

Los ensayos clínicos en el cuidado de la herida (I): ventajas y limitaciones de los distintos diseños de ensayos clínicos

Clinical trials in wound care (I): the advantages and limitations of different clinical trial designs

P. Venkatraman* *BTech, MSc, Investigador; Facultad de Tecnología.
S. Anand** **BSc, MSc, Tech PhD, C Text FTI, Director de Investigación, Facultad de Tecnología.
C. Dean*** ***PhD, profesor de la Facultad de Letras, Ciencias y Educación.
R. Nettleton**** ****BNurs, MSc, RN, RHV, NDCert, profesora de la Facultad de Letras, Ciencias y Educación.
 Instituto Bolton, Reino Unido. E-mail: dpvlmpo@bolton.ac.uk

RESUMEN

Los ensayos clínicos han de ser consistentes para que los médicos pongan en práctica sus recomendaciones. Este artículo describe los distintos diseños de investigación utilizados en el cuidado de la herida y los problemas que podrían afectar a la fiabilidad y validez de los resultados.

PALABRAS CLAVE

Enmascaramiento, ensayos clínicos, diseño, dispositivos médicos, análisis de potencia, aleatorización, tamaño muestral.

SUMMARY

Clinical trials need to be robust if practitioners are to apply their recommendations into practice. This article describes the various research designs used in wound care and the pitfalls that may compromise the reliability and validity of the results.

KEY WORDS

Blinding, clinical trials, designs, medical devices, power analysis, randomisation, sample size.

Los ensayos clínicos precisan de una planificación y de un diseño meticuloso. Un ensayo planificado y ejecutado correctamente puede proporcionar indicios suficientes para informar a los accionistas de las empresas del ámbito sanitario acerca de las ventajas asociadas a los nuevos materiales o tratamientos. Este artículo revisa los distintos diseños de investigación, junto con sus ventajas e inconvenientes relativos, utilizados en ensayos clínicos sobre el cuidado de la herida. Se destacan algunos problemas comunes que pueden aparecer en diversas etapas del diseño y realización de un estudio.

FASES DE INVESTIGACIÓN

En el marco del cuidado de la herida, un ensayo clínico es un trabajo de investigación en el que parti-



158 cipan voluntarios humanos con el fin de determinar la seguridad y eficacia de un nuevo producto. Tan sólo se puede estudiar a dichos voluntarios cuando se han llevado a cabo estudios analíticos y en animales, convincentes y prácticos. El objetivo principal de un ensayo clínico sobre dos o más productos es identificar la magnitud de la mejora asociada a un tratamiento en comparación con el otro, o bien demostrar su equivalencia o, como mínimo, mostrar que no existen diferencias entre ellos.

Los productos atraviesan cuatro fases de ensayo antes de poder ser recetados.

- Fase I: estudio inicial de un número pequeño de voluntarios sanos (20-80) para evaluar su seguridad y determinar sus efectos secundarios.

- Fase II: realizada sobre un número elevado de enfermos (100-300) con el fin de establecer su eficacia y valorar la seguridad.

- Fase III: utiliza un número grande de pacientes (> 1.000) para confirmar la eficacia. Esto se logra mediante la comparación con tratamientos o productos utilizados frecuentemente. Se recoge información sobre la posibilidad de utilizar el tratamiento de forma segura.

- Fase IV: realizado tras la incorporación del producto al mercado con el fin de estudiar su efecto a largo plazo en diversas poblaciones de estudio (1).

Los ensayos de fase I y de fase II constituyen evaluaciones tem-

pranas que investigan minuciosamente las propiedades de un producto en condiciones "ideales" (2). La fase I puede obviarse cuando se dispone de datos suficientes. Para que se puedan valorar y poner en práctica los resultados de un ensayo, es preciso incluir componentes básicos, como la información previa, el tipo de diseño, el número de individuos necesarios para dotar de la necesaria potencia estadística, las técnicas de aleatorización, los planes de supervisión de datos y los análisis estadísticos.

DISEÑOS PRINCIPALES DE ENSAYOS CLÍNICOS

Ensayo aleatorizado controlado

Los participantes aptos se asignan a un grupo de intervención o a un grupo control utilizando un procedimiento de aleatorización como las técnicas simples, en bloque o estratificadas. Todos los individuos tienen la misma probabilidad de ser incluidos en el grupo de intervención o en el control.

