

Manual de usuario para Análisis Secuencial de Ensayos

**Kristian Thorlund, Janus Engstrøm, Jørn Wetterslev, Jesper Brok,
Georgina Imberger, y Christian Gluud**

Copenhagen Trial Unit

Centre for Clinical Intervention Research

Department 3344, Rigshospitalet

DK-2100 Copenhagen

Denmark

Tel. +45 3545 7171 Fax +45 3545 7101

E-mail: tsa@ctu.dk

Contenido

Manifiesto de responsabilidad limitada	4
Roles de los miembros del equipo y contribuciones	5
Prefacio.....	6
1 Conceptos y racionalidad del Análisis Secuencial de los Estudios....	7
1.1 El error aleatorio en el metanálisis	7
1.2 Definición de la solidez de la evidencia – Tamaño necesario de la muestra	9
1.3 Evaluación de la significación estadística antes de que se haya alcanzado el tamaño necesario de la muestra.....	10
1.4 Pruebas de futilidad antes de que el tamaño de la información haya sido alcanzado	11
1.5 Resumen	13
2 Metodología del TSA	15
2.1 Métodos para combinar los resultados de los ensayos clínicos	15
2.1.1 Medidas de efecto para los datos dicotómicos y continuos	15
2.1.2 Generalidades del modelo de efectos fijos y de efectos aleatorios.....	17
2.1.3 Estrategias para el metanálisis con el modelo de efectos aleatorio.....	19
2.1.4 Métodos para el manejo de ensayos clínicos sin eventos	22
2.2 Examinando la significación ajustada y la futilidad en el metanálisis acumulativo .	24
2.2.1 El tamaño de la información necesaria para un metanálisis concluyente	26
2.2.2 La prueba estadística acumulada (curva Z).....	35
2.2.3 Los problemas con las pruebas de significación en el metanálisis	36
2.2.4 La función del gasto de alfa y los límites de monitoreo secuencial de los ensayos.....	38
2.2.5 Intervalos de confianza ajustados tras el análisis secuencial de los ensayos....	45
2.2.6 La ley del logaritmo iterado	47
2.2.7 La función del gasto de β y los límites de futilidad.....	48
3 Instalación e inicio del programa TSA.....	54
3.1 Requisitos previos.....	54
3.2 Instalación.....	54
3.3 Iniciando el TSA.....	54
3.3.1 ¿Por qué no funciona el TSA?	55
3.4 Uso RM5 Converter	56
3.4.1 ¿Por qué no funciona?.....	56
4 Cómo utilizar el software TSA	56
4.1 Iniciando	56
4.1.1 Creación de un nuevo metanálisis	57
4.1.2 Guardando un archivo TSA y abriendo un archivo TSA existente	59
4.1.3 Importando los datos del metanálisis desde Review Manager v.5.....	59
4.2 Añadiendo, editando y eliminando los ensayos clínicos	63

4.2.1	Añadiendo ensayos clínicos.....	64
4.2.2	Edición y eliminación de ensayos.....	65
4.3	Estableciendo las configuraciones del metanálisis.....	66
4.3.1	Elección de la medida de asociación.....	66
4.3.2	La elección de su modelo estadístico.....	67
4.3.3	Elección de un método para el manejo de datos con ausencia de eventos.....	67
4.3.4	La elección del tipo de intervalo de confianza.....	68
4.4	Aplicación de las pruebas ajustadas de significación (aplicando TSA).....	69
4.4.1	Adición de una prueba de significación.....	70
4.4.2	Edición y eliminación de una prueba de significación.....	77
4.4.3	Adición y recuperación de las plantillas de pruebas de significación.....	78
4.4.4	Realizando los cálculos de las pruebas de significación.....	79
4.5	Opciones de gráficos en el TSA.....	81
4.6	Explorando la diversidad entre los ensayos.....	86
5	Ejemplos de aplicaciones del TSA.....	87
5.1	Los conjuntos de datos.....	87
5.2	Cómo evitar los falsos positivos.....	87
5.3	Confirmación de un resultado positivo.....	89
5.3.1	Confirmación de la "answer is in" Desconozco una traducción adecuada para esta expresión.....	90
5.3.2	Evitando sobreestimaciones precoces.....	93
5.4	Pruebas de inutilidad.....	96
5.5	La estimación del tamaño de la muestra de un nuevo ensayo clínico.....	97
5.6	Otras aplicaciones publicadas del Análisis Secuencial de los Ensayos.....	100
6	Apéndices.....	102
6.1	Medidas de efecto para metanálisis de los datos dicotómicos y continuos.....	102
6.2	Estrategia de efectos aleatorio.....	103
6.2.1	Fórmulas del método Biggerstaff-Tweedie.....	103
6.3	Análisis secuencial de los ensayos.....	103
6.3.1	Incremento del error tipo I debido a la repetición de pruebas de significación.....	103
6.3.2	Los métodos alternativos no implementados en el software TSA.....	104
7	Lista de abreviaturas y notaciones estadísticas.....	107
7.1	Abreviaturas generales.....	107
7.2	Notaciones estadísticas.....	107
7.2.1	Símbolos con letras minúsculas.....	107
7.2.2	Símbolos con letras mayúsculas.....	108
7.2.3	Símbolos en letras griegas.....	110
8	Referencias bibliográficas.....	111

Manifiesto de responsabilidad limitada

EL SOFTWARE ES PROPORCIONADO "TAL CUAL", SIN GARANTÍA DE NINGÚN TIPO, EXPRESA O IMPLÍCITA, INCLUYENDO PERO NO LIMITADO A LAS GARANTÍAS DE COMERCIALIZACIÓN, IDONEIDAD PARA UN PROPÓSITO PARTICULAR Y NO INFRACCIÓN. EN NINGÚN CASO, LOS AUTORES O TITULARES DEL DERECHO DE AUTOR SERÁN RESPONSABLES DE NINGUNA RECLAMACIÓN, DAÑO U OTRA RESPONSABILIDAD, YA SEA EN UNA ACCIÓN DE CONTRATO, AGRAVIO O DE OTRO TIPO, QUE SURJA DE O EN CONEXION CON EL SOFTWARE O EL USO U OTROS EN EL SOFTWARE.

BAJO NINGUNA CIRCUNSTANCIA Y BAJO NINGUNA TEORÍA LEGAL, YA SEA POR AGRAVIO, CONTRATO, O CUALQUIER OTRA FORMA, LA *CLINICAL TRIAL UNIT* SERÁ RESPONSABLE ANTE USTED O CUALQUIER OTRA PERSONA POR PÉRDIDA DE BENEFICIOS, PÉRDIDA DE FONDO DE COMERCIO, O CUALQUIER DAÑO INDIRECTO, ESPECIAL, INCIDENTAL O CONSECUENTES DAÑOS, O DAÑOS POR NEGLIGENCIA GRAVE DE NINGÚN TIPO, INCLUYENDO, SIN LIMITACIONES, DAÑOS POR PÉRDIDA DE FONDO DE COMERCIO, INTERRUPCIÓN DEL TRABAJO, FALLO O MAL FUNCIONAMIENTO, O POR CUALQUIER OTRO DAÑO O PÉRDIDA.

El software de análisis secuencial de los ensayos (en adelante TSA) a que se refiere este manual está en versión Beta. La *Copenhagen Trial Unit* ha evaluado ampliamente el software TSA; pero, aún pudieran ocurrir errores. La retroalimentación es una parte importante del proceso de corrección de errores y de la implementación de mejoras. Por lo que le animamos a que nos comunique acerca de sus experiencias con este software. Para ello, por favor visite la página Web desde donde descargó el TSA (ctu.rh.dk/TSA), y haga clic en el enlace de formulario de comentarios.

Roles de los miembros del equipo y contribuciones

El TSA se desarrolló en la *Copenhagen Trial Unit*, Copenhague, Dinamarca. El equipo está formado por Kristian Thorlund (KT), Janus Engstrøm (JE), Jørn Wetterslev (JW), Jesper Brok (JB), Georgina Imberger (GI), y Christian Gluud (CG). Las funciones y contribuciones de cada miembro del equipo se describen a continuación:

Jefe del proyecto: KT.

Principal desarrollador de las aplicaciones del software: JE.

Co-desarrolladores de las aplicaciones del software: KT, JW, JB, CG.

Programador estadístico: KT.

Evaluadores internos de la versión Beta: JB, GI, JW, KT, CG.

Autores del manual: KT (principal), GI, JW, JB, JE, CG.

Supervisores del proyecto: JW y CG.

Prefacio

Este manual proporciona una guía, teórica y práctica, para el uso del programa 'Análisis Secuencial de los Estudios' (TSA, por sus siglas en inglés) realizado en la *Copenhagen Trial Unit*. El capítulo 1 se refiere a los conceptos y racionalidad; el capítulo 2 aporta una descripción técnica de las metodologías implementadas, y los capítulos 3 al 5 se refieren a cómo instalar, utilizar y aplicar el software.

El software TSA puede ser descargado desde www.ctu.dk/tsa. Le invitamos a utilizarlo en sus análisis y publicaciones de metanálisis acumulados, con la debida referencia del software y de algunos de nuestros artículos que describen la metodología.

En caso de que necesite ayuda con el software TSA, por favor, comuníquese con nosotros a través del correo electrónico: tsa@ctu.dk.

1 Conceptos y racionalidad del Análisis Secuencial de los Estudios

1.1 *El error aleatorio en el metanálisis*

Algunos hallazgos metanalíticos «positivos» pueden ser debidos a la influencia del azar (errores aleatorios) más que a algún efecto «verdadero» subyacente de la intervención.^{1- 10} Del mismo modo, algunos hallazgos neutrales o «negativo» («no positivo») del metanálisis pueden, también, representar un «hallazgo casual», debido a la carencia de poder estadístico y precisión.⁹⁻¹³ Estos dos tipos de errores son comúnmente conocidos como errores falsos positivos (o errores tipo I) y errores falsos negativos (o errores tipo II). Los metanálisis son típicamente considerados «positivo» o «negativo» sobre la base de alguna prueba estadística (prueba estadística), reportada con el valor de p o con el intervalo de confianza correspondiente.

Cuando un metanálisis incluye un bajo número de ensayos clínicos y de pacientes, los errores aleatorios pueden causar conclusiones espurias.^{1, 2, 4-6, 9, 11, 12, 14,15} Por el contrario, cuando hay un gran número de pacientes, y cuando numerosos ensayos clínicos han confirmado los hallazgos de estudios previos, las pruebas estadísticas y los estimadores del efecto de la intervención típicamente convergen hacia la «verdad».^{1, 2, 4-6, 9, 11, 12, 14, 15} Las figuras 1 (A) y 1 (B) ilustran ejemplos de tal convergencia en las pruebas estadísticas. En ambas situaciones, las inferencias acerca de la significación estadística son erróneas en los momentos iniciales, pero finalmente convergen hacia el lado «verdadero» de la significación estadística.

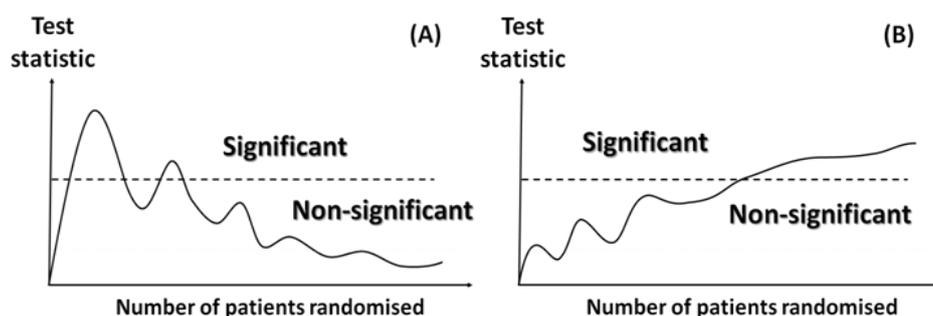


Figura 1. Ejemplos de la convergencia en las pruebas estadísticas cuando los pacientes son incluidos y seguidos en la medición de un desenlace (por ejemplo, muerte) en dos ensayos clínicos aleatorizados A y B.

El error aleatorio e imprecisión solo causan problemas si las pruebas estadísticas (y la estimación del efecto de la intervención) son utilizadas en etapas donde la magnitud del error aleatorio o imprecisión es «suficientemente extremo» para producir

inferencias estadísticas falsas. En la figura 2(A), las pruebas de significación, en los tiempos X_1 y X_3 se traduciría en una falsa declaración de significación estadística (es decir, un resultado falso positivo), mientras que las pruebas de significación en X_2 y X_4 no lo serían. De este modo, solo en los tiempos X_1 y X_3 , el impacto del error aleatorio es «suficientemente extremo» para producir resultados espurios estadísticamente significativos. En la figura 2(B), las pruebas de significación en X_1 y X_2 podría haber dado lugar a una declaración falsa de que las intervenciones objeto de la investigación no fueron significativamente diferentes (es decir, un resultado falso negativo), mientras que pruebas de significación en X_3 y X_4 no lo serían. Por lo tanto, solo en los momentos X_1 y X_2 es la imprecisión de una magnitud lo que provoca la ausencia espuria de significación estadística.

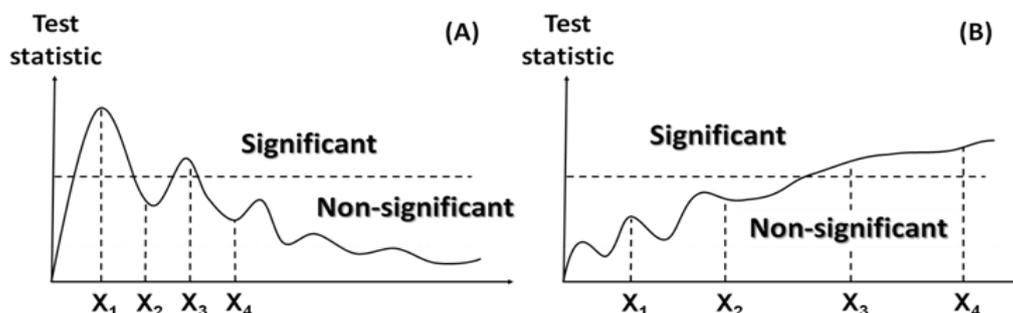


Figura 2. Ejemplos de resultados estadísticamente falsos positivos y falsos negativos durante la ejecución de dos ensayos clínicos aleatorizados A y B.

La mayoría de las pruebas estadísticas empleadas, a medida que se acumulan datos adicionales, incrementan la probabilidad de observar un resultado falso positivo o falso negativo. Este fenómeno se conoce comúnmente como «multiplicidad debido a pruebas repetidas de significación».^{10, 16-18}

En el metanálisis es importante minimizar el riesgo de realizar una conclusión falsa, positiva o negativa.³ En el metanálisis, los efectos combinados de la intervención suelen evaluarse sobre la base de los valores de p . Los autores del metanálisis deben decidir cuál es el umbral en el cual un valor de p es suficientemente pequeño, para justificar una conclusión «positiva». Por debajo de este umbral, una conclusión es considerada estadísticamente significativa. En un momento dado, cualquier umbral implica un equilibrio entre el riesgo de observar un resultado falso positivo (error tipo I) y el riesgo de observar un resultado falso negativo (error tipo II). Por ejemplo, si el umbral para la significación estadística en la Figura 2 (línea horizontal discontinua) hubiese sido desplazado hacia arriba, la posibilidad de observar un resultado falso positivo (figura 2 (A)) habría disminuido, mientras que el riesgo de observar un falso resultado «negativo» (figura 2 (B)) habría aumentado. Cuando las pruebas de significancia convencionales se realizan en etapas «tempranas» y/o en varias

ocasiones, estos riesgos máximos están distorsionados (como se ilustra en la figura 2).¹⁶⁻¹⁸ Por lo tanto, cualquier inferencia acerca de la significación estadística debería ser realizada en conjunto con la solidez de la evidencia, la cual debería ser medida utilizando el número acumulado de pacientes, número observado de eventos en los ensayos incluidos, y el impacto de la multiplicidad.^{1, 2, 4, 6, 10, 19-21}

1.2 **Definición de la solidez de la evidencia – Tamaño necesario de la muestra**

Los metanálisis de ensayos aleatorizados aumentan la potencia y precisión de los efectos estimados de la intervención.¹³ Cuando se incluyen todos los ensayos disponibles; las revisiones sistemáticas y los metanálisis son considerados como la mejor evidencia disponible.¹³ Sin embargo, «la mejor evidencia disponible» puede no ser sinónimo de «evidencias suficientes» o «evidencias sólidas».^{1; 2; 4; 6; 11; 12}

En un ensayo aleatorizado con desenlace binario se debe estimar el número de eventos y pacientes necesarios para permitir una inferencia estadística confiable. Es decir, se realiza el cálculo del tamaño de la muestra para asegurar que un número «suficiente» de eventos y de pacientes sean incluidos.²² Una estrategia similar, denominada ‘goal post’ se requiere para un metanálisis.^{1, 2, 6, 23} En el campo del metanálisis el ‘goal post’ ha sido denominado como el tamaño de la información (IS) requerida o el tamaño óptimo de información.^{1, 2, 4, 6, 11, 12,14,15, 19; 23-25}

La Figura 3 muestra dos escenarios metanalíticos típicos A y B, donde la estadística de prueba se ha estabilizado después de haber alcanzado el tamaño de la información requerida.

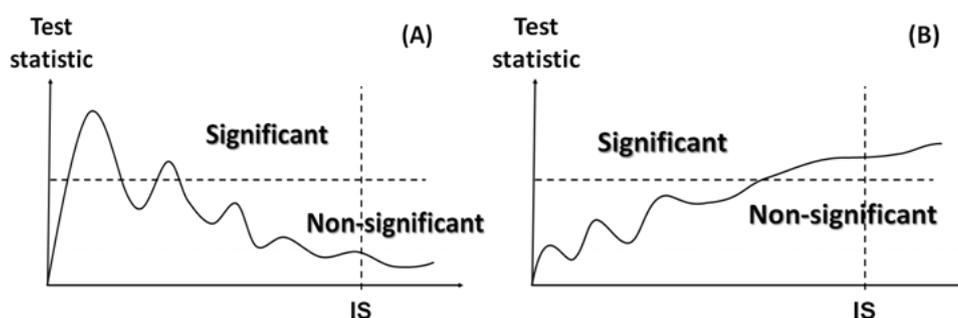


Figure 3. Ejemplos de cómo el tamaño de la información requerida asegura pruebas de significación confiables en dos metanálisis acumulados, A y B.

En un ensayo clínico, el cálculo del tamaño de muestra se basa generalmente en la proporción esperada del evento en el grupo control, la reducción relativa del riesgo esperado de la intervención experimental, y el riesgo máximo deseado de los errores tipo I y tipo II.²⁶ En un metanálisis es posible que haya heterogeneidad en las

poblaciones de los ensayos incluidos, las intervenciones y los métodos. Las consideraciones del tamaño de la muestra del metanálisis necesitan ser ajustadas -es decir, incrementadas- con el fin de permitir la varianza introducida por esta heterogeneidad.^{4, 6, 11, 12, 23} Estos ajustes son análogos a los ajustes por la variación a través de los centros en un ensayo multicéntrico.^{4; 6; 23}

Los métodos convencionales de metanálisis, tales como los que están disponibles en el *Review Manager* v.5.1²⁷ no toman en cuenta la cantidad de la evidencia disponible.¹³ En su lugar, la fiabilidad del efecto de la intervención estadísticamente significativa se da por descontada, independientemente del número acumulado de eventos y pacientes. Por el contrario, los efectos de la intervención que no son estadísticamente significativos no son considerados fiables comúnmente. Más bien, se supone que «se necesita más evidencia».²⁸

La evidencia empírica sugiere que los efectos de la intervención y valores de p , basados en un número limitado de eventos y pacientes, a menudo no son fiables.^{1, 2, 4-6, 9, 11, 12, 29}

Cerca del 25 % de los metanálisis convencionales que incluyen un número pequeño de eventos y de pacientes, pueden mostrar falsamente los efectos de la intervención como estadísticamente significativos.^{4, 5} La evidencia empírica también muestra, que los grandes efectos de la intervención observados en etapas tempranas o precoces de metanálisis positivos tienden a desaparecer a medida que se acumula más evidencia.^{4, 5, 9}

1.3 *Evaluación de la significación estadística antes de que se haya alcanzado el tamaño necesario de la muestra.*

El objetivo del metanálisis es identificar el beneficio o el daño de una intervención tan pronto y de manera tan fiable como sea posible.^{4, 11-13, 20} Por lo tanto, los metanálisis se actualizan frecuentemente cuando se publican nuevos ensayos. Por ejemplo, a los autores de revisiones Cochrane se les solicita actualizarlas por lo menos cada dos años.¹³ Cuando los metanálisis son actualizados, se exponen a una evaluación repetida de la significación estadística a lo largo del tiempo.

En los ensayos clínicos aleatorizados, las repetidas evaluaciones de la significación estadística incrementan el riesgo global del error tipo I.³⁰

Los estudios de simulación sugieren que si la repetición de pruebas de significación es realizada en los metanálisis y los valores de p menor que 0,05 son considerados como evidencia de «significación estadística», entonces el riesgo verdadero de error tipo I se situará entre el 10% y el 30%.^{7, 8, 10, 31} De acuerdo con esto, cuando se toman

decisiones para implementar la intervención como tratamiento, significa que entre 1 y 3 de cada 10 decisiones terapéuticas son probablemente inapropiadas.

Para hacer frente a este problema, se pueden ajustar los umbrales para determinar qué resultados son considerados estadísticamente significativos y cuáles no.^{1, 2, 4, 6, 11, 12, 14, 15, 24, 25} Alternativamente, se puede penalizar a la prueba estadística de acuerdo con la solidez de la evidencia y el número de pruebas de significación realizado (la «ley del logaritmo repetido»^{7,8}). El software TSA proporciona métodos para ambos enfoques, cada uno de ellos está construido sobre los teoremas de la teoría avanzada de probabilidades. El primer método utiliza la metodología desarrollada para las pruebas repetidas de significación en ensayos clínicos aleatorizados (es decir, límites de monitoreo estadístico).^{4, 6, 11, 12} El segundo enfoque penaliza, es decir- disminuye-, la prueba estadística de acuerdo con la solidez de la información disponible en el metanálisis y el número de pruebas de significancia realizadas.^{7, 8}

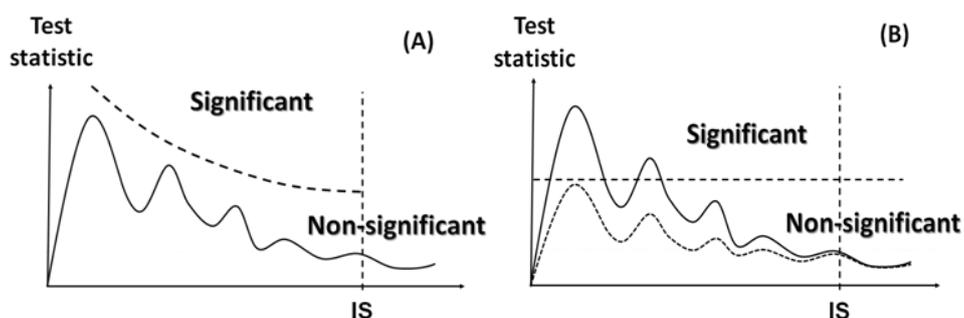


Figura 4. Ejemplos de ajuste del umbral de significación (límites predeterminados de monitorización) (A) y la prueba estadística penalizada (predeterminado) (B) para evitar resultados falsos positivos de la prueba estadística de dos metanálisis acumulados A y B.

Figura 4 (A) ilustra un ejemplo de un escenario de metanálisis, donde un resultado falso positivo se evita utilizando el ajuste del umbral para la significación estadística mediante el empleo de los límites de control. Figura 4 (B) ilustra un ejemplo donde un resultado falso positivo es evitado por la adecuada penalización de la prueba estadística.

1.4 ***Pruebas de futilidad antes de que el tamaño de la información haya sido alcanzado***

Es posible utilizar el software de TSA para evaluar cuando es poco probable que una intervención tenga algún efecto anticipado; también es posible su uso en el contexto clínico, para evaluar cuando una intervención tiene un efecto que es más pequeño que lo que sería considerado mínimamente importante para los pacientes.

Los metanálisis se utilizan a menudo para guiar la investigación futura. En la planificación de futuros ensayos, los investigadores necesitan tener un resumen preciso de los conocimientos actuales. Si un metanálisis ha encontrado que una determinada intervención no tiene efecto (importante), los investigadores necesitan saber si este resultado se debe a la falta de poder estadístico o si la intervención carezca de efecto. Utilizando el pensamiento convencional, un hallazgo de «ningún efecto» se considera que es debido a carencia de poder estadístico hasta cuando se haya alcanzado el tamaño adecuada de información. En algunas situaciones, sin embargo, es posible que podamos llegar a la conclusión anterior de que es poco probable que sea tan grande como se había previsto un efecto del tratamiento, y por lo tanto, prevenir a los investigadores del ensayo de gastar recursos en ensayos adicionales innecesarios. Por supuesto, el tamaño del efecto de la intervención prevista puede ser examinado de nuevo y la investigación adicional puede ser diseñada para investigar un tamaño del efecto más pequeño.

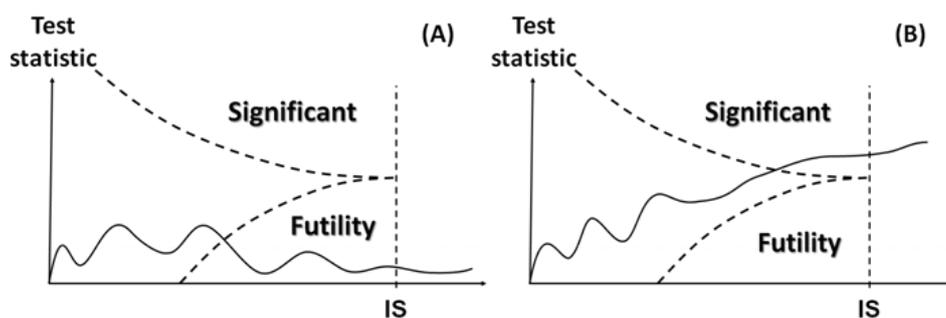


Figura 5. Ejemplos de límites de futilidad donde la intervención experimental no es superior a la intervención de control (y también se pueden haber llevado a cabo muchos ensayos) (A) y donde la intervención experimental es significativamente superior a la intervención de control (y demasiados ensayos pueden haber llevado a cabo) (B).

