión a sesión del paciente (π_{1ij}). El término aleatorio e_{tij} permite que el puntaje del paciente i

tratado por el terapeuta j varíe respecto del valor estimado para el tiempo t .

En el nivel dos, π_{0ij} es estimado a partir del puntaje promedio de bienestar en la sesión 4 en los pacientes tratados por el terapeuta j (β_{00j}), mientras que π_{1ij} es estimado a partir de la tasa de cambio promedio de pacientes tratados por el terapeuta j (β_{10j}). Los términos aleatorios r_{0ij} y r_{1ij} permiten que los puntajes del paciente i varíen respecto de los niveles promedios de bienestar en la sesión 4 y de la tasa de cambio sesión a sesión, de los pacientes del terapeuta j .

Finalmente, en el nivel 3 del modelo, β_{00j} es estimado a partir del promedio de bienestar poblacional en la sesión 4 (γ_{000}) y β_{10j} mediante el promedio poblacional de cambio entre sesiones (γ_{100}). Los términos aleatorios u_{00j} y u_{10j} permiten que los niveles de bienestar en la sesión 4 y el cambio sesión a sesión de los pacientes del terapeuta j varíen respecto de los promedios poblacionales.

Es importante señalar que la cantidad de niveles de un modelo no necesariamente se agota en tres. En datos en psicología clínica podrían incluirse modelos con mayor cantidad de niveles, representando un ejemplo paradigmático el de la supervisión. Si los terapeutas de un estudio se encuentran anidados en un conjunto de supervisores, podrían ajustarse modelos de cuatro niveles, incluyendo a la variabilidad entre supervisores como el último nivel del análisis. Esto permitiría estudiar efectos de los supervisores (por ejemplo, características relevantes) o de los procesos de supervisión (por ejemplo, la relación entre el terapeuta y el supervisor) sobre los resultados de la psicoterapia. Sin embargo, para realizar un estudio de esta naturaleza es necesaria una muestra sumamente grande de pacientes, que permita contar con una cantidad suficiente de observaciones en cada uno de los niveles de agrupación de los datos.

Aplicación de los modelos HLM en una base de datos clínica

Descripción de la base de datos usada

Como se adelantó previamente, para poder ilustrar cómo se aplican y reportan los modelos HLM para investigar los efectos de la psicoterapia, vamos a basarnos en una base de datos de un centro clínico. Esta institución orientada a proveer servicios de tratamiento en salud mental,

construyó en conjunto con un equipo de investigación, un cuestionario breve para evaluar la evolución del bienestar general del paciente, llamado Sistema de Evolución de Tratamiento (SET), que envían en forma sistemática a sus pacientes, luego de cada sesión, para monitorear su progreso (Ver [Areas et al., 2018](#)). El SET está compuesto por cuatro ítems puntuados en una escala de 1 a 10, que exploran el grado de bienestar del paciente en diferentes áreas de la vida (individual, relacional, conductual y global). Los cuatro puntajes de los ítems se suman en un índice general de bienestar que será nuestra variable dependiente en nuestra ejemplificación de los modelos. En este trabajo usaremos los resultados obtenidos durante las primeras cuatro sesiones de tratamientos. El cuestionario fue enviado a un total de 1999 pacientes tratados por 171 terapeutas en el periodo comprendido entre el 6 de abril del 2016 al 31 de agosto del 2017. De esos pacientes, únicamente 1154 (57.72%; tratados por 148 terapeutas) completaron el SET al menos una vez. Debido a que los HLM permiten analizar casos que presentan datos faltantes, se incluyeron en los análisis todos aquellos participantes que completaron al menos una de las medidas repetidas de la variable dependiente ([Raudenbush & Bryk, 2002](#)).

En resumen, esta base de datos cuenta con cuatro medidas repetidas del cuestionario que analiza el bienestar (nivel 1 de análisis) en 1154 pacientes (nivel 2 de análisis) tratados por 148 terapeutas (nivel 3 de análisis).

