

Evaluación de Valores Perdidos en Modelos de Crecimiento Latente

Assessment of Missing Data in Latent Growth Models

Silvia Nieva¹, Jesús M. Alvarado² y Marta Gràcia³

Resumen

Los modelos de crecimiento latente han sido un avance para el estudio del cambio en investigación longitudinal. La presencia de datos incompletos durante el proceso de medición es un problema permanente en estos estudios. Se plantea una simulación para explorar el papel de distintos mecanismos de pérdida de datos (MCAR, MAR y MNAR) y en la recuperación de parámetros en los modelos de crecimiento latente. Se esperaba encontrar diferencias en la estimación, el ajuste y el sesgo en función de los diferentes escenarios y del tratamiento de datos perdidos. Se han generado muestras de diferentes tamaños, con valores perdidos bajo distintas condiciones. Algunos mecanismos muestran desajustes en la estimación al aplicar procedimientos de eliminación de sujetos. No obstante, otros tratamientos permiten compensarlo. Los cambios en la parametrización afectan al ajuste, al porcentaje de sesgo y aumentan la variabilidad entre réplicas, mostrándose diferencias entre los distintos métodos aplicados.

Palabras clave: modelos de crecimiento latente, valores perdidos, sesgo, investigación longitudinal

Abstract

Latent growth models have been a move forward for the study of change in longitudinal research. Missing data presence through measuring process is a ubiquitous problem in this research. A simulation is set to exploring the role of different missing data mechanisms (MCAR, MAR y MNAR) on latent growth models' parameter recovering. Estimation, model adjustment and bias divergencies were expected according to the different scenarios and missing data treatment applied. Different size samples have been generated with missing data under different conditions. Some mechanisms show estimate malfunction when implementing subject deletion methods, nevertheless other treatment solution allows compensation. Parametrization changes affect adjustment, bias proportion as well as they increase variability within replications between different methods applied.

Keywords: latent growth models, missing data, bias, longitudinal research

Agradecimientos: Esta publicación es parte de los proyectos de I+D+i PID2019-105177GB-C21 y PID2019-105177GB-C22 financiado por MCIN/ AEI/10.13039/501100011033/

¹Doctora en Psicología. Profesora Contratada Doctora. Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, España. Tel.: +34 91393116. Correo: silnieva@ucm.es (Autora de correspondencia)

²Doctor en Psicología. Catedrático de Universidad. Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, España. Tel.: +34 913943055. Correo: jmalvara@ucm.es

³Doctora en Psicología. Profesora Titular, Facultad de Psicología, Universidad de Barcelona. Passeig de la Vall d'Hebrón. Barcelona, España. Tel.: +34 933125832. Correo: mgraciag@ub.edu

Introducción

Se han desarrollado diversos modelos estadísticos que permiten analizar el cambio en estudios en los que se recogen distintas medidas en el tiempo (longitudinales). A continuación, se describen sus fundamentos, principales ventajas e inconvenientes.

Los diseños longitudinales tienen ventajas respecto a los transversales, ya que generalmente cumplen un supuesto necesario para hacer inferencias causales entre variables (la causa es anterior al efecto en el tiempo) (Duncan et al., 2006).

Su aplicación se traduce, por ejemplo, en investigaciones que tienen como objetivo estudiar la cognición, competencias o conductas en desarrollo, o probar la eficacia de programas de intervención con varias sesiones. Al introducir varios momentos temporales en la recogida de datos para una o varias mediciones, se permite plantear estudios más robustos, reduciendo el error de medida, aumentando la fiabilidad y, por consiguiente, el tamaño del efecto (Byers-Heinlein et al., 2021).

Una vez recogidos los datos, determinar qué procedimiento estadístico aplicar para realizar los análisis en un estudio longitudinal es un elemento clave dentro del proceso de toma de decisiones. Evidentemente, los diferentes estudios longitudinales tienen que encontrar un método estadístico adecuado para sus preguntas de investigación, no hay un único método para todos. En ocasiones se aplican métodos estadísticos “mecánicamente”, sin tener en cuenta las particularidades de la investigación que se está llevando a cabo (Singer & Willett, 2003). Además, muchos investigadores parecen no tener clara la relación entre los diferentes tipos de modelos y qué modelo estadístico serviría para responder a sus preguntas de investigación (Geiser, 2021).

Históricamente los análisis sobre modelos de trayectorias se estimaban con Análisis de Varianza (ANOVA) de medidas repetidas (Bock, 1979) y con otras técnicas que lo combinan con el análisis de regresión: el Análisis de Covarianza (ANCOVA) y el Análisis Multivariante de Varianza (MANOVA) (para un ejemplo de su aplicación, ver Solano & Casullo, 2005). Esta

práctica fue extensiva hasta los años 50, cuando empezaron a estimarse los modelos de trayectorias desde el marco de variables latentes (Baker, 1954). Posteriormente, Tucker (1958) y Rao (1958) lo ampliaron, derivando a su vez en los modelos actuales de variables latentes (*latent growth / latent curve models*), que se sirven de los Modelos de Ecuaciones Estructurales (*Structural Equation Modeling*) (SEM) (para un ejemplo de su aplicación, ver Miranda et al., 2021). Estos Modelos de Crecimiento Latente (en lo sucesivo LCM) no son una única técnica, sino un conjunto de procedimientos de análisis que sirven para evaluar una teoría/hipótesis sobre relaciones causales entre variables (observadas o latentes) (ver Kline, 2015, para una descripción detallada).

El alcance de los modelos de análisis del cambio ha estado condicionado por el desarrollo computacional y en la actualidad existen otro tipo de modelos más complejos y completos. La decisión sobre el modelo y el software a utilizar puede impactar en los resultados. En concreto, puede afectar a la estimación y especificación de parámetros (McNeish & Matta, 2018). A pesar de estos avances computacionales y de que existen alternativas, el uso de modelos tradicionales, como el ANOVA, es una práctica aún vigente.

Tanto el ANOVA como los LCM son parametrizaciones de distintos aspectos del cambio. Así, en el ANOVA, se focaliza en las diferencias entre observaciones, mientras que en los LCM se trata de describir el cambio individual en el tiempo (Rovine & McDermott, 2017).

Como puede verse en la Figura 1, el LCM para trayectorias lineales se define por dos variables latentes (efectos fijos): 1) intersección media (η_0) (punto de partida), y 2) pendiente media (η_1) (tasa de cambio); y por sus respectivas varianzas (ζ) y covarianzas ($\zeta(0)$) (efectos aleatorios, además de las variables observadas o simuladas (y_1 a y_4) (medidas repetidas) y sus errores de medida (ε_1 a ε_4)).

Este modelo tiene varias extensiones, pudiendo añadirse predictores y considerar trayectorias no lineales (Bollen & Curran, 2006). Para una descripción más detallada de estos modelos y sus posibles ampliaciones, ver (Grimm et al., 2017).

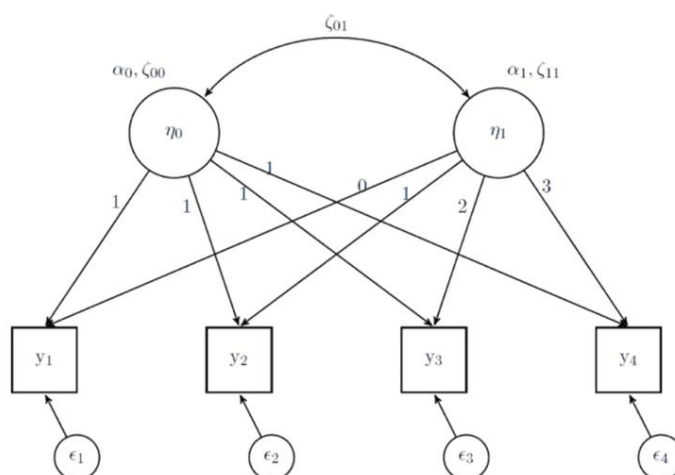


Figura 1. Esquema del modelo de crecimiento latente (Extraído de Cernat, 2021)

Datos incompletos

El problema de los datos incompletos o valores perdidos es frecuente (Allison, 2001). Sus consecuencias pueden generar costes en una investigación o desaprovechamiento del tiempo y los recursos (Lawes et al., 2020). Las razones de fondo pueden ser: respuestas parciales, inconsistentes o el abandono del estudio por parte de uno o varios sujetos (Graham, 2012).