La aleatorización previene cualquier sesgo de asignación por el investigador o el sujeto y aumenta la posibilidad de que las características de ambos grupos estén equilibradas de forma uniforme. Cualquier cambio observado puede atribuirse, por tanto, a la intervención. Si las características de los grupos no están bien equilibradas, la aleatorización estratificada puede facilitar la comparación de los grupos. Por

ejemplo, si es probable que los enfermos y las enfermas respondan de modo distinto al tratamiento, será necesario realizar una estratificación consecuente, con una lista de aleatorización distinta para varones y mujeres.

La aplicación de las pruebas estadísticas a los datos aportará niveles de significación válidos para cualquier diferencia respecto a las respuestas observadas entre los grupos de tratamiento. Sin embargo, al igual que sucede con cualquier ensayo, ello dependerá de si se ha utilizado un tamaño muestral adecuado.

Puede que no se realice aleatorización cuando el investigador es reacio a asignar a los participantes o a los sujetos a este procedimiento. En tales casos, se recomienda utilizar el mejor producto habitual como control e intentar demostrar que no existe ninguna diferencia estadística entre los dos productos de cuidado de la herida.

Los ensayos multicéntricos se emplean cuando la prevalencia de una enfermedad es pequeña y el número de individuos reclutados a partir de una única población de estudio es insuficiente (2, 3).

Estudios no aleatorizados controlados simultáneos

En ellos, la distribución de los participantes no se realiza mediante una técnica de aleatorización. Tanto el grupo control como el de intervención se tratan de forma simultánea; por ejemplo, un grupo de enfermos recibe

un nuevo producto en un centro, mientras que otro grupo de pacientes con características similares recibe un tratamiento control o habitual simultáneamente en otro centro.

Los investigadores asumen con frecuencia que la selección de grupos con pocas características ajustadas respecto a edad, sexo y cronología de la herida aporta con mayor probabilidad información acerca de los factores pronósticos que influyen en la respuesta. Por tanto, la selección de los individuos con características similares es muy poco práctica en la mayoría de los casos. Existe también el riesgo de un sesgo de distribución, mientras que las valoraciones realizadas por las enfermeras de distintos centros pueden variar, lo que debilitaría la credibilidad de los resultados (4, 5).

Estudios históricos de control

Son aquellos que comparan las respuestas de un grupo de pacientes sometidos a un nuevo tratamiento con las de sujetos comparables de un estudio anterior. Se utilizan cuando el investigador cree que el producto tiene efectos beneficiosos y contribuye a formar una opinión inicial acerca del producto o del tratamiento (3).

Una ventaja de esta estrategia es que es más probable que los enfermos participen en estos estudios cuando se les asegura que recibirán un tratamiento determinado. El período de estudio es más corto y se obtienen resultados con mayor rapidez.

Deben existir similitudes en las características, diagnósticos y tratamiento de las dos poblaciones de estudio, así como en la medición, análisis e interpretación de los datos. En caso contrario, las respuestas presentarán sesgo.

Diseño cruzado

En este tipo de diseño cada paciente actúa como su propio control. Se administra el nuevo tratamiento al enfermo una vez (primer período) seguido de un control (segundo período). Otro enfermo puede recibir el control en primer lugar y la intervención a continuación. La asignación al grupo de tratamiento o control puede efectuarse mediante aleatorización, en la que aproximadamente el 50% de los enfermos reciben una secuencia intervención-control y el 50% restante una secuencia control-intervención. Por otra parte, el tamaño necesario para este tipo de diseño es 0,4 veces el utilizado en diseños paralelos (4). Las variaciones se mantienen a niveles mínimos, puesto que el efecto medido tan sólo se debe a la respuesta del enfermo al tratamiento control y de intervención.

Para que este diseño sea eficaz, la enfermedad debe permanecer estable durante un cierto tiempo. Por tanto, los pacientes que se recuperaron tras recibir la intervención no pueden continuar en el segundo período del estudio, a menos que vuelvan al estado inicial de la enfermedad. En otras

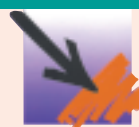
palabras, el trastorno del enfermo no debe ser tratado por la primera intervención.