El TSA proporciona una técnica para la búsqueda de una conclusión sin efecto tan pronto como sea posible. Se construyen y se usan «Límites de Futilidad» para proporcionar un umbral de «ningún efecto», que fueron desarrollados originalmente para el análisis intermedio en ensayos clínicos aleatorios.³⁰

Si la intervención experimental es realmente superior a la intervención control, se esperaría que la prueba estadística fluctúe en torno de algunas líneas rectas de pendiente positiva, generando, eventualmente, la significación estadística (cuando el metanálisis es suficientemente potente). Si un metanálisis de una intervención experimental verdaderamente eficaz incluye solo un pequeño número de eventos y de pacientes, será baja la probabilidad de obtener un resultado estadísticamente significativo, debido a la falta de poder estadístico. Sin embargo, a medida que se acumule más evidencia, el riesgo de obtener una conclusión negativa disminuye. Los límites de futilidad son un conjunto de umbrales que reflejan la probabilidad de la

incertidumbre para obtener un resultado negativo en relación con la fuerza de la evidencia disponible (por ejemplo, el número acumulado de pacientes). Por encima de los umbrales, la prueba estadística puede no haber generado significación estadística debido a la falta de poder estadístico, pero todavía hay posibilidad de que un efecto estadísticamente significativo sea encontrado antes de que el metanálisis sobrepase el tamaño de la información. Por debajo del umbral, la prueba estadística es tan baja que la probabilidad se torna insignificante para encontrar un efecto significativamente importante. En este último caso, aleatorización de pacientes es vana; la intervención carece del efecto postulado.

La figura 5 (a) ilustra un ejemplo donde la intervención experimental no es superior a la intervención de control. La prueba estadística cruza los límites de futilidad (la curva cóncava de pendiente positiva) antes de que se supere el tamaño de la información requerida. La figura 5 (b) ilustra un ejemplo donde la intervención experimental es estadísticamente significativamente superior a la intervención de control. En este ejemplo, la estadística de prueba se mantiene por encima de la curva de la futilidad (porque hay un efecto subyacente) y, finalmente, produce la significación estadística.

1.5 **Resumen**

El análisis secuencial de los ensayos (TSA, por sus siglas en inglés) es una metodología que utiliza una combinación de técnicas. Las pruebas necesarias se cuantifican, proporcionando un valor para el tamaño requerido de la información. Los umbrales de significación estadística son ajustados y estas modificaciones se realizan de acuerdo con la fuerza cuantitativa de la evidencia y el impacto de multiplicidad.^{4; 6; 1; 12} Los umbrales para futilidad también se pueden construir, utilizando un marco estadístico similar.

En resumen, el TSA puede proporcionar un IS (por su siglas en inglés), un umbral para un efecto de tratamiento estadísticamente significativo, y el umbral de insignificante utilidad. Las conclusiones obtenidas utilizando el TSA muestran que la fiabilidad de éstas es superior a las obtenidas mediante las técnicas tradicionales de metanálisis. La evidencia empírica sugiere que las consideraciones de tamaño de la información y umbrales de importancia ajustados pueden eliminar precozmente los resultados falsos positivos, que se originan por la imprecisión y la prueba de significación repetida en los metanálisis.^{4; 6; 11; 12}

Alternativamente, se puede penalizar a la prueba estadística de acuerdo con la solidez de la evidencia y el número de pruebas de significación realizadas (la «ley del logaritmo iterado»).^{7; 8} Los estudios de simulación han demostrado que la penalización

de las pruebas estadísticas puede permitir un adecuado control del error de tipo I en los metanálisis.^{7; 8}

El siguiente manual contiene una guía -teórica y práctica- para el uso del software TSA de la *Copenhagen Trial Unit*. El capítulo 2 proporciona, en un nivel intermedio, un resumen técnico de todas las metodologías incorporadas en el software del TSA. Los capítulos 3 al 5 son apartados prácticos que describen cómo instalar, utilizar y aplicar este software.

2 Metodología del TSA

El TSA combina la metodología convencional del metanálisis con los fundamentos metanalíticos del tamaño de la muestra (es decir, el tamaño de la muestra necesaria) y los métodos ya desarrollados para las pruebas de significación repetida sobre los datos acumulados en los ensayos clínicos aleatorizados.^{1, 2, 4, 6, 11, 12} En el capítulo 2, lo primero que se describe es la metodología de metanálisis usado para combinar los datos de una serie de ensayos clínicos. La descripción en la sección 2.1 se refiere a las medidas del efecto para los datos dicotómicos y continuos, modelos estadísticos de metanálisis (el modelo de efecto fijo y algunas variantes del modelo de efectos aleatorio), y métodos para el manejo de datos de eventos cero. En la sección 2.2, se describen los métodos para ajustar la significación cuando existe un mayor riesgo de error aleatorio (debido a una evidencia débil y pruebas de significación repetida). En esta sección, no se describe, en detalle, componentes más avanzados de esta metodología. Más bien, este capítulo está diseñado para aportar conocimientos a los usuarios, con un nivel conceptual intermedio, sobre los temas abordados en el capítulo 1.

2.1 *Métodos para combinar los resultados de los ensayos clínicos*

2.1.1 **Medidas de efecto para los datos dicotómicos y continuos**

El programa TSA facilita el metanálisis de los datos dicotómicos (binario) y de los datos continuos. Los datos dicotómicos se definen por una de las dos categorías (por ejemplo, muerte o supervivencia). Los datos continuos son datos que se miden en una escala numérica (por ejemplo, la presión arterial o las puntuaciones de calidad de vida). Para cada tipo de datos, hay varias medidas disponibles para comparar la efectividad de una intervención de interés.¹³

Medidas de efecto para los datos dicotómicos

Asumamos que tenemos k ensayos independientes que compararon dos intervenciones (intervención A vs intervención B) con un resultado dicotómico. Dichos ensayos (por lo general) reportarán el número de eventos observados (por ejemplo, muertes) en los dos grupos de intervención, EA y EB y el número total de participantes, NA y NB , en los dos grupos de intervención. Para los datos dicotómicos, el efecto de la intervención entre las dos intervenciones se puede medir como la diferencia de

riesgo (DR), riesgo relativo (RR), y odds ratio (OR).¹³ Las estimaciones del efecto de intervención basadas en estas medidas se calculan utilizando las siguientes fórmulas:

$$RD = \frac{e_A}{n_A} - \frac{e_B}{n_B}$$

$$RR = \frac{(e_A/n_A)}{(e_B/n_B)}$$

$$OR = \frac{e_A/(n_B - e_B)}{e_B/(n_A - e_A)}$$

Las razones de riesgo relativo y odds ratios típicamente serán expresadas en escala logarítmica debido a que la transformación genera ciertas propiedades estadísticas deseables (tales como la simetría y la normalidad aproximada).¹³ Los errores estándar, varianzas y pesos de los "efectos de la intervención relación" son, por lo tanto, también obtenidas en escala logarítmica. Las fórmulas de los errores típicos de la DR, log(RR) y log(OR) se muestran en el apéndice 6.1.

Cuando las proporciones de eventos en los dos grupos son bajos (datos de eventos raros), una alternativa preferida al OR es el odds ratio de Peto.¹³ Este OR se calcula con la fórmula:

$$OR_{Peto} = \exp\left(\frac{(e_A - E(e_A))}{v}\right)$$

Donde E (eA) es el número esperado de eventos en el grupo de intervención A, y v es la varianza hipergeométrica de eA. Las fórmulas para E (eA) y v se muestran en el apéndice 6.1.

Medidas de efecto para datos continuos

Asumamos que tenemos k ensayos independientes que compararon dos intervenciones (intervención A contra intervención B) con un desenlace continuo. Tales ensayos informan a menudo la respuesta promedio (por ejemplo, la media de la puntuación de calidad de vida) en los dos grupos de intervención, m_A y m_B , las desviaciones estándar de los dos grupos de intervención significan respuestas, SDA y SDB, y el número total de participantes en los dos grupos de intervención, n_A y n_B . Cuando la respuesta promedio se mide en la misma escala para todos los ensayos,

la eficiencia comparativa se mide con la diferencia de medias (DM), que viene dada por $m_A - m_B$. El error estándar de la diferencia de media está dada por

$$SE(MD) = \sqrt{\frac{sd_A^2}{n_A} + \frac{sd_B^2}{n_B}}$$

Cuando la media de la respuesta no es medida en la misma escala, aquellas deben ser estandarizadas a la misma escala, permitiendo la combinación de los ensayos clínicos.¹¹ El método convencional consiste en dividir la media de la respuesta en cada ensayo por su desviación estándar estimada; proporcionando así, una estimación del efecto medido en unidades de desviación estándar. Las diferencias de medias dividido por su desviación estándar se conocen como las diferencias de medias estandarizadas (DME).¹³

El programa TSA no facilita el metanálisis con DME. Las pruebas de significación ajustada para metanálisis utilizando DME requeriría el cálculo del tamaño de la muestra necesaria sobre la base de las diferencias de medias esperadas reportadas en unidades de desviación estándar. Esta medida del efecto carece inapropiadamente de sentido para la mayoría de los médicos y, por lo tanto, es propenso a producir información poco realista de tamaños requeridos de información

2.1.2 Generalidades del modelo de efectos fijos y de efectos aleatorios

Asumamos que tenemos k ensayos independientes. Permítame decir que Y_i es el efecto de la intervención observada en el ensayo i -th. Para metanálisis de datos dicotómicos, Y_i será o bien la diferencia estimada del riesgo, el riesgo relativo de registro, el log odds ratio, o log de OR de Peto para el ensayo i -th. Para el metanálisis de datos continuos, Y_i será la diferencia de media estimada para el ensayo i -th. Asumamos que μ_i sea el verdadero efecto del ensayo i -th y μ es el verdadero efecto de la intervención (para toda la población del metanálisis). Suponga que σ_i^2 denotan la varianza (error de muestreo) del efecto de la intervención observada en el ensayo i .

En el modelo de efecto fijo, las características de los ensayos incluidos (criterios de inclusión y exclusión de pacientes, las variantes administradas de la intervención, el diseño del estudio, la calidad metodológica, la duración del seguimiento, etc.) se supone que son similares.¹³ Esto es formulado matemáticamente como $\mu_1 = \mu_2 = \dots = \mu_k = \mu$. Los efectos de la intervención observados de los ensayos individuales se supone que satisfacen la relación distributiva $Y_i \sim N(\mu, \sigma_i^2)$. El peso de un ensayo, w_i , se define como el inverso de la varianza del ensayo, y por lo tanto, los pesos del

ensayo, en un modelo de efecto fijo, se convierten $w_i = \sigma_i^{-2}$. El efecto combinado de la intervención, $\hat{\mu}$, se obtiene como un promedio ponderado de los efectos de la intervención observada de los ensayos individuales

$$\hat{\mu} = \frac{\sum w_i Y_i}{\sum w_i}$$

Y tiene varianza

$$\text{Var}(\hat{\mu}) = \frac{1}{\sum w_i}$$

En el modelo de efectos aleatorio, se asume que los efectos de la intervención varían entre los ensayos, pero con un verdadero efecto subyacente, μ . Permítanos τ^2 significa la varianza entre los ensayos, el modelo de efectos aleatorio se define de la siguiente manera

$$Y_i = \mu_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$\mu_i = \mu + E_i, \quad E_i \sim N(0, \tau^2)$$

Donde ε_i es el error residual (muestreo) para el ensayo i , y E_i es la diferencia entre el "verdadero" efecto global y el "verdadero" efecto del ensayo subyacente. Si se unen la estructura jerárquica en las ecuaciones anteriores, se puede que Y_i satisface la relación de distribución $Y_i \sim N(\mu, \sigma^2 + \tau^2)$. Una vez más, los pesos de los ensayos se definen como el inverso de la varianza, y así el peso de los ensayos clínicos en un modelo de efectos aleatorio se convierten $W_i^* = (\sigma^2 + \tau^2)^{-1}$. El efecto combinado de la intervención $\hat{\mu}$, se obtiene como un promedio ponderado de los efectos de la intervención observados de los ensayos individuales.

$$\hat{\mu} = \frac{\sum w_i^* Y_i}{\sum w_i^*}$$

y tiene varianza

$$\text{Var}(\hat{\mu}) = \frac{1}{\sum w_i^*}$$

El estudio de la significación estadística es realizada con la prueba estadística de tipo Wald, la cual es igual al efecto metanalizado de la intervención (escala log de los riesgos relativos y odds ratios) dividido por su error estándar:

$$Z = \frac{\hat{\mu}}{\sqrt{\text{Var}(\hat{\mu})}}$$

Esta prueba estadística es la conocida típicamente como el estadístico Z o el valor Z. Bajo el supuesto de que dos intervenciones investigadas no difieren del valor Z, aproximadamente seguirán una distribución normal estándar (una distribución normal con media 0 y desviación estándar 1). Este supuesto también se conoce como la *hipótesis nula* y se denota H_0 . El correspondiente valor de p de dos colas se puede obtener usando la siguiente fórmula:

$$P = 2 \cdot (1 - \Phi(|Z|))$$

Donde $|Z|$ denota el valor absoluto del valor Z y Φ denota la función de distribución de la probabilidad estándar normal acumulada.¹³ El valor P es la probabilidad de observar un valor Z al menos tan «extremo» como la que se observa debido a la intervención del azar. Cuanto menor sea el valor P, menor será la probabilidad de que la diferencia observada entre los dos grupos de intervención no sea más que un hallazgo casual, y por lo tanto, mayor es la probabilidad de que la diferencia observada se deba a algún efecto del tratamiento «verdadero» subyacente.

2.1.3 Estrategias para el metanálisis con el modelo de efectos aleatorio

Como se explicó, el modelo de efectos aleatorio intenta incluir una cuantificación de la variación a través de los ensayos.¹³ El enfoque común es estimar la varianza entre ensayos, τ^2 , con un poco de variación estimada entre ensayos.¹³

El método de DerSimonian-Laird

El estimador de la varianza entre los ensayos que se ha utilizado con mayor frecuencia en la práctica de metanálisis (y es la única opción en el software *Review Manager* de la Colaboración Cochrane) es el estimador propuesto por DerSimonian y Laird (DL)^{13; 27; 32} El estimador DL se calcula mediante

$$\tau_{DL}^2 = \max(0, (Q - k + 1) / (S_1 - (S_2 / S_1)))$$

Donde Q es el estadístico de la prueba de homogeneidad Cochran dada por $Q = \sum w_i (Y_i - \hat{\mu})^2$, donde $S_r = \sum w_i^r$, para $r = 1, 2$, y donde k es el número de ensayos incluidos en el metanálisis.^{13; 32}

Debido a que el estimador DL es propenso a subestimar la varianza entre ensayos,³³⁻⁴⁰ el TSA tiene, además, dos alternativas al modelo de efectos aleatorio, el Sidik y Jonkman (SJ) y los métodos Biggerstaff y Tweedie (BT)^{33; 34; 41}.

El método de Sidik-Jonkman (SJ)

El modelo de efectos aleatorio SJ utiliza un estimador sencillo (no iterativo) de la varianza entre los ensayos basado en una re-parametrización de la varianza total del efecto estimado de la intervención observada Y_i .^{33;34} Esto viene dado por la expresión:

$$\tau_{SJ}^2 = \sum v_i (Y_i - \mu_0)^2 / (k-1)$$

en la cual $v_i = r_i + 1$, $r_i = \sigma_i^2 / \tau_0^2$, y τ_0^2 es una primera estimación de la variación entre los ensayos, que se puede definir, por ejemplo, como

$$\tau_0^2 = \sum (Y_i - \mu_{uw})^2 / k$$

μ_{uw} es la media no ponderada de las estimaciones de los efectos observados, y μ_0 es el estimador de los efectos aleatorios ponderados utilizando τ_0^2 como la estimación de la varianza entre ensayos. Los estudios de simulación han demostrado que el estimador SJ proporciona estimaciones menos sesgadas hacia la baja de la variación entre ensayos que el estimador DL.^{34; 37} Es decir, el método SJ es menos probable que subestime la heterogeneidad entre los ensayos. Este es el caso particular para metanálisis donde existe heterogeneidad moderada o alta. Los intervalos de confianza basados en el estimador SJ tienen una cobertura cercana al nivel deseado (por ejemplo, intervalos de confianza del 95 % tendrán el verdadero efecto en aproximadamente el 95 % de todos los metanálisis).^{34; 37} Por el contrario, la cobertura comúnmente reportada de los intervalos de confianza basado en el estimador DL está a menudo por debajo del nivel deseado.^{33; 35-38} Por ejemplo, muchos estudios de simulación que han investigado la cobertura de los intervalos de confianza del 95 % basados en DL han encontrado una cobertura real de 80 % a 92 %.^{34; 37} El tamaño de estos intervalos de confianza es equivalente a una proporción de falsos positivos del 8 % al 20 %, lo cual es evidentemente mayor que el 5 % convencionalmente aceptado.

El método Biggerstaff - Tweedie

Debido, a que la mayoría de los metanálisis contienen solo un número limitado de ensayos clínicos, la estimación de la varianza entre los ensayos es a menudo objeto de error aleatorio.⁴¹ Por lo tanto, la presencia de la incertidumbre en estimación de la varianza entre los ensayos en el modelo de efectos aleatorios debe ser advertida. Biggerstaff y Tweedie (BT) propusieron un método para lograr tal incorporación.⁴¹ Ellos derivaron una distribución de probabilidad aproximada, F_{DL} , para la estimación de DL de τ^2 . Definiendo los pesos de los ensayos, según $w_i(t) = ((\sigma_i^2 + t)^{-1})$, donde t es una variable que puede asumir todos los valores posibles para τ^2 , ellos utilizaron F_{DL} y obtuvieron el peso de los ensayos que toman en cuenta la incertidumbre del estimado τ^2 . Generalmente, esto crea un esquema de ponderación lo cual, relacionado con el enfoque de DL, atribuye más peso a ensayos de mayor tamaño y menos peso a ensayos más pequeños. Biggerstaff y Tweedie también propusieron una fórmula ajustada por la variación del efecto metanalizado de la intervención, facilitando así los intervalos ajustados de confianza (véase el apéndice, sección 6.2.1).

¿Cuál estrategia de los efectos aleatorio puede ser la mejor?

Los métodos SJ y BT presentan ventajas relativas sobre el enfoque DL. Sin embargo, estos métodos tienen sus propias limitaciones y es improbable que sean superiores en todos los casos. El estimador SJ puede sobreestimar la varianza entre los ensayos en los metanálisis con heterogeneidad leve, produciendo de esta manera, artificialmente, amplios intervalos de confianza.^{34, 37} Se ha demostrado que el enfoque de BT proporcionan una cobertura similar a los intervalos de confianza del enfoque DL en los metanálisis con ensayos no sesgados de bajo tamaño de muestra.³⁵ Sin embargo, cuando los ensayos incluidos difieren en tamaño y algunos ensayos pequeños están sesgados, el enfoque de BT pondrá apropiadamente altos pesos en los ensayos más grandes, mientras que continua considerando la heterogeneidad. Este punto es importante porque una crítica común del modelo de efectos aleatorios DL es que a los ensayos pequeños a menudo se asignan artificialmente grandes pesos en metanálisis heterogéneos. Una solución generalmente aplicada, y poco satisfactoria, es utilizar en este caso el modelo de efecto fijo. De esta manera, el estimador combinado puede estar menos sesgado por el uso de un esquema de ponderación inadecuada, pero los intervalos de confianza también serán artificialmente estrechos porque no toman en cuenta la heterogeneidad. El enfoque de BT reduce el sesgo incurrido por la ponderación inapropiada del modelo de efectos aleatorio mientras que continúa considerando la heterogeneidad.

La elección del modelo de efectos aleatorios debería incluir un análisis de sensibilidad comparando cada estrategia. Si DL, SJ, y BT mantienen inferencias estadísticas

similares (es decir, las estimaciones puntuales y los intervalos de confianza), sería razonable utilizar el enfoque DL y tener confianza en que la estimación de la varianza entre ensayo es confiable.

Si dos (o todos) de las tres estrategias son diferentes, se debe llevar a cabo un metanálisis con ambos (o todos) estrategias y considerar los resultados de acuerdo con las propiedades subyacentes de cada enfoque. Por ejemplo, si los enfoques DL y SJ producen resultados diferentes, dos posibles explicaciones deben ser consideradas: 1) el metanálisis está sujeto a una heterogeneidad moderada o sustancial y por lo tanto el estimador DL subestima la varianza entre los ensayos y produce intervalos de confianza artificialmente estrechos, y 2) el metanálisis está sujeto a heterogeneidad leve y por lo tanto el estimador SJ sobreestima la varianza entre ensayos y produce intervalos de confianza artificialmente anchos. En esta situación, se debe entonces llevar a cabo metanálisis con los dos enfoques y considerar las implicaciones de cada uno de los dos escenarios siendo «verdad».

2.1.4 Métodos para el manejo de ensayos clínicos sin eventos

En los ensayos dicotómicos, el desenlace de interés puede ser raro. Por ejemplo, la incidencia de una enfermedad cardíaca por el uso de hormona de reemplazo hormonal es muy baja.⁴² Algunas veces hay ausencia de eventos de un desenlace «eventos Cero» en un grupo. En esta situación la medida de la razón del efecto (RR y OR) no dará una estimación útil del efecto de la intervención.⁴² Una solución a este problema es añadir alguna constante al número de eventos y no eventos de cada grupo de intervención. Esta estrategia es conocida como *corrección de continuidad*.⁴² Numerosas estrategias de corrección continua han sido propuestas en la literatura metanalítica.

Corrección constante de continuidad

La corrección constante de continuidad es un método simple y es el más comúnmente utilizado en la literatura metanalítica.⁴² El método involucra la adición de un factor de corrección continuo (una constante) al número de eventos y no eventos en cada grupo de intervención.

Grupo	Eventos	No eventos	Total
Intervención	0	20	20
Control	5	20	25

Tabla 1. Ejemplo de un ensayo sin eventos

Considere el ejemplo del ensayo con evento cero que se muestra en la tabla 1. Si, por ejemplo, el método de corrección constante de continuidad utiliza un factor de corrección de 0,5, el número de eventos en el grupo de intervención se transformará en $0 + 0,5 = 0,5$, el número de no eventos en el grupo de intervención será ahora igual a $20 + 0,5 = 20,5$, el número de eventos en el grupo control resulta en $5 + 0,5 = 5,5$; para finalizar, el número de no eventos en el grupo control será $20 + 0,5 = 20,5$. Debido a que el número total de pacientes es el número de eventos más el número de no eventos, el número total de pacientes (después de la corrección constante de continuidad con la constante 0,5) será $20,5 + 0,5 = 21$ en el grupo de intervención y $20,5 + 5,5 = 26$ en el grupo control.

Si, por ejemplo, se utiliza un factor de corrección de 0,1, el número de eventos y el número total de pacientes (después de la corrección de continuidad) serían entonces 0,1 y 20,2 en el grupo de intervención y 5,1 y 25,2 en el grupo control.

La versión 5 del *Review Manager* utiliza corrección constante de continuidad utilizando un factor de corrección de 0,5.^{13, 27} Los estudios de simulación han demostrado problemas con el uso de esta constante, esto produce estimaciones inexactas cuando la relación de asignación aleatorizada no es 1:1, y produce intervalos de confianza que son demasiado estrechos.⁴²

Corrección de continuidad utilizando la recíproca del opuesto del grupo de intervención

Otro método potencial de corrección de continuidad es agregar el recíproco del número total de pacientes en el opuesto al grupo de intervención al número de eventos y no eventos.⁴² Este tipo de corrección de continuidad es también comúnmente conocido como corrección de continuidad del «grupo de tratamiento».⁴² En el ejemplo de la tabla 1, el factor de corrección para el grupo de intervención sería $1/25 = 0,04$, y el factor de corrección para el grupo control sería $1/20 = 0,05$. Este método de corrección continua provoca 0,04 eventos y 20,04 pacientes en el grupo de intervención y 5,05 eventos y 25,05 pacientes en el grupo control.

Corrección de continuidad empírica

Tanto el método de corrección de continuidad constante y el método de corrección de continuidad «grupo de tratamiento» impulsan los estimadores del efecto de la intervención hacia el efecto nulo (es decir, hacia 0 para las diferencias de riesgo y

hacia 1 para medidas de razón).⁴² Una alternativa de corrección continua es la corrección de continuidad *empírica* que arrastra el estimador del efecto de la intervención hacia el efecto metanálisis.⁴² Por ejemplo, asuma que $\hat{\theta}$ sea el odds ratio del metanálisis que no incluye los ensayos de eventos cero, y sea R la relación de aleatorización en el ensayo clínico que necesita corrección de continuidad. El factor de corrección de continuidad para el grupo de intervención, CF_I , y la corrección de continuidad para el grupo control, CF_C , puede ser calculada con las siguientes fórmulas:

$$CF_I = \frac{R}{R + \hat{\theta}} \cdot C$$

$$CF_C = \frac{\hat{\theta}}{R + \hat{\theta}} \cdot C$$

bajo la restricción de que dos correcciones de continuidad se suman a algunos constantes C.⁴²

2.2 ***Examinando la significación ajustada y la futilidad en el metanálisis acumulativo***

Examinar la significación ajustada en el metanálisis acumulativo tiene dos objetivos: debe medir y tomar en cuenta la fortaleza de la evidencia disponible y debe controlar el riesgo de errores estadísticos (error tipo I y tipo II) cuando se producen repetidamente pruebas de significación sobre los datos que se van acumulando.

La cuantificación de la fortaleza de la evidencia disponible exige la definición de un '*goal post*'.^{1, 2, 4, 6, 11, 12, 23} En el programa de TSA se mide la fortaleza de la evidencia disponible, y es considerada, para calcular el tamaño de información requerida. Esta información del tamaño es análoga al tamaño requerido de la muestra en un único ensayo clínico aleatorizado.^{1, 2, 4, 6, 11, 12, 23}

Controlar el riesgo de error tipo I implica una alteración en la forma en que medimos la significación estadística. Si un metanálisis se somete a pruebas de significación antes de que haya superado el tamaño requerido de la información, el umbral para la significación estadística se puede ajustar para tener en cuenta el alto riesgo del error aleatorio.^{1, 2, 4, 6, 11, 12, 23} Por otra parte, la propia prueba estadística puede ser penalizada en conjunto con la fortaleza de la evidencia disponible. El TSA ofrece la opción de utilizar ambos enfoques para controlar el error de tipo 1.

Controlar el riesgo de error tipo II antes de que un metanálisis supere su tamaño requerido implica el establecimiento de umbrales (reglas) para cuando la intervención experimental pueda ser considerada no superior (y/o no inferior) a la intervención de control.

Los métodos para ajustar los umbrales de significación (por ejemplo, controlando el error tipo I) basados en los métodos introducidos por Armitage y Pocock; se les conoce como 'análisis secuencial de grupo',^{18, 43, 44} En el análisis secuencial de los grupos de Armitage y Pocock, es necesario conocer el número aproximado de pacientes aleatorizados entre cada análisis intermedio de los datos.³⁰ En los ensayos clínicos aleatorizados, los análisis intermedios sobre la acumulación de datos suelen ser pre-planificados y por lo tanto es posible definir el tamaño de los grupos conocidos entre cada análisis intermedio.³⁰ En el metanálisis, el análisis intermedio de los datos ocurre cuando hay una actualización, agregando datos de nuevos ensayos clínicos. Las actualizaciones en el metanálisis se producen a un ritmo arbitrario, rara vez son regulares, y el número de pacientes que agregamos es variada e impredecible. Los métodos propuestos por Armitage y Pocock, por lo tanto, son inaplicables para el metanálisis.

Lan y DeMets extendieron la metodología propuesta por Armitage y Pocock, permitiendo análisis intermedios más flexibles y no planificados. Lan y DeMets pretendieron esta metodología para la evaluación de significación repetida en un solo ensayo aleatorizado.^{16,17,30} Debido a la flexibilidad del momento de análisis intermitente, esta metodología es aplicable a los metanálisis. Por lo tanto, la estrategia de Lan y DeMets es la metodología usada en el TSA, esto implica la construcción de límites de control que facilitan la definición de umbrales sensibles para "significación estadística" en el metanálisis.