Para interpretar si la probabilidad de encontrar valores perdidos en un conjunto de datos determinado depende de alguna/as de la/s variable/s de las que se toma/n medidas, se interpretan los patrones que presentan a partir de mecanismos de datos perdidos (que tratan de representar matemáticamente esta hipotética relación, esto es, su causa) (Enders, 2010). Estos mecanismos de posibles causas de pérdida de datos han sido propuestos por Rubin (1976), y son tres: *Missing Completely at Random* (MCAR), *Missing at Random Data* (MAR) y *Missing not at Random* (MNAR). El supuesto MCAR corresponde a la posibilidad de encontrar valores perdidos no depende de ninguna variable observada. MAR se refiere a que la pérdida de datos puede explicarse por variables que están registradas, y MNAR a que la probabilidad de los datos perdidos depende solo de la propia variable en la que se encuentran, cuando se han controlado otras variables. Existen métodos para determinar el mecanismo de pérdida de datos (ver Fielding et al., 2009, para un resumen de los más utilizados).

Las consecuencias de obviar y/o eliminar de la muestra estos datos incompletos son relevantes, por la pérdida de eficiencia y el sesgo (Carpenter

& Kenward, 2013). Todo ello afectaría a la precisión de la inferencia de los resultados a la población. No obstante, el impacto real que tiene esta pérdida dependería del objeto de la investigación. En contextos clínicos es especialmente relevante, por ejemplo, cuando la incidencia de algunas condiciones clínicas es baja. Por tanto, la proporción de pérdida de datos asumible no es la misma para todos los casos, y dependerá de si otras variables (variables auxiliares) pueden recoger la información que se está perdiendo.

Tradicionalmente se han probado técnicas para solventar estas situaciones sobrevenidas, por lo que se encuentran multitud de métodos más o menos eficientes (ver resumen en Enders, 2010) desde el procedimiento más sencillo y menos eficiente de ignorar los casos con datos perdidos, hasta los métodos más sofisticados de imputación múltiple o la estimación de información completa.

El análisis con datos completos o eliminación por lista LW (*listwise deletion*) es el estándar en la mayoría de programas y paquetes estadísticos (Rioux & Little, 2021b). En LW se eliminan de la muestra los casos incompletos (Allison, 2001) con la consiguiente disminución del tamaño muestral y pérdida de potencia (ver McNeish, 2017, Shi et al., 2021). En los últimos años, gracias a los avances en el software, se observa un intento de incorporar métodos en los análisis para el tratamiento de los datos perdidos (Rioux & Little, 2021b). Los métodos considerados como más eficientes son la estimación de máxima verosimilitud con información completa (*Full-Information Maximum Likelihood*, FIML), que

consiste en estimar el logaritmo de verosimilitud para cada caso utilizando las variables especificadas en el modelo (datos observados) (Graham, 2009), y la imputación múltiple (*Multiple Imputation*, MI), que genera una base de datos que sustituye los valores perdidos con valores simulados (Lee & Shi, 2021).

Estos dos métodos, de calidad similar, no tienen un uso extendido, entre otras causas, debido a la complejidad de la literatura publicada y de su tardía puesta a disposición en los paquetes estadísticos que hace que o bien se ignoren o bien se utilicen métodos antiguos para tratar valores perdidos (Enders, 2010). Según fuentes de la *American Psychological Association* (APA) (en 2018 y 2019) el 15% de estudios no indica cómo han sido tratados los datos, y, de los que lo hacen, el 38% usa el método FIML, el 31% eliminación y el 19% imputación múltiple. Solo el 12% que informaba de datos perdidos explicitaba los mecanismos de datos perdidos (Rioux & Little, 2021b).

El problema está en que, LW (o una sencilla imputación de datos perdidos por la media), si bien pueden, en ocasiones, producir estimaciones insesgadas (cuando la pérdida de datos se debe a mecanismos MCAR), bajo mecanismos MAR, producen producir estimaciones sesgadas. Debido a que el investigador aplicado difícilmente conocerá el mecanismo responsable de la pérdida de datos es relevante conocer las ventajas que ofrece la estimación FIML o MI (de calidad similar) frente a LW para los distintos mecanismos de pérdida de datos en estudios longitudinales (ver Enders, 2010; Little & Rhemtulla, 2013).

Existen escasas investigaciones metodológicas que se centren exclusivamente en estudiar el efecto de la pérdida de datos en estudios longitudinales. A pesar de que algunos estudios disponibles incluyen información sobre el porcentaje de datos perdidos, incluso se incluye entre las condiciones manipuladas, no se comparan los diferentes mecanismos de pérdida de datos. Por ejemplo, en una revisión de 262 estudios sobre epidemiología, solo 14 informaron del tipo de mecanismo concreto y siempre fue MAR (Eekhout et al., 2012). Asimismo, en McNeish (2017), uno de los escasos estudios sobre MCL y tratamiento de datos perdidos con diferentes tamaños muestrales, solo se simula el mecanismo MAR.

La literatura ha mostrado que estos tipos de mecanismos no funcionan igual en la recuperación de parámetros (Enders, 2010, Little & Rhemtulla, 2013), por lo que se hace necesario nueva evidencia que tenga en cuenta esta variable. Concretamente, con estudios de simulación, que permitan hacer un ensayo con mínimo coste de las diferentes condiciones combinando mecanismos y tratamiento de datos perdidos con diferentes muestras para contribuir a poner luz en las cuestiones que no están claras.

Una contribución necesaria es explorar cómo afectan los datos incompletos a las investigaciones con estudios longitudinales basados en MCL, al ser los modelos más flexibles. Para ello es necesario responder cuestiones como las que se abordan en el presente trabajo: ¿cómo afectan los diferentes mecanismos de pérdida de datos a la estimación de los parámetros de los MCL? ¿Es relevante el tamaño muestral? ¿Cómo afectan al sesgo las diferentes soluciones?

Objetivos e hipótesis

En este trabajo se estudia cómo los distintos mecanismos de pérdida de datos afectan a la recuperación de parámetros en los MCL y en qué medida mejora la estimación al comparar el tratamiento clásico LW frente a uso de la estimación por información completa FIML, en función del mecanismo de pérdida de datos.

Para ello se han generado conjuntos de datos a partir de técnicas de simulación en los que se han manipulado condiciones de datos incompletos que se han tratado con diferentes métodos.

Se espera un efecto de los datos incompletos en la recuperación de los parámetros simulados y en la bondad de ajuste del modelo, y que la estimación de parámetros sea distinta dependiendo de cuál sea el mecanismo de pérdida. En concreto, la pérdida de datos con un mecanismo MAR afectaría más a la recuperación de parámetros que la pérdida de datos MNAR, pero, tal y como se menciona en la literatura (Enders, 2010; Little & Rhemtulla, 2013), con un escenario MAR, se podrían compensar estos desajustes en la estimación, aplicando un método de tratamiento de datos perdidos (en este caso FIML). En un escenario MCAR, no se esperan desajustes en la recuperación de parámetros independientemente del método de tratamiento de datos perdidos.

Tabla 1. Diseño de simulación

	Modelo 1	Modelo 2
1. Intersección media (μ_0)	0.5	1
2. Pendiente media (μ_s)	0.5	1
3. Varianza Intersección (σ_0^2)	1	0.5
4. Varianza pendiente (σ_s^2)	1	0.5
5. Covarianza entre variables latentes (σ_{0s}^2)	0.5	0.5
6. Tamaño muestral (n)	100, 200, 500, 1000	
7. Número de observaciones temporales	4	
8. Mecanismo de valores perdidos	MAR, MNAR, MCAR	
9. Método de tratamiento de valores perdidos	Eliminación por lista (LW), Máxima verosimilitud con información completa (FIML)	

Nota. Se realizarán los análisis con la media y la desviación típica de los resultados de las estimaciones de las 1000 réplicas simuladas

En relación con el tamaño muestral, se espera que, a mayor tamaño muestral, mayor precisión en la recuperación de parámetros.

Se estudiará la precisión de las estimaciones (sesgo) y el ajuste de los diferentes modelos, para profundizar en la interpretación de los resultados.

Método

Se han aplicado las recomendaciones de Singer y Willet (2003) para la aplicación del modelo estadístico de crecimiento a datos simulados ad hoc. En los siguientes apartados se exponen las condiciones que se han simulado, el procedimiento y el software que se ha usado para el análisis.

Generación de los datos

Los datos fueron generados con el software *R*, mediante la ID RStudio (RStudioTeam, 2020).

Condiciones simuladas

Se especificaron dos MCL en los que se fijaron diferentes valores para los parámetros poblacionales. Los efectos fijos: variable latente intersección (μ_0) y variable latente pendiente (μ_s), los efectos aleatorios: varianza de la intersección (σ_0^2) y varianza de la pendiente (σ_s^2) y la covarianza entre intersección y pendiente (σ_{0s}^2). El Modelo tipo 1 representaba MCL con pendientes moderadas, mientras que el Modelo tipo 2 supone pendientes abruptas, como se especificará más adelante.