Este diseño es adecuado para aquellos ensayos en los que se estudia, por ejemplo, el efecto del fármaco sobre la disminución de la presión sanguínea. Incluso en ese caso, los efectos de un fármaco no deberían continuar en el segundo período (efecto remanente). Por consiguiente, se debería estudiar el período comprendido entre ambas pautas, administrando el tratamiento correspondiente cuando el paciente vuelve al estado inicial de la enfermedad (2, 3).

Estudios de retirada

Dichos estudios determinan el tiempo necesario para que un enfermo interrumpa un tratamiento o para que se produzca una modificación de la frecuencia de su administración. Por ejemplo, ¿cuánto tiempo puede soportar un paciente con una úlcera venosa un determinado grado de compresión durante una semana? En ocasiones, este tipo de ensayos puede ayudar a evaluar cualquier intervención o material que no ha mostrado aportar efectos significativos.

Una limitación importante asociada a los estudios de retirada es que los investigadores podrían seleccionar a los enfermos que en su opinión se beneficiarán en mayor grado del tratamiento, haciendo que la muestra sea muy homogénea. En tales casos, el estudio infra-estimaría cualquier



160 toxicidad y exagerará los efectos beneficiosos. El resto de las características del ensayo, como la aleatorización, enmascaramiento y utilización de análisis estadísticos adecuados, siguen siendo las mismas (3).

Diseño factorial

Se utiliza para evaluar dos o más intervenciones experimentales por separado (M sólo o N sólo) y en combinación (M y N). Su diseño responde al objetivo de evaluar los efectos de la interacción de las intervenciones. La interacción es el efecto de M, dependiendo de la presencia o ausencia de N, y viceversa. Es probable que suceda cuando las respuestas de ambas intervenciones M y N son semejantes; por ejemplo, si ambas reducen el olor de la herida.

Se aleatorizan por bloques cifras idénticas de enfermos a cada una de las células del diseño. Por tanto, 40 pacientes y un bloque de 4 permiten incluir 10 individuos en cada célula. En este caso, se puede valorar a 10 pacientes de las células M y N, y 10 de la célula M con respecto al efecto de la intervención de M. El diseño debe contar con un tamaño muestral adecuado para ser eficaz e informativo. Sin embargo, si no existen efectos de interacción entre las intervenciones, los dos experimentos se pueden realizar en un único diseño siempre y cuando se produzca un aumento moderado del tamaño muestral.

En la mayor parte de los casos, es necesario probar estadísticamente el efecto de la interacción sobre la respuesta. Con frecuencia no es posible utilizar todas las células o combinaciones terapéuticas debido a motivos éticos -por ejemplo, el control (ausencia de M y ausencia de N)- en cuyo caso las células pueden quedar vacías. Esto se denomina diseño factorial incompleto.

En un estudio de tres intervenciones existen ocho células, con un aumento correspondiente del tamaño muestral total (2, 3).

PROBLEMAS EN EL DISEÑO DE ENSAYOS CLÍNICOS

Elección del diseño

La selección de un diseño del estudio adecuado depende de la cuestión que el investigador pretenda abordar. Por ejemplo, ni un ensayo comparativo simple (es decir, de vendajes A y B) ni un ensayo paralelo serán capaces de detectar ningún efecto de interacción sobre los vendajes. No obstante, un diseño factorial de cuatro células sí sería capaz de hacerlo. Por otra parte, si un enfermo no responde al vendaje A, existe la posibilidad de que se pudiera beneficiar de la aplicación del vendaje B. Si éste es el caso, se utilizará un diseño cruzado.

De modo semejante, no se puede aleatorizar a enfermos reacios a confiar en la elección al azar de su tratamiento. En este caso, será necesario emplear un diseño no aleatorizado, aunque las características de los grupos de trata-

miento no estén equilibradas uniformemente. 49

Explicación insuficiente de la población

Se debe describir a la población estudiada, puesto que de lo contrario resulta imposible evaluar la respuesta o aplicar los resultados de la investigación a la práctica. Se debería también especificar lo siguiente: los criterios de inclusión y de exclusión, el procedimiento de aleatorización, los fracasos en la aleatorización, cualquier retirada o cambio entre los grupos de tratamiento y los análisis realizados.