Además del cuestionario, los pacientes completaron un ítem extra, orientado a explorar “¿En qué medida ha experimentado un progreso respecto del motivo de consulta hasta ahora?”, con una escala de 4 puntos de 1 (“nada”) a 4 (“mucho”). Este ítem adicional, junto con el género de los participantes, se utilizaron como ejemplos de predictores en los modelos.

Cabe destacar que los pacientes incluidos en la muestra utilizada para ejemplificar los análisis contaban con una amplia variabilidad diagnóstica (trastornos de ansiedad, trastornos del estado de ánimo, trastornos adaptativos, trastornos alimentarios, entre otros) que no fue registrada para este estudio ni incluida en los análisis. Además, en todos los casos el tratamiento realizado fue un abordaje no manualizado basado en la terapia breve de resolución de problemas ([de Shazer et al., 1986](#); [de Shazer & Berg, 1997](#)),

que caracteriza al centro de donde fueron recabados los datos, sin contar con un tratamiento alternativo o grupo de control, que permita estudiar el efecto diferencial del abordaje.

Softwares estadísticos para aplicar modelos HLM

En la actualidad hay disponible una amplia variedad de softwares que permiten ajustar modelos HLM. Existe un programa específico para correr estos análisis que se denomina HLM y que se basa en el armado de las ecuaciones del modelo (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). Este software permite estimar el efecto de predictores en cada uno de los niveles, a partir de su inclusión en la ecuación multinivel del modelo que se está ajustando. Si bien el programa tiene una licencia paga, cuenta con una versión para estudiantes que puede ser descargada en forma gratuita. Otra alternativa para correr estos modelos, altamente recomendable, es el programa de acceso abierto R (R Core Team, 2018). Dicho programa tiene un paquete denominado "lme4" (Bates, Mäechler, Bolker, & Walker, 2015) que incluye la función lmer(), la cual permite un alto grado de flexibilidad en el ajuste de modelos HLM.

Identificando el modelo HLM que mejor se ajusta a los datos

Al realizar un análisis HLM, el primer paso es ajustar distintos modelos incondicionales (sin ningún predictor) con diferente cantidad de niveles, para observar cuál de ellos es el que mejor se ajusta a los datos. Para esto es necesario determinar a partir de qué nivel de agrupamiento de la información empieza a observarse una variación significativa entre los participantes (Bliese, 2016). Posteriormente, una vez identificada la cantidad de niveles del modelo, se ajustarán modelos con el tiempo como único predictor, introduciendo distintas trayectorias posibles (más específicamente lineales y cuadráticas), nuevamente para observar cuál es la que mejor describe los datos obtenidos.

Debido a que en la base de datos que utilizaremos, esperamos tres fuentes de variación (efectos intra-pacientes, efectos entre pacientes y efectos entre terapeutas), en primer lugar debemos correr un modelo incondicional de tres niveles (Raudenbush & Bryk, 2002), para analizar la variabilidad en torno a la constante estimada, que en este caso representará a los niveles de

bienestar de los pacientes durante las primeras cuatro sesiones (Bliese, 2016). La falta de una variabilidad significativa entorno a esa media, implicaría que no es necesario incluir efectos aleatorios a ese nivel y, por ende, que el nivel puede ser excluido.

Para poder establecer si existe una variabilidad significativa en el nivel superior de agrupamiento (el nivel de los efectos de los terapeutas), siendo adecuado utilizar un modelo de tres niveles, existen dos métodos complementarios. Una de las estrategias consiste en identificar el porcentaje de varianza de la variable dependiente, explicada por el tercer nivel de agrupamiento, calculando ICCs. Para calcular estos coeficientes, se toma a la varianza observada en el nivel 3 (τ_β) y se la divide por la suma de la varianza a nivel 1 (σ^2), nivel 2 (τ_π) y nivel 3 (τ_β) (Raudenbush & Bryk, 2002).