Para valorar la recuperación de los parámetros se simularon diferentes modelos: modelo completo (sin valores perdidos), tres mecanismos de pérdida de datos: MAR, MNAR y MCAR. Cada uno de estos mecanismos de pérdida de datos se divide en dos, en función del método de

tratamiento (método con eliminación por lista y método de máxima verosimilitud con información completa).

Se generaron bases de datos en *R* simulando datos longitudinales extraídos de cuatro mediciones/observaciones longitudinales con cuatro tamaños muestrales (n=100, 200, 500 y 1000) y con 1000 réplicas para cada modelo (ver Kim et al., 2018, Whittaker & Khojasteh, 2017), suficientes para poder evaluar muestras pequeñas de solo 100 casos y que son habituales en este tipo de estudios (Curran et al., 2010), si bien para un funcionamiento adecuado de FIML se recomiendan muestras mayores e idealmente de 1000 casos (Maydeu-Olivares, 2017).

Por tanto, se simularon 7 condiciones x 2 modelos poblacionales, un total de 14 condiciones para los 4 tamaños muestrales (diseño 7x2x4). En la Tabla 1 se presenta un resumen de las condiciones de simulación y los parámetros poblacionales de los dos modelos principales.

Procedimiento

Se ha llevado a cabo un estudio de simulación de MCL tomando cuatro observaciones temporales, en línea con la propuesta de Graham, 2012, y teniendo en cuenta que para conseguir una estimación de parámetros sin constricciones, hace falta un mínimo de tres observaciones/mediciones (Whittaker & Khojasteh, 2017). Se especificaron dos posibles modelos poblacionales (Modelo tipo 1 de crecimiento moderado y Modelo tipo 2 de crecimiento abrupto) con dos variables latentes cada uno (una intersección y una pendiente aleatorias). Además de la covarianza entre ambas.

En cuanto a la especificación de los dos modelos en *R* en el Modelo tipo 1, los efectos fijos se establecieron en 0.5 para la intersección

Tabla 2. Valores perdidos en cada modelo

	nMAR		nMNAR		nMCAR	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2	Modelo 1	Modelo 2
n=100	50.06	49.99	50.01	49.93	49.94	49.94
n=200	100.15	100.19	100.3	100.32	100.34	100.34
n=500	249.82	249.81	249.66	249.66	249.85	249.85
n=1000	500.39	499.88	499.945	499.87	499.55	499.55

media (μ_0) y para la pendiente media (μ_s). Los efectos aleatorios se establecieron en 1 para la varianza de la intersección σ_0^2 y para la varianza de la pendiente σ_s^2 . La covarianza entre la intersección y la pendiente σ_{0s}^2 se estableció en .5.

En el Modelo tipo 2, los efectos fijos se fijaron en 1, los efectos aleatorios en 0.5 y la covarianza en 0.5. Estos valores, aunque arbitrarios, se eligieron considerando dos situaciones: una de crecimiento moderado de medio punto entre observaciones para una varianza de uno y otro escenario de gran crecimiento con incremento de un punto entre observaciones para una varianza de 0.5.

Respecto a la pérdida de datos, se simula la condición de datos completos (sin valores perdidos) y otras tres para los mecanismos que causantes de los valores perdidos, produciéndose estos en la tercera medición (y_3) para los mecanismos MAR y MNAR y en la cuarta medición (y_4) para el MCAR.

En la Tabla 2 se expone el número de valores perdidos por cada modelo para cada muestra (aproximadamente 50% en cada caso). La media de datos perdidos en estudios longitudinales del ámbito educativo y psicológico es entre el 15- 20% (Enders, 2003), con lo que la pérdida simulada se situaría aproximadamente en esta media.

En cada tipo de modelo se computaron las estimaciones medias sobre 1000 réplicas para los diferentes escenarios: modelo completo (C), MAR, MNAR y MCAR.

Para la recuperación de los MCL se utilizó el paquete *lavaan* (Rosseel, 2012). Para cada mecanismo de pérdida de datos (MAR, MNAR y MCAR) se evaluó la bondad de ajuste y la calidad de la recuperación de los parámetros intersección, pendiente y covarianza, así como la dispersión (desviación típica), con el tratamiento de valores perdidos por defecto en *R* (eliminación por lista, LW) y con el método de estimación de parámetros máxima verosimilitud con información completa (FIML). Se incluyó la recuperación de parámetros

con LW al ser el que los paquetes estadísticos incluyen por defecto

Con el fin de evaluar la precisión de las estimaciones de los parámetros, se calcularon el porcentaje de sesgo y el error cuadrático medio (*Root Mean Square Error*, RMSE) para todos los modelos simulados. Para ambos índices, los valores bajos muestran precisión de la estimación.

El porcentaje de sesgo se define como la diferencia entre los valores medios en las réplicas y el valor estimado poblacional, y se calcula con la fórmula (1).

$$\% \text{sesgo} = \frac{\sum(\hat{p}-p)}{Nr} \times 100$$

Siendo \hat{p} el valor estimado para cada coeficiente, p el valor simulado y Nr el número de réplicas. Por tanto $(\hat{p} - p)$ son las diferencias entre el valor estimado y el simulado (se calcula para cada valor estimado). Dichos valores pueden ser positivos o negativos.

Por otro lado, el RMSE se refiere a la distribución de los datos en torno a la línea del mejor ajuste. En concreto es la desviación típica de los residuos (o errores de predicción) e indica el error medio de predicción.

Se calcula mediante (2).

$$RMSE = \sqrt{\sum(\hat{p} - p)^2 / Nr}$$

Los valores recomendados para una buena recuperación son: un % de sesgo <5% y un RMSE $\leq .05$ (Hoogland & Boomsma, 1998).

Tanto RMSE, como el porcentaje de sesgo se calcularon automáticamente en *R* para las 1000 réplicas para todos los escenarios.

Se calculó el ajuste del modelo a los datos simulados a través varios índices de ajuste: χ^2 (a un nivel de significación de 5%) y *Root Mean Square Error of Approximation* (RMSEA) (Steiger & Lind, 1980), *Comparative Fit Index* (CFI) (Bentler, 1990) y *Tucker-Lewis Index* (TLI) (Tucker & Lewis, 1973). Finalmente, se valoró el ajuste con los valores de referencia publicados en la guía de Hooper, Coughlan y Mullen (2008).

Esto es, para un buen ajuste se consideró: $RMSEA \leq .07$ (Steigler, 2007) y $< .03$, como excelente y $CFI \geq .95$ y $TLI \geq .95$.

Análisis de datos

Los análisis se realizaron en *R* con el paquete *lavaan* (Rosseel, 2012) que permite la estimación de MCL y realiza por defecto el tratamiento de los valores perdidos con el método de eliminación por lista (LW). Para el tratamiento de los datos incompletos con máxima verosimilitud con información completa (FIML) se incluye, en un solo paso, el argumento *missing="ML"*, que se puede usar con la función de ajuste *growth*. Para obtener los índices de ajuste se usó el argumento *fit.measures=TRUE*.

Resultados

Tanto los MCL tipo 1 como el MCL tipo 2 convergieron bajo las distintas condiciones simuladas. A continuación, se presentan los resultados de la recuperación de los parámetros de cada modelo, junto con el porcentaje de sesgo y el RMSE, además de los principales índices de bondad de ajuste.

Recuperación de parámetros y precisión de las estimaciones

MCL tipo 1

En la Tabla 3 pueden verse los valores medios de las réplicas simuladas en función del tamaño muestral simulado para MCL tipo 1 (Modelo 1).

Bajo MAR y MNAR, se recuperan mal todos los parámetros con el tratamiento de eliminación por lista. Con el tratamiento FIML se corrige el problema en todos los casos, aunque en FIML no todas las estimaciones alcanzan el valor fijado de .5 para RMSE.

Bajo MCAR se recuperan bien los parámetros con eliminación por lista y en algunos casos peor con FIML. No existen diferencias en la recuperación de parámetros para los diferentes tamaños muestrales. No obstante, según aumenta el tamaño muestral, va disminuyendo la dispersión en cuanto a los valores de las estimaciones para las distintas réplicas. Esto se hace más evidente en $n=500$ y 1000 . FIML siempre reduce la dispersión respecto al tratamiento de eliminación por lista.

En cuanto al porcentaje de sesgo, es elevado en todos casos (bastante mayor de 5%) bajo MAR y MNAR con el tratamiento por eliminación por lista. Con FIML se reduce el sesgo, y en algunos casos resulta negativo. La dispersión del sesgo en las distintas réplicas disminuye considerablemente a partir de una muestra de $n=500$.