Sesgo de selección

Un investigador que ha elaborado el proceso de distribución podría asignar a un enfermo a un grupo concreto con el fin de conseguir una respuesta seleccionada. Por ejemplo, la distribución de pacientes con enfermedad de menor gravedad al grupo de tratamiento aumentará la probabilidad de una respuesta positiva.

Sesgo accidental

Existe la posibilidad, especialmente en estudios de tamaño pequeño, de la existencia de un desequilibrio accidental en las características o factores de riesgo. Ni siquiera una distribución aleatorizada sencilla carece de inconvenientes (2). La Tabla 1 recoge los desequilibrios que aparecerán con probabilidades de al menos 0,05 y de al menos 0,01 en varios tamaños de ensayo.

Tabla 1. Desequilibrio en la aleatorización sencilla en ensayos sobre dos tratamientos (2)

Nº total enfermos	Diferencia numérica	
	$p \geq 0,05$	$p \geq 0,01$
10	2:8	1:9
20	6:14	4:16
50	18:32	16:34
100	40:60	37:63
200	86:114	82:118
500	228:272	221:279
1.000	469:531	459:541

Cuanto menor sea el ensayo ($n=50$), mayor será el número de desequilibrios que se podrían producir, especialmente si el cociente es 16:34 con una probabilidad de, como mínimo, 0,01. Esto sería difícil si, por ejemplo, se asignase un nuevo tratamiento a 16 pacientes y la pauta habitual a 34 enfermos, en especial cuando el investigador está estudiando los efectos del nuevo tratamiento.

Explicación insuficiente de la aleatorización

Para que una recomendación de un estudio se ejecute con confianza, es necesario explicar de forma precisa cómo se ha llevado a cabo la aleatorización. Altman y Dore descubrieron que el 30% de los artículos publicados como ensayos clínicos no aportaban ningún indicio de la existencia de aleatorización, el 10% de ellos se autodenominaban ensayos aleatorización pero utilizaban métodos de distribución no aleatorizado, y el 60% no refería el método de aleatorización empleado (5).

Pérdida de enmascaramiento

En un estudio enmascarado, se puede asignar de forma aleatoria a los pacientes a los grupos de estudio utilizando una lista que siga un orden determinado. El enmascaramiento se pierde si la persona que aplica el tratamiento averigua dicho orden. Cuando es así, el médico podría esperar una respuesta mejor de un enfermo cuyos valores son muy bajos o muy elevados, y podría comprobar o modificar las mediciones (7).

Algunas respuestas fisiológicas son indicativas de la eficacia del tratamiento. Su presencia puede revelar la identidad del mismo, especialmente cuando un paciente responde mejor a una intervención y posteriormente otro enfermo muestra una respuesta inferior. Esto provocará que el cuidador suponga subconscientemente las respuestas al tratamiento.

Duración insuficiente del estudio

En los ensayos sobre la etiología del proceso crónico de cicatrización de la herida, es necesario dis-

poner del tiempo suficiente para 161 que la herida evolucione a través de las diversas etapas de dicho proceso. Un estudio realizado por Kantor y Margolis (8) afirmó que la cicatrización de úlceras venosas no era un proceso lineal: la cicatrización era lineal durante las primeras 12 semanas, pero esta tendencia disminuía rápidamente durante las siguientes. A las 20 semanas, los autores refirieron que el 63% de las heridas habrían cicatrizado mediante el tratamiento habitual. No se han referido observaciones semejantes en un marco temporal más corto.

Tamaño muestral insuficiente

Un ensayo de tamaño pequeño dispone, inevitablemente, de menor potencia estadística para detectar diferencias entre distintos grupos de tratamiento. La potencia de un ensayo se define como la probabilidad de encontrar una diferencia cuando ésta existe. Cuanto menor sea la variabilidad muestral, mayor será la potencia del ensayo. Si el número de pacientes que participan en el ensayo es insuficiente, existe la probabilidad de que se cometa un error de tipo I (rechazar incorrectamente la hipótesis nula) o de tipo II (no rechazar la hipótesis nula) (9). Dicha hipótesis afirma que no existe ninguna diferencia entre los tratamientos.

Sesgo subconsciente

La mayor parte de los ensayos sobre cuidado de la herida son ensayos abiertos (10, 15) y no suelen ser mono-enmascarados; es



162 decir, o bien el enfermo es consciente del producto terapéutico o la enfermera conoce el material de tratamiento.