Basado en los componentes de varianza del modelo incondicional de tres niveles, ajustados a los datos clínicos utilizados, la fórmula del ICC en este caso sería: $.003 [.30 / (.30 + 62.11 + 25.57)]$. Para facilitar la comprensión del coeficiente, el ICC puede ser multiplicado por 100 e interpretado como porcentaje de varianza explicada por el nivel 3 (es decir, por el efecto de los terapeutas). En este caso el efecto de los terapeutas en esta muestra explica el .3% (ICC = .003) de la varianza de bienestar en los pacientes. El efecto de un nivel se considera significativo cuando explica al menos un 1% de la varianza en la variable dependiente. En caso contrario, como sucede en los análisis realizados, se concluye que no existe un efecto significativo de los terapeutas sobre el bienestar de los pacientes, por lo que debería eliminarse el tercer nivel del modelo.

El segundo método para analizar la adecuación de ajustar un modelo de tres niveles consiste en observar si los efectos aleatorios del nivel 3 (u_0) son significativos, lo que sugeriría una variación relevante entre los pacientes de los distintos terapeutas, en su grado de bienestar. La falta de una variación significativa entre terapeutas en los puntajes promedios de sus pacientes implicaría que no es necesario incluir un tercer nivel en el modelo. En la muestra utilizada en este estudio, de manera consistente con los resultados del ICC, no se observó una variación significativa entre terapeutas en los puntajes de bienestar de sus pacientes, $u_{00} = .55$, $\chi^2(143) = 144.22$, $p = .46$.

Debido a que tanto los bajos valores del ICC como la falta de una variación significativa entre terapeutas sugieren no ajustar un modelo de tres niveles, a continuación se debe correr un modelo de dos niveles, excluyendo el nivel de variabilidad a partir de los terapeutas. De esta forma se ajusta un modelo incondicional de dos niveles (sin ningún predictor), que da cuenta de las medidas repetidas de bienestar (nivel 1) agrupadas en pacientes (nivel 2).

Al ajustar este modelo incondicional de dos niveles, se encontró que el 30% de la varianza del puntaje de bienestar se explica a partir de diferencias entre pacientes (ICC = .30). A su vez, se observó una variación significativa en torno a la media de los puntajes de bienestar de los pacientes, $r_0 = 5.13$, $\chi^2(1131) = 2166.14$, $p < .001$. Estos dos indicadores sugieren conservar el modelo de dos niveles como modelo de base para los análisis.

Posteriormente, es necesario identificar cuáles son las trayectorias que mejor describen el cambio en la variable de bienestar a lo largo del tiempo. De esta forma, se ajustó un modelo lineal y luego un modelo cuadrático. Con fines ilustrativos, en la [Figura 2](#) se muestra una representación gráfica de ambos modelos (hipotéticos) de trayectoria. Para seleccionar el modelo más adecuado en función de los datos, se utiliza *el test de comparación de modelos* para determinar si la inclusión de nuevos términos mejora el porcentaje de varianza explicada por los modelos. Este test basado en el estadístico *chi cuadrado* (χ^2), usa tanto información de la variabilidad del modelo como de la cantidad de parámetros a estimar. Un resultado significativo en este test implica que la inclusión de nuevos parámetros mejora el porcentaje de varianza explicada por el modelo.

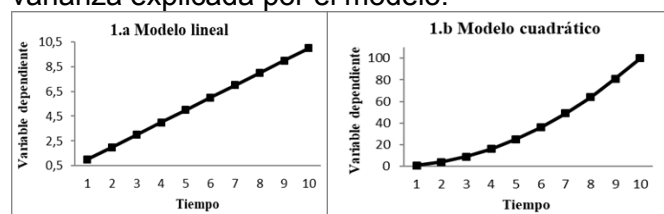


Figura 2. Representación de la evolución temporal de una variable en un Modelo Lineal (1.a) y un Modelo Cuadrático (1.b).