Del mismo modo, los límites de futilidad pueden ser construidos, lo que facilita la definición de umbrales sensibles para 'inutilidad' en metanálisis.³⁰ Las secciones 2.2.1. a 2.2.5. aportan la descripción de la metodología subyacente y las consideraciones teóricas para estos métodos.

Los métodos de control de error tipo II son una extensión de la metodología de Lan-DeMets que permite evaluar la no superioridad y no inferioridad. Es decir, en lugar de construir umbrales ajustados para la significación estadística, el método construye umbrales ajustados para no superioridad y no inferioridad (o ninguna diferencia). Juntos, los límites de no superioridad y de no inferioridad ajustados constituyen lo que se conoce como los límites de futilidad o límites de la «cuña interna» (*inner wedge*). Las secciones 2.2.7 proporcionan una descripción de la metodología y consideraciones teóricas subyacentes de este método.

Tal como fue descrito, un enfoque alternativo a la alteración de los umbrales es penalizar a la prueba estadística en sí. El método para penalizar a las pruebas estadísticas empleadas es relativamente una nueva estrategia, que se basa en los teoremas de la teoría de la probabilidad avanzada. En particular, la técnica utiliza el teorema conocido como "la ley del logaritmo iterado".^{7; 8} En las secciones 2.2.2 y 2.2.6 se ofrece una descripción de la metodología subyacente y consideraciones teóricas para este método.

2.2.1 El tamaño de la información necesaria para un metanálisis concluyente

La determinación del tamaño de la información requerida (por ejemplo, el número necesario de pacientes) de un metanálisis concluyente y fiable es un requisito previo para la construcción de los umbrales ajustados para 'significación estadística' usando TSA.^{1, 2, 4, 6, 11, 12} Los niveles de los umbrales deben ser construidos de acuerdo con la fuerza de la evidencia.^{1, 2, 4, 6, 11, 12} La metodología estadística subyacente TSA se basa en el supuesto de que los datos se acumularán hasta que el tamaño de la información requerida sea superada.³⁰ Para mayor explicación sobre esta suposición, por favor refiérase a los documentos metodológicos anteriores sobre esta cuestión.^{16, 17, 30, 43, 44}

Consideraciones convencionales para determinar el tamaño de la información

Se ha argumentado que el tamaño de la muestra necesaria para que un metanálisis sea concluyente y confiable, debería ser, por lo menos, tan grande como el tamaño de la muestra necesaria para detectar el efecto verdadero de la intervención en un ensayo clínico con un tamaño de muestra suficiente.^{1,2,4,6,11,12} En concordancia con esta constructo, el tamaño mínimo necesario de información (número de pacientes) en un metanálisis puede ser derivado usando esta conocida fórmula:

$$IS_{Patients} = 2 \cdot (Z_{1-\alpha/2} + Z_{1-\beta})^2 \cdot 2 \cdot \sigma^2 / \delta^2$$

donde α es el riesgo máximo deseado de obtener un resultado falso positivo (error tipo I) y β es el riesgo máximo deseado de obtener un resultado falso negativo (error tipo II), y donde $Z_{1-\alpha/2}$ y $Z_{1-\beta}$ son los $(1 - \alpha/2)$ y $(1 - \beta)$ distribución estándar normal de los cuantiles.^{1,2,4,6,11,12} Observe que el uso de $\alpha/2$ en lugar de α significa que el tamaño de la información es construida asumiendo una prueba estadística bilateral. Para los datos binarios, $\delta = P_C - P_E$ denota una estimación *a priori* para un efecto de la intervención realista o mínimamente importante (P_C y P_E siendo la proporción con un resultado en el grupo control y el del grupo de intervención, respectivamente), donde

$\sigma^2 = P^* (1 - P^*)$, que es la varianza asociada, y suponiendo que $P^* = (PC + PE) / 2$ (es decir, que los grupos de intervención y de control son iguales en tamaño). Para los datos continuos, δ denota, una estimación *a priori* de la diferencia entre las medias de los dos grupos de intervención y σ^2 indica la varianza asociada.

Alternativas al número de pacientes acumulados

En el metanálisis de datos binarios, la información y la precisión en un metanálisis dependen principalmente del número de eventos de un desenlace. Por lo tanto, se puede argumentar que en el contexto de las consideraciones de tamaño de la información de un metanálisis, el número requerido de eventos es una medida más apropiada que el número necesario de pacientes. Bajo el supuesto de que un número igual de pacientes son aleatorizados a las dos intervenciones investigadas en todos los ensayos, el número requerido de eventos puede ser determinado de la siguiente manera:

$$IS_{Events} = P_C * IS/2 + P_E * IS/2$$

donde IS_{Events} es el número requerido de eventos para un metanálisis concluyente y fiable, y P_C y P_E son como se definen en el párrafo anterior.

La *información estadística* (información Fischer) es una medida estadística de la información contenida en un conjunto de datos (dado el modelo estadístico que sea asumido).⁴⁵ En el metanálisis estándar comparando dos intervenciones, la información estadística es simplemente el recíproco de la varianza combinada.⁴⁶ En un metanálisis, la información estadística es una medida teóricamente ventajosa, ya que combina tres factores en una sola medida: número de pacientes, número de eventos y el número de ensayos. Esta medida proporciona una propuesta simple para las consideraciones del tamaño de la muestra de un metanálisis. Los datos metanalíticos se consideran como algo análogo a la acumulación de datos en un único ensayo y la información estadística requerida viene dada por:

$$IS_{Statistical} = (Z_{1-\alpha/2} + Z_{1-\beta})^2 / \delta^2$$

Donde $IS_{Statistical}$ es la información estadística verdadera alcanzada en el metanálisis, α es el riesgo máximo deseado de error tipo I, $Z_{1-\alpha/2}$ es el percentil estándar normal $(1-\alpha/2)$, β es el riesgo máximo deseado de error tipo II, $Z_{1-\beta}$ es el percentil estándar normal $(1-\beta)$ y δ es efecto pre-especificado (mínimamente relevante) de alguna intervención.^{30, 45}

El factor de ajuste de la heterogeneidad

Los ensayos incluidos en un metanálisis, a menudo, incluyen pacientes de un amplio abanico de grupos de población, utilizan diferentes regímenes de una intervención y con diferentes diseños de estudio, y varían en la calidad metodológica (es decir, el riesgo de sesgo o «error sistemático»). Por todas estas razones, es natural esperar un mayor grado de variación en los datos del metanálisis en comparación con los datos de un solo ensayo.^{13, 47} Tal variación adicional se conoce como la heterogeneidad (o variación entre ensayos).^{13, 47} Debido a que el aumento de la variación puede disminuir la precisión de los resultados, las consideraciones de tamaño de la información deben incorporar todas las fuentes de variación en un metanálisis, incluyendo heterogeneidad.^{4,6,11,12} Uno de los enfoques para la incorporación de la heterogeneidad en consideraciones de tamaño de la información es multiplicar el tamaño de la información requerida en un metanálisis por algún factor de ajuste de heterogeneidad.^{6,23} Recientemente, un factor similar de ajuste de heterogeneidad ha sido propuesto para la estimación del tamaño de la muestra en un solo ensayo clínico.⁴⁸

El factor de ajuste de la heterogeneidad se conceptualiza a través de los supuestos subyacentes que hacemos para nuestro modelo de metanálisis. En el modelo de efecto fijo, se supone que todos los ensayos incluidos pueden ser vistos como repeticiones del mismo ensayo (con respecto al diseño y conducción). Por lo tanto, el tamaño de la información requerida para un metanálisis de modelo de efecto fijo sea concluyente y efectivo puede calcularse de la misma manera que el tamaño de muestra necesario para un ensayo clínico individual. En el modelo de efectos aleatorio, se supone que los ensayos incluidos provienen de una distribución de posibles ensayos (con respecto al diseño y conducción). Por definición, la varianza en un modelo de efectos aleatorios es siempre mayor que en un modelo de efecto fijo. Por lo tanto, un ajuste del factor de heterogeneidad debe tener en cuenta el aumento en la variación en que un metanálisis incurre desde pasar de la hipótesis de efecto fijo para el supuesto de efectos aleatorio. Un ajuste preciso se puede lograr haciendo que el factor de ajuste de heterogeneidad sea igual a la razón de la varianza total en el metanálisis de modelo de efectos aleatorio y de la varianza total en el metanálisis de modelo de efecto fijo.^{6, 23} Por consiguiente, el factor de ajuste de heterogeneidad es siempre igual o mayor que 1. Suponiendo IS_{Fixed} denote el tamaño de la información necesaria para un metanálisis de modelo de efecto fijo dado por la ecuación (1), v_R denota la varianza total en el metanálisis con modelo de efectos aleatorio y v_F denota la varianza total en el modelo de metanálisis de efecto fijo, el tamaño de la información ajustada por heterogeneidad se puede calcular usando la siguiente fórmula:

$$IS_{Random} = \frac{V_R}{V_F} IS_{Fixed}$$

Dado que los efectos anticipados de la intervención en los modelos de efecto fijo - (δ_F) y de modelos de efectos aleatorios (δ_R) - son aproximadamente iguales (es decir, dado $\delta_R = \delta_F$), se puede demostrar matemáticamente que en el caso especial donde todos los ensayos en un metanálisis tengan el mismo peso, el factor de ajuste de heterogeneidad (AF) se expresa de la siguiente manera:

$$AF = \frac{V_R}{V_F} = \frac{1}{1 - I^2}$$

Donde I^2 es el factor de inconsistencia comúnmente utilizado para medir la heterogeneidad en un metanálisis.⁴⁷

Es importante recordar que, en cualquier caso donde los pesos de los ensayos no sean iguales, utilizando I^2 dará lugar a una subestimación del factor de ajuste, y por lo tanto, una subestimación del tamaño de la información requerida.²³ En esta situación, se puede definir una medida de la diversidad (D^2) como la cantidad necesaria para satisfacer la ecuación:

$$AF = \frac{V_R}{V_F} = \frac{\sum_{i=1}^k w_i}{\sum_{i=1}^k w_i^*} = \frac{1}{1 - D^2}$$

donde w_i denota los pesos de los ensayos en el modelo de efecto fijo y w_i^* indica los pesos de los ensayos en el modelo de efectos aleatorio. La solución de la ecuación con respecto a D^2 , se obtiene:

$$D^2 = \frac{V_R - V_F}{V_R} = 1 - \frac{V_F}{V_R} = 1 - \frac{\sum_{i=1}^k w_i^*}{\sum_{i=1}^k w_i} = \frac{\sum_{i=1}^k (w_i^{-1} + \tau^2)^{-1}}{\sum_{i=1}^k w_i}$$

Donde τ^2 indica la variación entre ensayos. Una propiedad ventajosa de la medida de diversidad, D^2 , es que las derivaciones anteriores son generalizables a cualquier modelo de metanálisis. Por lo tanto, si queremos realizar un metanálisis de algunos ensayos mediante un modelo de efectos aleatorios alternativo con varianza total V_R , la medida de diversidad y el correspondiente factor de ajuste simplemente se toma la expresión:

$$D^2 = \frac{V_R - V_F}{V_R} \quad \text{and} \quad AF = \frac{V_R}{V_F}$$

Las estimaciones de la variabilidad, y, en particular, la variabilidad entre ensayos, pueden estar sujetos al error aleatorio y al sesgo.^{41, 47, 49, 50} Por esta razón, y en algunas situaciones, el uso de D^2 o I^2 basado en los datos disponibles puede ser inadecuada. En el metanálisis que incluye solamente un número limitado de ensayos (por ejemplo, menos de 10 ensayos), las estimaciones de la heterogeneidad y de la varianza entre ensayos pueden ser tan poco fiables como las estimaciones del efecto de intervención de ensayos clínicos aleatorizados pequeños (por ejemplo, ensayos con menos de 100 pacientes). Cuando un metanálisis está sujeto a sesgo de tendencia temporal (es decir, cuando los ensayos, la mayoría con resultados positivos, han sido publicados), por lo general se subestima la varianza entre los ensayos. Esta subestimación ocurre porque el conjunto de ensayos incluidos «tempranamente» es probable que reproduzcan una estimación del efecto similar a la de la intervención («positiva»)⁵⁰. Los metanálisis posteriores- actualizaciones-, es probable que incluyan más ensayos con resultados neutros o incluso negativos, en cuyo caso las estimaciones de la heterogeneidad serán mayores.

Para los metanálisis con un esperado número pequeño de ensayos, se sugiere hacer una estimación «*a priori*» sobre el grado previsto de heterogeneidad. Si se permite que H muestre una estimación conceptual de D^2 , podemos utilizar la siguiente fórmula en un cálculo *a priori*:

$$AF = \frac{1}{1-H}$$

Por ejemplo, si se espera que un metanálisis incluya un grado leve de heterogeneidad, sobre la base de lo que sabemos sobre el tema clínico, las diferencias observadas entre los ensayos incluidos, las diferencias esperadas entre presente y futuro, y el alcance de la revisión- uno puede optar por definir H como el 25 %. En este caso, la AF se estima en 1,33. Si se espera un moderado grado de heterogeneidad, se puede optar por definir H como el 50 %, y AF sería entonces estimada en 2,00. Si se espera gran heterogeneidad, entonces H puede llegar a ser del 75 % y el AF se estima en 4,00.

Debido a que el grado de la heterogeneidad esperado puede ser difícil estimar cuando un metanálisis solo incluye pocos ensayos, se recomienda que los usuarios del TSA realicen análisis de sensibilidad para esta variable. Por ejemplo, se podría concebir grados máximos y mínimo realistas o aceptables de heterogeneidad para un metanálisis dado. Como un ejemplo, se podría especular que el grado mínimo

aceptable de heterogeneidad estadística sería 20 %. También se podría decidir que si la heterogeneidad estadística supera el 60 %, entonces la medición del efecto en los subgrupos, en lugar de la estimación del efecto del tratamiento estimado mediante la combinación de los ensayos, sería más apropiada. En este caso, no debería ser realizado el metanálisis. En este ejemplo, se podría usar el promedio de los dos, $(60\% + 20\%)/2 = 40\%$, para el cálculo del tamaño de la información primaria, pero reconociendo que el tamaño de la información requerida puede ser tan grande como el basado en el 60 % de ajuste de la heterogeneidad o tan baja como el basado en el 20 %. Como otro ejemplo, se podría concebir la construcción de un número de «mejores» y «peores» escenarios de casos (cualesquiera que sean) mediante la adición de futuros ensayos «imaginario» para el metanálisis actual.

Este enfoque permitiría evaluar la solidez y la fiabilidad de la estimación D^2 y construir un espectro de grados realistas o aceptables de heterogeneidad que fácilmente podrían ser utilizados para el análisis de sensibilidad.

Estimación de la proporción de eventos del grupo control y un supuesto efecto de la intervención

La estimación de la proporción de eventos del grupo control y un probable efecto de la intervención son determinantes importantes para calcular el tamaño de la información necesaria al usar TSA. Por lo tanto, debe hacerse todo lo posible para que estas estimaciones sean lo más precisa y verdadera posible.

Para los datos binarios, la proporción de eventos del grupo control se puede estimar mediante el uso de la experiencia clínica y la evidencia de áreas relacionadas. Una estimación *a priori* de un efecto objetivo de la intervención se expresa generalmente como una reducción relativa del riesgo (RRR). Cuando la disponibilidad de la evidencia, acerca de la intervención bajo investigación, es limitada se puede estimar un efecto clínicamente relevante de la intervención mediante el uso de la experiencia clínica y las pruebas de las áreas relacionadas. Un ejemplo se puede encontrar en un artículo de Pogue y Yusuf, en el que la proporción de eventos del grupo control, P_c , y un RRR *a priori* se basaron en las experiencias de las áreas relacionadas en cardiología.^{1,2} Pogue y Yusuf aplicaron consideraciones del tamaño de la información a dos metanálisis conocidos bien destacados en cardiología: «La estreptoquinasa intravenosa en infarto agudo de miocardio» y «magnesio intravenoso en infarto agudo de miocardio». Ellos postularon que para la mayoría de los principales resultados vasculares, como la muerte, pudiera ser sensato esperar una mortalidad del 10% en el grupo control. Pogue y Yusuf, además, consideraron un ejemplo de una intervención

teórica para la prevención de la mortalidad después de un infarto al miocardio. Señalaron que los verdaderos tratamientos eficaces para reducir el riesgo de eventos cardiovasculares, tales como la muerte, habían generado unos RRRs de 10 %, 15 %, o -en el mejor de los casos- 20 %.

Para cualquier pregunta clínica, la decisión tiene que ser realizada sobre qué valores son adecuados para el P_c y RRR. La proporción anticipada de eventos en el grupo de intervención (experimental), P_E , a continuación, se pueden obtener usando la fórmula $P_E = P_c (1 - RRR)$. Posteriormente, el P_E hipotética y P_c pueden ser introducidos en la fórmula para el tamaño de la información requerida.

La inferencia realizada sobre los efectos de intervención verdadera prevista en un área de intervención a otra puede ser problemática porque una estimación *a priori* puede a menudo representar aproximaciones pobres de la «verdad». La literatura de ensayos clínicos está llena de ejemplos de cálculos del tamaño de la muestra realizados sobre la base de los efectos sobreestimados de las intervenciones previstas. No hay ninguna razón de por qué esto debería ser diferente para los cálculos del tamaño de información para un metanálisis.

Si los ensayos aleatorizados ya han investigado el efecto de una intervención, entonces, una recolección de tales estimados se puede usar para cuantificar mejor el efecto de la intervención prevista. Sin embargo, no todos los ensayos aportan estimaciones válidas, y se debe tener cuidado en asegurarse de la validez de los estimadores de los efectos de la intervención utilizados para el cálculo del efecto anticipado de una intervención dada.

Muchos ensayos producen sobreestimación de los efectos de la intervención investigada debido al informe selectivo de resultados y riesgos de sesgo (es decir, los errores sistemáticos debido a la inadecuada generación de la secuencia de asignación, ocultamiento inadecuado de la asignación, el enmascaramiento inadecuado, pérdida durante el seguimiento, u otros mecanismos).^{13,51-58} Tales ensayos se pueden clasificar en ensayos con *alto riesgo de sesgo*.¹³ Al contrario, los ensayos que generen estimaciones válidas del efecto de la intervención pueden ser clasificados como ensayos con *bajo riesgo de sesgo*.¹³ Si la evidencia sobre el efecto de la intervención investigada está disponible a partir de ensayos con bajo riesgo de sesgo, sería apropiado establecer un efecto esperado *a priori* de la intervención utilizando un metanálisis de estos ensayos.^{6,11,12} Sin embargo, las situaciones metanalíticas que requieren cálculos del tamaño de la información a menudo se producen cuando la evidencia es escasa. Aunque una serie de ensayos con bajo riesgo de sesgo estén disponibles para la aproximación de un efecto de verdadera la intervención prevista, la estimación combinada de estos ensayos todavía puede estar

sujeto a considerable error aleatorio, el sesgo de desfase del tiempo, y el sesgo de publicación. Por lo tanto, un efecto esperado *a priori* de la intervención sobre la base de la estimación del efecto combinado de metanálisis de los ensayos con bajo riesgo de sesgo solo es fiable en la medida en que este metanálisis puede considerarse libre de grandes errores aleatorios. Además, solo es válido en la medida en que puede considerarse libre de sesgo de desfase del tiempo y de publicación.

No es posible recomendar una técnica para definir los efectos de intervención para los cálculos del tamaño de la información. Más bien, las consideraciones del tamaño de la información deben basarse en las fluctuaciones de proporciones creíbles de eventos del grupo control, efectos de la intervención y errores de tipo I y II. Las consideraciones adecuadas para el cálculo del tamaño de la muestra para un ensayo clínico no solo se basan en un solo número. En su lugar un rango de tamaños de muestras aceptable se produce a partir de una gama de efectos posibles de tratamiento, la proporción de eventos del grupo de control, y los errores tipo I y tipo II, lo que proporciona un razonable intervalo de tamaños de muestra en el cual se encuentre un número apropiado de pacientes para obtener un ensayo clínico concluyente. Del rango producido de tamaños de muestra, se seleccionaría una primaria y dejar que el tamaño de la muestra actúe como cálculo de sensibilidad (potencia). Se recomienda que las consideraciones de tamaño de la información para metanálisis sigan la misma construcción. Las RRR y P_c obtenidos de estudios con bajo riesgo de sesgo fácilmente podrían combinarse *a priori* con un rango de efectos de intervención utilizando la metodología del mejor y peor escenario «realistas», lo que proporciona una gama de información en la cual sustentarse el metanálisis para generar inferencias metanalíticas concluyentes.

Limitaciones

El tamaño de la información requerida para un metanálisis (sea determinada como el número requerido de pacientes, eventos o información estadística) tiene limitaciones. En los ensayos clínicos aleatorizados, es razonable asumir que la distribución de los factores de pronóstico en los pacientes aleatorizados se asemeja a la de la población de referencia. En las revisiones sistemáticas con metanálisis, los ensayos se incluyen normalmente sobre la base de unos criterios de inclusión que se deciden *a priori* en el protocolo de la revisión sistemática. Debido a que los criterios de inclusión (y exclusión) en los ensayos clínicos casi nunca son idénticos y debido a que los ensayos suelen variar en tamaño de las muestras, es poco probable que los metanalistas y autores de las revisiones sistemáticas tengan control sobre la distribución de los factores pronósticos. Si en una actualización de la revisión sistemática se modifican

los criterios de inclusión, los autores no podrán predecir con precisión la distribución de los factores pronósticos entre los ensayos recientemente publicados. Los factores pronósticos basales pueden tener un considerable impacto sobre las tasas de incidencia en el grupo de control. En esta situación, puede ser apropiado hacer un intento *previo* de cuantificar la diferencia entre la incidencia basal en la población del metanálisis y, en la población de referencia, y luego realizar una serie de análisis de sensibilidad *post hoc* si es necesario.

Los efectos comparativos de las intervenciones mínimamente importantes (también conocidos como diferencias mínimamente importantes) no siempre pueden ser similares en todos los ensayos incluidos. Por ejemplo, si la población de pacientes investigados entre los ensayos experimenta diferentes riesgos de eventos adversos, la diferencia mínimamente importante también puede diferir entre los ensayos. Esta variación es el resultado del propósito clínico. Para cualquier intervención médica, la posibilidad de beneficio debe ser superior a cualquier aumento en el riesgo de daño. Una población con mayor riesgo de daño necesitará una mayor posibilidad de beneficio para hacer un tratamiento que valga la pena. Cuando las diferencias mínimamente importantes varían entre los ensayos, las consideraciones de tamaño de la información todavía pueden ser sensibles. Sin embargo, es importante recordar que la inferencia acerca de las conclusiones de un metanálisis solo se puede generalizar a la población de pacientes para los que se aplican *a priori* la diferencia mínimamente importante.

Cuando el tamaño de la información requerida es definida por el número requerido de pacientes o eventos, el problema de la heterogeneidad impredecible puede ser tratado mediante la previsión de algún grado máximo apropiado de la heterogeneidad y en consecuencia ajustar el tamaño de la información requerida.⁴ La aparente limitación de este enfoque es que el grado de heterogeneidad esperada es a la vez difícil de adivinar y estimar cuando solo unos pocos ensayos clínicos están disponibles. Aunque se recomienda el análisis de sensibilidad sobre el grado de ajuste de la heterogeneidad, estos análisis pueden todavía ser inapropiados si el grado esperado(s) de la heterogeneidad no refleja el verdadero grado de heterogeneidad en el cual los metanálisis incurrirán a medida que se publiquen nuevos ensayos clínicos.

Cuando el tamaño de la información requerida se define por la información estadística necesaria, la fórmula para el tamaño de la información requerida no requiere una estimación del grado previsto de heterogeneidad. Por el contrario, la información real en el metanálisis (la información estadística estimada) incorpora directamente la heterogeneidad a través de la variación estimada entre los ensayos.

Esto, sin embargo, presenta una limitación en que la información estadística acumulada solo es fiable en la medida en que la estimación de la varianza entre ensayos sea fiable. Las posibles soluciones a este problema implican el uso de una metodología más compleja para ajustar la incertidumbre asociada a la estimación de la variación entre ensayos. Una opción es utilizar el enfoque de efectos aleatorios por Biggerstaff-Tweedie, que incorpora la incertidumbre asociada a la estimación de la varianza entre los ensayos cuando se utiliza el estimador convencional de DerSimonian-Laird (véase la sección 2.1.3).⁴¹ Otra opción es utilizar metanálisis Bayesiano, donde una distribución *a priori* es delineada para la varianza entre ensayos.

2.2.2 La prueba estadística acumulada (curva Z)

Como se mencionó en la sección 2.1.2., la prueba del metanálisis para «significación estadística» utiliza una prueba estadística tipo Wald. Este estadístico está dado por el logaritmo del efecto combinado de la intervención dividido por su error estándar¹³ y que comúnmente se conoce como el estadístico Z o el valor Z. Bajo el supuesto de que las dos intervenciones investigadas no difieren (la hipótesis nula) el valor Z será aproximadamente una distribución estándar normal (una distribución normal con media 0 y desviación estándar 1). Cuanto mayor sea el valor absoluto de un valor Z observado, más fuerte es la evidencia estadística de que las dos intervenciones investigadas sean diferentes. Si el valor Z absoluto observado es sustancialmente mayor que 0, es habitual concluir que la diferencia observada entre el efecto de las dos intervenciones no solo se explica por la influencia del azar. En esta situación, la diferencia entre las dos intervenciones se describe como «estadísticamente significativa». Por definición, un valor P es la probabilidad de encontrar diferencias observadas, o uno más extremo, si la hipótesis nula fuera verdadera. En la práctica, el valor P es el valor que se utiliza para evaluar la significación estadística. El valor P se obtiene a partir del valor de Z (véase la sección 2.1.2 para los detalles matemáticos); estas dos medidas representan dos maneras diferentes de comunicar la misma información, y son intercambiables. Por ejemplo, un valor de p bilateral menor que el 5 % es lo mismo que un valor absoluto de Z mayor que 1,96 y viceversa.

Cada vez que un metanálisis es actualizado, se calcula un nuevo valor Z. Por tanto, una serie de valores consecutivos Z, procede de una serie de actualizaciones del metanálisis. Para inspeccionar la evolución de las pruebas de significación, la serie de valores Z puede ser trazada con respecto a la información acumulada (pacientes acumulados, eventos o información estadística), produciendo de este modo una curva que se conoce comúnmente como la curva Z.^{1, 2, 4, 6, 11, 12}

2.2.3 Los problemas con las pruebas de significación en el metanálisis

Como se mencionó en el capítulo 1, las pruebas de significación convencional en el metanálisis fallan en relacionar las pruebas estadísticas observadas y los valores de p con la solidez de la evidencia disponible y el número de pruebas de significación repetidas.^{1-4, 6, 11, 12} La consecuencia de esta omisión es un aumento en el riesgo de obtener un resultado metanalítico falso positivo. Esta sección proporciona descripciones conceptuales y estadísticas, con un nivel básico a intermedio, de las pruebas de significación en el metanálisis y los problemas que resultan de no incorporar la solidez de la evidencia y el número de pruebas de significación repetidas en el proceso.