Los pacientes que saben que están participando en un ensayo pueden exagerar o infrareferir los sucesos, el llamado efecto Hawthorne.

Con frecuencia, la identidad del tratamiento en los ensayos sobre cuidado de la herida no puede enmascararse, a diferencia de los ensayos farmacológicos en los que se administran pastillas placebo. Este sesgo es inevitable debido al consentimiento informado.

Si la enfermera que atiende conoce la asignación, podría prestar un mayor cuidado y atención al enfermo, especialmente cuando se utiliza un producto control o habitual. Esto podría aumentar cualquier duda existente en el paciente acerca de la eficacia del producto. Otra posibilidad sería que la enfermera exagerase la gravedad del trastorno, en especial cuando el enfermo recibe un nuevo tratamiento, puesto que el éxito de un ensayo depende generalmente de su mejoría.

Abandonos

Los enfermos pueden retirarse o abandonar el estudio debido a la pérdida de confianza en el ensayo o a los efectos secundarios. Si se ignoran los datos de dichos pacientes, se genera un sesgo, puesto que el resultado favorecerá al nuevo tratamiento. En tales casos, los datos del ensayo deben analizarse utilizando el procedi-

miento de intención de tratar; por ejemplo, los abandonos del grupo control deberían incluirse en el análisis.

Ello no es necesario en estudios de eficacia cuyo objetivo es mostrar el efecto de la intervención sobre los pacientes que realmente reciben el tratamiento (16).

ELECCIÓN DE PRUEBAS ESTADÍSTICAS

Grupos comparables en la situación inicial

Las características de los dos grupos de tratamiento deben ser comparables en el momento inicial. Si existen desequilibrios, cualquier conclusión obtenida a partir de dichas pruebas podría inducir a errores graves.

Elección de una prueba unilateral o bilateral

La prueba unilateral asume que la respuesta a un tratamiento viene especificada direccionalmente; es decir, que el tratamiento aumentará o disminuirá dicha respuesta. Una prueba bilateral asume que existirá un cambio porcentual en la respuesta entre grupos I y II, pero la dirección no se especifica. No sería adecuado realizar una prueba unilateral sobre un nuevo tratamiento si existen indicios insuficientes para ignorar el riesgo de efectos secundarios.

Selección de pruebas estadísticas adecuadas

Todas las pruebas estadísticas se clasifican dentro de dos tipos

amplios: métodos paramétricos y métodos no paramétricos (5). La utilización de métodos estadísticos adecuados aporta información más fiable acerca de los datos reunidos e identifica el nivel de significación.

Las pruebas paramétricas asumen la selección aleatoria de los datos a partir de una variable de distribución normal, así como la uniformidad de las variancias de los grupos comparados. Las pruebas no paramétricas o estadísticas de distribución libre carecen de tales suposiciones.

Los datos analizados utilizando métodos paramétricos son numéricos, de modo que es posible realizar operaciones aritméticas como sumas y divisiones. Como ejemplos cabe citar el tamaño del área ulcerada (2,83 cm²) y el peso de un individuo (75 kg).

Cuando se asigna un número a una propiedad, como en el índice del dolor en el que 1= dolor, 2= menor dolor, 3= dolor moderado y 4= dolor intenso, no se pueden sumar dichos números ni calcular la media, puesto que no existe una diferencia fija entre dos unidades cualquiera (como dolor moderado e intenso). Se trata de variables ordinales que deberían analizarse mediante pruebas no paramétricas. Sin embargo, es posible expresar dichas variables como porcentajes; por ejemplo, el 45% de los participantes presentó una puntuación menor de 2. En este caso se debería referir el tamaño muestral.

Por tanto, es importante conocer cuál es la distribución de los datos antes de decidir qué método estadístico se utilizará.

Intervalos de confianza

Un estudio de investigación que refiere que el vendaje A es significativo ($p < 0,05$) en comparación con el vendaje B respecto a la absorción de exudados durante un período de ocho semanas, no muestra en qué medida aquel vendaje es superior a éste. Tan sólo demuestra que existen suficientes indicios a partir de los datos obtenidos para afirmar que el vendaje A es significativamente mejor que el B.