Bajo MAR y MNAR con tratamiento de eliminación por lista, RMSE está por encima de los límites recomendados ($\geq .05$) en $n=100$, 200 para todos los parámetros. En $n=500$ y $n=1000$, mejora en el caso de la covarianza (fijada a .5 en el modelo). La dispersión entre estimaciones es pequeña en estas muestras grandes.

MCL tipo 2

En la Tabla 4 pueden verse los valores medios de las réplicas simuladas para cada tamaño muestral para el MCL tipo 2 (Modelo 2).

Independientemente del tamaño muestral, en todos los casos no se recuperan bien los parámetros intersección y pendiente (fijados a 1 en el modelo poblacional simulado) bajo MAR y MNAR por eliminación por lista, el tratamiento con FIML soluciona el problema y hace que el ajuste sea perfecto (1.0) en todos los casos MAR y aproximado (.98-.99) para algunos casos MNAR. No obstante, se manifiestan dificultades para estimar la covarianza en todos los casos y escenarios (modelo completo y mecanismos MAR, MNAR, MCAR), independientemente del tamaño muestral y del tratamiento de valores perdidos).

En cuanto a la dispersión de estas estimaciones en las distintas réplicas, existe menor variabilidad con el tratamiento FIML en general. La variabilidad desciende drásticamente a partir de un tamaño muestral de $n=500$.

En todas las muestras, en los efectos fijos (intersección y pendiente), el porcentaje de sesgo es nuevamente enorme bajo MAR y MNAR con el tratamiento por eliminación por lista, y disminuye bajo FIML. Se observan, por tanto, sesgos menores con el tratamiento FIML, especialmente bajo MAR. En el caso de la covarianza, tanto para el modelo completo como para el caso de modelos los sesgos son generalmente negativos y bastante altos.

Tabla 3. Modelo 1. Estimación de parámetros, sesgo y RMSE

		i						s						c											
		C		MAR		MNAR		MCAR		C		MAR		MNAR		MCAR		C		MAR		MNAR		MCAR	
		LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML
Parámetros	M	.49	.04	.49	.14	.48	.49	.48	.50	.17	.50	.06	.49	.50	.49	.50	.33	.50	.32	.50	.49	.50	.49	.50	
n=1000	SD	.13	.20	.13	.19	.13	.19	.13	.11	.15	.11	.15	.11	.15	.11	.16	.27	.17	.27	.17	.27	.17	.24	.17	
%Sesgo	M	0.68	46.42	0.74	35.82	1.92	1.01	1.92	0.08	33.23	0.12	44.13	1.31	0.14	1.31	0.32	17.06	0.39	18.18	0.42	0.56	0.42	0.56	0.47	
	SD	13.19	19.79	13.25	19.08	13.26	19.09	13.26	1.71	15.02	1.76	15.20	1.84	15.21	1.84	16.21	26.96	16.68	27.44	16.71	24.21	16.71	24.21	16.71	
RMSE	M	.02	.25	.02	.16	.02	.04	.02	.01	.13	.01	.22	.01	.02	.01	.03	.10	.03	.11	.03	.06	.03	.06	.03	
	SD	.02	.20	.02	.15	.02	.05	.02	.02	.11	.02	.14	.02	.03	.02	.04	.15	.04	.15	.04	.08	.04	.08	.04	
Parámetros	M	.50	.04	.50	.14	.49	.49	.49	.50	.17	.50	.06	.49	.50	.49	.50	.33	.50	.32	.50	.50	.50	.50	.50	
n=200	SD	.09	.13	.09	.13	.09	.13	.09	.08	.11	.08	.11	.08	.11	.08	.11	.18	.11	.18	.11	.16	.11	.16	.11	
%Sesgo	M	0.14	45.61	0.10	35.57	1.32	0.54	1.32	-0.08	33.43	-0.11	44.10	1.06	0.30	1.06	0.37	17.41	0.44	17.52	0.37	0.34	0.37	0.34	0.37	
	SD	9.35	13.48	9.40	13.20	9.41	13.14	9.41	7.55	1.72	7.55	1.78	7.61	1.74	7.61	11.16	17.65	11.29	18.28	11.42	15.83	11.42	15.83	11.42	
RMSE	M	.01	.23	.01	.14	.01	.02	.01	.01	.12	.01	.21	.01	.01	.01	.01	.06	.01	.06	.01	.03	.01	.03	.01	
	SD	.01	.13	.01	.10	.01	.02	.01	.01	.07	.01	.10	.01	.02	.01	.02	.08	.02	.08	.02	.04	.02	.04	.02	
Parámetros	M	.50	.04	.50	.15	.49	.50	.49	.50	.17	.50	.06	.49	.50	.49	.50	.34	.50	.34	.50	.51	.50	.50	.50	
n=500	SD	.06	.08	.06	.08	.06	.08	.06	.05	.07	.05	.07	.05	.07	.05	.07	.11	.07	.12	.07	.10	.07	.10	.07	
%Sesgo	M	-0.23	45.59	-0.25	35.39	0.94	-0.04	0.94	0.00	33.26	-0.02	43.91	1.19	-0.06	1.19	-0.07	15.57	-0.07	16.02	-0.11	-0.54	-0.11	-0.54	-0.11	
	SD	5.76	7.63	5.77	7.93	5.77	8.10	5.77	4.91	6.91	4.92	6.86	4.92	7.07	4.92	7.13	11.44	7.20	11.61	7.20	1.28	7.20	1.28	7.20	
RMSE	M	.00	.21	.00	.13	.00	.01	.00	.00	.12	.00	.20	.00	.00	.00	.01	.04	.01	.04	.01	.01	.01	.01	.01	
	SD	.00	.17	.00	.06	.00	.01	.00	.00	.05	.00	.06	.00	.01	.00	.01	.04	.01	.04	.01	.01	.01	.01	.01	
Parámetros	M	.50	.04	.50	.15	.49	.50	.49	.50	.16	.50	.06	.49	.50	.49	.50	.34	.50	.34	.50	.50	.50	.50	.50	
n=1000	SD	.04	.06	.04	.06	.04	.06	.04	.04	.05	.04	.05	.04	.05	.04	.05	.08	.05	.08	.05	.07	.05	.07	.05	
%Sesgo	M	-0.13	45.54	-0.14	35.32	1.06	-0.16	1.06	0.01	33.75	0.00	44.27	1.20	0.18	1.20	-0.26	15.89	-0.25	16.23	-0.26	-0.08	-0.26	-0.08	-0.26	
	SD	3.88	5.61	3.90	5.57	3.90	5.54	3.90	3.56	5.04	3.57	5.02	3.59	5.12	3.59	4.86	7.95	4.94	8.04	4.95	6.95	4.95	6.95	4.95	
RMSE	M	.00	.21	.00	.13	.00	.00	.00	.00	.12	.00	.20	.00	.00	.00	.00	.03	.00	.03	.00	.00	.00	.00	.00	
	SD	.00	.05	.00	.04	.00	.00	.00	.00	.03	.00	.04	.00	.00	.00	.00	.03	.00	.03	.00	.01	.00	.01	.00	

i: intersección poblacional media simulada modelo 1 (0.5); s: pendiente poblacional media simulada modelo 1 (0.5) ; c: covarianza poblacional media simulada modelo 1 (1); n: tamaño muestral; C: modelo completo; MAR: *Missing at Random*; MNAR: *Missing Not at Random*; MCAR: *Missing Completely at Random*; LW: *listwise deletion* (eliminación por lista); FIML: *Full Information Maximum Likelihood* (máxima verosimilitud con información completa); M: media; SD: desviación típica; %sesgo $\geq |5\%$; RMSE $\geq .05$: error cuadrático medio (*Root Mean Square of Error*).