Mediante este procedimiento primero comparamos un *modelo incondicional* de dos niveles con un *modelo lineal* de dos niveles incluyendo a la variable tiempo en semanas como el único predictor a nivel 1. Al realizar este análisis

en la base de datos clínica utilizada, el resultado mostró que la inclusión del término lineal mejoró significativamente el ajuste del modelo, $\chi^2(3) = 38.40$, $p < .001$. A continuación comparamos el modelo lineal con un modelo cuadrático de dos niveles. Este modelo cuadrático tiene como predictores al tiempo en semanas (término lineal) y al tiempo en semanas elevado al cuadrado (término cuadrático). No obstante, el test de comparación de modelos, determinó que la inclusión del nuevo término cuadrático no mejoró significativamente el porcentaje de varianza explicada por el modelo, $\chi^2(4) = 3.43$, $p > .50$. Por lo tanto, la comparación de todos los modelos sugiere que el modelo que mejor se ajusta a los datos (modelo final del análisis) es un modelo lineal de dos niveles, con el tiempo en semanas como el único predictor (Ver Modelo de Ejemplo A). Los resultados de los distintos modelos ajustados se encuentran en la [Tabla 1](#).

Cómo reportar e interpretar los parámetros de modelos lineales (tiempo como único predictor)

Los resultados del modelo muestran que en promedio los pacientes presentaron una media de bienestar de 23.59 al finalizar la sesión 4 ($\beta_{00} = 23.59$; en un rango teórico de 0 a 40), y un incremento promedio de .83 unidades de bienestar ($\beta_{10} = .83$) sesión a sesión. La forma apropiada de reportar estos resultados es presentar los coeficientes del parámetro, el error estándar (EE), el intervalo de confianza (CI), el estadístico de la prueba *t* y el valor *p* de significancia (en ese orden). Todos estos elementos aparecen en los resultados de los distintos programas estadísticos que permiten estimar modelos lineales jerárquicos (HLM, R, etc.), con excepción de los CI, que programas como HLM no lo reportan. En este caso, los intervalos pueden calcularse a partir de la fórmula: $\text{Coeficiente} \pm \text{EE} * 1.96$.

Por lo tanto, para los promedios de puntajes de bienestar al finalizar la sesión 4, los resultados se presentan de la siguiente manera: $\beta_{00} = 23.59$, $EE = .34$, $CI_{95} [22.92, 24.26]$, $t(1153) = 68.961$, $p < .001$. Mientras que, para el promedio de cambio semanal de bienestar los resultados se presentan de esta forma: $\beta_{10} = .83$, $EE = .16$, $CI_{95} [.52, 11.13]$, $t(1153) = 5.333$, $p < .001$. Para la prueba de significancia de la constante (β_{00}), la hipótesis nula es que los puntajes de bienestar en la sesión

Tabla 1

Resumen de los modelos ajustados en base a la base de datos clínica utilizada

Efectos Fijos	Modelo de Tres Niveles			Modelo de Dos Niveles		
	Coefficiente	EE	Componente de Varianza	Coefficiente	EE	Componente de Varianza
<i>Modelo Incondicional</i>						
Constante	22.24***	.23	.30	22.22***	.22	26.20***
<i>Modelo Lineal</i>						
Constante				23.59***	.34	36.30***
Semanas				.83***	.16	3.17**
<i>Modelo Cuadrático</i>						
Constante				23.46***	.38	36.15***
Semanas				.46	.52	15.04
Semanas al Cuadrado				-.12	.16	1.40*
<i>Modelos condicionales con cambios en motivo de consulta como predictor a nivel 1</i>						
Constante				22.57***	.34	26.45***
Semanas				.23	.16	3.12**
Motivo				2.12***	.20	6.70**
<i>Modelos condicionales con género como predictor a nivel 2</i>						
Constante				23.57***	.34	36.27***
Semanas				.81***	.16	3.15**
Efecto del género en la constante				.29	.71	
Efecto del género en semanas				.29	.33	

Nota. EE = Error Estándar, Semanas = Cambio en semanas, Motivo = cambios en relación al motivo de consulta, *** $p < .001$, ** $p < .01$, * $p < .05$.

4 son iguales a 0. Mientras que para el término lineal (β_{10}) la hipótesis nula del contraste es que el cambio sesión a sesión en el bienestar es igual a 0. Por ende, los resultados muestran que los pacientes presentan un nivel de bienestar superior a 0 al finalizar la sesión 4 (algo que desde el punto de vista conceptual no tiene relevancia en este caso, pero podría tenerlo en otros análisis) y que los pacientes tuvieron una mejoría significativa entre sesiones durante los cuatro primeros encuentros con sus terapeutas.

