

Recomendaciones para la prevención y el tratamiento de los datos incompletos en ensayos clínicos

Maytee Robaina,^{1*} Rolando Uranga,¹ Martha María Fors,¹ Carmen Viada²

¹Centro Nacional Coordinador de Ensayos Clínicos (CENCEC). Calle 200 y 21. Atabey. La Habana. Cuba. CP 11600. Teléfono: 2717345.

²Centro de Inmunología Molecular (CIM).

email: mrobaina@infomed.sld.cu

Múltiples investigaciones en estadística aplicada evidencian un amplio desarrollo teórico y diversidad de perspectivas para abordar el problema de los datos incompletos en los ensayos clínicos y se han establecido pautas a seguir para mitigar los potenciales sesgos que provocan cuando se analizan los datos. En este artículo se presentan de forma resumida algunas recomendaciones para abordar la prevención y el tratamiento de datos incompletos en los ensayos clínicos; se tomó como referencia fundamental un informe emitido en el 2010 por expertos internacionales y las disposiciones de las entidades regulatorias. Existe consenso en que la estrategia principal es la prevención desde el diseño y conducción del estudio. También se establecen recomendaciones para el tratamiento en el momento del análisis, donde se enfatiza en la necesidad de utilizar toda información disponible de todos los sujetos aleatorizados y aplicar métodos estadísticos más robustos a las ausencias e incluir siempre la evaluación de la sensibilidad de los resultados. En el 2011 se publicó una primera guía metodológica para datos incompletos adoptada por la Agencia Europea de Medicamentos (EMA), donde se dispone de la información necesaria, desde la perspectiva regulatoria, para avalar la calidad de los resultados en ensayos clínicos confirmatorios. El problema de tener datos incompletos es a veces inevitable y no debe ser obviado, por lo que es necesario aplicar estrategias desde el diseño para su prevención y análisis más adecuado. La calidad en la presentación de los resultados es indispensable para evitar conclusiones e interpretaciones sesgadas del estudio.

Palabras clave: ensayos clínicos, análisis estadístico, datos incompletos, datos faltantes.

Introducción

Los ensayos clínicos son investigaciones de elevada complejidad y con implicaciones éticas que requieren de gran cantidad de recursos y de numerosos procedimientos que consumen un tiempo prolongado e implican elevados costos.

Las agencias regulatorias de medicamentos deciden el autorizo o no de la comercialización de un producto (tratamiento, vacuna, dispositivo médico u otra), luego de haber sido evaluado en las diversas fases de desarrollo mediante múltiples ensayos clínicos. La autoridad sanitaria determina si los nuevos productos son idóneos para su uso en la población general en términos de beneficio y riesgo y regula su comercialización y condiciones de utilización a partir de estos estudios. Es por tanto indispensable la calidad de los resultados que se presentan.

Desde el punto de vista de su diseño los ensayos clínicos son estudios longitudinales, lo que hace que los datos estén propensos a omisiones. En las fases avanzadas de desarrollo del producto se confirma la eficacia mediante ensayos clínicos controlados y aleatorizados (ECCA), estudios paradigmáticos dentro de las investigaciones clínicas por su carácter

experimental y el rigor metodológico con el que se conducen. No obstante, con bastante frecuencia durante la ejecución de la investigación se deja de registrar información, lo que genera los llamados datos incompletos, ausentes o faltantes. Los datos incompletos son los datos de las variables de interés que no pudieron ser registrados a pesar de que estaba planificado obtener su valor (1) y son necesarios para el análisis estadístico y la evaluación de los resultados del producto en investigación.

Cuando hay datos incompletos la credibilidad científica de las conclusiones acerca de la causalidad de la intervención en la respuesta evaluada se debilita, porque su ausencia es una fuente potencial de sesgos y por tanto, los resultados del proceso inferencial (generalización de resultados) pueden no ser confiables (2).

En los ensayos confirmatorios, donde se evalúa la eficacia y es indispensable que los resultados observados sean generalizables a la población a la que va dirigida la intervención (validez externa) el análisis de los datos observados debe realizarse según el principio de "intención

* Dra en Medicina, Especialista de 1er grado en Bioestadística, Investigador agregado.

de tratar". Este principio consiste en analizar los datos de todos los sujetos aleatorizados, según el grupo al que fueron asignados inicialmente, independientemente de la adhesión al protocolo (3). No incluir los datos de todos ellos puede provocar un sesgo de selección y la reducción de la muestra disminuye la potencia en el estudio. Pero el análisis por intención de tratar no puede adoptarse directamente cuando hay datos faltantes en la respuesta, sino que hay que cuidar la manera con que se "rellenan" o se toman en cuenta esas "ausencias" porque podemos incurrir entonces en un sesgo durante el análisis (3, 4).