Crterios generales para las pruebas de significación

Las pruebas de significación convencional trabajan con un riesgo máximo de error tipo I, α , que también funciona como el umbral para cuando los valores de p se consideran pruebas de significación estadística. Los valores p y valores de Z son intercambiables en la evaluación de la significancia estadística. Como se mencionó, para cada umbral de valor p , α , existe un valor umbral Z correspondiente, $Z\alpha$. Por ejemplo, si deseamos un riesgo de error tipo I bilateral de un máximo del 5 % solo deberíamos considerar valores absolutos Z superiores a 1,96 como prueba de significación estadística. Pero si deseamos un error máximo de tipo I bilateral del 1 % solo debemos considerar valores absolutos de Z superiores a 2,58 como evidencia de significación estadística.

Asuma que $\Pr(X|Y)$ indica la probabilidad de que el evento X se produce debido a que el evento Y es verdad (o se ha producido), y que $|Z|$ representa el valor absoluto de Z . En general, nos enfrentamos al reto de determinar adecuadamente el umbral, c , que hará que las siguientes ecuaciones sean verdaderas

$$\Pr(|Z| \geq c \mid H_0 \text{ is true}) \leq \alpha \quad (2)$$

$$\Pr(|Z| = c \mid H_0 \text{ is true}) = \alpha \quad (3)$$

Para las secciones teóricas restantes sobre pruebas de significación repetidas (secciones 2.2.2 a 2.2.5), asumiremos que todas las pruebas estadísticas son bilaterales. También asumiremos que todos los valores estadísticos de pruebas, Z , son valores absolutos. Asumimos esto último debido a que el álgebra involucrada hace más sencillo realizar ese proceso. Por ejemplo, en la definición de los umbrales

bilaterales de una prueba estadística no absoluta, habría que considerar la probabilidad de que $Pr(Z \leq -c \text{ or } Z \geq c \mid \dots)$ en vez de $Pr(|Z| \geq c \mid \dots)$.

Los problemas con las pruebas de significación repetida

Las pruebas sencillas convencionales de significación pueden considerarse fiables si «suficientes» datos se han acumulado. En el metanálisis, una sola prueba de significación puede ser considerada confiable una vez que el tamaño de la información requerida es superada.^{1-4,6,11,12,20,59} Si realizamos una sola prueba para la significación estadística en o después de un metanálisis que ha superado su tamaño de información requerida, las pruebas de significación estadística solo implican la determinación de un umbral adecuado, c , que hará que las ecuaciones (2) y (3) sean ciertas. Por ejemplo, para $\alpha = 5\%$ consideraríamos $c = 1,96$ apropiado si los datos del metanálisis anteriormente no habían sido sometidos a pruebas de significación.

Cuando un metanálisis acumulado es sometido a pruebas de significación más de una vez (antes superando su tamaño de información requerido), la situación se vuelve más compleja. Considerando el ejemplo donde un metanálisis se actualiza una primera vez y donde se utiliza el error máximo, convencional, tipo I de 5%. En esta situación, el primer metanálisis producirá un valor de Z , Z_1 , y la actualización del metanálisis producirá otro, Z_2 . Si el primer metanálisis genera un valor Z mayor que 1,96, las dos intervenciones investigadas se declaran significativamente diferentes. Sin embargo, si el primera metanálisis no es significativo (es decir, $Z_1 < 1,96$), las dos intervenciones pueden todavía ser declaradas estadísticamente significativa si la actualización del metanálisis genera un valor Z mayor que 1,96 (es decir, si $Z_2 \geq 1,96$). Por las leyes de la teoría básica de la probabilidad, la probabilidad de que las dos intervenciones serán declaradas estadísticamente significativas bajo la hipótesis nula es:

$$\begin{aligned} Pr(H_0 \text{ rejected}) &= Pr(|Z_1| \geq 1.96 \text{ or } |Z_2| \geq 1.96) \\ &= Pr(|Z_1| \geq 1.96) \cdot Pr(|Z_2| \geq 1.96 \mid |Z_1| < 1.96) \end{aligned}$$

Se puede mostrar que esta expresión es siempre mayor que la deseada 5% (véase el apéndice A.3.1). En general, las pruebas de significación repetidas utilizando los umbrales de ensayo individuales siempre dará lugar a una exageración del error de tipo I, y cuanto mayor sea el número de (repeticiones) pruebas de significación empleadas en los datos que se acumulan, lo peor es la exageración del error tipo I.³⁰ Para los datos del metanálisis, los estudios de simulación han demostrado que las pruebas de significación repetidas generan un error de tipo I, de 10% a 30% cuando

se utiliza un umbral α convencional = 5 %, 1,96, para probar la significación estadística en cada actualización.^{7, 8,10,31}

2.2.4 La función del gasto de alfa y los límites de monitoreo secuencial de los ensayos

Una solución al problema descrito en el punto 2.2.3. es ajustar los umbrales de los valores de Z , permitiendo que el riesgo de error tipo I sea restaurado al riesgo máximo deseado^{1,2,6,17} En el ejemplo de esas dos pruebas, por lo tanto, necesitamos encontrar dos umbrales, c_1 y C_2 , para lo cual

$$\Pr(|Z_1| \geq c_1 \text{ or } |Z_2| \geq c_2) \leq \alpha$$

se cumple bajo la hipótesis nula. Esto es equivalente a encontrar dos riesgos máximos de error de tipo I, α_1 y α_2 , que se suman a α y donde

$$\Pr(|Z_1| \geq c_1) \leq \alpha_1$$

$$\Pr(|Z_2| \geq c_2 \mid |Z_1| < c_1) \leq \alpha_2$$

bajo la hipótesis nula. En la situación general en la que se emplea pruebas de significación repetida k veces (i.e., donde un metanálisis inicial es $k-1$ veces *actualizado*), necesitaríamos encontrar umbrales c_1, \dots, c_k para asegurar cada una de las pruebas de significación k

$$\Pr(|Z_1| \geq c_1 \text{ or } |Z_2| \geq c_2 \text{ or } \dots \text{ or } |Z_k| \geq c_k) \leq \alpha$$

bajo la hipótesis nula. Esto es equivalente a encontrar k riesgos máximos de error de tipo I que, α, \dots, α_k , que suma a α y donde

$$\Pr(|Z_1| \geq c_1) \leq \alpha_1$$

$$\Pr(|Z_2| \geq c_2 \mid |Z_1| < c_1) \leq \alpha_2$$

$$\Pr(|Z_3| \geq c_3 \mid |Z_1| < c_1 \text{ and } |Z_2| < c_2) \leq \alpha_3$$

⋮

$$\Pr(|Z_k| \geq c_k \mid |Z_1| < c_1 \text{ and } \dots \text{ and } |Z_{k-1}| < c_{k-1}) \leq \alpha_k$$

bajo la hipótesis nula.

El registro de los umbrales para la curva Z se conoce como límites de control, o límites de monitorización secuencial de grupo (una serie de límites aplicados al desarrollo de las pruebas en grupos acumulativos de los pacientes aleatorizados en un ensayo clínico).^{17,30,44} En el metanálisis, tales límites se aplican a una serie de ensayos, y por lo tanto nos referimos a ellas como límite de monitorización secuencial de los ensayos.⁶ La combinación de metanálisis y los límites de monitoreo secuencial ensayos se conoce como análisis secuencial de los ensayos.⁶

Los límites de monitoreo secuencial de los ensayos requieren pre-especificación de los riesgos de máximo k error tipo I, $\alpha_1, \dots, \alpha_k$, así como también la integración numérica intensiva para su aplicación.⁶⁰ Un método simple para la asignación de valores para la $\alpha_1, \dots, \alpha_k$ riesgos de error tipo I es el método de gasto de alfa (o la función gasto de alfa).^{1,2,17,30} Este método está incorporado en el programa TSA. La función de gasto de alfa es una función monótonamente creciente de tiempo que se puede utilizar para asignar apropiadamente los riesgos de error máximo tipo I $\alpha_1, \dots, \alpha_k$ en cada prueba de significación de acuerdo con la cantidad de información acumulada.^{16,17} La variable independiente se define por la fracción de información (IF); la cual se calcula dividiendo la información acumulada por el tamaño de la información necesaria (por ejemplo, el número acumulado de pacientes dividido por el número necesario de pacientes).^{6,15,17} La variable dependiente (la función) es el error acumulativo tipo 1, lo que da la cantidad de error que debería ser considerado el máximo al definir la significación dada en el IF. Dado que la IF se incrementa - es decir, como aumenta la cantidad de información acumulada - el tamaño del error tipo 1 "aceptable" también aumenta. La función proporciona una manera de cuantificar el riesgo de error aleatorio permitido en cualquier IF determinado, con el fin de garantizar que el riesgo global de error aleatorio - después que se haya alcanzado el IS - se mantenga por debajo del 5 %. La función monótonamente creciente corresponde a un umbral monótonamente decreciente para la significación estadística medida por la estadística de prueba de Z.

La función gasto de alfa se define a partir de 0 a 1 (0 es el punto donde 0 pacientes han sido asignados al azar, y 1 es el punto donde la información acumulada es igual al tamaño de la información requerida).^{16,17} La función de gasto de alfa 0 es siempre igual a 0: $\alpha(0) = 0$; en ese momento, ninguna información se ha acumulado. La función gasto de alfa de 1 es siempre igual a α : $\alpha(1) = \alpha$, en este punto, toda la información requerida ha sido acumulada y la cantidad total de error alfa es todo lo que se define como el total global aceptable del error tipo 1 (generalmente 5 %). En cualquier punto entre 0 y 1 (por la fracción de la información en el momento de una prueba de significación i (IF $_i$)) la función gasto de alfa es igual al riesgo total máximo de error tipo I que ha surgido de los umbrales elegidos para todas las pruebas de significancia hasta e incluyendo la prueba de significación de orden i -th. En otras palabras, la

función de gasto de alfa es igual a la cantidad de error tipo 1 que ha sido 'gastado'. En la notación: $\alpha(IF) = \alpha_1 + \alpha_2 + \dots + \alpha_i$, y por lo tanto

$$\begin{aligned} \Pr(|Z_1| \geq c_1) &\leq \alpha_1 = \alpha(IF_1) \\ \Pr(|Z_2| \geq c_2 \mid |Z_1| < c_1) &\leq \alpha_2 = \alpha(IF_2) - \alpha(IF_1) \\ \Pr(|Z_3| \geq c_3 \mid |Z_1| < c_1 \text{ and } |Z_2| < c_2) &\leq \alpha_3 = \alpha(IF_3) - \alpha(IF_2) \\ &\vdots \\ \Pr(|Z_k| \geq c_k \mid |Z_1| < c_1 \text{ and } \dots \text{ and } |Z_{k-1}| < c_{k-1}) &\leq \alpha_k = \alpha(IF_k) - \alpha(IF_{k-1}) \end{aligned}$$

La verdadera función gasto de alfa utilizada puede ser cualquier función monótonamente creciente.^{16,17} Un ejemplo bien conocido es $\alpha(t) = t \cdot \alpha$.^{16,17, 30} Cuando todas las pruebas de significación se realizan a la misma distancia (con respecto a la escala de fracción de la información), esta función gasto de alfa producirá iguales umbrales para los valores Z (es decir, $c_1 = c_2 = \dots = c_k$). Este ajuste fue propuesto por primera vez por Pocock. Un enfoque más general de gasto de alfa es la función gasto de alfa de *Power Family* definida como $\alpha(t) = t^p \cdot \alpha$.^{16,17,30} Las funciones gasto de alfa *Power Family*, donde $p > 1$ y todas las pruebas de significación se realizan a la misma distancia, producirán más umbrales conservadores para las pruebas de significación realizadas al inicio que para las pruebas de significación posteriores. En general, los umbrales para (los valores absolutos) la curva Z serán monótonamente decrecientes cuando la función de gasto de alfa es convexa y todas las pruebas de significación se llevan a cabo a igual distancia.^{16,17,30} Los umbrales decrecientes monótonamente (que resultan de las funciones monótonamente crecientes) son de forma deseable ya que el impacto del error aleatorio es típicamente inversamente proporcional a la cantidad de información acumulada. Aunque existe una infinita combinación de umbrales decrecientes, algunos grupos de umbrales pueden ser preferibles.

A partir de la teoría de probabilidad avanzada, la función gasto de alfa que genera umbrales teóricamente óptimos viene dada por la expresión

$$\alpha(IF) = 2 - 2\Phi\left(Z_{\alpha/2} / \sqrt{IF}\right)$$

donde Φ es la función de distribución acumulada normal estándar.^{16,17,30} El tipo de límites producidos por esta función gasto de alfa se propuso por primera vez para incrementos iguales de IF por O'Brien y Fleming.⁶¹ Posteriormente, Lan y DeMets propusieron la función gasto de alfa mencionada para permitir incrementos flexibles en IF.^{16, 17,30} Por esta razón, la función gasto de alfa mencionada se refiere típicamente como la implementación de LAN-DeMets de la función gasto de alfa de O'Brien-

Fleming. A menudo, los límites de vigilancia o de supervisión¹ producidos por esta función gasto de alfa se refieren simplemente como los límites de supervisión Lan-DeMets o los límites de supervisión O'Brien-Fleming. Para el resto de este manual, nos referiremos a ellos como los límites de supervisión O'Brien-Fleming. Actualmente, la función de gasto de alfa O'Brien-Fleming es la única función gasto de alfa implantada en el software TSA.

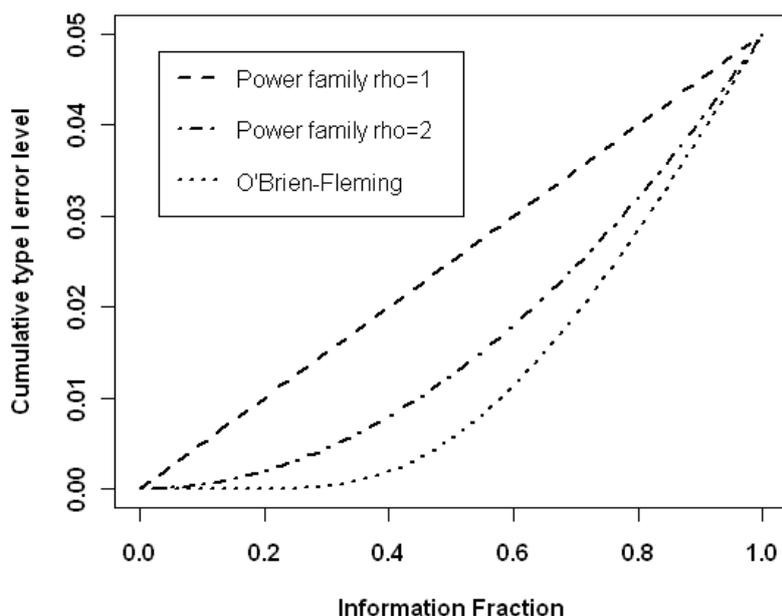


Figura 6. La forma de las funciones gasto de alfa *power family* $\rho=1$ and $\rho=2$ y la función gasto de alfa O'Brien-Fleming.

Como se muestra en la figura 6, la función gasto de alfa O'Brien-Fleming es una función exponencialmente creciente. Produce límites conservadores en las primeras etapas donde solo una limitada cantidad de datos ha sido acumulada, y a medida que más datos son acumulados los límites serán más flexibles.

Los límites O'Brien-Fleming han sido recomendados por expertos metodológicos como la opción preferida en la mayoría de los ensayos clínicos aleatorizados donde se realizan las pruebas de significación repetida sobre datos que se van acumulando.^{30,62} En el metanálisis, donde el riesgo de error aleatorio (y los sesgos de tendencia temporal) es motivo de especial preocupación en las primeras etapas (es

¹ NT: Preferimos utilizar 'vigilancia o supervisión' para evitar los anglicismos "monitorización" o "monitoreo", tal como sugiere el Diccionario crítico de dudas inglés-español de medicina, de Navarro.

decir, en el metanálisis que incluye un pequeño número de pacientes y de eventos), los límites de O'Brien-Fleming asimismo han sido la opción preferida.^{1,2,4,6,11,12}

Hay dos razones para esta preferencia. En primer lugar, si el ajuste de la heterogeneidad del tamaño de la información requerida se basa en una razonable estimación *a priori* del grado previsto de la heterogeneidad, los límites de O'Brien-Fleming, naturalmente, tendrán en cuenta el grado de fluctuaciones en las que las inferencias metanalíticas incurrirán debido al error aleatorio y la heterogeneidad. En segundo lugar, siempre y cuando las pruebas de significación posteriores se realicen a una distancia razonable sobre el eje de información (por ejemplo, al menos 1 % del tamaño de la información requerida), los límites O'Brien-Fleming siguen siendo relativamente poco afectados por el número previo de pruebas de significación. Esta segunda característica es deseable en el contexto de un metanálisis porque no siempre está claro con qué frecuencia un metanálisis ha sido sometido a pruebas de significación, como resultado de la actualización. Por ejemplo, algunos metanálisis pueden incluir datos diferentes pero con solapamiento elevado porque los criterios de inclusión han sido modificados en relación con cambios en la actualización de una revisión sistemática. Otros límites de supervisión, tales como un conjunto de los límites de control basados en la función gasto de alfa *power family* con $\rho=2$, podría producir inferencias discrepantes acerca de la significación estadística si, por ejemplo, los límites de supervisión contaron 2 actualizaciones previas en vez de 4.

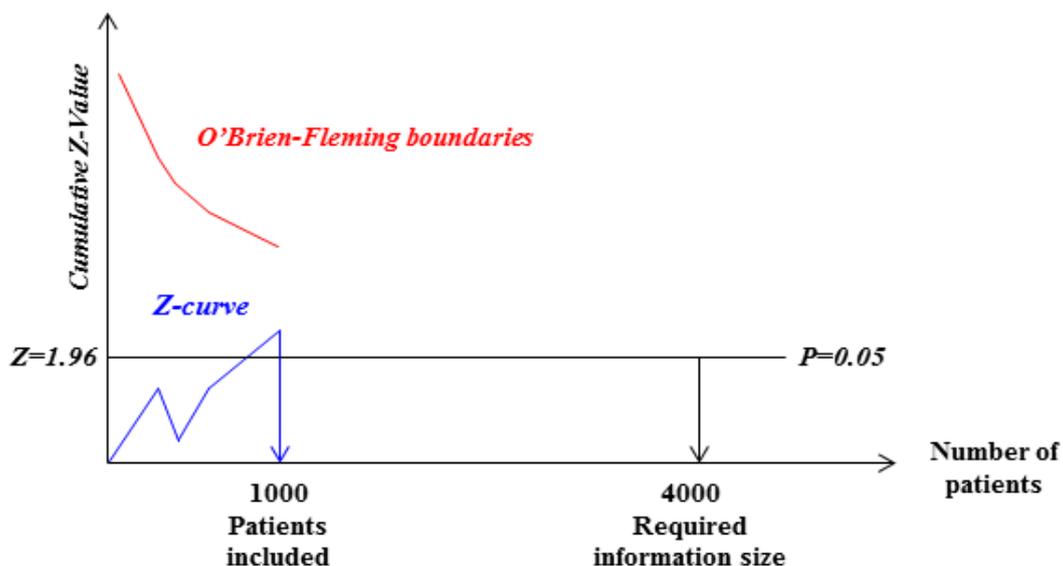


Figura 7. Ejemplo de un metanálisis no concluyente después de cuatro metanálisis acumulativos.

La figura 7 muestra un ejemplo de la utilización de los límites O'Brien- Fleming. En este metanálisis, el tamaño de la información requerida es de 4000 pacientes, pero la información obtenida es de solo 1000 pacientes. El valor Z final es mayor que 1,96. Utilizando el umbral de la prueba única convencional, este valor Z habría conducido a una conclusión de la significación estadística. El uso de los límites O'Brien-Fleming, requiere un valor mayor de Z en este tamaño de información para concluir la significación estadística. Los límites no se cruzan y, por tanto, el metanálisis no es concluyente.

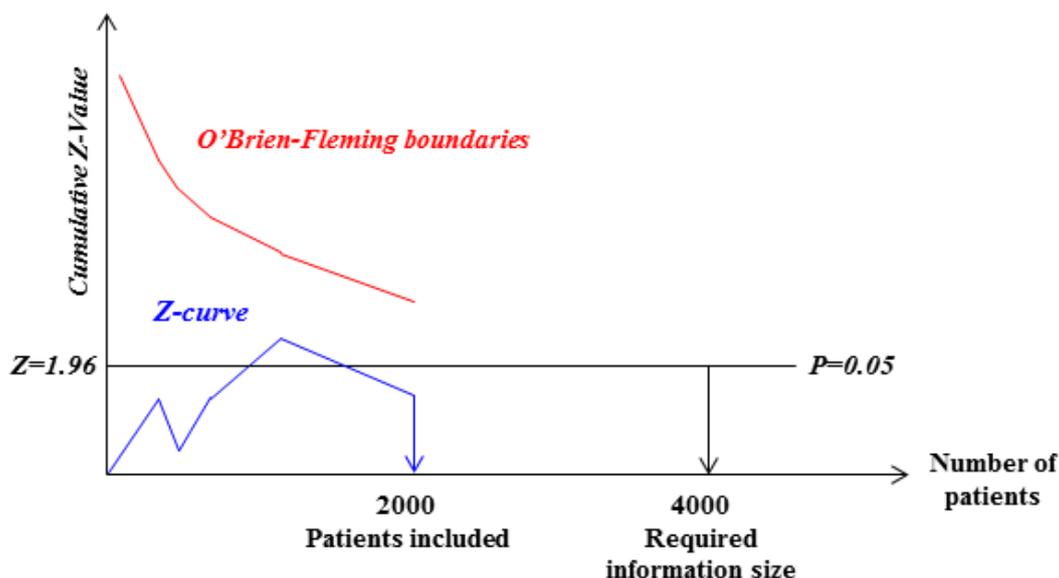


Figura 8. Ejemplo de un metanálisis que incluye un valor Z de falsos positivos en la quinta prueba de significación acumulativa.

En el ejemplo dado en la figura 8, el tamaño de la información requerida es de nuevo 4000 pacientes y la información obtenida es ahora 2000 pacientes. El valor final de Z es menor que 1,96; este resultado habría sido no concluyente utilizando tanto técnicas límites como convencionales. Hay, sin embargo, valores previos de Z que se habían calculado en el proceso acumulativo, incluyendo uno con un valor superior a 1,96. Este ejemplo ilustra cómo una curva Z acumulada pudo cruzar el umbral convencional de significación en un metanálisis inicial, solo se declaró no significativo en un metanálisis posterior. Los límites O'Brien-Fleming pueden prevenir tales conclusiones prematuras falsas positivas.

En el ejemplo dado en la figura 9, el tamaño de la información requerida y el tamaño de la información obtenida son los mismos que los de la figura 8. Aquí, el valor calculado de Z en la quinta prueba de significación es "suficientemente extremo"; la curva Z cruza los límites O'Brien-Fleming, y el metanálisis se puede declarar como

concluyente en relación con el efecto de la intervención prevista que lleva a la información del tamaño requerido.

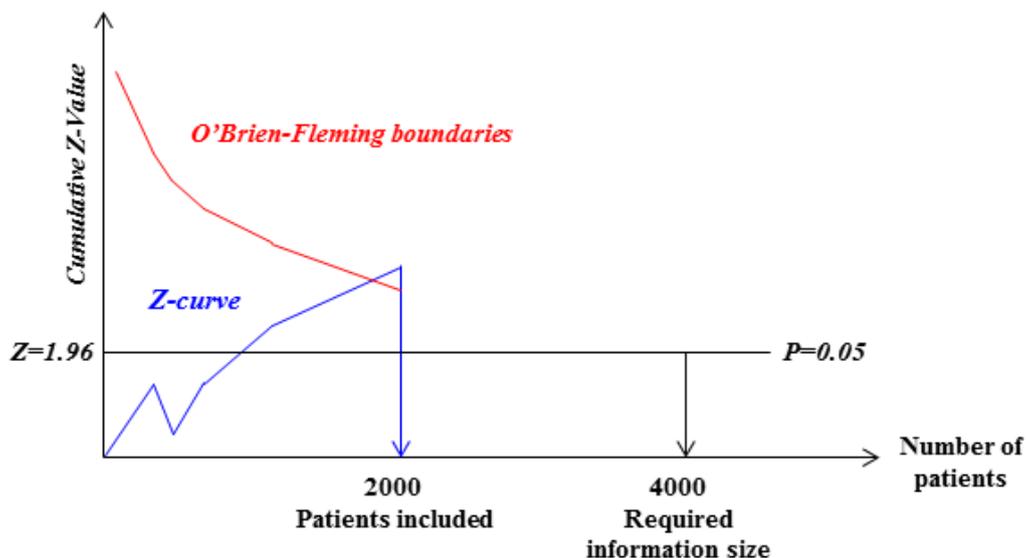


Figura 9. Ejemplo de un metanálisis que se hace concluyente según los límites O'Brien-Fleming después de la quinta prueba de significación acumulativa.

En los ejemplos anteriores (Figura 7-9), los límites de control se construyen solo para la mitad positiva del eje y. Los límites de pruebas de significancia simétricos bilaterales se pueden construir tanto para las mitades negativa y positiva del eje y. El programa TSA permite pruebas de significación tanto unilateral como bilateral. Cuando la medida de resultado para metanálisis de datos binarios es definida como falla (por ejemplo, la muerte o la recaída), los valores Z en la mitad superior del eje de las y indicarán beneficio de la intervención experimental, mientras que los valores Z en la mitad inferior indicarán daño.

Los valores de los límites de supervisión para la curva Z son una función de la función gasto de alfa; se calculan por integración numérica recursiva según Reboussin et al.⁶⁰ Aunque todos los valores límite son puntos discretos calculados para cada actualización acumulativa de los metanálisis, el programa TSA conecta estos puntos y crea una línea límite continua para una mejor interpretación visual.

2.2.5 Intervalos de confianza ajustados tras el análisis secuencial de los ensayos

Así como las pruebas de significación repetidas afecta el error de tipo I global, también afecta la construcción de intervalos de confianza. Por ejemplo, cuando se supone que nuestra estimación combinada del efecto se distribuye normalmente (como normalmente hacemos en el metanálisis), formamos un intervalo de confianza 'nuevo' del 95% simétrica, $\hat{\mu} \pm 1.96 \cdot se(\hat{\mu})$ donde $\hat{\mu}$ denota nuestro efecto de la intervención estimada metanalizada y $se(\hat{\mu})$ denota su error estándar asociado. Sin embargo, si un metanálisis se somete a evaluación estadística repetida, y por lo tanto, produce una serie de intervalos de confianza con el tiempo, la probabilidad de que todos estos intervalos de confianza contengan el «verdadero» efecto global es sin duda inferior a 95 %. Eso quiere decir, que si construimos una serie de nuevos intervalos de confianza simétricos $(1-\alpha) \%$, $\hat{\mu} \pm z_{1-\alpha/2} \cdot se(\hat{\mu})$ la probabilidad de que todos estos intervalos de confianza contengan el efecto global «verdadero» es ciertamente menor que $(1-\alpha) \%$. Por lo tanto, cuando un metanálisis se somete a evaluación estadística repetida, existe un riesgo exagerado que los «nuevos» intervalos de confianza generarán inferencias ficticias. Cuando existe algún efecto subyacente «verdadero» de la intervención, pueden ocurrir inferencias falsas basadas en los intervalos de confianza tal como muestran los dos escenarios que se ilustran en la figura 10.