Tabla 4. Modelo 2. Estimación de parámetros, sesgo y RMSE

		i						s						c											
		C		MAR		MNAR		MCAR		C		MAR		MNAR		MCAR		C		MAR		MNAR		MCAR	
		LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML	LW	FIML
Parámetros	M	1.00	.63	1.00	.68	.98	1.00	.98	1.00	.74	1.00	.68	1.00	1.00	1.00	.68	.53	.64	.48	.64	.68	.64	.68	.64	
n=100	SD	.11	.16	.11	.15	.11	.16	.11	.08	.12	.09	.12	.09	.12	.09	.22	.41	.21	.42	.21	.32	.21	.32	.21	
%Sesgo	M	0.24	36.90	0.19	32.33	1.54	0.47	1.54	0.27	25.64	0.30	32.29	0.43	0.38	0.43	18.17	-3.32	-14.16	2.13	-14.19	-17.95	-14.19	-17.95	-14.19	
	SD	11.03	15.90	11.08	15.49	11.08	15.67	11.08	8.42	12.14	8.52	12.19	8.63	11.88	8.63	21.72	41.24	21.21	42.43	21.19	31.81	21.19	31.81	21.19	
RMSE	M	.01	.16	.01	.13	.01	.02	.01	.01	.08	.01	.12	.01	.01	.01	.08	.17	.06	.18	.07	.13	.07	.13	.07	
	SD	.02	.12	.02	.11	.02	.03	.02	.01	.07	.01	.08	.01	.02	.01	.09	.22	.08	.24	.08	.16	.08	.16	.08	
Parámetros	M	1.00	.62	1.00	.67	.99	.99	.99	1.00	.75	1.00	.68	1.00	1.00	1.00	.67	.54	.63	.50	.63	.67	.63	.67	.63	
n=200	SD	.08	.11	.08	.11	.08	.11	.08	.06	.08	.06	.08	.06	.08	.06	.14	.28	.14	.30	.14	.21	.14	.21	.14	
%Sesgo	M	0.10	37.62	0.02	32.85	1.37	0.61	1.37	-0.03	25.23	-0.04	31.76	0.03	0.19	0.03	-16.91	-4.18	-12.89	-0.50	-12.83	-16.98	-12.83	-16.98	-12.83	
	SD	7.67	11.12	7.71	1.99	7.71	1.75	7.71	5.80	8.38	5.87	8.33	5.90	8.38	5.90	14.14	27.59	13.82	3.06	13.82	2.84	13.82	2.84	13.82	
RMSE	M	.01	.15	.01	.12	.01	.01	.01	.00	.07	.00	.11	.00	.01	.00	.05	.08	.04	.09	.04	.07	.04	.07	.04	
	SD	.01	.09	.01	.07	.01	.02	.01	.00	.04	.00	.05	.00	.01	.00	.05	.11	.04	.13	.04	.09	.04	.09	.04	
Parámetros	M	1.00	.63	1.00	.67	.99	1.00	.99	1.00	.75	1.00	.68	1.00	1.00	1.00	.68	.57	.64	.54	.64	.68	.64	.68	.64	
n=500	SD	.05	.07	.05	.07	.05	.07	.05	.04	.05	.04	.05	.04	.05	.04	.09	.17	.09	.18	.09	.13	.09	.13	.09	
%Sesgo	M	-0.13	37.16	-0.23	32.62	1.11	0.07	1.11	-0.03	25.27	-0.06	31.60	0.08	-0.09	0.08	-18.05	-7.26	-14.11	-3.84	-14.07	-18.27	-14.07	-18.27	-14.07	
	SD	4.89	7.16	4.91	6.90	4.91	6.99	4.91	3.67	5.17	3.69	5.19	3.71	5.32	3.71	8.92	16.70	8.71	18.13	8.74	12.67	8.74	12.67	8.74	
RMSE	M	.00	.14	.00	.11	.00	.00	.00	.00	.07	.00	.10	.00	.00	.00	.04	.03	.03	.03	.03	.05	.03	.05	.03	
	SD	.00	.05	.00	.05	.00	.01	.00	.00	.03	.00	.03	.00	.00	.00	.03	.04	.03	.05	.03	.05	.03	.05	.03	
Parámetros	M	1.00	.63	1.00	.67	.99	1.00	.99	1.00	.75	1.00	.68	1.00	1.00	1.00	.68	.56	.64	.53	.64	.68	.64	.68	.64	
n=1000	SD	.03	.05	.03	.05	.03	.05	.03	.03	.04	.03	.04	.03	.04	.03	.06	.12	.06	.13	.06	.09	.06	.09	.06	
%Sesgo	M	-0.07	37.16	-0.15	32.55	1.22	0.04	1.22	-0.01	25.40	-0.03	31.91	0.09	0.04	0.09	-17.64	-6.45	-13.71	-3.20	-13.67	-17.60	-13.67	-17.60	-13.67	
	SD	3.47	5.15	3.49	4.97	3.49	5.08	3.49	2.62	3.67	2.63	3.72	2.64	3.68	2.64	6.30	11.80	6.25	12.79	6.25	8.87	6.25	8.87	6.25	
RMSE	M	.00	.14	.00	.11	.00	.00	.00	.00	.07	.00	.10	.00	.00	.00	.04	.02	.02	.02	.02	.04	.02	.04	.02	
	SD	.00	.04	.00	.03	.00	.00	.00	.00	.02	.00	.02	.00	.00	.00	.02	.02	.02	.02	.02	.03	.02	.03	.02	

i: intersección poblacional media simulada modelo 2 (1); s: pendiente poblacional media simulada modelo 2 (1) ; c: covarianza poblacional media simulada modelo 2 (0.5); n: tamaño muestral; C: modelo completo; MAR: *Missing at Random*; MNAR: *Missing Not at Random*; MCAR: *Missing Completely at Random*; LW: *listwise deletion* (eliminación por lista); FIML: *Full Information Maximum Likelihood* (máxima verosimilitud con información completa); M: media; SD: desviación típica; %sesgo $\geq |5\%$; RMSE $\geq |5\%$: error cuadrático medio (*Root Mean Square of Error*).

La dispersión del sesgo en las distintas réplicas disminuye considerablemente a partir de una muestra de n=500, aunque en menor medida que en el modelo 1 y no para la estimación de la covarianza, donde en MAR y MNAR la variabilidad es alta en el caso de eliminación por lista, en concreto.

Los valores de RMSE están por encima de los valores recomendados en MAR y MNAR con eliminación por lista en todos los casos para la estimación de la interacción y la pendiente. Para la covarianza, los valores están también por encima de los recomendados en todos los escenarios, completo y de pérdida de datos, y con

todos los tratamientos para n=100. En n=200, mejoran con FIML. Con n=500 los valores son ajustados y RMSE está en el límite solo en MCAR con tratamiento por eliminación por lista. Con n=1000 los valores son adecuados. Por tanto, en la estimación de la covarianza, RMSE es sensible a los cambios en el tamaño muestral. La dispersión es pequeña en todos los casos, aunque con el aumento de la muestra se acerca a cero.

Bondad de ajuste

Como se muestra en la Tabla 5 para el modelo 1 en los diferentes mecanismos de pérdida de

Tabla 5. Valores medios de Índices de Ajuste al Modelo 1

	C	MAR		MNAR		MCAR	
		LW	FIML	LW	FIML	LW	FIML
n=100							
χ^2	7.28	1.89	7.36	9.96	1.22	7.46	1.22
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.48	.29	.47	.32	.30	.46	.30
CFI	.99	.96	.99	.97	.98	.99	.98
TLI	1.00	.97	1.00	.98	.99	1.00	.99
RMSEA	.03	.09	.03	.08	.06	.04	.06
n=200							
χ^2	7.07	14.24	7.34	11.73	12.32	7.20	12.32
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.50	.17	.48	.24	.21	.48	.21
CFI	1.00	.97	1.00	.98	.99	.99	.99
TLI	1.00	.98	1.00	.98	.99	1.00	.99
RMSEA	.02	.09	.02	.07	.05	.03	.05
n=500							
χ^2	6.98	23.08	6.97	17.85	19.60	7.03	19.60
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.51	.03	.50	.09	.07	.50	.07
CFI	1.00	.98	1.00	.98	.99	1.00	.99
TLI	1.00	.98	1.00	.99	.99	1.00	.99
RMSEA	.01	.09	.01	.07	.06	.02	.06
n=1000							
χ^2	6.92	38.42	6.94	28.56	32.21	7.21	32.21
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.50	.00	.50	.01	.01	.49	.01
CFI	1.00	.98	1.00	.98	.99	1.00	.99
TLI	1.00	.98	1.00	.99	.99	1.00	.99
RMSEA	.01	.09	.01	.08	.06	.01	.06

Nota. n: tamaño muestral; c: modelo completo, MAR: *Missing at Random*; MNAR: *Missing Not at Random*; MCAR: *Missing Completely at Random*; LW: listwise *deletion* (eliminación por lista); FIML: *Full Information Maximum Likelihood* (máxima verosimilitud con información completa); $\chi^2 p \geq .05$: Chi cuadrado; df: grados de libertad, pvalue: pvalores; $\leq .07/.03$: *Root Mean Square Error of Approximation*; CFI: *Comparative Fit Index*; $TLI \geq .95$: *Tucker-Lewis Index*.

datos y en todos los tamaños muestrales se observó bondad de ajuste en los distintos índices, excepto con n=1000, donde con un nivel de significación de .05, el modelo no ajusta en MAR con eliminación por lista ($\chi^2=38.42$, $p < .001$), aunque se corrige con el tratamiento FIML ($\chi^2=6.94$, $p=.5$). Con este tamaño muestral, MNAR no presenta bondad de ajuste con ninguno de los métodos: eliminación por lista ($\chi^2=28.56$, $p=.01$) y FIML ($\chi^2=32.21$, $p=.01$). MCAR no presenta bondad de ajuste ni con eliminación por lista ($\chi^2=.21$, $p=.49$), aunque se aproxima, ni con FIML ($\chi^2=32.21$, $p=.01$).