Los datos incompletos o ausentes se pueden producir por diversas causas, desde el olvido o imposibilidad del sujeto de asistir a una consulta planificada hasta su decisión de abandonar el estudio, cuyas razones pueden estar relacionadas o no con la intervención. También puede ocurrir que no se registre por problemas de disponibilidad de recursos u otras causas no relacionadas con el paciente ni con el efecto del producto investigado. Esta relación entre el dato faltante y los otros disponibles o no de la variable en cuestión o de otras variables del estudio, interviene ahora de manera determinante en el análisis en lo que se llama "mecanismo de pérdidas"; pero establecer esta relación no es trivial y muchas veces es imposible de determinar (5).

El mecanismo que genera las pérdidas clásicamente se clasifica en tres tipos: completamente aleatorio, aleatorio y no aleatorio (5, 6). Usualmente se asume uno de ellos cuando se realiza el análisis principal de un ensayo y las propiedades de los métodos de análisis dependen del mismo. Un problema añadido es que, a partir de los datos disponibles, no es posible identificar con certeza ese mecanismo; el análisis se hace entonces bajo supuestos inestables. Se hace necesario por tanto examinar la robustez de las conclusiones mediante el llamado análisis de sensibilidad (4, 5, 7).

Ningún método de análisis puede suplir la falta de los valores de una variable. Múltiples son las propuestas ante la necesidad de enfrentar y atenuar este problema, pero no hay uno que sea universalmente recomendable ante la diversidad de posibles situaciones (5, 7, 8). Los expertos recomiendan cuatro posibilidades cuando hay datos faltantes: análisis de los casos completos, aplicar métodos simples de imputación o de ecuaciones estimadoras, o los basados en modelos estadísticos; este último incluye los modelos basados en la verosimilitud como, por ejemplo, los lineales mixtos, también las imputaciones múltiples y los bayesianos (7).

En la práctica es frecuente que los investigadores pasen por alto la presencia de datos faltantes y analicen solo los sujetos que tienen los datos completos o realicen imputaciones simples, como es arrastrar el último valor hacia las observaciones faltantes (9, 10). Sin embargo, estos no son recomendados actualmente para realizar el análisis principal

(7, 8, 11). Los expertos recomiendan los métodos basados en modelos y los que utilizan ecuaciones estimadoras (7).

Como estrategia general hay autores que proponen la siguiente secuencia para el análisis cuando hay datos faltantes (4): Primero, procurar un seguimiento de todos los sujetos aleatorizados aún si estos descontinúan el tratamiento, para obtener el valor de las mediciones de respuesta; posteriormente efectuar el análisis inferencial principal que sea válido bajo un mecanismo de pérdidas que se asume plausible, que supone la exploración detallada de los datos disponibles del estudio, valorando en conjunto las características del problema en cuestión y lo planificado. Finalmente, realizar siempre un análisis de sensibilidad que evalúe si los resultados se modifican cuando cambia el supuesto asumido del mecanismo de pérdidas.

La falta de información se hace evidente en el momento en que se analizan los datos, pero este problema puede minimizarse si se prevén y, con sumo detalle, se planifican acciones contra la posible ocurrencia de esas pérdidas. Este es el momento crítico para abordar los datos incompletos, por lo que el equipo de investigadores que planifica ensayos clínicos debe tener conciencia de que la perspectiva de enfrentamiento mejor es más táctica-preventiva que analítica (7, 8, 11, 12).

El tratamiento de datos incompletos es complejo. Los detalles de cómo enfrentarlo es objeto de un amplio desarrollo teórico. La literatura científica es vasta y de gran complejidad y no es propósito de este trabajo abordarla. En este trabajo se propone en esencia alertar a los investigadores en la necesidad de no obviar el problema y presentar algunos principios generales aplicables a todos los ensayos utilizando básicamente el informe emitido recientemente por los expertos y sus recomendaciones para la prevención y tratamiento en ensayos clínicos (11). Se evidencia la existencia y disponibilidad de documentación regulatoria actualizada, todo lo cual puede orientar a los investigadores para enfrentar mejor la práctica de la investigación.