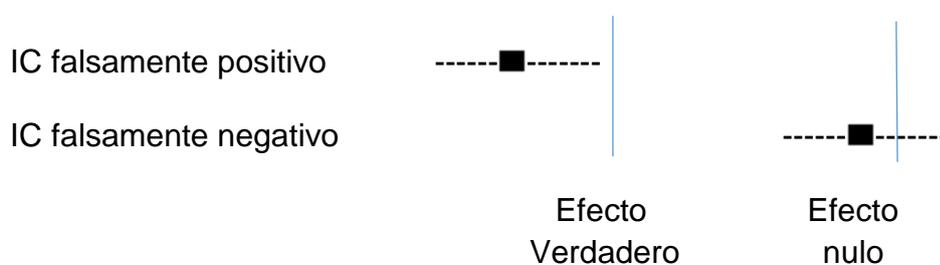


Figure 10. Ejemplos de inferencias de intervalos de confianza falsamente positivo y negativo

Cuando no hay un efecto de la intervención, los intervalos de confianza producirán inferencias falsas si excluyen el efecto nulo. Esta situación es idéntica a una prueba de significación falsa positiva (véase la sección 2.2.4).

Tal como sucede para las pruebas de significación repetida, los intervalos de confianza se pueden ajustar de acuerdo con la solidez de la información disponible (por ejemplo, el número de pacientes) y el número de evaluaciones estadísticas. Si asumimos que l y u marquen el límite inferior y superior de algún nuevo intervalo de confianza que incluya $1-\alpha$, sabemos que

$$\Pr(l \leq \mu \leq u) = 1 - \alpha$$

Cuando un metanálisis se somete a la evaluación estadística repetida, los nuevos intervalos de confianza repetidos, no producirá la cobertura deseada. Por lo tanto, tenemos que establecer una serie de intervalos que permitan alcanzar la cobertura deseada. Supongamos que un metanálisis se somete a evaluación estadística k veces hasta el punto en que supera su tamaño de información requerida. Asumamos que l_1, l_2, \dots, l_k y u_1, u_2, \dots, u_k denotan los límites inferior y superior de los intervalos de confianza para cada una de las k veces que el metanálisis fue sometido a una evaluación estadística. Para mantener la cobertura deseada, estos límites tendrían que satisfacer:

$$\Pr(l_1 \leq \mu \leq u_1, l_2 \leq \mu \leq u_2, \dots, l_k \leq \mu \leq u_k) \geq 1 - \alpha$$

Y así, uno solo de estos k intervalos, digamos j , tendría que satisfacer:

$$\Pr(l_j \leq \mu \leq u_j) \geq 1 - \alpha$$

Se desprende de lo anterior, que el nivel α para cada intervalo de confianza repetido no puede superar el máximo global α . Además, los niveles α respectivos para cada uno de los intervalos de confianza repetidos deberían sumar el máximo global α . De este modo, mediante el control del nivel α global, podemos controlar la cobertura global. El marco de trabajo para controlar el nivel α global ya ha sido desarrollado en la sección anterior (2.2.4), y se aplica fácilmente a los intervalos de confianza repetidos. Los nuevos intervalos de confianza se obtienen utilizando la fórmula $\hat{\mu} \pm z_{1-\alpha/2} \cdot se(\hat{\mu})$ porque sabemos que $z_{\alpha/2} \leq \hat{\mu} / se(\hat{\mu}) \leq z_{1-\alpha/2}$ con aproximadamente $(1 - \alpha)$ % de probabilidad (bajo la hipótesis nula), y por lo tanto:

$$z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}$$

donde Z representa el valor Z para la prueba de significación estadística. Mediante la sustitución de $z_{\alpha/2}$ y $z_{1-\alpha/2}$ por los umbrales que constituyen los límites de seguimiento estadístico, C_1, C_2, \dots, C_K , y único para $\hat{\mu}$, hemos construido una expresión sencilla para los intervalos de confianza repetidos que mantengan un buen control de la cobertura. Para uno solo de los k intervalos de confianza digamos j , la expresión para el intervalo de confianza es:

$$\hat{\mu} \pm c_j \cdot se(\hat{\mu})$$

Y tenemos

$$\Pr(\hat{\mu} - c_1 \cdot se(\hat{\mu}) \leq \mu \leq \hat{\mu} + c_1 \cdot se(\hat{\mu}), \dots, \hat{\mu} - c_k \cdot se(\hat{\mu}) \leq \mu \leq \hat{\mu} + c_k \cdot se(\hat{\mu})) \geq 1 - \alpha$$

Todo lo anterior fácilmente se generaliza para intervalos de confianza unilaterales.

El software TSA aporta la opción de calcular el intervalo de confianza para el último de una serie de evaluaciones estadísticas (véase el capítulo 4).

2.2.6 La ley del logaritmo iterado

Otra solución al problema a la repetición de pruebas de significación, se indica en la sección 2.2.3. es penalizar los valores de Z de acuerdo con la solidez de la evidencia disponible y el número de pruebas de significación repetida.^{7,8} En probabilidad avanzada, existe un teorema, *la ley de los logaritmos iterados*, que nos dice que si tenemos un variable con distribución estándar normal, tal como el valor de Z, y se divide por el logaritmo del logaritmo del número de observaciones en los datos, habrá un 100 % de probabilidad de que esta fracción asumirá un valor ubicado entre $-\sqrt{2}$ y $\sqrt{2}$. En el contexto del estudio estadístico, esta ley puede ser utilizada para controlar la exageración de error de tipo 1 en el metanálisis debido al estudio repetido de la significación. Dividiendo una prueba estadística con distribución estándar normal por el logaritmo del logaritmo de la información disponible, siempre que existan suficientes datos acumulados, puede proporcionar un buen control del "comportamiento" de la prueba estadística empleada. Lan *et al* aplicaron esta teoría, introduciendo una sanción para los valores de Z obtenidos en cada prueba de significación y la creando valores Z ajustados (penalizados), Z^* , dado por

$$Z_j^* = \frac{Z_j}{\sqrt{\lambda \ln(\ln(I_j))}}$$

donde Z_j es el valor Z convencional, I_j es la información estadística acumulada en la prueba de significación j -th (ver sección 2.2.1. denominada alternativas al número de paciente acumulados), y λ es alguna constante que asegurará un adecuado control del error tipo I máximo.⁸ Lan *et al* emplearon una simulación para estimar las opciones adecuadas de esta constante, λ , para el metanálisis de datos continuos,⁸ y Hu *et al* hicieron lo mismo para los metanálisis de datos dicotómicos.⁷ Para el metanálisis de datos continuos, Lan *et al* encontraron que $\lambda=2$ exhibiría, generalmente, un adecuado control del error tipo I, cuando se utiliza un error máximo deseado tipo I de $\alpha=5\%$ para una prueba estadística bilateral (es decir, $\alpha=2,5\%$ para cada lado).⁸ Eso es, cuando Z^* se evaluó basado en los criterios convencionales para la significación estadística (es decir, $|Z^*| \geq 1,96$ representa significación estadística bilateral $\alpha = 5\%$). Para el metanálisis de datos dicotómicos, Hu *et al* estimaron elecciones adecuadas de λ para

diferentes niveles de error máximo tipo I y diferentes medidas de efecto.⁷ Los resultados de su simulación llevaron a recomendar los valores λ que se presentan en la tabla 2.

Medida del efecto	Error máximo tipo I (umbral correspondiente)		
	$\alpha=0,01(c=2,33)$	$\alpha=0,025(c=1,96)$	$\alpha=0,05(c=1,65)$
Diferencia de riesgo	$\lambda=3$	$\lambda=1,5$	$\lambda=1,5$
Riesgo relativo	$\lambda=3,5$	$\lambda=2$	$\lambda=2$
Odds ratio	$\lambda=3,5$	$\lambda=2$	$\lambda=2$

Tabla 2. Valores λ recomendados para valores Z penalizados con la ley del logaritmo iterado.

Estos valores λ pertenecen solo a los rangos de tamaño de los estudios, proporción de eventos en el grupo de control, y la variación entre ensayos utilizados en las simulaciones, y por lo tanto pueden no ser aplicables a todos los escenarios del metanálisis.^{7,8} Por ejemplo, la proporción mínima de eventos en los grupos control utilizados en las simulaciones era 0,05. Muchas condiciones clínicas importantes producen proporciones de eventos del grupo control inferiores a 0,05. Además, ninguna de las simulaciones incorporó el sesgo de tendencia temporal, tales como el sesgo de desfase del tiempo y el sesgo de publicación. Estos sesgos tienen un impacto considerable sobre las pruebas de significación en los metanálisis. Además, como se señaló anteriormente (sección 2.2.1 - Limitaciones), la información estadística se basa en una estimación precisa y fiable de la variación entre ensayos. Si la variación entre ensayos es subestimada (por ejemplo, debido al sesgo de desfase del tiempo), el estadístico Z penalizado será artificialmente grande. Por las razones anteriores, es razonable suponer que los valores λ recomendados en la tabla 2, constituyen lo más mínimo de un rango de opciones apropiadas. Valores λ apropiadas para los metanálisis de datos dicotómicos incluyendo solo un número pequeño de ensayos, los pacientes y/o eventos son probablemente más altos que los recomendados por Hu *et al.*

2.2.7 La función del gasto de β y los límites de futilidad

Cuando se encuentra el resultado de un metanálisis que es no significativo, es importante evaluar si esta no significación se debe a la falta de potencia o si se debe a la equivalencia subyacente entre las intervenciones.

El ejercicio estadístico para examinar la equivalencia - es decir, estudios, tanto para no superioridad y no inferioridad de una determinada intervención - es comúnmente conocida como estudios de inutilidad.³⁰ Los umbrales de las pruebas estadísticas que surgen de este ejercicio se denominan límites de futilidad. Cuando una curva Z cruza los límites de futilidad, podemos aceptar que las dos intervenciones no se diferencian más que el efecto previsto de la intervención.

Los metanálisis que ya han superado su información requerida deben tener suficiente potencia para demostrar la superioridad de una intervención sobre la otra. Para esta sub-sección, vamos a considerar solo los metanálisis no significativos que no han superado su tamaño requerido de información. Además, ya no consideramos todos los valores de Z como absoluta. En su lugar hacemos la distinción de los valores positivos de Z indican que la intervención experimental es superior a la intervención de control y los valores negativos de Z indican que la intervención experimental es inferior a la intervención de control. La siguiente sección trata primero con pruebas de no superioridad, seguido por la prueba de no inferioridad y pruebas de inutilidad en general.

En cualquier momento, un metanálisis puede producir un valor de Z que no es estadísticamente significativo a favor de la intervención experimental. Sin embargo, solo cuando este valor Z se encuentra 'suficientemente por debajo' del umbral de significación estadística (a favor de la intervención experimental) podemos estar seguros de que la intervención experimental no es superior al control. Para dar sentido a lo mencionado, debemos primero definir lo que entendemos por superior y 'lo suficientemente por debajo'.

En el marco de trabajo del examen repetido de significación estadística, la definición de la superioridad está vinculada a la suposición subyacente hecha para el tamaño de la información requerida. Al calcular el tamaño de la información requerida, se supone, *a priori*, un efecto de la intervención, δ . La magnitud de este efecto representa lo que creemos que es una *diferencia mínima importante* entre las dos intervenciones. Idealmente, el tamaño de δ debe definirse de tal manera que cualquier cosa más pequeña sería considerada clínicamente, o prácticamente, poco importante y por lo tanto no merece la pena investigar. El valor de δ depende del contexto del estudio. Por ejemplo, una RRR del 10 % por lo general se considera importante si el desenlace es mortalidad, pero no se puede considerar importante si el desenlace es náusea.

Antes de definir qué se entiende por «suficientemente bajo» en el contexto del estudio repetido de pruebas de significación estadística, en primer lugar hay que considerar la situación en la que la información contenida en un metanálisis es igual al tamaño requerido de la información y donde la prueba estadística se realiza por primera vez.

En primer lugar, sea H_δ que significa que la hipótesis de que el efecto es igual a δ – esta es la hipótesis alternativa (en contraste con la hipótesis nula). Bajo el supuesto de que H_δ es cierto, la probabilidad de que el metanálisis será estadísticamente significativa (con el nivel α elegido) es igual a la potencia elegida, $1-\beta$. Cuando se ha alcanzado el tamaño de la información, la probabilidad de que el metanálisis sea un falso negativo es igual a β . En esta situación, nuestro umbral para la significación estadística, c , que satisface:

$$Pr(|Z| \geq c \mid H_0 \text{ is true}) \leq \alpha$$

implícitamente se convierte en nuestro umbral para no superioridad porque c también satisface:

$$Pr(Z < c \mid H_\square \text{ is true}) \leq \beta.$$

Cuando la prueba estadística repetida se produce antes de que un metanálisis supere su tamaño requerido de información, también es posible probar la no superioridad. Este examen se puede realizar mediante la definición de umbrales que, bajo la hipótesis alternativa, no generan una inflación del riesgo total de error tipo II. Por ejemplo, si examinamos dos veces la no-superioridad, tenemos que encontrar umbrales, c_1 y c_2 , para los dos valores emergentes de Z subsiguientes, Z_1 y Z_2 ,

$$Pr(Z_1 < c_1 \text{ or } Z_2 < c_2) \leq \beta$$

En esta situación, los valores Z_1 inferiores a c_1 y valor Z_2 más pequeño que c_2 se considerará «suficientemente bajos» el umbral de significación estadística para justificar la conclusión de no-superioridad. En un contexto más general, donde podríamos probar no-superioridad k veces, necesitaríamos encontrar umbrales c_1, \dots, c_k que satisfaga

$$Pr(Z_1 < c_1 \text{ or } Z_2 < c_2 \text{ or } \dots \text{ or } Z_k < c_k) \leq \beta$$

bajo la hipótesis alternativa, H_δ . Esto es equivalente a encontrar el máximo k riesgos de error tipo II, β_1, \dots, β_k , que suman β y donde

$$\begin{aligned}
& \Pr(Z_1 < c_1) \leq \beta_1 \\
& \Pr(Z_2 < c_2 \mid Z_1 \geq c_1) \leq \beta_2 \\
& \Pr(Z_3 < c_3 \mid Z_1 \geq c_1 \text{ and } Z_2 \geq c_2) \leq \beta_3 \\
& \vdots \\
& \Pr(Z_k < c_k \mid Z_1 \geq c_1 \text{ and } \dots \text{ and } Z_{k-1} \geq c_{k-1}) \leq \beta_k
\end{aligned}$$

bajo la hipótesis alternativa.

Este deseo de controlar el error tipo II en el contexto del estudio de pruebas repetidas es análogo al deseo de controlar el error tipo I. Los múltiples exámenes aumentan la cantidad real de error y tenemos que encontrar una técnica para controlar este aumento. Tal como es causado por el mismo fenómeno, el problema de un aumento del error tipo II se puede manejar mediante una solución similar. En la sección 2.2.3, la función de gasto alfa fue descrita como una técnica que se puede utilizar para crear límites razonables para las pruebas de significación. Del mismo modo, el problema de encontrar los umbrales para pruebas repetidas de no-superioridad, lo que no asegurará un adecuado control del error tipo II, se puede resolver mediante la introducción de la función del gasto de β . La función del gasto de β es monótonamente creciente de tiempo que se utiliza para asignar apropiadamente los riesgos máximos para error de tipo II β_1, \dots, β_k en cada prueba de no-superioridad de acuerdo con la cantidad de información acumulada. La función del gasto de β es una función de la fracción de información, IF (la información acumulada dividida por el tamaño de la información requerida), y solo se define de 0-1. La función de gasto de β de 0 es siempre igual a 0: $\beta(0)=0$, y la función de gasto de β de 1 es siempre igual a β : $\beta(1)=\beta$. En cualquier punto entre 0 y 1, la función de *gasto de β* es igual al riesgo de error máximo total de tipo II que ha surgido de los umbrales elegidos para todas las pruebas de no-superioridad, hasta e incluyendo la prueba de orden i -th. En otras palabras, la función de *gasto de β* es igual a cuánto error tipo II ha sido 'gastado'. En la notación: $\beta(IF_i)=\beta_1+\beta_2+\dots+\beta_i$.

Por las mismas razones que se describen en la sección 2.2.4, la función O'Brien-Fleming también puede constituir la mejor opción para la función de gasto de β . En el TSA versión 0.8, la única función de gasto de β disponible es la función del gasto O'Brien-Fleming.

La figura 11 muestra un ejemplo de un metanálisis que incluye tanto las pruebas repetidas de no-superioridad y de significación. En este metanálisis, el tamaño de la información requerida es de 4000 pacientes. Con 2000 pacientes, el metanálisis no es concluyente, ya que aún no ha cruzado el límite (superior) para la significación

estadística o los límites (inferior) para no-superioridad. Las extensiones discontinuas de la curva Z ilustran ejemplos de cómo el metanálisis podría llegar a ser concluyente con 3000 pacientes.

En el ejemplo (A), la curva Z cruza los límites de no-superioridad (los límites más bajos), en cuyo caso, se infiere que la intervención experimental no es superior a la intervención de control. En el ejemplo (B), la curva Z cruza los límites de significación O'Brien- Fleming para la superioridad, en cuyo caso, se infiere que la intervención experimental es superior a la intervención de control.

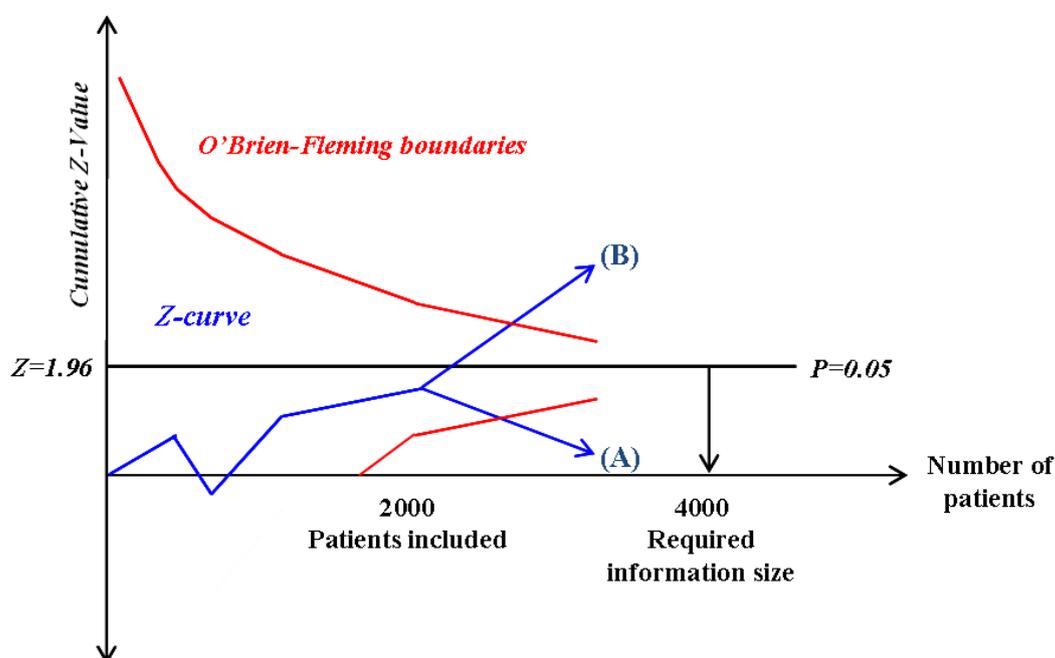


Figura 11. Ejemplo de un metanálisis que incluye estudios repetidos de no-superioridad (línea roja) y significación (línea marrón). La curva Z acumulada de los primeros cuatro ensayos alcanza la mitad del tamaño de la información requerida. Dos nuevos ensayos son añadidos al metanálisis - (A) mostrando que no hay efecto (y la puntuación Z acumulada ahora llega a futilidad) y (B) que muestra un beneficio significativo de la intervención (y el puntaje Z acumulado ahora alcanza significación ya que cruza tanto los límites convencional, y de O'Brien-Fleming).

Con el fin de evaluar la equivalencia entre dos grupos, los límites de no-superioridad y de no-inferioridad deben ser utilizados simultáneamente. Imagine un metanálisis comparando dos grupos: el grupo A y grupo B. Si un valor acumulado de Z cae por debajo del umbral de no-superioridad, entonces el grupo A no es mejor que el grupo B. Pero puede ser peor. Si el mismo valor acumulado de Z también cae por encima del umbral de no-inferioridad, entonces, el grupo A no es peor que el grupo B. En esta situación, se puede concluir que el grupo A y B son equivalentes. Gráficamente, esta

"área en equivalencia" es el área dentro de los dos límites después de cruzar - también llamada la *inner wedge* (ver figura 12).

La figura 12 muestra un ejemplo de un metanálisis que incluye todos los componentes del TSA que se han discutido: el tamaño requerido de la información, los límites de las pruebas de significación bilateral, los límites de futilidad tanto de no-superioridad como de no-inferioridad. En este ejemplo, el tamaño de la información requerida es 4000. Con aproximadamente 3000 pacientes, el valor de Z está ubicado dentro del área de equivalencia (*inner wedge*) y una conclusión se puede hacer: el efecto de la intervención no es mayor que el previsto.

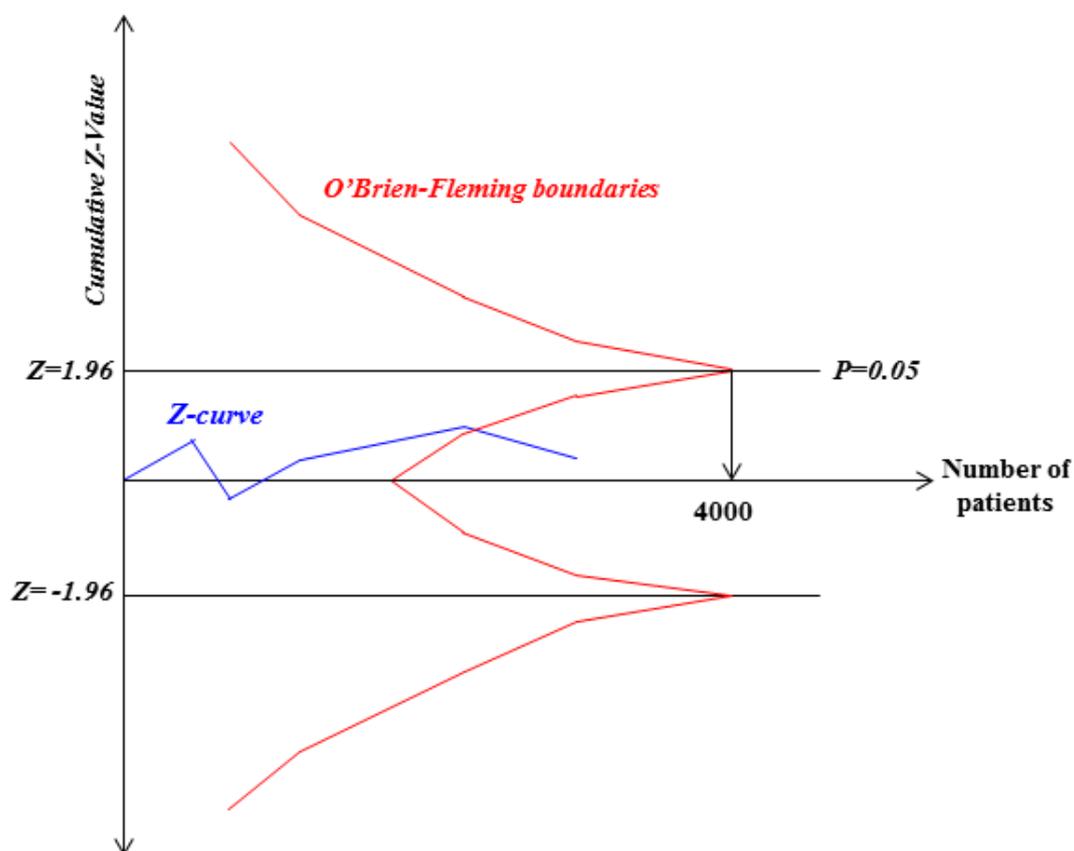


Figura 12. Ejemplo de un metanálisis con límites de evaluación de significación repetida, no-superioridad y no-inferioridad.

3 Instalación e inicio del programa TSA

3.1 *Requisitos previos*

El software de Análisis Secuencial de los Ensayos (TSA, por sus siglas en inglés) es un programa Java™ y, por tanto, funciona en cualquier sistema operativo que utilice Java™ (Microsoft Windows, Mac OS, UNIX, Linux, etc.). El software TSA requiere la versión más reciente (o, al menos, una reciente) de *Java Runtime Environment* (JRE) instalado en su ordenador. Puede descargar JRE de forma gratuita desde el sitio www.java.com.

Para el momento de escribir este manual (agosto 2011), la última versión de JRE es de 1.6. El software TSA funciona bien con esta versión.

3.2 *Instalación*

El software TSA se entrega en un archivo ZIP. Utilice cualquier herramienta de archivo, como WinRAR o GZIP, para descomprimir el archivo. En el archivo se encuentran tres archivos denominados *TSA.jar*, *RM5Converter.jar*, y *TEMPLATES.TPL* junto con dos carpetas con el nombre *lib* y *samples*.

El *TSA.jar* es un archivo Java que contiene la aplicación *Trial Sequential Analysis*. *RM5Converter.jar* es un archivo Java que contiene el programa para la conversión de datos de los ensayos (en la actualidad, sin embargo, solo para los resultados dicotómicos) exportados desde versión 5 del *Review Manager* en el formato de datos adecuado para TSA. *TEMPLATES.TPL* contiene plantillas de límites de supervisión que puede utilizar cuando se está realizando el análisis secuencial de los ensayos en los datos de su metanálisis. El contenido del archivo de plantillas (*templates*) es controlado a través del programa TSA. La carpeta 'lib' contiene varios paquetes externos utilizados por el programa TSA. La carpeta "muestras" (*samples*) contiene archivos '.TSA' para los ejemplos suministrados en este manual (véase el capítulo 5).

Para instalar el programa, descomprima todo el archivo en una carpeta de su elección en su disco duro. No se requieren pasos adicionales.

3.3 *Iniciando el TSA*

Para comenzar a utilizar el software TSA, haga doble clic en el archivo *TSA.jar*.

Como alternativa, el software TSA se puede iniciar en un símbolo del sistema. Para iniciar el software TSA de manera rápida, primero inicie un símbolo, vaya a la carpeta en la que ha descomprimido el software TSA, y el tipo "java-jar TSA.jar".