Con n=100, el modelo completo tiene un buen ajuste para CFI, TLI y RMSEA medios (CFI=.90, TLI=1.00 y RMSEA=.03). Bajo MAR, con eliminación por lista, RMSEA toma un valor de .09, mayor de lo recomendado. Con el tratamiento FIML se corrige el problema (RMSEA=.03). En MNAR, con eliminación por lista RMSEA=.08, con FIML no se corrige tan bien, se queda dentro de lo recomendado, pero al límite (RMSEA=.06).

El ajuste en RMSEA mejora con el aumento del tamaño muestral. Así, con n=200 el ajuste es excelente en RMSEA, en el modelo completo (RMSEA=.02), el mismo que con MNAR con FIML (.02). Con n=500 el ajuste en el modelo completo en RMSEA=.01. No obstante, con pérdida de datos MAR con el tratamiento de eliminación por lista, RMSEA está por encima de los valores recomendados (RMSEA=.09), recuperándose con FIML (RMSEA=.01). Con MCAR el ajuste también es excelente (RMSEA=.02), y más ajustado con FIML (RMSEA=.06), con este tipo de pérdida de datos FIML no es tan efectivo. Con n=1000 el modelo tiene un ajuste excelente para el modelo completo (RMSEA=.01), presenta mayores desajustes, que se observan en χ^2 , y en RMSEA, para MAR con eliminación por lista (RMSEA=.09), que se recupera con el tratamiento FIML (.01).

En la Figura 2, se muestra un ejemplo para el Modelo 1 con n=1000. Este ha sido seleccionado por ser en el que la diferencia entre condiciones es más sensible. Si el mecanismo es MNAR, los

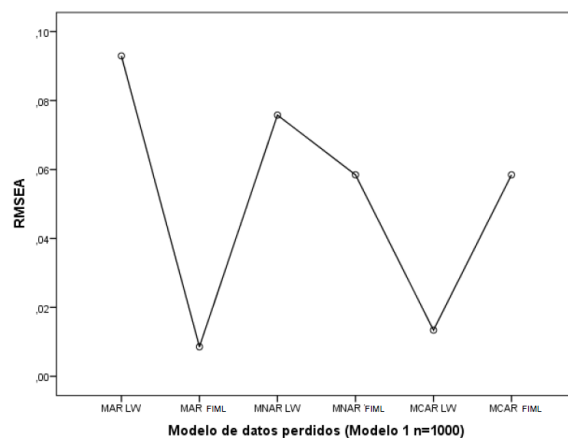


Figura 2. Ajuste al modelo mecanismos de pérdida de datos con diferentes tratamientos

Tabla 6. Valores medios de Índices de Ajuste al Modelo 2

	C	MAR		MNAR		MCAR	
		LW	FIML	LW	FIML	LW	FIML
n=100							
χ^2	22.23	22.59	21.24	22.04	25.01	15.24	25.01
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.03	.03	.04	.03	.02	.12	.02
CFI	.94	.85	.92	.85	.90	.93	.90
TLI	.95	.87	.93	.87	.91	.94	.91
RMSEA	.14	.20	.14	.20	.16	.14	.16
n=200							
χ^2	36.23	37.29	34.18	36.13	41.48	21.96	41.48
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.00	.00	.00	.00	.00	.03	.00
CFI	.94	.86	.93	.86	.90	.94	.90
TLI	.95	.88	.94	.88	.92	.95	.92
RMSEA	.14	.20	.14	.20	.15	.14	.15
n=500							
χ^2	78.94	8.49	73.97	78.27	92.12	43.27	92.12
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.00	.00	.00	.00	.00	.00	.00
CFI	.94	.87	.93	.87	.91	.94	.91
TLI	.95	.89	.94	.89	.92	.95	.92
RMSEA	.14	.20	.14	.20	.16	.14	.16
n=1000							
χ^2	15.56	154.19	14.45	15.08	177.57	79.75	177.57
df	7.00	7.00	7.00	7.00	7.00	7.00	7.00
p value	.00	.00	.00	.00	.00	.00	.00
CFI	.95	.87	.93	.87	.91	.94	.91
TLI	.95	.89	.94	.89	.92	.95	.92
RMSEA	.14	.20	.14	.20	.16	.14	.16

Nota. n: tamaño muestral; c: modelo completo, MAR: *Missing at Random*; MNAR: *Missing Not at Random*; MCAR: *Missing Completely at Random*; LW: *listwise deletion* (eliminación por lista); FIML: *Full Information Maximum Likelihood* (máxima verosimilitud con información completa); $\chi^2 p \geq .05$: Chi cuadrado; df: grados de Libertad, pvalue: p valores; $RMSEA \leq .07/.03$: *Root Mean Square Error of Approximation*; $CFI \geq .95$: *Comparative Fit Index*; $TLI \geq .95$: *Tucker-Lewis Index*.

valores de RMSEA son menores que con MAR (RMSEA=.08) que se recupera, más al límite de los valores recomendados, con FIML (RMSEA=.06).

El ajuste de MCAR es excelente, y nuevamente, mejor con eliminación por lista (RMSEA=.01) que con FIML (RMSEA=.06), a pesar de haber perdido la mitad de la muestra (la media de nMCAR=499.55 en las 1000 réplicas simuladas). El procedimiento FIML parece

contraproducente, al mostrar un ajuste más bajo que el de eliminación por lista.

Para CFI y TLI, en todas las muestras, FIML mejora el ajuste en MAR y MNAR (mejor en MAR), y lo empeora en MCAR o mantiene los mismos valores. Conforme aumenta el tamaño muestral, la corrección FIML consigue un ajuste perfecto en MAR, mejorando los ajustes con eliminación por lista (que van de .96 a .98

progresivamente aumentando según aumenta el tamaño muestral).

En el Modelo 2, debido a lo abrupto de las pendientes simuladas se observaron mayores desajustes que en el Modelo 1 (ver Tabla 6). En relación con el índice χ^2 , se observó bondad de ajuste en todos los mecanismos, excepto MCAR para $n=100$ con tratamiento *listwise* ($\chi^2=15.24$, $p=.12$), para un nivel de significación de .05. El índice CFI solo ajusta para el modelo completo $n=1000$ (CFI=.95), el índice TLI ajusta para el modelo completo en $n=100$, 200, 500 y 1000 (TLI=.95), para MCAR *listwise* con $n=200$, 500 y 1000 (TLI=.95). RMSEA no ajusta en ninguno de los escenarios.

Discusión

En los últimos años las instituciones y sociedades científicas han impulsado iniciativas para asegurar el uso de métodos de investigación adecuados para los problemas a los que se aplican y la publicación rigurosa de sus resultados, informando de las prácticas y decisiones. Esto se traduce en una proliferación de guías de buenas prácticas, y en números especiales en revistas especializadas. Por ejemplo, en los estudios de desarrollo (Rioux & Little, 2021a), en los que han destacado dos métodos infrutilizados en los estudios longitudinales: la especificación de la temporización y las unidades de cambio y el tratamiento de los valores perdidos (Rioux & Little, 2020). En nuestro estudio se han combinado estos dos enfoques, el temporal y el de la pérdida de datos.

Para el problema de los datos incompletos, en estudios previos se han propuesto medidas como los diseños de datos faltantes planeados (*planned missing data designs*), como se muestra en la reciente actualización publicada por Wu & Jia (2021). Esta solución, ha resultado previamente en una reducción de costes, mayor validez y menor cantidad de valores perdidos no planeados (Little & Rhemtulla, 2013). Una solución más económica pasa por realizar estudios de simulación que permitan prever los modelos y las condiciones más favorables para alcanzar los objetivos de cada investigación concreta y así hacer un diseño más parsimonioso antes de su implementación. Nuestro estudio está enfocado en esta línea. Se

discutirán a continuación los resultados de este estudio en relación con sus objetivos e hipótesis planteadas sobre el efecto de los diferentes mecanismos de pérdida de datos y métodos de tratamiento en la recuperación de parámetros de los MCL.