Materiales y Métodos

Se realizó una investigación descriptiva de tipo documental. Se revisaron artículos y libros sobre metodología de ensayos clínicos, estadística y métodos de análisis. Se utilizó estrategia de búsqueda en PUBMED y Google Académico, utilizando los descriptores en inglés: *missing data*, *clinical trials*, *statistical analysis*, y se hizo una selección cuidadosa de artículos claves en el tema. Para las referencias de tipo regulatorio también se consultaron los sitios web de la Conferencia Internacional de Armonización (ICH) y de la Agencia Europea de Medicamentos (EMA), la Administración de Medicamentos y Alimentos de los Estados Unidos (FDA).

Desarrollo

En el año 2010 se publica el informe *The Prevention and Treatment of Missing Data in Clinical Trials* (11), donde 15 expertos internacionales en el tema de datos incompletos, convocados por la Academia Nacional de Ciencias de Estados Unidos, hacen recomendaciones generales que, a manera de consenso, acotan la diversidad de criterios y establecen principios aplicables a todos los ensayos clínicos para abordar la falta de datos. Ese documento está dirigido a todos los investigadores involucrados, desde promotores hasta investigadores y estadísticos a cargo del estudio.

Para abordar los datos incompletos los expertos diferencian dos momentos de intervención: uno durante el diseño y conducción de la investigación, con estrategias cuidadosas que mitiguen las posibles fuentes de pérdidas, desde la planificación (enfoque preventivo) y otro encaminado a los métodos estadísticos y estrategias para el análisis de los datos (enfoque de tratamiento).

El enfoque preventivo se refiere a las acciones que se deben planificar para obtener el valor de las variables de interés durante la ejecución del estudio y disponer de estas en el análisis. Si se obtienen todos los datos previstos (o con solo un mínimo de ausencias) las implicaciones de la ausencia en el análisis son menores, es por tanto la estrategia preferible (11, 12) y serán pocos los esfuerzos realizados si se compara con el beneficio para la calidad de la investigación.

Los expertos recomiendan durante la planificación: definir correctamente en el protocolo los criterios de selección de los sujetos y de la(s) variable(s) que evaluará(n) el efecto de tratamiento, tiempo del estudio y momentos en que se registrarán las variables, frecuencia de seguimiento y cómo realizarlo.

Es indispensable elegir cuidadosamente tanto los sitios donde se llevará a cabo la investigación como a los investigadores. Todos los participantes deben sensibilizarse en la necesidad de limitar las pérdidas, particularmente los investigadores como “ejes” en la conducción del estudio. Los monitores intervienen también realizando estrategias para obtener datos con el máximo de calidad y para favorecer la retención de los sujetos incluidos.

Se sugiere que los investigadores proporcionen a los sujetos incluidos recordatorios frecuentes de las visitas programadas. Además, desde la planificación deben quedar especificadas medidas para limitar las dificultades que pudieran presentar los sujetos en investigación para asistir a las evaluaciones. Cada estudio debe prever y adoptar sus propias acciones (11, 12).

Un elemento fundamental de esta etapa es distinguir claramente en el protocolo cuáles serán las condiciones que se considerarán “no adherencia” (en ese caso, deja de administrarse la intervención que se evalúa pero los datos de las variables se continúan registrando) de las condiciones que serán consideradas: “abandono del estudio” o de “no retención” (en este caso dejan de registrarse los datos del sujeto). Muchas veces se definen causas de “abandono” que en verdad debieran ser de “no adherencia”, como es el caso de la intolerancia a la intervención, toxicidad, decisión del investigador clínico de interrumpir el tratamiento y necesidad de otros tratamientos. Se sugiere, siempre que los sujetos no hayan retirado su consentimiento de participación, realizar un seguimiento a los que descontinuaron el estudio para obtener las causas de abandono y los datos de las principales variables del ensayo, incluso, si los sujetos interrumpieron el tratamiento al que fueron aleatorizados o iniciaron otras intervenciones (3, 11).

Otra estrategia que se sugiere para prevenir datos ausentes es el uso de terapias de rescate o regímenes alternativos de tratamiento en los sujetos en que se decide interrumpir el tratamiento asignado. Estos tratamientos deben quedar definidos de antemano. Estas y otras medidas se describen por Little y cols. en el 2012 (7).

Por su parte, el enfoque de tratamiento se refiere a las estrategias encaminadas al abordaje de los datos incompletos en la etapa de procesamiento y análisis. Aunque el tema se ha desarrollado principalmente para ensayos controlados y aleatorizados, sus principios son aplicables a la generalidad de estos estudios.