Si está utilizando el sistema operativo Microsoft Windows, puede abrir una ventana de *dos prompt*, haciendo primero clic en el botón 'Iniciar' (Start) (ubicada en la esquina inferior izquierda de la pantalla), luego haga clic en "Ejecutar ..." (Run). Cuando la ventana 'Ejecutar' (Run) aparece, escriba "cmd" (sin comillas) en el campo de texto y pulse Aceptar. Debe aparecer el indicador *dos prompt*. Utilice el comando cd (cambiar directorio) para buscar la carpeta en la que haya descomprimido el software TSA. Por ejemplo, si ha creado una carpeta llamada TSA dentro de la carpeta *Archivos de Programa* (Program Files) en la unidad C y desempaquetado el programa TSA en esta carpeta, primero debe cambiar el directorio a la carpeta del TSA en el indicador *dos prompt*. Esto se puede hacer escribiendo 'cd C:\Program Files\TSA' (sin comillas). Después de que el directorio en el indicador *dos prompt* se ha cambiado, escriba 'java -jar TSA.jar'.

3.3.1 ¿Por qué no funciona el TSA?

Si usted está teniendo problemas para iniciar el TSA, existen varias razones posibles para explicar esto. A continuación se muestra una lista de verificación para ayudar a identificar las posibles razones:

- ¿Está el JRE instalado en el sistema?
- ¿Está instalada, al menos, la versión 1.6 del JRE?
- ¿Están extraídos todos los archivos del archivo ZIP?
- ¿Cambió el nombre, movió o eliminó cualquiera de los archivos descomprimidos o carpetas?

Si un programa diferente (que no sea TSA) se inicia al hacer doble clic en el archivo TSA.jar, esto significa que la extensión de nombre de archivo jar. no está asociado a Java (JRE). Si esto sucede, usted puede tratar de iniciar el programa manualmente mediante un símbolo (véase más arriba), o usted puede tratar de cambiar la asociación de nombre de archivo. Si está utilizando Windows, puede cambiar la asociación siguiendo los pasos a continuación:

- Abrir una ventana del Explorador (por ejemplo, haga doble clic en mi PC) y haga clic en el menú 'Herramientas'
- Seleccione "Opciones de carpeta ..." y vaya a la pestaña "Tipos de archivo"
- Encontrar la extensión JAR en la lista
- Haga clic en 'Cambiar'
- Seleccione "Java (TM) Plataforma SE binary" de la lista y haga clic en Aceptar
- Si 'Java (TM) Plataforma SE binary' no está en la lista, haga clic en "Examinar" y busque la javaw.exe en la carpeta bin del JRE's. Su ruta predeterminada es: C:\Program Files\Java\jre6\bin.

Si su sistema operativo no es Microsoft Windows, consulte el manual del usuario *ad hoc*.

3.4 **Uso RM5 Converter**

Para iniciar el programa convertidor de datos del Review Manager versión 5 (en la actualidad, sin embargo, solo para datos dicotómicos), haga doble clic en el archivo *RM5Converter.jar* file.

3.4.1 **¿Por qué no funciona?**

El *RM5Converter.jar* tiene los mismos requisitos básicos tal como el *TSA.jar*, así que si usted está teniendo dificultad para abrirlo, por favor consulte la sección 3.3.1. También, *RM5 Converter.jar* requiere el archivo *TSA.jar* para poder funcionar. Por esta razón, *TSA.jar* tiene que estar ubicado en la misma carpeta que *RM5Converter.jar*.

4 **Cómo utilizar el software TSA**

4.1 **Iniciando**

Cuando se inicia el TSA, debería aparecer una ventana similar a la figura 13. La ventana de inicio debe contener una barra de menú con los menús *Archivo* (File), *Configuración* (Settings) y *Ayuda* (Help), así como cinco en fichas (tabs) sombreadas en gris (no seleccionables): *Metanálisis* (Metanalysis), *Ensayos* (Trials), *TSA*, *Gráficos* (Graphs) y *Diversidad* (Diversity).

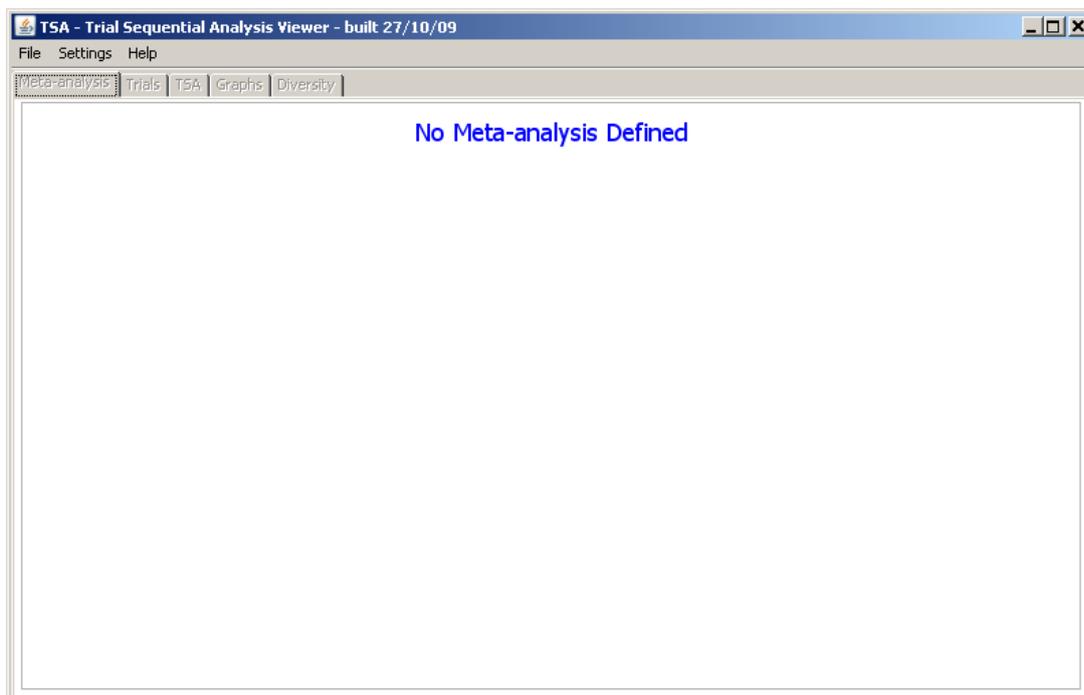
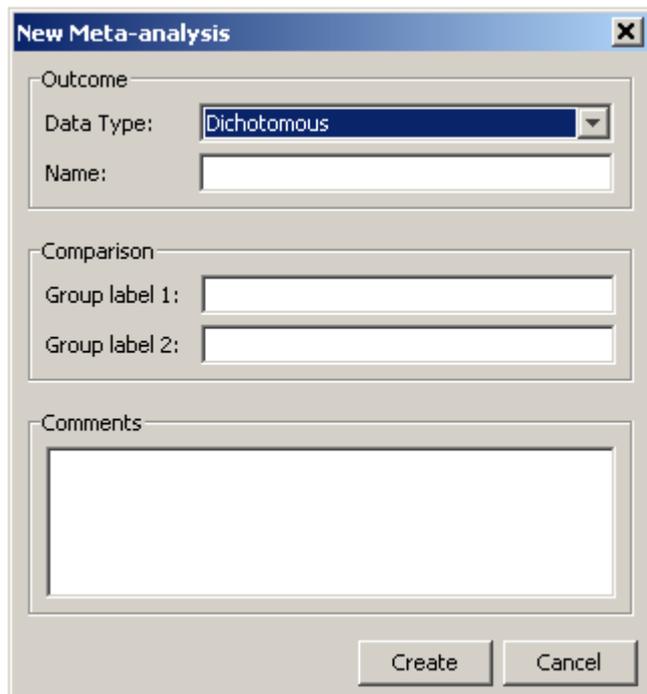


Figura 13. La ventana de inicio del TSA.

4.1.1 Creación de un nuevo metanálisis

Para crear un nuevo metanálisis, vaya a la barra de menú y seleccione Archivo (File) > (Nuevo Metanálisis)). Aparecerá un cuadro de diálogo (figura 14), lo que le permitirá dar un nombre al metanálisis, elegir el tipo de los datos que serán combinados (datos dicotómicos o continuos), especificar cuáles son las dos intervenciones que serán comparadas, clasificar el tipo de desenlace como `positivo` o `negativo` y añadir comentarios. Pulse `Crear` (Create) para generar el nuevo metanálisis. Pulse Cancelar (Cancel) para deshacer el paso previo. Si desea editar el nombre del metanálisis, las intervenciones, o sus comentarios, vaya a la barra de menú y seleccione Archivo (File) > Editar (Edit) metanálisis. El cuadro de diálogo que se muestra en la figura 14 debería reaparecer.



The image shows a software dialog box titled "New Meta-analysis". It contains the following fields and controls:

- Outcome section:**
 - Data Type:** A dropdown menu currently showing "Dichotomous".
 - Name:** An empty text input field.
- Comparison section:**
 - Group label 1:** An empty text input field.
 - Group label 2:** An empty text input field.
- Comments section:** A large, empty text area for notes.
- Buttons:** "Create" and "Cancel" buttons at the bottom right.

Figura 14. Cuadro de diálogo para la creación de un nuevo metanálisis.

Nota: Con respecto a los datos binarios, ejemplos de desenlaces negativos serían mortalidad, ictus, o incidencia de nuevos cánceres; mientras que tipos de desenlaces positivos son supervivencia, aclaramiento viral o cese del hábito de fumar. Para datos continuos, ejemplos de desenlaces negativos son un incremento en el promedio de la respuesta de un suceso desfavorable (incremento en el puntaje de depresión) y un ejemplo de desenlace positivo son incrementos en el promedio de la respuesta de un acontecimiento favorable (cuenta plaquetaria).

Después de crear el nuevo metanálisis, una serie de opciones deberían aparecer en el lado izquierdo de la ventana de inicio. (Estas opciones se describen en la sección 4.3. *Definición de la configuración de los metanálisis.*)

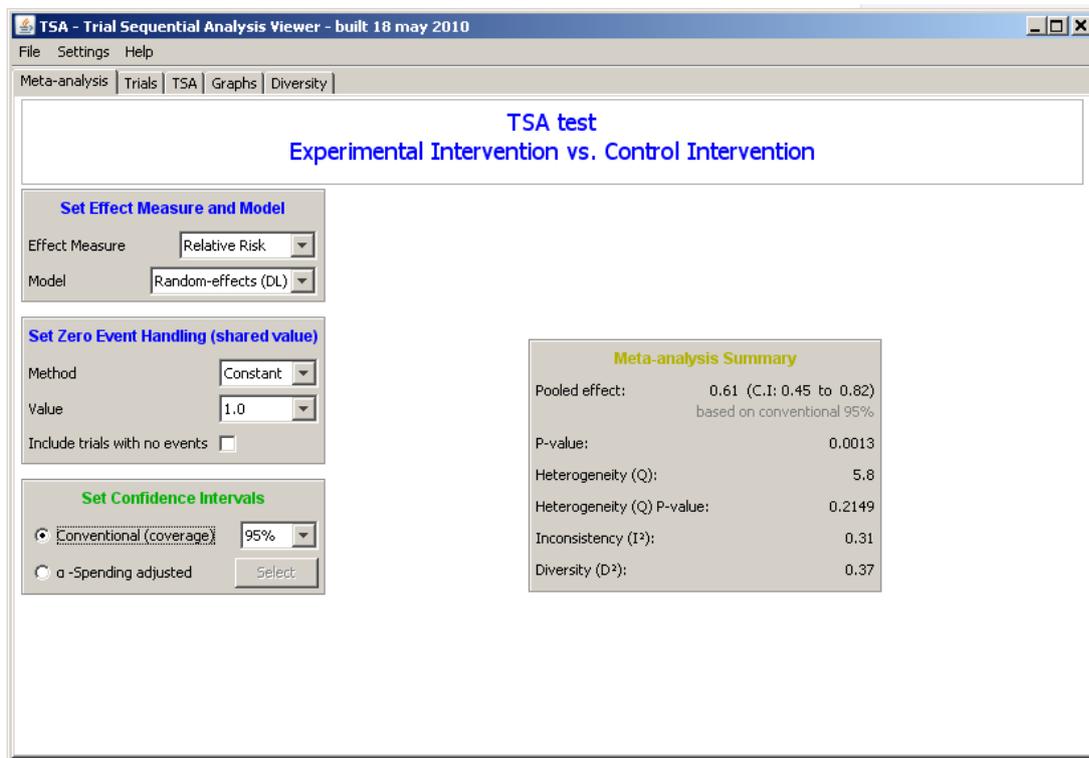


Figura 15. Ventana de inicio después de que un nuevo metanálisis ha sido creado y los datos han sido introducidos.

Un cuadro titulado *Resumen del Metanálisis* debería aparecer en el centro de la ventana.

4.1.2 Guardando un archivo TSA y abriendo un archivo TSA existente

Si desea guardar su trabajo, vaya a la barra de menú y seleccione Archivo (File)> *Guardar como* (Save as)...

Si desea continuar trabajando con un archivo TSA preexistente, vaya a la barra de menú, seleccione Archivo (File) > Abrir (Open) y localice el archivo TSA en el que desea seguir trabajando.

4.1.3 Importando los datos del metanálisis desde Review Manager v.5

Para importar los datos del metanálisis guardados en un archivo Review Manager v.5 (*.rm5) (en la actualidad, sin embargo, solo para los resultados dicotómicos), necesitará utilizar el software RM5Converter, que se incluye en el archivo Zip que ha descargado antes de instalar TSA (véase el capítulo 3).

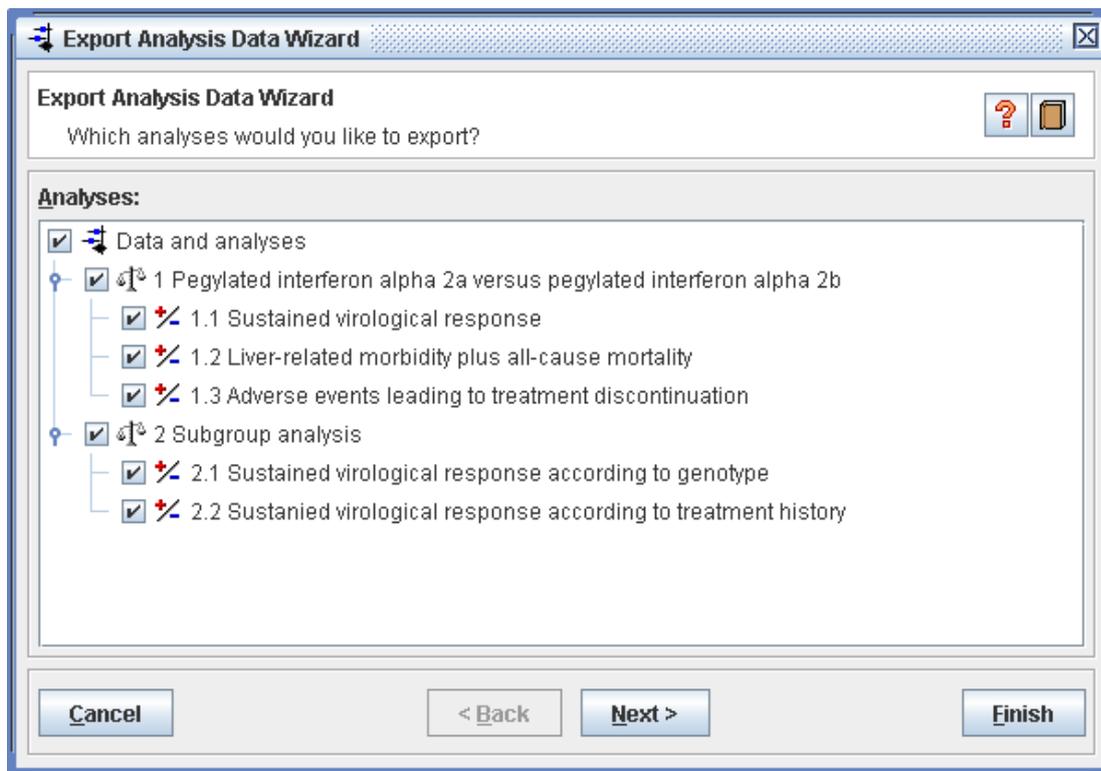


Figura 16. Ventana que muestra los datos para exportar en el Review Manager v.5, que le permite seleccionar los metanálisis que desea exportar (en la actualidad, sin embargo, solo para los resultados dicotómicos) como un archivo *.csv. El archivo RevMan a convertir es el de la revisión Cochrane titulada " alfa-2a versus pegylated interferon alfa-2b for treating chronic hepatitis C ".⁶³

El RM5Converter puede leer archivos separados por comas (*.csv). Lo primero que hay que hacer, por lo tanto, es convertir su archivo RevMan en un archivo separado por comas. Abra el archivo RevMan en el Review Manager v.5. en la barra de menú, seleccione Archivo (File) > Exportar (Export) > Datos y análisis (Data and analyses). Una ventana pop-up con casilla de verificación con forma de árbol se mostrará (figura 16). Compruebe los datos del metanálisis que desea exportar como un archivo separado por comas, y haga clic en el botón *Next*.

En la siguiente ventana chequee las tres primeras casillas de verificación: Comparación (Comparison), Número (Number), Número de desenlace (Outcome Number) y Número de Subgrupo (Subgroup Number). Luego, pulse *Finish*.

Nota, si el botón *Next* es pulsado dos veces, se le presentará la opción de elegir un campo denominado *delimiter* (el cual separa las celdas en los datos). Es importante destacar que este delimitador es una coma (lo cual está predeterminado).

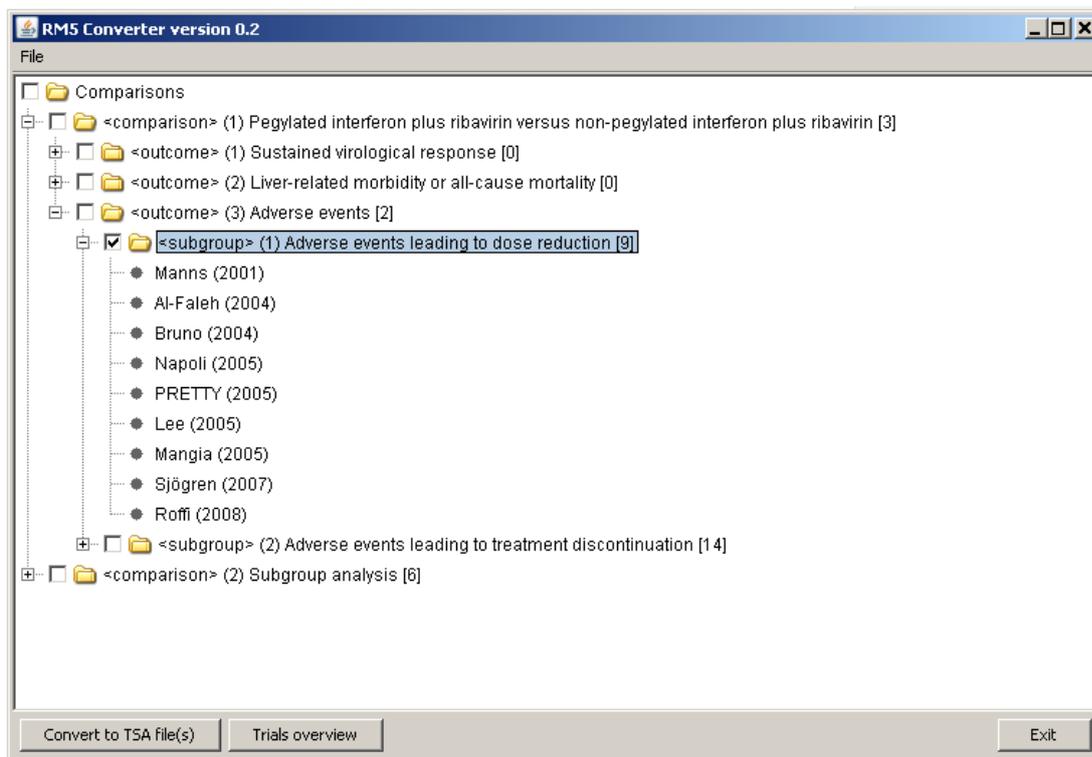
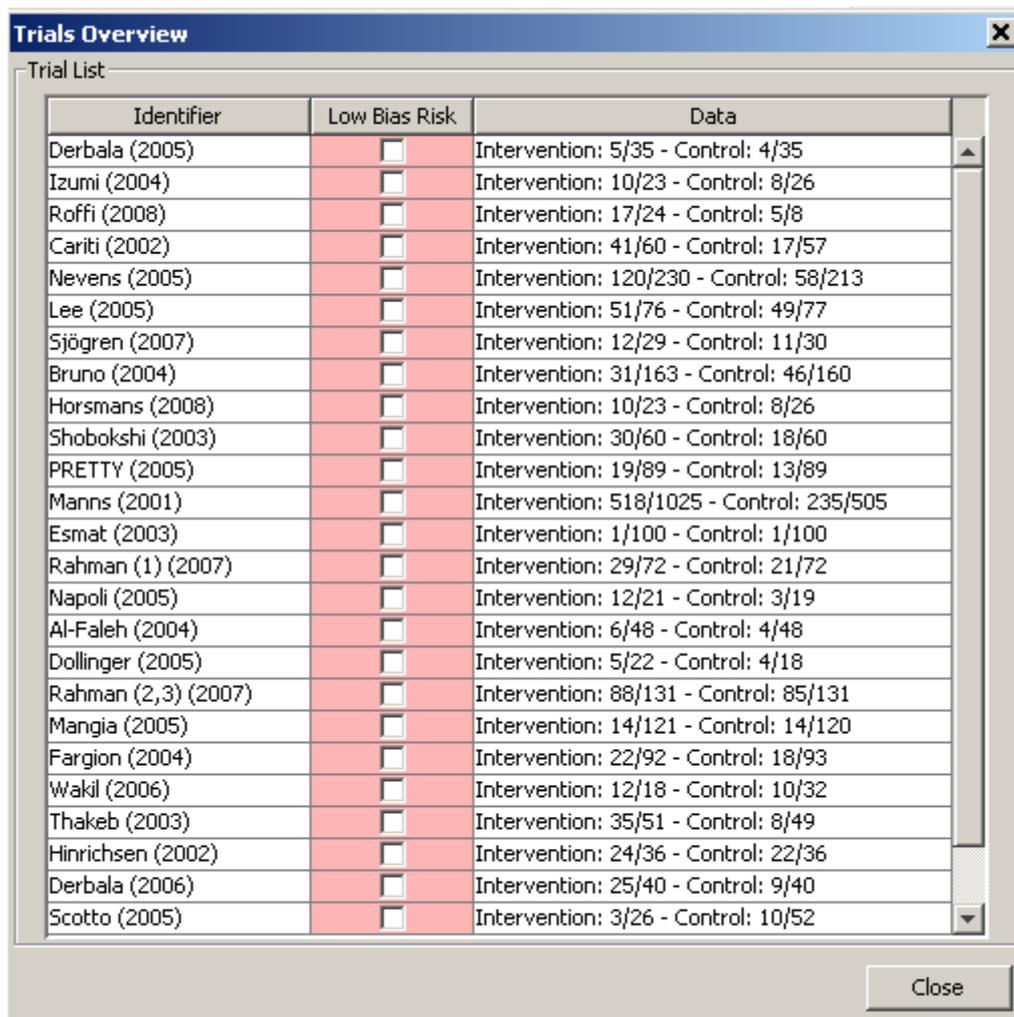


Figura 17. Comprobación en la caja con estructura de árbol en RM5Converter.

Después de exportar los datos al archivo *.CSV, abra el RM5Converter haciendo doble clic en el icono. Ir a la barra de menú y seleccione Archivo (File) > Abrir (Open). Una caja con estructura de árbol aparecerá en la ventana de la aplicación (figura 17). Los datos están estructurados de la misma manera que en Review Manager, versión 5. Para cada comparación, puede haber múltiples resultados y cada resultado representa un metanálisis. Si un metanálisis contiene análisis de subgrupos, éstos se enumeran bajo cada resultado. Si una comparación es seleccionada, todos los desenlaces de esa comparación se marcarán automáticamente. Además, si un desenlace es marcado, todos los subgrupos bajo ese desenlace, automáticamente, serán seleccionados. Todos los ensayos clínicos en cada comparación, desenlace, o subgrupo se incluyen automáticamente. Usted no tiene que convertir todo lo que aparece en una comparación dada. Usted puede 'desmarcar' las comparaciones, desenlaces y/o subgrupos que no desee convertir. Si los ensayos clínicos en los subgrupos son poco común usted tendrá la opción de combinar esos subgrupos en un único análisis.



Identifier	Low Bias Risk	Data
Derbala (2005)	<input type="checkbox"/>	Intervention: 5/35 - Control: 4/35
Izumi (2004)	<input type="checkbox"/>	Intervention: 10/23 - Control: 8/26
Roffi (2008)	<input type="checkbox"/>	Intervention: 17/24 - Control: 5/8
Cariti (2002)	<input type="checkbox"/>	Intervention: 41/60 - Control: 17/57
Nevens (2005)	<input type="checkbox"/>	Intervention: 120/230 - Control: 58/213
Lee (2005)	<input type="checkbox"/>	Intervention: 51/76 - Control: 49/77
Sjögren (2007)	<input type="checkbox"/>	Intervention: 12/29 - Control: 11/30
Bruno (2004)	<input type="checkbox"/>	Intervention: 31/163 - Control: 46/160
Horsmans (2008)	<input type="checkbox"/>	Intervention: 10/23 - Control: 8/26
Shobokshi (2003)	<input type="checkbox"/>	Intervention: 30/60 - Control: 18/60
PRETTY (2005)	<input type="checkbox"/>	Intervention: 19/89 - Control: 13/89
Manns (2001)	<input type="checkbox"/>	Intervention: 518/1025 - Control: 235/505
Esmat (2003)	<input type="checkbox"/>	Intervention: 1/100 - Control: 1/100
Rahman (1) (2007)	<input type="checkbox"/>	Intervention: 29/72 - Control: 21/72
Napoli (2005)	<input type="checkbox"/>	Intervention: 12/21 - Control: 3/19
Al-Faleh (2004)	<input type="checkbox"/>	Intervention: 6/48 - Control: 4/48
Dollinger (2005)	<input type="checkbox"/>	Intervention: 5/22 - Control: 4/18
Rahman (2,3) (2007)	<input type="checkbox"/>	Intervention: 88/131 - Control: 85/131
Mangia (2005)	<input type="checkbox"/>	Intervention: 14/121 - Control: 14/120
Fargion (2004)	<input type="checkbox"/>	Intervention: 22/92 - Control: 18/93
Wakil (2006)	<input type="checkbox"/>	Intervention: 12/18 - Control: 10/32
Thakeb (2003)	<input type="checkbox"/>	Intervention: 35/51 - Control: 8/49
Hinrichsen (2002)	<input type="checkbox"/>	Intervention: 24/36 - Control: 22/36
Derbala (2006)	<input type="checkbox"/>	Intervention: 25/40 - Control: 9/40
Scotto (2005)	<input type="checkbox"/>	Intervention: 3/26 - Control: 10/52

Figura 18. Revisión de ensayos con casillas de verificación de riesgo de sesgo.