Ajustando modelos condicionales

Además de los modelos que nos permiten estimar valores de la variable dependiente y su evolución a lo largo del tiempo, una de las aplicaciones más relevantes de los modelos HLM reside en la posibilidad de incluir predictores en cada uno de los niveles de organización de la información, analizando su efecto sobre la variable dependiente, mediante

el ajuste de modelos condicionales.

A modo de ejemplo, ajustamos dos modelos lineales condicionales, de dos niveles cada uno. En el primer modelo, se incluyó la percepción de cambios en relación al motivo de consulta como un predictor a nivel 1. Debido a que esta variable tiene medidas repetidas en los mismos momentos que la variable dependiente, representa una covariable temporal en el modelo. Para facilitar la interpretación de los parámetros, centramos este predictor a partir de su media muestral. Luego, ajustamos un modelo incluyendo al género de los pacientes (codificado como Femenino = 1, Masculino = 0) como predictor del nivel 2. En este caso, la estrategia que se usa para incluir al género como predictor es la misma que se utilizaría para estudiar los efectos de un tratamiento en comparación con un grupo de control o de dos grupos diagnósticos diferentes.

Reportando e interpretando modelos condicionales con predictores a nivel 1. Los resultados de los modelos muestran un efecto significativo de la mejoría en relación al motivo de consulta sobre los niveles de bienestar del paciente, $\beta_{10} = 2.12$, $EE = .20$, $CI_{95} [1.73, 2.51]$, $t(1153) = 10.741$, $p < .001$. Los resultados del efecto de los predictores se interpretan en términos de unidades de aumento en la variable dependiente, a partir del aumento de una unidad en el predictor. Por lo tanto, estos hallazgos muestran que una unidad de aumento en los puntajes de mejoría asociada al motivo de consulta, se relacionó con un aumento de 2.12 unidades en los puntajes de bienestar de los pacientes. Se observó un efecto significativo en el componente aleatorio r_1 , lo que implica que los pacientes variaron significativamente en torno al efecto medio de la mejoría en relación al motivo de consulta sobre el bienestar de los sujetos, $r_1 = 2.59$, $\chi^2(326) = 406.01$, $p = .002$. Este resultado sugiere la posibilidad de incluir predictores (a nivel 2) para explicar la variabilidad de ese efecto. Finalmente, la inclusión del predictor mejoró significativamente el ajuste del modelo, $\chi^2(4) = 150.99$, $p < .001$, por lo que se recomienda conservar este modelo condicional.

Reportando e interpretando modelos condicionales con predictores a nivel 2. Los resultados de los modelos condicionales con el género como predictor a nivel 2, muestran que esta variable no presenta un efecto significativo sobre los niveles de bienestar de los pacientes al finalizar la sesión 4 (la constante), $\beta_{01} = .29$, $EE = .71$, $CI_{95} [-1.10, 1.68]$, $t(1130) = .404$, $p = .69$, ni sobre su tasa de cambio (pendiente lineal), $\beta_{11} = .29$, $EE = .33$, $CI_{95} [-.36, .94]$, $t(1130) = .880$, $p = .38$. Nuevamente, en estos modelos los coeficientes se interpretan en términos de cantidad de unidades de incremento/reducción en la constante (β_{01}) y la pendiente lineal (β_{11}), a partir de una unidad de aumento en el predictor. Como el predictor en este caso es una variable dicotómica (género), el coeficiente representa la diferencia que existe entre el grupo de referencia, codificado como 1 (género femenino) y la otra condición codificada como 0 (género masculino). De esta forma, en la sesión 4 las personas de género femenino presentaron en promedio un nivel de bienestar

.29 unidades mayor en comparación con el de las personas de género masculino ($\beta_{01} = .29$) y también un aumento .29 unidades mayor en su cambio sesión a sesión en los niveles de bienestar ($\beta_{11} = .29$). Los términos aleatorios muestran que los pacientes variaron significativamente en torno al efecto medio del género sobre los puntajes de bienestar en la sesión 4, $r_0 = 6.06$, $\chi^2(717) = 1147.69$, $p < .001$, y sobre su cambio sesión a sesión, $r_1 = 1.82$, $\chi^2(717) = 841.01$, $p < .001$. Esto implica que se pueden incluir otros predictores en el modelo para los niveles de bienestar de los pacientes al finalizar la sesión 4 y su cambio sesión a sesión. Finalmente, incluir al género como predictor no mejoró el ajuste del modelo, $\chi^2(2) = .94$, $p > .50$, por lo cual no se recomienda tomar este modelo como definitivo.