Cuando se detectan datos incompletos en la variable de respuesta, el análisis estadístico debe permitir utilizar la información disponible de todos los sujetos aleatorizados y prestar cuidadosa atención en los supuestos que se asumen del mecanismo de pérdidas. Los expertos establecen como principio la necesidad de comunicar de forma clara a los investigadores qué es lo que se asume así como valorar en conjunto la validez de estas suposiciones. Como estrategia principal de análisis no se recomiendan métodos como el que utiliza solo a los sujetos con datos completos (13), porque provoca sesgos y viola el principio de análisis por intención de tratar; esencialmente implica reducción de la potencia del estudio, además de que asume un mecanismo de pérdida completamente aleatorio, lo cual es muy poco probable que suceda en la práctica. Tampoco se recomienda aplicar métodos de imputación simples, como es el arrastre del último valor hacia el resto de las mediciones faltantes (11). Este análisis asume que la respuesta en estos sujetos no cambia después que dejan de obtenerse sus mediciones, lo cual es difícil que suceda en verdad, además de que este método subestima la incertidumbre asociada a las estimaciones.

Los métodos de inferencia basados en la maximización de la verosimilitud, los bayesianos, los basados en ecuaciones estimadoras generalizadas ponderadas y las imputaciones múltiples se consideran en general más adecuados cuando hay datos incompletos, ya que incorporan en el análisis la información auxiliar registrada en el estudio (11). Estos métodos estadísticos, aplicados de manera adecuada, hacen que potencialmente se puedan reducir los sesgos aunque aún con estos, los supuestos que se asumen no pueden ser probados. Por ello, en la presentación de los resultados que derivan del análisis primario con datos incompletos es imprescindible realizar un análisis de sensibilidad (11, 13).

Se han realizado varias propuestas sobre las posibles formas de realizar un análisis de sensibilidad, que todavía es un área de investigación activa (5, 11). Informalmente puede definirse cuando se consideran varios modelos estadísticos de manera simultánea o donde un modelo determinado se analiza utilizando herramientas especializadas (ejemplo, medidas diagnósticas). Esta es una definición muy general que abarca diversas posibilidades dentro de lo que es realmente un análisis de sensibilidad (11). Hay consenso en que los métodos que asumen mecanismo de pérdida no aleatorio son útiles dentro del mismo.

En este escenario, los métodos basados en verosimilitud requieren de la especificación de la distribución conjunta del modelo de los datos y del mecanismo de pérdida. Esta especificación se realiza según tres tipos de modelos: de selección, de patrones mixtos y de parámetros compartidos (11). Las estrategias de modelación más complejas, al igual que el análisis de sensibilidad, son áreas de creciente desarrollo dentro del tema de los datos ausentes.

Disposiciones de las agencias regulatorias

La falta de rigor en la presentación de los resultados de los ensayos clínicos pudiera llegar a invalidar la aprobación de la comercialización del producto. En este punto, las agencias regulatorias también se pronuncian ante el problema de los datos incompletos a la luz de los actuales conocimientos.

Las agencias regulatorias de medicamentos son autoridades sanitarias que tienen la responsabilidad legal de aprobar (o no) la comercialización del producto o intervención evaluada a partir de criterios estrictos de calidad, seguridad y eficacia, avalados por la presentación de los resultados. Sus guías metodológicas contienen recomendaciones generales sobre la información a presentar, necesaria para documentar la calidad de los resultados que se ofrecen. Estas guías abordan diversos temas sobre ensayos clínicos y están disponibles para el uso de toda la comunidad científica en los sitios correspondientes, tanto de la Conferencia Internacional para la Armonización de ensayos clínicos como de las agencias correspondientes (ICH, EMA, FDA).

EMA emitió en el 2001 la primera versión de las recomendaciones regulatorias para el manejo de datos ausentes en ensayos clínicos (14) y en el 2011, debido a la expansión del uso de nuevos métodos y a la incompleta aplicación de algunos conceptos básicos como los análisis de sensibilidad, realizó una profunda renovación y emite la guía *Guideline on missing data in confirmatory clinical trials* (13).

Este documento proporciona una visión de las normas regulatorias vigentes para la evaluación de ensayos clínicos confirmatorios con datos incompletos y las recomendaciones generales de la información a presentar. Plantea documentar por grupos tratados el número de datos faltantes, el momento en que se produjeron y las causas. Además, se debe describir la frecuencia de los patrones de pérdidas, especificar si ocurrió desviación de lo previsto en el protocolo y las posibles implicaciones que tienen los datos faltantes en la evaluación que se realiza de la eficacia y seguridad. El documento también expone, de forma general, diferentes posibilidades de análisis y se pronuncia respecto a que, desde la perspectiva regulatoria, es de interés que el enfoque de análisis que se realice no sesgue los resultados a favor del producto en investigación. Establecen, además, la necesidad de evaluar la sensibilidad de los resultados que se presentan.