Si pulsa el botón *Trials overview*, se abrirá una nueva ventana con una lista de todos los ensayos en el archivo csv. Cada estudio tiene una casilla de verificación asociada, que indica si el estudio se considera o no como un ensayo de "bajo riesgo de sesgo" (de manera predeterminada es alto riesgo de sesgo). Usted puede revertir la calificación del riesgo de sesgo del ensayo *ad hoc* mediante el cliqueo (o descliqueo) de la casilla respectiva de verificación del sesgo. Haga clic en el botón Cerrar (Close) una vez que haya terminado la selección del riesgo de sesgo.

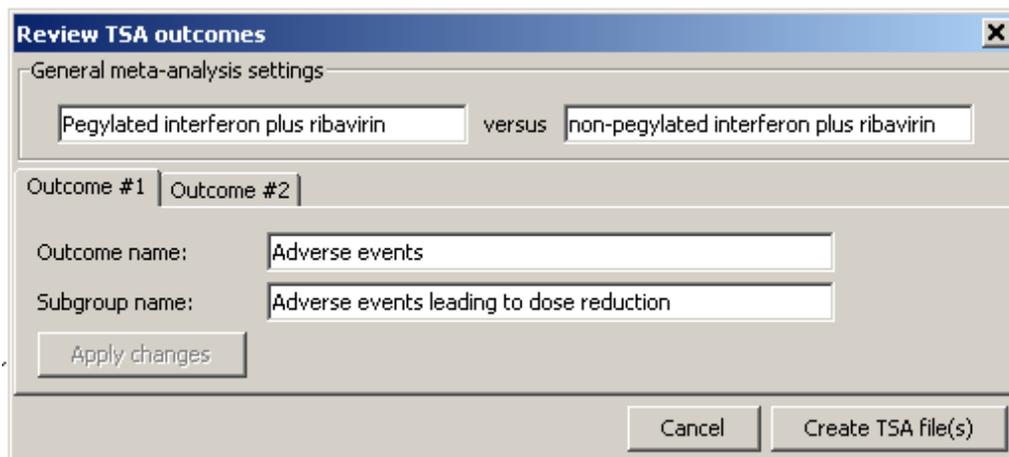


Figura 19. Revisando los desenlaces TSA en el RM5Converter.

Una vez que haya revisado todas las comparaciones, desenlaces y los subgrupos que desee incluir para su análisis secuencial de los ensayos, haga clic en el botón en la parte inferior de la ventana titulado *Create TSA file(s)*. Observará que una ventana pop-up, que le permitirá revisar los nombres de las intervenciones, los desenlaces y subgrupos. Una vez que haya verificado las selecciones, haga clic en el botón *Create TSA file(s)* y guarde los metanálisis seleccionados como archivos TSA en la carpeta seleccionada.

4.2 ***Añadiendo, editando y eliminando los ensayos clínicos***

En la barra de menú del programa TSA, hay cinco pestañas: *Meta-analysis*, *Trials*, *TSA*, *Graphs* y *Diversity*. Para agregar, editar o eliminar cualquier ensayo en el metanálisis, primero seleccione la pestaña *Trials* (ver figura 20).

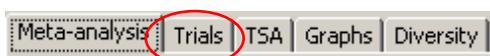


Figura 20. Haga clic en la pestaña *Trials* cuando desee agregar, editar o eliminar las pruebas en el metanálisis.

A la izquierda de la ventana (en la pestaña *Trials*) debería haber tres zonas: *Añadir Ensayos Dicotómicos/Continuos*, *Editar/Eliminar Ensayos*, e *Ignorar Ensayos*.

4.2.1 Añadiendo ensayos clínicos

Para añadir un nuevo ensayo clínico, rellene los campos respectivos en la zona llamada *Añadir Dicotómicos/Continuos*. Sin importar el tipo de datos que se están combinando, se requiere proporcionar algún nombre o título para el estudio en el campo de entrada: 'Estudio' (usualmente es el acrónimo del estudio o el apellido del primer autor). También es necesario indicar el año en que el estudio fue publicado en el campo de entrada: 'Año'. Existe la opción para calificar al ensayo como de bajo riesgo de sesgo.

Si está trabajando con datos dicotómicos, se le pedirá que introduzca el número de eventos y el número total de pacientes en el grupo de intervención (experimental) y en el grupo control (ver figura 21).

Add Dichotomous Trial

Study :

Year :

	Event	Total
Intervention	<input type="text"/>	<input type="text"/>
Control	<input type="text"/>	<input type="text"/>

Low Bias Risk

Comment :

Add Trial

Add Continuous Trial

Study :

Year :

	Mean Response	Standard Deviation	Group Size
Intervention	<input type="text"/>	<input type="text"/>	<input type="text"/>
Control	<input type="text"/>	<input type="text"/>	<input type="text"/>

Low Bias Risk

Comment :

Add Trial

Figura 21. Áreas donde debe introducir los datos requeridos cuando se añaden los datos de un nuevo ensayo dicotómico (izquierda) o ensayos de datos continuos (derecha).

Si está trabajando con datos continuos, se requiere suministrar la media, la desviación estándar y el tamaño del grupo (número de pacientes) para el grupo de intervención (experimental) y el grupo control (ver figura 21). También es posible (pero no necesario) agregar un comentario acerca de los datos introducidos. Para enviar los datos introducidos, haga clic en el botón *Agregar Ensayo (Add Trial)*.

En el lado derecho de la ventana, debe encontrar cuatro columnas denominadas: Estudio (*Study*), Riesgo de sesgo (*Bias risk*), Ignorar (*Ignore*), y Datos (*Data*). Si ha agregado ensayos, una lista de estos ensayos debería aparecer como en la figura 22. Los nombres y años de publicación de los ensayos añadidos deberían aparecer en la

primera columna (de la izquierda) en el formato '(año) Título'. Los riesgos de sesgo asignados a los respectivos ensayos deberían aparecer en la segunda columna. El riesgo de sesgo de un ensayo puede ser o bien 'Bajo' (letras verdes) o 'Alto' (letras rojas). La tercera columna le da la opción de ignorar uno o más ensayo (s) añadidos para cuando está realizando los metanálisis. Solo tiene que marcar la casilla de verificación 'Ignorar' para no tomar en cuenta un ensayo. La cuarta columna debería mostrar los datos de los ensayos. Para los ensayos clínicos de datos dicotómicos el formato es 'Intervención: Eventos/Total. Control: Eventos/Total'. Para los datos continuos el formato es 'Intervención: Promedio de la respuesta/Desviación Estándar/Tamaño de la muestra del Grupo. Control: Promedio de la respuesta/Desviación Estándar/Tamaño de la muestra del Grupo'.

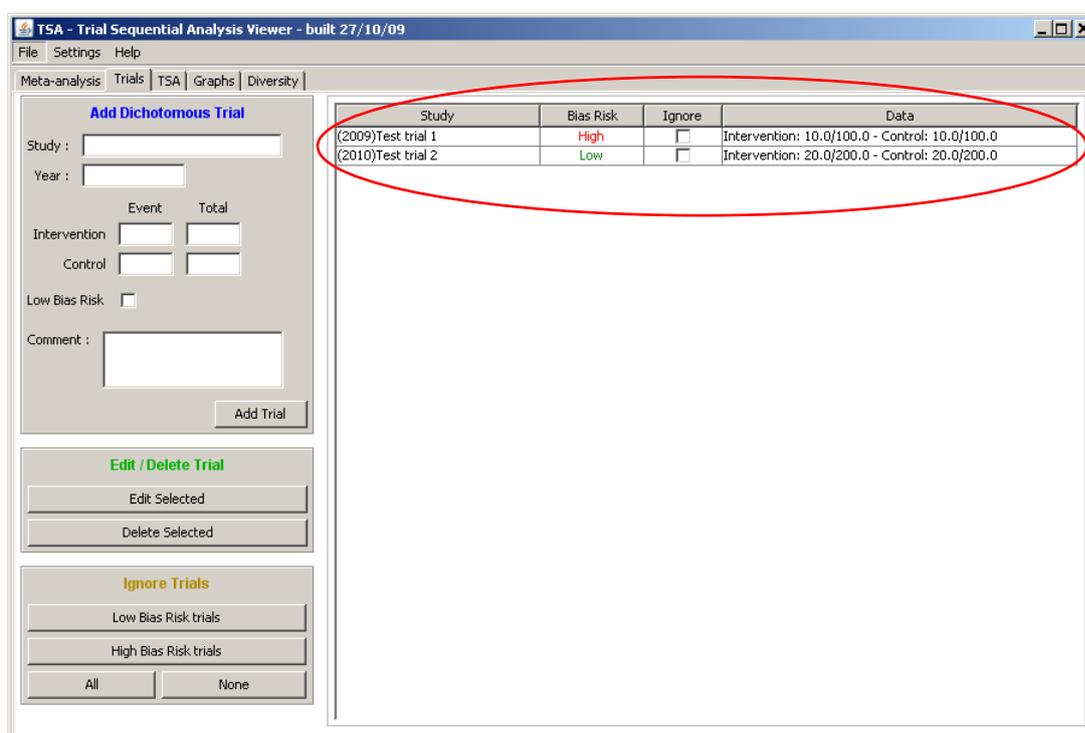


Figura 22. Lista de los ensayos clínicos añadidos marcados dentro de la elipse roja.

4.2.2 Edición y eliminación de ensayos

Para editar los datos del ensayo, primero seleccione la fila del ensayo que desea editar y haga clic en el botón *Edit Trial* en el área *Edit/Delete Selected* (figura 23). Como alternativa, puede hacer doble clic en la fila del ensayo que desea editar.

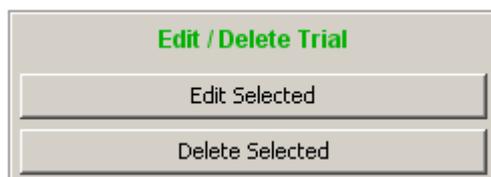


Figura 23. El área *Editar/Eliminar* del ensayo seleccionado.

Los datos del ensayo aparecen en la misma zona donde se escribió nuevos datos de los ensayos y ahora puede editar los datos. Este espacio contiene un botón denominado *Edit Trial*, en lugar de un botón *Add Trial*. Para editar los datos de los ensayos, cambia el contenido de los campos que desea editar y haga clic en el botón *Edit Trial*.

Si desea eliminar un ensayo, seleccione la fila del ensayo que desea borrar, y presione el botón *Delete Trial* en el área *Edit/Delete Selected*. Alternativamente, usted puede seleccionar la fila para el ensayo que desea eliminar y pulse el botón <Suprimir> en el teclado.

4.3 **Estableciendo las configuraciones del metanálisis**

El programa de TSA proporciona un número de opciones para realizar metanálisis. Se puede elegir entre un número de medidas de efecto, los modelos estadísticos, los métodos de manejo de datos con ausencia de eventos (para los datos dicotómicos), y los niveles de los intervalo de confianza. Todas estas opciones se pueden configurar en la pestaña *Meta-analysis* ubicada a la izquierda de la pestaña *Trials* (figura 24).



Figura 24. Haga clic en la ficha *Meta-análisis* cuando se quiere configurar la medida del efecto, el modelo estadístico, o método para el manejo de estudios dicotómicos con ausencia de eventos.

En la parte izquierda de la ventana encontrará el área denominada *Set Effect Measure and Model*, otra zona llamada *Set Zero Event Handling*, y el área de *Set Confidence Intervals* (figuras 25-28). En el centro de la ventana, encontrará el área *Meta-analysis Summary*.

4.3.1 **Elección de la medida de asociación**

El programa TSA proporciona las mismas medidas de efecto como el *Review Manager* versión 5 (ver sección 2.1.1. Para una descripción de estas medidas). Para seleccionar una medida de efecto, primero pulse en el cuadro desplegable llamado *Effect Measure* en el área denominada *Set Effect Measure and Model* con el fin de

visualizar las medidas del efecto que están disponibles (figura 25, área marcada en la imagen izquierda), a continuación, haga clic en la medida del efecto que desea utilizar para el metanálisis.

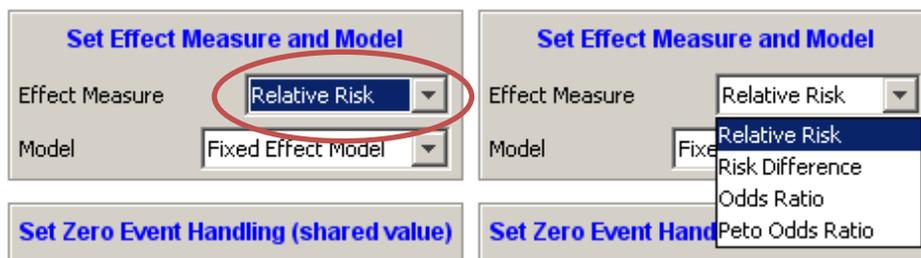


Figura 25. Selección de la medida del efecto haciendo clic en el cuadro desplegable *Effect Measure*.

4.3.2 La elección de su modelo estadístico

El programa TSA ofrece cuatro modelos estadísticos para combinar los datos del metanálisis - tres de los cuales son variantes del modelo de efectos aleatorio (ver sección 2.1.2). Para establecer el modelo estadístico, primero haga clic en el cuadro desplegable *Model* para visualizar las medidas disponibles de efectos (Figura 26, marcado en el área en la imagen izquierda) y, a continuación, haga clic en el modelo que desea utilizar para el metanálisis.

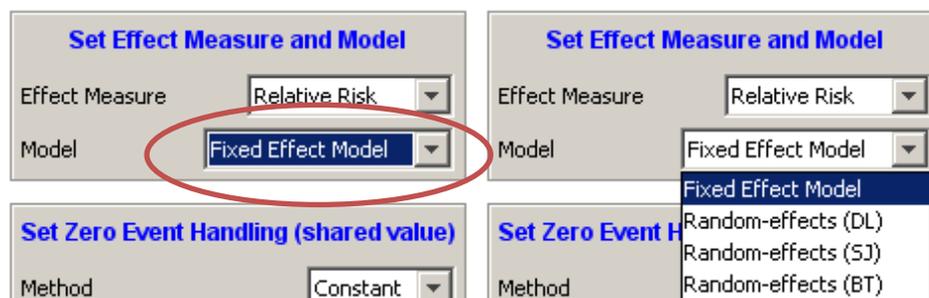


Figure 26. Elección de la medida del efecto en el cuadro desplegable *Model*.

4.3.3 Elección de un método para el manejo de datos con ausencia de eventos

El programa TSA ofrece tres métodos para el manejo de datos con ausencia de eventos. (véase la sección 2.1.4). Para seleccionar el método que desea utilizar para el manejo de esta situación, primero haga clic en el cuadro desplegable *Method* para mostrar los métodos disponibles de corrección de continuidad y, a continuación, haga clic en el método que desea utilizar (figura 27).

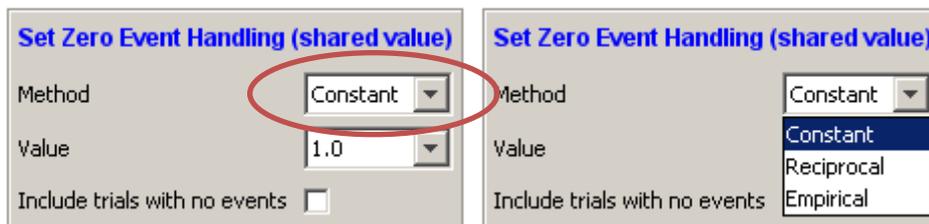


Figura 27. Selección del método de corrección de continuidad haciendo clic en el cuadro desplegable *Method*.

También es necesario establecer el factor de corrección de continuidad. En el programa TSA, los factores de corrección se derivan de la suma de los factores de corrección en los dos grupos (también conocida como el "Value"). Por ejemplo, la suma de los factores de corrección en la corrección de continuidad utilizada en el Review Manager es $1 = 0,5 + 0,5$ - porque 0,5 se añade a la serie de eventos en ambos grupos. Para establecer la suma de los dos factores de corrección, primero haga clic en el cuadro desplegable *Value*, seleccione la suma que desea para los dos factores de corrección que se añaden. Además, tiene la opción de aplicar la corrección de continuidad en los ensayos que no tienen eventos en ambos grupos. Para ello, marque la casilla *'Included trials with no events'*.

4.3.4 La elección del tipo de intervalo de confianza

El TSA ofrece una variedad de opciones para el tipo de intervalo de confianza que desea emplear (figura 28). Si está empleando intervalos de confianza convencionales puede elegir entre los niveles de cobertura del 95 %, 99 %, 99,5 % y 99,9 %. Para ello, active el botón de *'Conventional (coverage)'* botón ubicado a la izquierda en el área *Set Confidence Intervals*, haga clic en el menú desplegable a la derecha y seleccione su cobertura deseada.

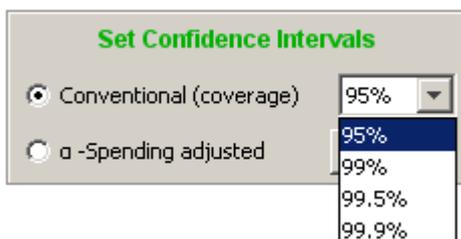


Figura 28. Elección de la cobertura de los intervalos convencionales de confianza.

Si ya ha construido límites ajustados de significación utilizando una función de gasto de alfa (véase la sección 2.2.4 y 4.4.1), también tendrá la opción de obtener el intervalo de confianza ajustado de gasto de alfa (véase la sección 2.2.5). Para ello, primero

haga clic en el botón de opción en el área ' α -Spending adjusted' y posteriormente, haga clic en el botón 'Select'.



Figura 29. Selección de los intervalos de confianza ajustados de la función gasto de alfa.

Una ventana emergente con una lista de los límites de gasto de alfa que han sido añadidos, deberían aparecer en el centro de la pantalla. Seleccione cuál de los límites de gasto de alfa en el cual debería basarse el ajuste y haga clic en el botón *Select* (figura 30). Tenga en cuenta que la cobertura acumulada de los intervalos de confianza ajustados del gasto de alfa corresponderá al nivel alfa establecido para la función de gasto de alfa seleccionada.

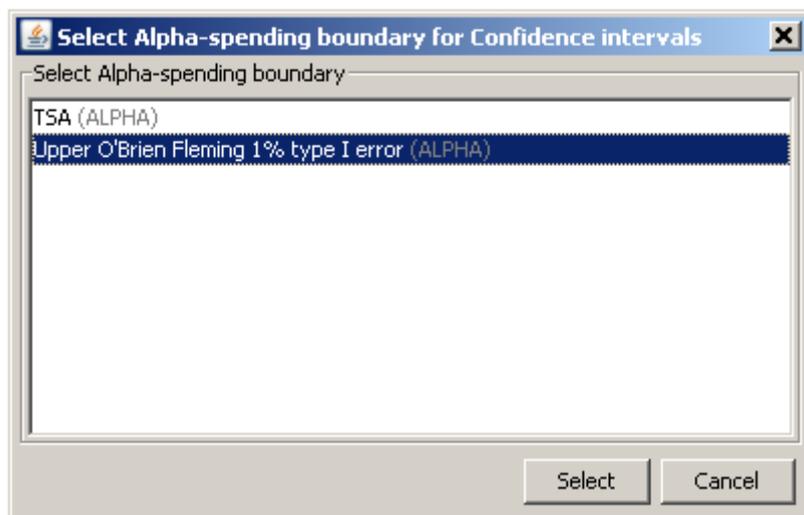


Figura 30. Elección de los límites de gasto de alfa sobre los cuales debe basarse el ajuste.

También tenga en cuenta que solo los límites de gasto de alfa que se han calculado y no han sido 'ignorados' serán incluidos en la lista (ver sección 4.4.1).

4.4 ***Aplicación de las pruebas ajustadas de significación (aplicando TSA)***

El TSA actualmente ofrece dos métodos para evaluar el ajuste de la significación ellos son el método de gasto de alfa de O'Brien-Fleming, descrito en la sección 2.2.4. y la ley del método del logaritmo iterado, descrito en el apartado 2.2.6. El TSA también

ofrece la opción de combinar el método O'Brien-Fleming con las pruebas de futilidad, como se describe en el apartado 2.2.7. Para utilizar estos métodos, haga clic en la pestaña TSA (a la derecha de la pestaña *Trials*), como se muestra en la figura 31.



Figura 31. Haga clic en la pestaña TSA cuando se quiere emplear los métodos para la evaluación de significación corregida.

4.4.1 Adición de una prueba de significación

En la parte superior izquierda de la ventana, se encuentra el área denominada *Add* (ver figura 32), que contiene los botones *Conventional Test Boundary*, *Alpha-spending Boundaries* y *Law of the Iterated Logarithm*. Al hacer clic en cualquiera de estos tres botones debería aparecer una nueva ventana en el centro de la ventana del programa TSA. Esta ventana debería contener una serie de campos, lo que permitirá definir los ajustes para el tipo de prueba de significación que desea utilizar.

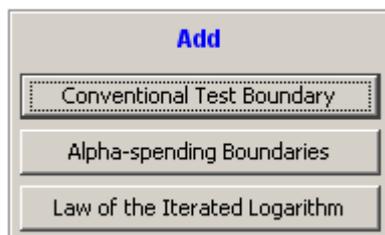


Figura 32. Haga clic en uno de los botones para añadir una nueva prueba de significación.

El límite convencional de significación

La opción *Conventional* permite añadir un límite para la curva Z que corresponde a una única prueba de significación con algún riesgo de máximo error tipo I, α . Por ejemplo, un límite convencional para una prueba de significación bilateral $\alpha = 5\%$ producirá dos líneas horizontales en 1,96 y -1,96. Cuando haga clic en el botón *Conventional* debería aparecer una ventana similar a la mostrada en la figura 33.

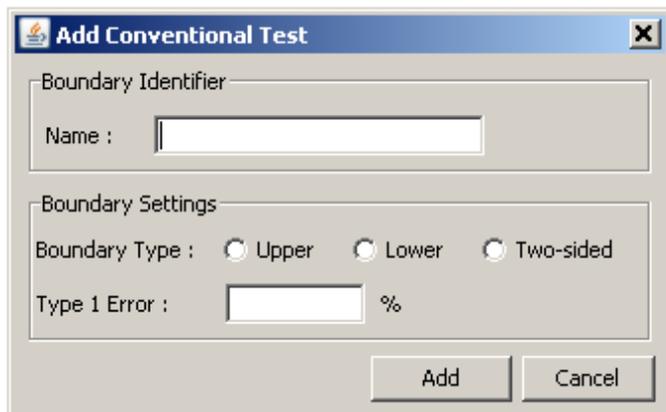


Figura 33. Ventana emergente que muestra la prueba convencional que aparece al hacer clic en el botón *Conventional Test Boundary*.

Usted tendrá que dar un nombre a la prueba convencional, por ejemplo, «Umbral de una prueba de 5 %», define si la prueba es bilateral (simétrica) o unilateral y lo que será, globalmente, el error máximo tipo I. Para las pruebas unilaterales, la prueba superior solo examinará la superioridad de la intervención experimental, mientras que el valor más bajo solo evaluará la superioridad de la intervención control. Para el metanálisis de datos binarios, cabe señalar que cuando el resultado se define como «beneficioso» en lugar de «perjudicial,» (véase sección 4.1.1) las funciones superior e inferior se invierten. Cuando usted haya dado un nombre al límite convencional y definido los ajustes, pulse el botón Añadir (Add) para añadir el límite.

Los límites de gasto de α

La opción de *gasto de alfa*- permite agregar límites de significación ajustados para la curva Z con el método de -gasto de α descrito en la sección 2.2.4. Debido a que el método de gasto de α no se puede aplicar sin determinar algún tamaño de la información requerida del metanálisis, los cálculos del tamaño de la información deben ser definidos de forma simultánea. Por lo tanto, la ventana de configuración de los límites de gasto de alfa para metanálisis de datos dicotómicos será diferente del metanálisis de datos continuos con respecto a los ajustes para el cálculo del tamaño de la información. Para el metanálisis de datos dicotómicos la ventana de configuración de los límites de gasto de alfa que aparecen al pulsar en el botón de gasto de α debería ser similar a la mostrada en la figura 34.

Add Dichotomous Alpha-spending Boundary

Boundary Identifier

Name:

Hypothesis Testing

Boundary Type: One-sided Upper One-sided Lower Two-sided

Type 1 Error: %

α -Spending Function:

Information Axis: Sample Size Event Size Statistical Information

Inner Wedge

Apply Inner Wedge:

Power: %

β -Spending Function:

Required Information Size

Information Size: User Defined Estimate

Type 1 Error: %

Power: %

Relative Risk Reduction: % User Defined Low Bias Based

Incidence in Intervention arm: % User Defined

Incidence in Control arm: %

Heterogeneity Correction: % User Defined Model Variance Based

Add Cancel

Figura 34. Ventana de configuración de los límites de gastos de alfa para el metanálisis de datos dicotómicos que aparece al pulsar en el botón de *alfa-spending*.

En el caso de metanálisis de datos continuos, la ventana de configuración de los límites de gasto de alfa- que aparecen al pulsar en el botón de gasto de alfa debería ser similar a la mostrada en la figura 35.

Add Continuous Alpha-spending Boundary

Boundary Identifier

Name:

Hypothesis Testing

Boundary Type: One-sided Upper One-sided Lower Two-sided

Type 1 Error: %

α -spending Function:

Information Axis: Sample Size Event Size Statistical Information

Inner Wedge

Apply Inner Wedge:

Power: %

β -spending Function:

Required Information Size

Information Size: User Defined Estimate

Type 1 Error: %

Power: %

Mean Difference: User Defined Empirical Low Bias

Variance: User Defined Empirical Low Bias

Heterogeneity Correction: % User Defined Model Variance Based

Add Cancel

Figura 35. Ventana de configuración de los límites de gastos de alfa para el metanálisis de datos continuos que aparece al pulsar en el botón de *alfa-spending*

En primer lugar, se requiere dar un nombre a la prueba basada en el gasto de α - (por ejemplo, «5% simétrica O'Brien-Fleming»). A continuación, tendrá que definir si desea emplear una prueba bilateral (simétrica) o unilateral, lo que será su error global máximo Tipo I, qué tipo de gasto de α desea emplear (en la actualidad solo está disponible la función O'Brien-Fleming). A continuación, tendrá que decidir si desea definir la información en su metanálisis como el número acumulado de pacientes

(sample size), el número acumulado de eventos (event size²), o la información estadística acumulada. Una vez más, para las pruebas unilaterales

la prueba unilateral superior examinará la superioridad de la información experimental, mientras que la prueba unilateral inferior solo examinará la superioridad de la intervención del grupo control. Para el metanálisis de datos binarios, cabe señalar que cuando el resultado se define como beneficioso en lugar de un resultado perjudicial (véase sección 4.1.1), las funciones superior e inferior se invierten.

Para evaluar la prueba de la futilidad revise la casilla '*Apply inner wedge*'. El error de tipo II (potencia) Para los límites de futilidad se ajustarán automáticamente cuando introduzca los ajustes para el cálculo del tamaño de la información (véase más adelante). Actualmente, la única función de gasto de β - disponible en el TSA es la función de O'Brien-Fleming.