Calculando tamaños de efectos para modelos condicionales. Al observar un efecto significativo, es posible estimar medidas de tamaño de efecto, en términos de proporción de varianza explicada en la variable de resultado a partir del predictor, computando coeficientes *Pseudo R²*. Estos coeficientes se calculan comparando la diferencia entre la varianza del modelo incondicional y la del modelo condicional, dividido por la varianza del modelo incondicional (Hofmann, 1997). Por ende, $Pseudo R^2 = (Varianza\ del\ modelo\ incondicional - Varianza\ del\ modelo\ condicional) / Varianza\ del\ modelo\ incondicional$. Los coeficientes *Pseudo R²* pueden ser transformados a porcentaje de varianza explicada por el predictor, a partir de multiplicar sus valores por 100.

Para el modelo condicional, con la mejoría en relación al motivo de consulta como predictor a nivel 1, calculamos *Pseudo R²* basados en la varianza de nivel 1 de dicho modelo y del modelo lineal (en reemplazo del incondicional). En este caso se utilizó el modelo lineal ya que sobre ese modelo de estimación del cambio (sin ningún otro predictor más que el tiempo) se incluyó el predictor del modelo condicional. El cálculo, entonces, fue el siguiente: $Pseudo R^2 = .07 [(55.77-51.82) / 55.77]$. El resultado sugiere, por lo tanto, que las mejorías en relación al motivo de consulta, explican un 7% de la varianza de los puntajes de bienestar en los pacientes. Para el modelo condicional con el

género de los pacientes incluido como predictor de nivel 2, calculamos coeficientes *Pseudo R²* tanto para la constante como para la pendiente lineal, usando los componentes de varianza de ambos elementos. *Pseudo R²* de la constante = .002 [(36.80 - 36.74) / 36.80]. *Pseudo R²* de la pendiente lineal = .007 [(3.31 - 3.29) / 3.31]. Este resultado implica que el género de los pacientes explica únicamente el .2% de la varianza de su bienestar al finalizar la sesión 4 (constante) y un .7% de la varianza en la tasa de cambio semanal (pendiente lineal). A pesar de que hemos presentado el cálculo de ambos tamaños de efecto con fines ilustrativos, si las pruebas de hipótesis correspondientes a los coeficientes no arrojan resultados estadísticamente significativos, desde el punto de vista teórico no tendría valor calcular estas medidas de tamaño de efecto.

Síntesis y conclusiones

Complementando esfuerzos previos realizados en la bibliografía internacional (Atkins, 2005; Baldwin et al., 2014), este trabajo intenta resaltar el valor de los modelos HLM como un recurso estadístico de importancia fundamental para los estudios en psicoterapia orientados a determinar sus efectos y explicarlos mediante predictores de distinta naturaleza. Estos modelos, en primer lugar, permiten lidiar con el ubicuo anidamiento de los datos en el campo de la investigación en psicoterapia. En segundo lugar, facilitan una estimación robusta del cambio de los pacientes, foco principal de los estudios en psicología clínica, analizando tanto trayectorias promedio en una muestra como patrones de evolución individual. En tercer lugar, brindan la posibilidad de incluir predictores, en diferentes niveles de la organización de los datos, para explicar la varianza del cambio. Finalmente, nos permiten determinar la contribución de la figura del terapeuta a la mejoría de los pacientes.