Por su parte, la agencia estadounidense FDA tiene identificado el tema de los datos incompletos como prioritario, publicó en el 2008 dos documentos regulatorios actualmente en uso y tomará en cuenta los criterios y las sugerencias del panel de expertos para desarrollar una guía sobre el tema, aún sin publicar (11).

Conclusiones

En la mayoría de los ensayos clínicos es casi inevitable la pérdida o no disponibilidad de una proporción variable de los datos de los sujetos aleatorizados. En los últimos años este problema se aborda intensamente por la literatura científica que refleja la complejidad y las dificultades de la inferencia cuando los datos no están completos. Un grupo de expertos internacionales establecieron un conjunto de principios generales aplicables a todos los ensayos y la EMA adoptó una guía metodológica para poder exigir y velar por la calidad de los resultados de los ensayos confirmatorios. Todos apuntan en la misma dirección: implementar estrategias de diseño y de conducción que minimicen los posibles datos ausentes y la necesidad de un mejor abordaje estadístico que comprenda el uso de toda la información de todos los sujetos aleatorizados, además de la necesidad de evaluar el impacto de tener datos incompletos en las principales inferencias mediante el análisis de sensibilidad.

Referencias

1. Carpenter JR, Kenward MG. Missing data in randomised controlled trials—a practical guide. Birmingham: National Institute for Health Research, Publication RM03/JH17/MK. 2008. Disponible en: http://www.haps.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml
2. International Conference on Harmonisation (ICH-E9): Statistical Principles for Clinical Trials. Tripartite Harmonised Guideline, ICH 1998. Disponible en: <http://www.ich.org/cache/compo/276-254-1.html>.
3. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167-89.
4. White IR, Carpenter J, Nicholas, HJ. Including all individuals is not enough: Lessons for intention-to-treat analysis. *Clin Trials* 2012;9:396-407.
5. Molenberghs G, Kenward M. *Missing Data in Clinical Studies*. New York: Springer-Verlag; 2007.
6. Rubin DB. Inference and missing data. *Biometrika* 1976;63(3):581-92.
7. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The Prevention and Treatment of Missing Data in Clinical Trials. *N Engl J Med* 2012;367:1355-60.
8. O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical Pharmacology & Therapeutics* 2012;91:550-4.
9. Wood MA, White RI, Thompson GS. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004;1:368-76.
10. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004;5:445-64.
11. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies; 2010.
12. Fleming TR. Addressing Missing Data in Clinical Trials. *Ann Intern Med* 2011;154:113-7 .
13. Agencia Europea de Medicamentos (EMA). Guideline on missing data in confirmatory clinical trials. (CPMP/EWP/1776/99); 2009. Disponible en: [ww.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf](http://www.ema.europa.eu/pdfs/human/ewp/177699endraft.pdf).
14. Agencia Europea de Medicamentos (EMA). Points to consider on missing data; 2001. Disponible en: www.ema.europa.eu/docs/en_GB/document.../09/WC500003641.pdf.

Recommendations for the prevention and treatment of incomplete data in clinical assays

Abstract

Several researches in applied statistics show a broad theoretical development and diversity of perspectives to address the problem of incomplete data in clinical trials. Recent publications show consensus and establish guidelines to mitigate potential biases that causes this problem when data are analyzed. This article will summarize some of the key recommendations to better address the prevention and treatment of incomplete data in clinical trials presented in a report by international experts in 2010 and also referred to the provisions of the regulatory bodies with emphasis on the most recent (2011). There is consensus that the main strategy is to prevent possible incomplete data from the design and conduct of the study. Recommendations are also set to the time of analysis, which basically emphasizes the need of using any available information of all randomized subjects, and implement more sensitive strategies for the inference, that always includes assessing the sensitivity of the results. In 2011 was published a methodological guide -adopted by the European Medicines Agency (EMA)- which provides necessary information from the regulatory perspective to ensure the quality of the results in confirmatory clinical trials. The problem of incomplete data is sometimes unavoidable and should not be ignored, it is necessary to apply appropriate design and analysis strategies. The quality in the presenting results is essential, to avoid biased conclusions and interpretations of the study.

Key words: clinical trials, statistical analysis, incomplete data, missing data.