Estimación de parámetros. Se esperaba encontrar diferencias entre los diferentes mecanismos de pérdida de datos en estudios longitudinales. Con este objetivo se especificó un MCL con variantes en los parámetros poblacionales (Modelo 1 y 2) y se estimaron submodelos para los diferentes escenarios con distintos mecanismos. Se encontraron diferencias en los modelos bajo los MAR y MNAR frente a MCAR.

Así, la recuperación de parámetros muestra similitudes en ambos modelos (Modelo 1 y 2). Tanto en uno como en el otro, los valores perdidos bajo mecanismos MAR y MNAR muestran una peor ejecución con el método de eliminación por lista, que mejora bajo la solución del tratamiento de máxima verosimilitud con información completa (FIML) para la estimación de los efectos fijos (interacción y pendiente) (aunque es menos precisa bajo MNAR). Bajo MCAR, en el Modelo 1 la estimación con eliminación es adecuada, y el tratamiento FIML empeora la estimación; no obstante, en el Modelo 2 la estimación es adecuada en ambos casos. Para la covarianza, en el Modelo 2 se recupera mal en todos los mecanismos. En cuanto a MCAR, al tratarse de un mecanismo de pérdida aleatoria, tal y como se esperaba, no se encuentran diferencias con respecto al modelo completo más allá de la pérdida de precisión en la recuperación de parámetros por el menor tamaño muestral

Enders (2010), referente sobre estos temas, consideraba un problema el escenario MNAR, por su dificultad para identificarlo. En realidad, ni MAR ni MNAR son contrastables, pero, en todo caso, se puede saber que no son MCAR y sería suficiente, ya que MCAR es el único mecanismo que funciona diferente, y entre MAR y MNAR las diferencias en la recuperación de parámetro no son importantes. Respecto al método de tratamiento, nuestros resultados confirman, en consonancia con los reportados por Little & Rhemtulla (2013), la ventaja del método FIML frente al método de eliminación en función de la

causa (mecanismo) de pérdida de datos (excepto para MCAR). Cuando se producen desajustes en MAR o MNAR se compensan con FIML, aunque no en todas las condiciones (los resultados para la covarianza en el Modelo 2 lo ponen en duda).

Tamaño muestral y variabilidad de las estimaciones. Respecto al tamaño muestral, en ninguno de los modelos la recuperación promedio de parámetros fue sensible al tamaño muestral en esta simulación (n=100, 200, 500 y 1000).

Si tenemos en cuenta la estimación media de las distintas réplicas (en este caso, 1000), existe una mayor dispersión entre los resultados para cada réplica en muestras hasta n=500 (esta situación se da en ambos modelos, aunque con menor descenso en el Modelo 1 a partir de este punto y no para la covarianza en el Modelo 2, con alta variabilidad en eliminación por lista en MAR y MNAR). El tratamiento FIML reduce esta dispersión en todos los casos, por lo que quedarse solo con la recuperación de parámetros y no tener en cuenta la variabilidad sería subestimar otros indicadores.

Precisión de las estimaciones. Un indicador importante de la precisión es el porcentaje de sesgo, que es elevado bajo MAR y MNAR en todos los casos en ambos modelos, precisamente bajo MAR y MNAR con eliminación por lista, que se resuelve con FIML, incluso se vuelve negativo. En el Modelo 2 destacan sesgos negativos muy altos en la estimación de la covarianza. También RMSE es sensible en estos casos. En el Modelo 2, RMSE es sensible al tamaño muestral en la estimación de la covarianza.

El sesgo se corrige con el uso de métodos modernos (FIML), tanto en MAR como en MNAR. Si se hubiera usado otro método moderno, el método de estimación por imputación múltiple (*Multiple Imputation*, MI), es posible que hubiera sido más pequeño, ya que existe evidencia sobre esta característica en diseños planeados de datos faltantes. Aunque el método FIML y el MI muestren convergencias similares, MI minimiza el sesgo, aunque requiere muestras más grandes (Jia et al., 2014).

La dispersión del sesgo en las distintas réplicas también disminuye a partir de n= 500 en ambos modelos, por tanto, el tamaño muestral afecta a la dispersión de la estimación y del porcentaje de sesgo.

Como indicaron Carpenter & Kenward (2013), existen consecuencias al eliminar los casos con valores perdidos, y esto afecta a la eficiencia y al sesgo. Se confirma que omitir los datos incompletos, comparado con el método FIML (bajo normalidad multivariada) puede aumentar el sesgo, tal y como ya afirmaron Baraldi & Enders (2010). La inferencia poblacional se ve comprometida, como se ha mostrado en los resultados de este estudio, en los problemas para la recuperación de parámetros.

Bondad de ajuste. El cambio de los valores poblacionales en el Modelo 1 y 2 produce cambios en el ajuste. El Modelo 1 presenta un buen ajuste del modelo con χ^2 en general para todos los tamaños muestrales, excepto para n=1000 bajo MAR con eliminación por lista (se resuelve bien con FIML) y bajo MNAR y MCAR que no ajustan con ningún método. Es posible que la falta de ajuste esté mostrando un error tipo I (rechazar H_0 siendo verdadera).

El ajuste en CFI y TLI es bueno en todos los casos. El ajuste en RMSEA mejora con el aumento del tamaño muestral. Este aspecto se había encontrado en investigaciones previas, como la de Hancock & Freeman (2001), donde encuentran que la potencia estadística de RMSEA no es aceptable con muestras inferiores a 150. En resumen, el tamaño muestral favorece al ajuste, en la mayor parte de los escenarios y FIML mejora el ajuste en MAR y MNAR, no en MCAR.

En el Modelo 2 se dan mayores desajustes para χ^2 , RMSEA no ajusta en ningún escenario. Por lo tanto, RMSEA podría ser un índice de ajuste más sensible en estos casos. Se ha observado que RMSEA con pequeños grados de libertad puede conducir a un mal ajuste y no está recomendado su uso en estos casos (Kenny et al., 2015). Así, la probabilidad de que en RMSEA se obtengan valores más bajos es proporcional al tamaño muestral y al número de grados de libertad, esto es, a mayor tamaño muestral, y/o mayores grados de libertad, menores valores de RMSEA y mejor ajuste. En nuestro caso, los grados de libertad eran pequeños (df=7), por lo que es posible que esté condicionando los valores altos de RMSEA, no así el tamaño muestral, ya que no ajusta con ningún tamaño muestral en este modelo.

Finalmente, los tamaños muestrales no tuvieron apenas efecto en la recuperación media

de los parámetros, lo que implica que, en los MCL simulados, el tamaño muestral no permitió compensar el problema subyacente asociado al mecanismo de pérdida de datos.

Entre las limitaciones de la investigación, indicar que se ha realizado bajo el supuesto de normalidad multivariada por lo que sería recomendable realizar simulaciones bajo el incumplimiento de este supuesto, especialmente si se desea comparar la imputación múltiple con la estimación por información completa. También interesaría evaluar el efecto del incremento de casos perdidos con las observaciones que en este estudio fue constante. A su vez, se recomienda estudiar el efecto de soluciones bajo MAR y MNAR, se podría compensar la información perdida incluyendo variables auxiliares. Algunos autores han manifestado que los métodos como FIML solo proporcionan resultados insesgados bajo MAR cuando las variables auxiliares están incluidas en el tratamiento. Es necesario mencionar las variables auxiliares y la mayor parte de los estudios no lo hacen (Rioux & Little, 2021b). En estudios reales, clínicos, se ha mostrado que incluir este tipo de variables puede mitigar el sesgo y disminuir los valores de RMSE (Ayilara et al., 2019).

Las futuras líneas de investigación que se abren a partir de este estudio pasan por ampliar las condiciones y/o explorar nuevas condiciones en estudios de simulación y extenderlas a investigaciones con datos reales. Una aplicación interesante sería aplicar estos modelos en estudios clínicos o educativos sobre la eficacia de programas de intervención, y explorar aspectos concretos de dosificación del tratamiento (número de sesiones) en el incremento de la potencia (Rast & Hofer, 2014), aprovechando la flexibilidad de los MCL para considerar la incorporación de participantes en distintos momentos del estudio. Otra aplicación sería en estudios de desarrollo, en este caso probando modelos no lineales, que describen mejor la naturaleza del aprendizaje y no suelen incluirse en los análisis de este tipo de estudios. Por otro lado, en el ámbito de la evaluación, se podría hacer un diseño de datos faltantes planeados, ya que sus características, al ser una asignación aleatoria son similares a las del mecanismo MCAR, que funciona bien. Con respecto al sesgo, Hamilton et al. (2003) hicieron

un estudio muy completo sobre las tasas medias en el intercepto y en la varianza, comparando estos parámetros y comparando también la variabilidad entre parámetros. Al encontrarse en nuestros datos diferencias entre la recuperación de los distintos parámetros estimados en el Modelo 2 (afectando a la estimación de la covarianza), se podría estudiar este aspecto de forma sistemática. Además de analizar el porcentaje de convergencia al modelo.