Si bien a nivel internacional, en los últimos años ha habido diversas investigaciones que aplicaron estos modelos para la investigación en psicoterapia, se han encontrado escasos ejemplos de su aplicación en autores argentinos y/o latinoamericanos, y las pocas excepciones han sido publicadas en revistas angloparlantes (Ver Behn, Errázuriz, Cottin, & Fischer, 2017; Gómez Penedo, Constantino, Coyne,

Bernecker, & Lotte Smith-Hansen, 2018; Mellado et al., 2017).

Poder realizar y entender estos modelos requiere de un alto grado de conocimiento técnico en dichos análisis, que para muchos estudiantes e investigadores en psicoterapia de la región permanece inaccesible, debido a la falta de guías de aplicación e interpretación de estos modelos en castellano.

Con este trabajo hemos intentado presentar una guía para realizar, entender y reportar modelos de análisis HLM, facilitando su aplicación e intentando amplificar su alcance a potenciales investigadores del área. La posibilidad de generalizar el uso de estos modelos a la investigación en psicoterapia va a representar una contribución valiosa tanto para la investigación como para la práctica en la región latinoamericana. Al superar las limitaciones de los métodos paramétricos clásicos, una mayor diseminación de estos modelos tiene el potencial de mejorar la calidad de los análisis para estudios en psicoterapia. Esto no sólo tendría implicancias metodológicas en términos de la solidez de los resultados (y su posible impacto en el ámbito académico), sino que también podría impactar a nivel de la práctica, a partir de la creación de un conocimiento más robusto y cercano a la clínica, con validez cultural para la región latinoamericana, que ayudará a mejorar los procesos y resultados de las terapias, haciendo que los pacientes obtengan un mayor provecho de las mismas.

Agradecimientos

Los resultados de la base de datos utilizada con fines ilustrativos en este trabajo metodológico han sido parcialmente reportados en una investigación previa, orientada a presentar el instrumento de medición y su desarrollo (Areas et al., 2018).

Referencias

- Areas, M., Roussos, A., Hirsch, H., Hirsch, P., Becerra, P., & Gómez Penedo, J. M. (2018). Evaluación de un dispositivo de investigación orientada a la práctica en el desarrollo de un sistema de feedback en psicoterapia. *Revista Argentina de Clínica Psicológica*, 27(2), 229-249.
- Atkins, D. C. (2005). Using multilevel models to

- analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology*, 19(1), 98–110. doi: [10.1037/0893-3200.19.1.98](https://doi.org/10.1037/0893-3200.19.1.98)
- Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing multiple outcomes in clinical research using multivariate multilevel models. *Journal of Consulting and Clinical Psychology*, 82(5), 920–930. doi: [10.1037/a0035628](https://doi.org/10.1037/a0035628)
- Bates, D., Mäeçhler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Behn, A. J., Errázuriz, P. A., Cottin, M., & Fischer, C. (2017). Change in symptomatic burden and life satisfaction during short-term psychotherapy: Focusing on the role of family income. *Counselling & Psychotherapy Research*, 18(2), 133-142. doi: [10.1002/capr.12158](https://doi.org/10.1002/capr.12158)
- Bliese, P. (2016). *Multilevel modeling in R (2.5)*. Recuperado de https://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf
- Constantino, M. J., Westra, H. A., Antony, M. M., & Coyne, A. E. (2017). Specific and common processes as mediators of the long-term effects of cognitive-behavioral therapy integrated with motivational interviewing for generalized anxiety disorder. *Psychotherapy Research*, 5, 1-13. doi: [10.1080/10503307.2017.1332794](https://doi.org/10.1080/10503307.2017.1332794)
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62(1), 583–619. doi: [10.1146/annurev.psych.093008.100356](https://doi.org/10.1146/annurev.psych.093008.100356).
- de Shazer, S., & Berg, I. K. (1997). "What works?" Remarks on research aspects of Solution-Focused Brief Therapy. *Journal of Family Therapy*, 19(2), 121–124. doi: [10.1111/1467-6427.00043](https://doi.org/10.1111/1467-6427.00043)
- de Shazer, S., Berg, I. K., Lipchik, E., Nunnally, E., Molnar, A., Gingerich, W., & Weiner-Davis, M. (1986). Brief therapy: focused solution development. *Family Process*, 25(2), 207–221. doi: [10.1111/j.1545-5300.1986.00207.x](https://doi.org/10.1111/j.1545-5300.1986.00207.x)
- Etchebarne, I., O'Connell, M., & Roussos, A. (2008). Estudio de mediadores y moderadores en la investigación en Psicoterapia. *Anuario de Investigaciones*, 13(1), 33-56.
- Finch, W. H., & Bolin, J. E. (2017). *Multilevel modeling using Mplus*. Boca Raton: CRC Press.
- Gómez Penedo, J. M., Constantino, M. J., Coyne, A., Bernecker, S. L., & Lotte Smith-Hansen, L. (2018). Patient Baseline Interpersonal Problems as Moderators of Outcome in Two Psychotherapies for Bulimia Nervosa. *Psychotherapy Research*. Publicación anticipada en línea. doi: [10.1080/1050337.2018.1425931](https://doi.org/10.1080/1050337.2018.1425931)
- Gómez Penedo, J. M., Constantino, M. J., Coyne, A., Westra, H., & Antony, M. (2017). Markers for Context-Responsiveness: Client Baseline Interpersonal Problems Moderate the Efficacy of Two Psychotherapies for Generalized Anxiety Disorder. *Journal of Consulting and Clinical Psychology*, 85(10), 1000-1011. doi: [10.1037/ccp0000233](https://doi.org/10.1037/ccp0000233)
- Gómez Penedo, J. M., Juan, S., Manubens, R. T., & Roussos, A. (2018). El estudio del cambio en psicoterapia: desafíos conceptuales y problemas de investigación empírica. *Anuario de Investigaciones en Psicología*, 24, 15-24.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Publications.
- Hofmann, D. A. (1997). An overview of the logic and rationale of Hierarchical Linear Models. *Journal of Management*, 23(6), 723–744. doi: [10.1177/014920639702300602](https://doi.org/10.1177/014920639702300602)
- Hox, J., & Maas, C. (2005). Multilevel analysis. *Encyclopedia of Social Measurement*, 2, 785–793. doi: [10.1016/B0-12-369398-5/00560-0](https://doi.org/10.1016/B0-12-369398-5/00560-0)
- Mellado A., Suárez, N., Altimir, C., Martínez, C., Pérez J. C., Krause, M., & Horvath, A. (2017). Disentangling the change-alliance relationship: Observational assessment of the therapeutic alliance during change and stuck episodes. *Psychotherapy Research*, 27(5), 595-607. doi: [10.1080/10503307.2016.1147657](https://doi.org/10.1080/10503307.2016.1147657)
- Ogles, B. M. (2013). Measuring change in psychotherapy research. En M. J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (pp.134–166). New Jersey: Wiley.
- Power, M. J., & Freeman, C. (2012). A Randomized Controlled Trial of IPT versus CBT in primary care: with some cautionary notes about handling missing values in clinical trials. *Clinical Psychology & Psychotherapy*, 19(2), 159–169. doi: [10.1002/cpp.1781](https://doi.org/10.1002/cpp.1781)
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd Ed.)*. Thousand Oaks, California: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM7: Hierarchical Linear and Nonlinear Modeling*. Chicago, IL: Scientific Software International.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. Boca Raton: Chapman & Hall/CRC.
- R Core Team (2018). *R: A language and environment for statistical computing*. [programa informático]. Viena, Austria: R Foundation for

Statistical Computing. Recuperado de:
<https://www.R-project.org/>.

Wang, J., Xie, H., & Fisher, J. H. (2012). *Multilevel models: Applications using SAS*. Berlin: De Gruyter.

Westra, H. A., Constantino, M. J., & Antony, M. M. (2016). Integrating motivational interviewing with cognitive-behavioral therapy for severe generalized anxiety disorder: an allegiance-controlled randomized clinical trial. *Journal of Consulting and Clinical Psychology, 84*(9), 768–782. doi: 10.1037/ccp0000098

Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance–outcome association. *Journal of Counseling Psychology, 62*(4), 579–591. doi: 10.1037/cou0000103