Las pruebas de significación deberían ir seguidas de pruebas descriptivas como los intervalos de confianza (3), que aportan un límite superior e inferior de un valor verdadero, es decir, la variación esperada a partir del valor medio obtenido. Generalmente se

utiliza un intervalo de confianza del 95%. Suponiendo, en el ejemplo anterior, que la absorción media de exudados del vendaje A sea de 56,4 y el intervalo de confianza del 95% $\pm 9,28$, se puede afirmar con una seguridad del 95% que el valor se encuentra entre 47,12 y 65,68. Si el informe no refleja tales intervalos de con-

fianza, no se puede estar seguro del hallazgo ni de su variación a partir del valor medio.

- La parte II enumera formas y métodos de minimizar el sesgo. Se definen, explican e ilustran con ejemplos los elementos básicos del diseño de un ensayo -cálculo del tamaño muestral, potencia y efectos del tamaño-.

Tabla 2. Resumen de los hallazgos principales

Existen diversos diseños de estudio: los que permiten la comparación de las respuestas con aquellas de pacientes similares de un estudio previo, aquellos en los que cada enfermo actúa como su propio control, o en los que los investigadores evalúan cuánto tiempo tarda un enfermo en interrumpir el tratamiento o modificar su frecuencia, o comparan los efectos de interacción de dos intervenciones.
La aleatorización puede evitar el sesgo de distribución y aumenta la probabilidad de que las características de ambos grupos estén equilibradas uniformemente. Por tanto, cualquier cambio observado sería debido a la intervención.
Las pruebas estadísticas adecuadas aplicadas a los datos aportan niveles de significación válidos para cualquier diferencia en las respuestas observadas entre los grupos de tratamiento. Para que las recomendaciones de un estudio puedan ponerse en práctica, la población de estudio debe ser descrita adecuadamente, se deberá evitar el sesgo, explicar el proceso de aleatorización, utilizar un tamaño muestral y un período de estudio adecuados y tener en cuenta los abandonos o retiradas de los enfermos.

BIBLIOGRAFÍA

1. www.clinicaltrials.com (dated 03.01.01).
2. Pocock S. Clinical Trials: A practical approach. London: John Wiley and Sons, 1983.
3. Friedman LM, Furberg CD, De Mets DL. Fundamentals of Clinical Trials (3rd edn). New York: Springer Publications, 1998.
4. James KE, Forest WH Jr, Rose RL. Crossover and non-crossover designs in four-point parallel line analgesic assays. Clin Pharmacol Ther 1985; 37: 242-52.
5. Greenhalgh T. Statistics for non-statisticians. How to read a paper: different types of data need different statistical tests. BMJ 1997; 315: 7104, 364-6.
6. Altman D, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet 1990; 335: 149-53.
7. Chapat de Saintonage DM, Vere DW. Current Problems in Clinical Trials. Oxford: Blackwell Scientific, 1984.
8. Kantor J, Margolis DJ. Expected healing rates for chronic wounds. Wounds 2000; 12: 6, 155-8.
9. Ruckley CV. Evidence-based management of patients with leg ulcers. J Wound Care 1997; 6: 1, 442-4.
10. Butterworth RJ, Bale S, Harding KG. Comparing Allevyn cavity wound dressings and Silastic foam. J Wound Care 1992; 1: 1, 10-3.
11. Banks V, Harding KG. Comparing two dressings for exuding pressure sores in community patients. J Wound Care 1994; 3: 4, 175-8.
12. Thomas S, Banks V, Hagelstein A, Harding KG. A comparison of two dressings in the management of chronic wounds. J Wound Care 1997; 6: 8, 383-6.
13. Armstrong SH, Ruckley CV. Use of fibrous dressing in exuding ulcers. J Wound Care 1997; 6: 7, 322-4.
14. Poole M, Milward P. A randomised prospective crossover trial to investigate the use of an odour-absorbing dressing with a four-layer bandage system. In: Cherry, NN, Leaper DJ, Lawrence JC, Milward, P. (eds). Proceedings of 4th European Conference on Advances in Wound Management London: Macmillan Magazines, 1994.
15. Banks V, Bale S, Harding K. Evaluation of a new polyurethane foam dressing. J Wound Care 1997; 6: 6, 266-9.
16. Greenhalgh T. Assessing the methodological quality of published papers. BMJ 1997; 3 15: 7103, 305-8.