Conclusiones

La investigación realizada aporta evidencia a favor de que para los análisis de los datos en diseños longitudinales es fundamental el tratamiento de los datos perdidos. Cuando se ignora el problema de datos perdidos y se aplica una técnica estadística ciega de eliminación de casos, no solo se ve afectada la precisión de las estimaciones de los parámetros, sino que se corre el riesgo de realizar afirmaciones incorrectas, lo que afectaría a distintos aspectos de la validez, desde la validez estructural, pasando por las posibles evidencias en relación a otras variables y criterios, y finalmente a la validez consecuencial, especialmente si se toman decisiones erróneas o se llega a conclusiones equivocadas.

Referencias

- Allison, P. D. (2001). *Missing data*. Sage Publications.
- Ayilara, O. F., Zhang, L., Sajobi, T. T., Sawatzky, R., Bohm, E., & Lix, L. M. (2019). Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry. *Health and Quality of Life Outcomes*, 17(1), 1-9. <https://doi.org/10.1186/s12955-019-1181-2>
- Baker, G. A. (1954). Factor analysis of relative growth. *Growth*, 18, 137-143.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5-37. <https://doi.org/10.1016/j.jsp.2009.10.001>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>

- Bock, R. D. (1979). Univariate and multivariate analysis of variance with time-structured data. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*, 199-232. Academic Press.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons.
- Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3), 473-514. <https://doi.org/10.1214/06-ba117>
- Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research. *Infant and Child Development*, e2296. <https://doi.org/10.1002/icd.2296>
- Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. John Wiley & Sons.
- Cernat, A. (2021, 29 de Abril). Estimating and visualizing change in time using Latent Growth Models with R <https://www.alexcernat.com/estimating-and-visualizing-change-in-time-using-latent-growth-models-with-r/>
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, 11(2), 121-136. <https://doi.org/10.1080/15248371003699969>
- Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable*. Psychology Press. Taylor & Francis.
- Eekhout, I., de Boer, R. M., Twisk, J. W., de Vet, H. C., & Heymans, M. W. (2012). Missing data: A systematic review of how they are reported and handled. *Epidemiology*, 23(5), 729-732. <https://doi.org/10.1097/ede.0b013e3182576cdb>
- Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods*, 8(3), 322. <https://doi.org/10.1037/1082-989x.8.3.322>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: A comparison of approaches. *Health and Quality of Life Outcomes*, 7(1), 1-10. <https://doi.org/10.1186/1477-7525-7-57>
- Geiser, C. (2021). *Longitudinal structural equation modeling with Mplus: A latent state-trait perspective*. Guilford Press.
- Gottman, J. M., & Rushe, R. H. (1993). The analysis of change: Issues, fallacies, and new ideas. *Journal of Consulting and Clinical Psychology*, 61(6), 907-910. <https://doi.org/10.1037/0022-006x.61.6.907>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549-576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science & Business Media.
- Grimm, K. J., Ram, & N. Estabrook, R. (2017). *Growth Modeling. Structural Equation and Multilevel Modeling Approaches*. The Guilford Press.
- Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003). The effect of sample size on latent growth models. Presentado en el *Annual Meeting of the American Educational Research Association* (Chicago, IL, April, 21-25, 2003)
- Hancock, G. R., & Freeman, M. J. (2001). Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement*, 61(5), 741-758. <https://doi.org/10.1177/00131640121971491>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329-367. <https://doi.org/10.1177/0049124198026003003>
- Hooper, D., Coughlan, J., Mullen, M., & Hooper, D. (2008) Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6 (1). <https://doi.org/10.21427/D7CF7R>
- IBM Corp. (2012). *IBM SPSS Statistics for Windows*, Version 21.0. Armonk, NY: IBM Corp.

- Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *International Journal of Behavioral Development*, 38(5), 435-452. <https://doi.org/10.1177/0165025414531095>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486-507. <https://doi.org/10.1177/0049124114543236>
- Kim, M., Hsu, H. Y., Kwok, O. M., & Seo, S. (2018). The optimal starting model to search for the accurate growth trajectory in latent growth models. *Frontiers in Psychology*, 9, 349. <https://doi.org/10.3389/fpsyg.2018.00349>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Lawes, M., Schultze, M., & Eid, M. (2020). Making the most of your research budget: Efficiency of a three-method measurement design with planned missing data. *Assessment*, 27(5), 903-920. <https://doi.org/10.1177/1073191118798050>
- Lee, T., & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, 26(4), 466-485. <https://doi.org/10.1037/met0000381>
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, 7(4), 199-204. <https://doi.org/10.1111/cdep.12043>
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 383-394. <https://doi.org/10.1080/10705511.2016.1269606>
- McNeish, D. (2017). Missing data methods for arbitrary missingness with small samples. *Journal of Applied Statistics*, 44(1), 24-39. <https://doi.org/10.1080/02664763.2016.1158246>
- McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods*, 50(4), 1398-1414. <https://doi.org/10.3758/s13428-017-0976-5>
- Miranda, H., Saracostti, M., Miranda-Zapata, E., & Lara, L. (2021). Análisis longitudinal del rendimiento escolar según el compromiso escolar y factores contextuales: El caso chileno. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 2(59), 163-175. <https://doi.org/10.21865/ridep59.2.13>
- Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1), 1-17. <https://doi.org/10.2307/2527726>
- Rast, P., & Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychological Methods*, 19(1), 133. <https://doi.org/10.1037/a0034524>
- Rioux, C., & Little, T. D. (2020). Underused methods in developmental science to inform policy and practice. *Child Development Perspectives*, 14(2), 97-103. <https://doi.org/10.1111/cdep.12364>
- Rioux, C., & Little, T. D. (2021a). Introduction to the special section on developmental approaches to prevention science. *International Journal of Behavioral Development*, 45(1), 1-2. <https://doi.org/10.1177/0165025420943677>
- Rioux, C., & Little, T. D. (2021b). Missing data treatments in intervention studies: what was, what is, and what should be. *International Journal of Behavioral Development*, 45(1), 51-58. <https://doi.org/10.1177/0165025419880609>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rovine, M. J., & McDermott, P. A. (2018). Latent growth curve and repeated measures ANOVA contrasts: What the models are telling you. *Multivariate behavioral research*, 53(1), 90-101.

- <https://doi.org/10.1080/00273171.2017.1387511>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
<https://doi.org/10.1093/biomet/63.3.581>
- RStudio Team (2020). RStudio: Integrated development for R. Rstudio, PBC, Boston, MAC. URL: <http://www.rstudio.com>
- Shi, D., DiStefano, C., Zheng, X., Liu, R., & Jiang, Z. (2021). Fitting latent growth models with small sample sizes and non-normal missing data. *International Journal of Behavioral Development*, 45(2), 179-192.
<https://doi.org/10.1177/0165025420979365>
- Singer, J. D., & Willett, J. B. (2003). A framework for investigating change over time. En: J. D. Singer, J. B. Willet. *Applied longitudinal data analysis: Modeling change and event occurrence*. (pp. 1-17). Oxford University Press.
- Solano, A. C., & Casullo, M. M. (2005). Análisis del cambio en los estilos de liderazgo de un grupo de estudiantes militares. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 1(19), 105-120.
- Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. Presentado en *Annual Meeting of the Psychometric Society*, Iowa City, IA.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893-898.
<https://doi.org/10.1016/j.paid.2006.09.017>
- Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1), 19-23.
<https://doi.org/10.1007/bf02288975>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
<https://doi.org/10.1007/BF02291170>
- Vidal Arenas, J. (2015). La concepción del tiempo en Aristóteles. *Byzantion nea hellás*, (34), 323-340.
<https://doi.org/10.4067/s0718-84712015000100014>
- Whittaker, T. A., & Khojasteh, J. (2017). Detecting appropriate trajectories of growth in latent growth models: The performance of information-based criteria. *The Journal of Experimental Education*, 85(2), 215-230.
<https://doi.org/10.1080/00220973.2015.1123669>
- Wu, W., & Jia, F. (2021). Applying planned missingness designs to longitudinal panel studies in developmental science: An overview. *New Directions for Child and Adolescent Development*, 175, 35-63.
<https://doi.org/10.1002/cad.20391>