Se requerirá introducir los componentes necesarios para el cálculo del tamaño de la información requerida. Usted tendrá la opción de definir la información necesaria como cualquier número arbitrario que pueda haber obtenido independiente del software TSA. Para enviar su propio valor para IS, marque el botón de '*User Defined*' y escriba en la casilla '*Information Size*'. Usted también tiene la opción para calcular la IS requerida de acuerdo con los métodos delineados en la sección 2.2.1. Al utilizar TSA para el cálculo del tamaño de la información requerida, pulse el botón '*Estimate*'. La estimación requerida será automáticamente generada a partir del tipo de información que se está acumulando. Por ejemplo, si ha seleccionado "*Sample size*" en "*Information Axis*", el tamaño de la información requerida será el generado como el número necesario de pacientes en el metanálisis.

El cálculo del tamaño de la información se basa automáticamente en el máximo error tipo I que definió para el límite gasto de α -, pero tendrá que introducir el error máximo tipo II. o potencia mínima deseada (1-error de tipo II) en el campo de entrada "Power".

Existen dos opciones para ajustar el tamaño de la información requerida para la heterogeneidad en el metanálisis. La primera opción se basa en el ajuste de la heterogeneidad en la relación estimada entre la varianza en el modelo de efectos aleatorios seleccionados y la varianza en el modelo de efectos fijos (ver sección 2.2.1). Para utilizar esta opción, marque el botón de opción '*Model Variance Based*'. Tenga en cuenta que si ha seleccionado el modelo de efectos fijos, este factor de ajuste es siempre igual a 1, y por lo tanto, no se aplica ningún ajuste.

² NT: se ha mantenido el nombre original de los botones del software

La segunda opción es hacer una estimación de la heterogeneidad fijada por el usuario. Cuando el metanálisis incluye un número insuficiente de ensayos para calcular un estimado fiable del factor de ajuste, se puede ajustar *a priori* el tamaño de la información requerida para algún grado de heterogeneidad anticipada verosímil o máxima. Para utilizar esta opción, marque la opción "*User Defined*" y escriba la heterogeneidad máxima prevista en el campo de entrada a la izquierda. La heterogeneidad se define como el porcentaje de la variación total en el metanálisis, que se explica por la variación entre ensayos en lugar de la variación dentro del ensayo. De este modo, un ajuste definido por el usuario de 50%, por ejemplo, genera un tamaño de la información necesaria que permite la inferencia confiable cuando aproximadamente la mitad de la variación total entre los ensayos en el metanálisis se explica por la variación entre ensayos.

Para establecer las tasas de eventos previstos y efecto de la intervención de un metanálisis de datos dicotómicos, solo tiene que rellenar dos de los tres campos: "*Relative risk reduction*", "*Incidence in Intervention Group*", y "*Incidence in Control Group*". Si ha categorizado algunos de sus ensayos incluidos como ensayos con bajo riesgo de sesgo, es posible utilizar el estimado obtenido del metanálisis de estos ensayos como la reducción relativa del riesgo anticipada. Para utilizar esta opción, seleccione opción '*Low--bias Based*'.

Para establecer la diferencia prevista de media y la varianza de un metanálisis de datos continua, solo tiene que rellenar dos campos: "*Mean Difference*" y '*Variance*'. Si ha categorizado algunos de sus ensayos incluidos como ensayos de bajo riesgo de sesgo, es posible utilizar los estimados del metanálisis de esos ensayos como la prevista diferencia de medias y la varianza, seleccionando de nuevo la opción '*Low-bias Based*'. También tiene la opción de utilizar el estimado combinado de todos los ensayos incluidos (sin tener en cuenta el riesgo de sesgo) como la varianza esperada. Para utilizar todos los ensayos, seleccione la opción '*Empirical*'.

Cuando haya delimitado los límites del gasto de alfa, definido los ajustes de la prueba de hipótesis y los parámetros para el cálculo del tamaño de la información, pulse el botón "Add" para añadir los límites.

Después de añadir los límites de gasto de alfa- tendrá que definir cuándo el metanálisis fue sometido a pruebas de significación. Vaya a *Interim analyses* ubicado a la derecha de la lista de pruebas de significación ajustadas y active (o desactive) los ensayos después de que las pruebas de significación fueron realizadas. En la figura 36, los ensayos 2, 4, y 5 han sido activados, y los ensayos 1 y 3 han sido desactivados, lo que significa que tres metanálisis (incluyendo pruebas de significación) se realizaron a través del tiempo: uno que incluye el ensayo 1 y 2, uno los ensayos 1 a 4, y otro los

1 a 5 Tenga en cuenta que la última prueba en la lista siempre debe estar activado, ya que esto representa la prueba de significación que están empleando en todos los ensayos incluidos.

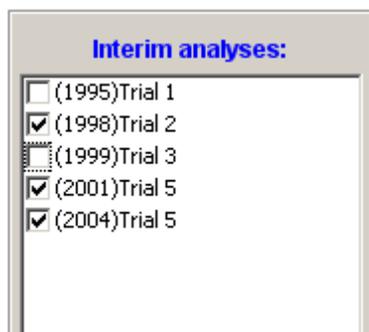


Figura 36. Ventana emergente de la configuración de los límites de gasto de alfa para el metanálisis de datos continuos que aparece al hacer clic en el botón de *alpha-spending*.

En algunos casos, puede desear activar o desactivar todos los ensayos para las pruebas de significación previas. Pulse el botón 'Select none' en la parte inferior del área *Interim analyses* para desactivar todos los ensayos, o pulse en el botón 'Select all' para activar todos los ensayos (figura 37). Además, usted tiene la opción de invertir la selección de análisis provisionales en el área *Interim analyses*.



Figura 37. Active o desactive todos los ensayos para las pruebas de significación anteriores.

La curva Z penalizada con la ley del logaritmo iterado

La opción de la ley de logaritmo iterado le permite realizar pruebas de significación ajustado por penalización de la curva Z con los métodos descritos en la sección 2.2.6. Al pulsar en el botón *Law of Iterated Logarithm*, debería aparecer una ventana similar a la mostrada en la figura 38.

Necesitará dar un nombre a la penalización de la curva Z (por ejemplo, '*5% symmetric LIL*'), definir si la prueba es de bilateral o unilateral, lo que será su error global máximo Tipo I, y establecer el parámetro de penalización, λ (véase la sección 2.2.6 y el cuadro 2). Para las pruebas unilaterales, la prueba unilateral superior examinará la

superioridad de la información experimental, mientras que la prueba unilateral inferior solo examinará la superioridad de la intervención del grupo control. Para el metanálisis de datos binarios, cabe señalar que cuando el resultado se define como beneficioso en lugar de un resultado perjudicial (véase sección 4.1.1), las funciones superior e inferior se invierten.

Figura 38. Ventana emergente estableciendo la penalización por la ley del logaritmo iterado que aparece pulsar el botón de la *Law of Iterated Logarithm*.

4.4.2 Edición y eliminación de una prueba de significación

Cada vez que se añade una prueba de significación, esta debe aparecer en el centro de la pantalla. Cada prueba de significación estará representada por una fila como se muestra en la figura 39.

Identifier	Ignore	Type
TSA (ALPHA)	<input type="checkbox"/>	Alpha-spending
Conventional 2-sided (CON)	<input type="checkbox"/>	Conventional
LIL 5% (LIL)	<input type="checkbox"/>	Law of Iterated Logarithm

Figura 39. Lista de las pruebas de significación agregada.

Para editar una prueba de significación, primero seleccione la fila para la prueba que desea editar; luego, pulse el botón '*Edit selected*' en la zona *Edit* (ver figura 40). Alternativamente, usted puede simplemente hacer doble clic en la fila para la prueba que desea editar.

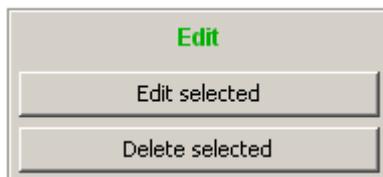


Figura 40. Zona para editar/eliminar las pruebas de significación seleccionadas

Ahora aparecerá la ventana emergente con la configuración de las pruebas. Realice los cambios y pulse en el botón 'Apply changes' en la esquina inferior derecha de la ventana emergente.

Si desea eliminar una prueba, seleccione la fila de la prueba que desea eliminar y pulse el botón *Delete selected* en el área de *Edit*. Alternativamente, usted puede seleccionar la fila para la prueba que desea eliminar y pulse el botón <Delete> en el teclado.

4.4.3 Adición y recuperación de las plantillas de pruebas de significación

El área *Templates* ubicada en la esquina inferior izquierda del TSA le ofrece la opción de guardar las pruebas de significación construidas, y recuperar las pruebas de significación previamente construidas (figura 41). Si desea reutilizar una prueba de significación de otros metanálisis, puede guardar esto en sus plantillas y recuperarla en cualquier otro momento. Para guardar una prueba de significación construida como una plantilla, seleccione la fila para la prueba que desea guardar y pulse el botón *Save as template*. Para recuperar una plantilla, primero pulse en el botón *Manager templates*.



Figura 41. Área de Plantilla donde se puede cargar y recuperarr (añadir) pruebas de significación construidas.

Una ventana emergente aparecerá en el centro de la ventana del programa TSA (figura 42). La lista de plantillas disponibles se muestra a la izquierda.

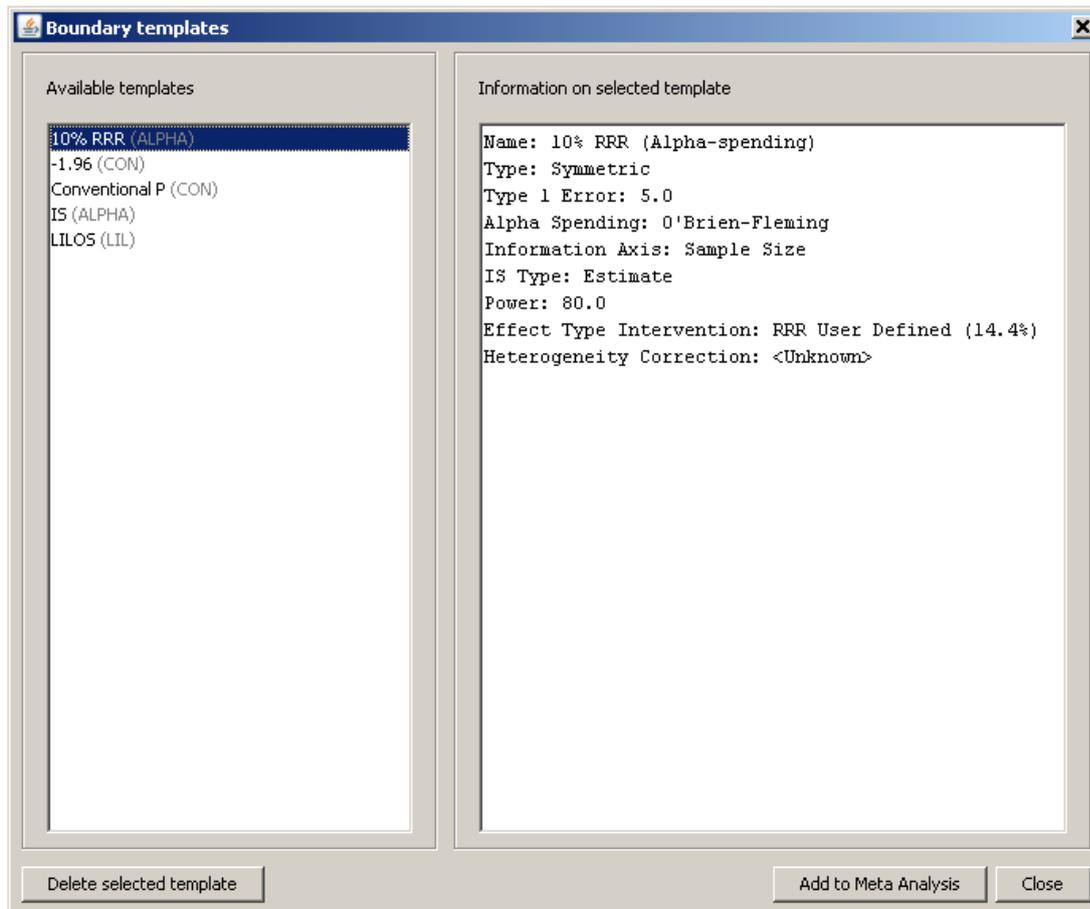


Figura 42. Ventana de plantillas. La prueba de significación “10% RRR “ha sido seleccionada y las configuraciones de esta prueba se muestran en” Information on selected boundary”.

Puede pulsar sobre un título de la plantilla para mostrar las configuraciones de las pruebas de significación disponibles en el lado derecho. Para recuperar la plantilla de una prueba de significación de un metanálisis, seleccione la plantilla que desea recuperar y pulse en el botón “Add to Meta-analysis”. Si desea eliminar de forma permanente una de las plantillas disponibles, seleccione la plantilla que desea eliminar y pulse el botón “Delete selected Template”.

4.4.4 Realizando los cálculos de las pruebas de significación

Una vez que haya agregado todas las pruebas de significación que desea utilizar, necesita el programa TSA para llevar a cabo los cálculos necesarios. Para ello, pulse en el botón “Perform calculation” en el área “Calculations” en zona *Edit*. Dependiendo del número de pruebas de significación que ha añadido, el programa TSA pudiera tardar unos segundos para completar los cálculos. Los límites O’Brien-Fleming (gasto de alfa) con muchos análisis intermedios pueden tardar 5-10 segundos por grupos de límites para completar los cálculos.

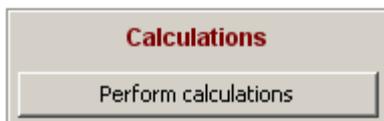


Figura 43. Botón para realizar los cálculos.

En algunos casos, hay un pequeño aumento relativo en la información entre dos análisis intermedios que los cálculos numéricos (integración numérica de probabilidades de una cola extremadamente pequeña) rompen los límites de gasto de alfa-. Por ejemplo, si el tamaño de la información requerida es de 20.000 pacientes y los análisis intermedios se realizan después de cada ensayo, la adición de un nuevo ensayo con 40 pacientes solo proporcionaría un incremento de 0,2 % en la fracción de información acumulada. Para evitar rupturas en los cálculos, el TSA elimina automáticamente (desactiva) los análisis intermedios que corresponden a un incremento de la fracción de información del 1 % o menos. Cuando esto sucede, una ventana emergente se abrirá automáticamente en el centro de la ventana del TSA para informar que se eliminaron análisis intermedios (figura 44). Los datos de estos ensayos son, sin embargo, retenidos en su metanálisis TSA y en el valor Z-acumulado.



Figura 44. Ventana emergente que informa cuáles fueron los análisis provisionales eliminados. Los datos de estos ensayos son, sin embargo, retenidos en metanálisis TSA y en el -valor Z acumulado.

Si ha agregado más de una prueba de significación y no desea llevar a cabo los cálculos de todos ellos, tiene la opción de ignorar las pruebas de significación. Para ignorar una prueba de significación, marque la casilla en la columna del medio de la fila correspondiente a la prueba de significación que desea ignorar.

Identifier	Ignore	Type
TSA (ALPHA)	<input type="checkbox"/>	Alpha-spending
Conventional 2-sided (CON)	<input checked="" type="checkbox"/>	Alpha-spending boundary ignored by user
LIL 5% (LIL)	<input type="checkbox"/>	Law of Iterated Logarithm

Figura 45. Ejemplo de una prueba de significación ignorada ('prueba bilateral convencional ignorada).

La -curva Z acumulada y los límites de significación (funciones de gasto de alfa) se pueden visualizar utilizando una de las tres variables en el eje x: tamaño de la muestra, tamaño de evento, o la información estadística. Las pruebas de significación definidas en diferentes escalas no se pueden mostrar simultáneamente en un gráfico, por lo que necesita seleccionar una de estas variables para el análisis total. Marque el botón en el área *Information Axis* debajo del área *Calculation* (figura 46).

Information axis

Sample size

Event size

Statistical information

Figura 46. Botones para elegir la escala de información en la cual se muestra la prueba de significación acumulada.

Si una o dos de las tres escalas (tamaño de la muestra, tamaño de evento o información estadística) no han sido seleccionados en cualquiera de los límites de -gasto de alfa añadidos, automáticamente aparecerán atenuados en el área *Information Axis*.

4.5 Opciones de gráficos en el TSA

La opción *Graph* en el programa TSA permite mostrar la -curva Z y las pruebas de significación construidas en relación con la fortaleza de la evidencia (es decir, el número acumulado de pacientes, eventos, o la información estadística). También ofrece una serie de opciones de edición de gráficos que pueden ser útiles cuando se están preparando los gráficos para los manuscritos de las investigaciones. Para ir a la opción *Graph*, pulse en la pestaña *Graph* (a la derecha de la pestaña *TSA*) como se muestra en la figura 47.



Figura 47. Pulse en la pestaña *Graph* para ver la -curva Z y las pruebas de significación construidos en relación con la fortaleza de la evidencia.

En el lado izquierdo de la ventana del programa TSA se encuentra el área de *Test and boundaries Layout*, el *Set Graph Layout*, y dos opciones de impresión ("Print current graph" y "Generate TSA Report"). A la derecha de estas áreas se encuentra el gráfico que muestra la -curva Z y las pruebas de significación construidas.

En el área *Test and boundaries layout*, encontrará una gran cantidad de opciones de edición de gráficos que le permiten cambiar la presentación de la -curva Z y las pruebas de significación construidas (límites). Las pruebas de significación construidas y la curva Z se mostrarán en el área blanca, véase la figura 49. Para cambiar la presentación de uno de estos, primero seleccione una de las pruebas (curvas) en la lista y edite de acuerdo con sus preferencias.

En el programa TSA, tendrá la opción de editar el color, el tipo de línea, el tipo y tamaño del icono que se muestra en cada ensayo o análisis intermedio, así como el tamaño y la fuente de la prueba asociada con una curva o una prueba. También tiene la opción de ocultar una curva o prueba de la gráfica.

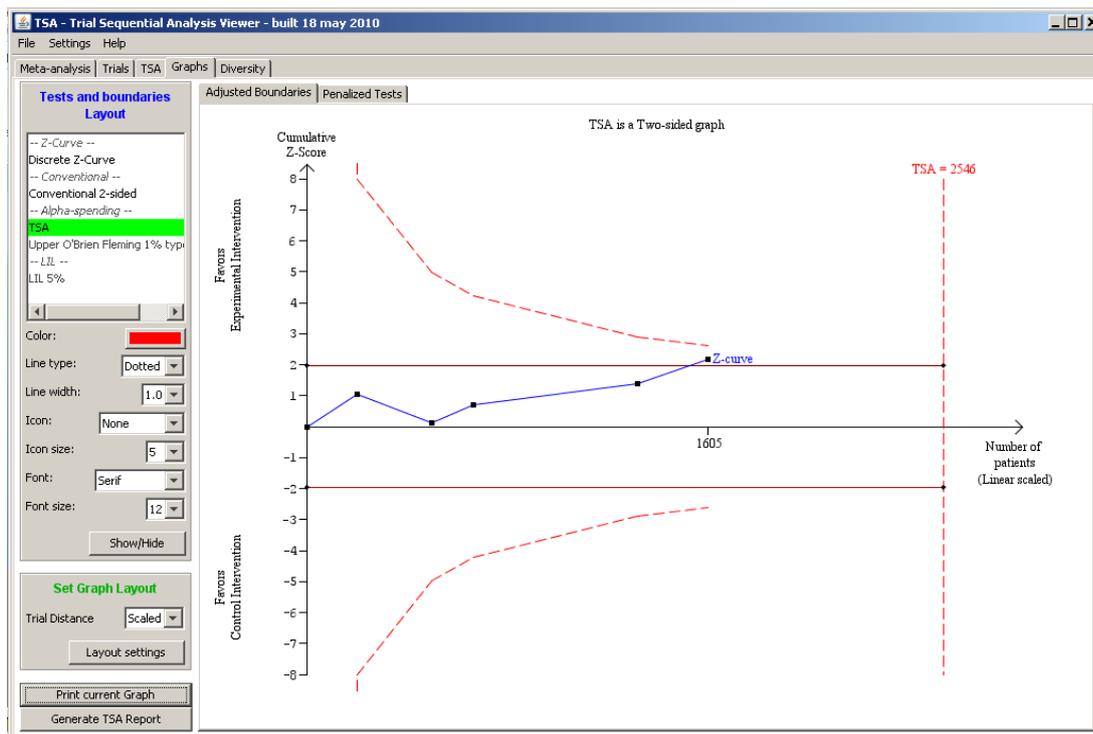


Figura 48. La ventana de gráficos.

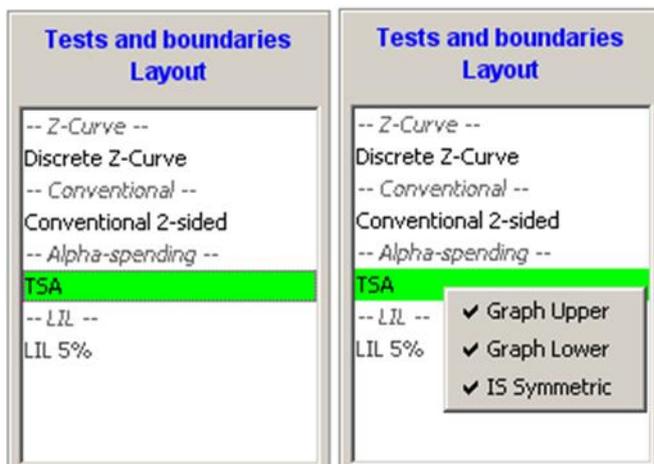


Figura 49. Área de la pruebas de significación y de la curva Z.

Los objetos gráficos que representan las pruebas de significación son una colección de varios objetos. Por ejemplo, un objeto gráfico que representa los límites bilaterales de gasto de alfa contendrá los límites superior e inferior y la línea vertical superior e inferior para el tamaño de la información requerida. Si desea ocultar algunos de estos objetos, pulse en la prueba de significación y desmarque los objetos que desea ocultar (figura 49, la imagen de la derecha).

En el área *Set Graph Layout*, encontrará un número de opciones para cambiar la presentación general del gráfico. Al pulsar el botón "Layout setting", aparecerá una ventana emergente (figura 50) la cual muestra las opciones de ajuste del grosor del eje X y el Y, el tamaño de las fuentes de las coordenadas y la fuente y tamaño de los componentes del texto suministrados en el gráfico.

En el área Conjunto Graph Layout, usted encontrará una serie de opciones para cambiar la presentación general gráfico. Si hace clic en el botón "Configuración del diseño", aparecerá una ventana emergente (figura 50), que le proporciona las opciones de ajuste de la anchura del eje X y el eje Y, la coordenada tamaño de fuente, o la fuente y el tamaño de los componentes de texto fijo en el gráfico.

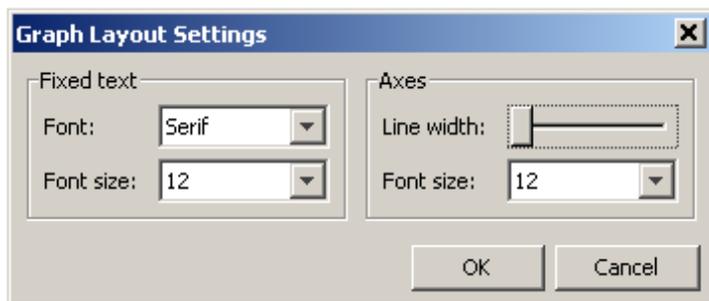


Figura 50. Configuración general del diseño de los gráficos para ajustar la fuente de los componentes fijos del texto y tamaño de la fuente, el ancho del eje X y el eje Y, y el tamaño de la fuente de las coordenadas.

En el eje de la información, la distancia entre los límites y la distancia entre los valores Z se muestran convencionalmente con respecto al aumento relativo de la información. El programa TSA muestra automáticamente estas distancias en escala.

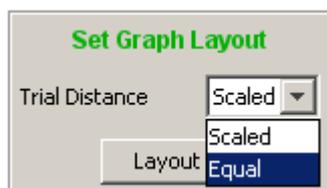


Figura 51. Seleccione 'Equal' para igualar la distancia entre los ensayos en el eje de las X.

En algunos casos, sin embargo, otras disposiciones pueden proporcionar una mejor base para la interpretación visual. El programa TSA también proporciona el formato de diseño utilizado en el documento de Pogue y Yusuf, que muestra los ensayos a igual distancia en el eje de la información y muestra los títulos de los ensayos en un ángulo de 45° por debajo del eje x. Para elegir este formato de diseño, haga clic en el desplegable 'Trial Distance' en la casilla ubicada en el área *Set Graph Layout* y seleccione "Equal" (figura 51).

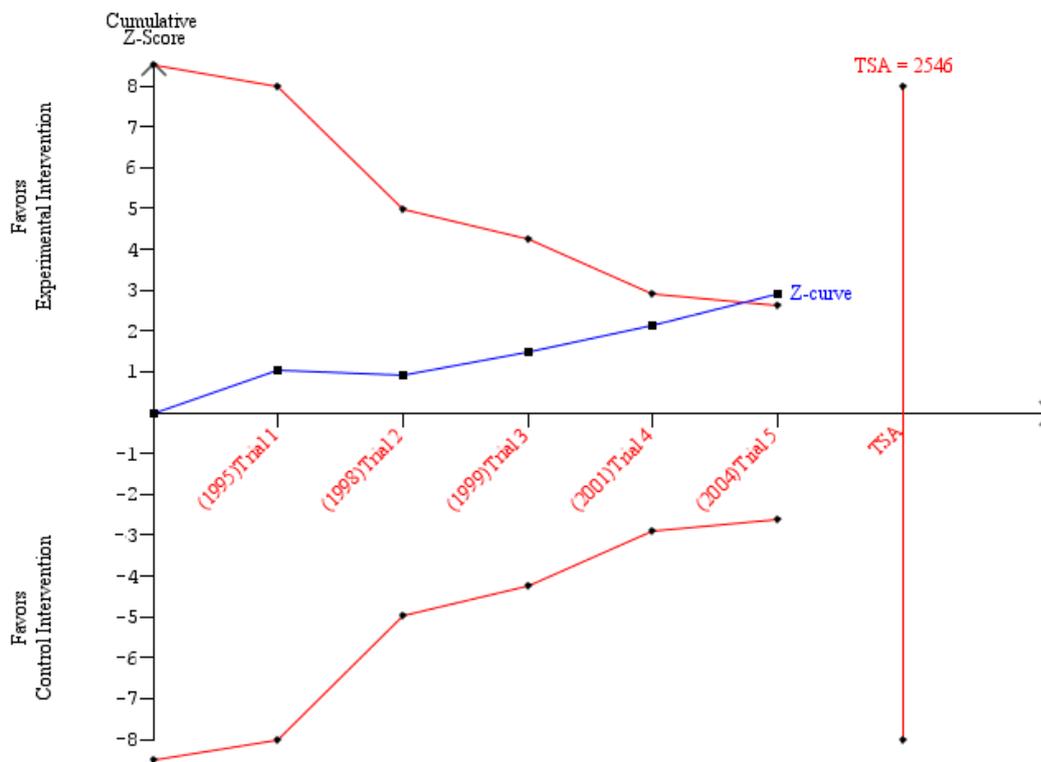


Figura 52. Gráfico que muestra el formato seleccionado para la información de los ejes utilizando la opción escala.

Las pruebas de significación ajustada basadas en las funciones de -gasto de alfa son, en efecto, los umbrales ajustados de la curva Z, mientras que las pruebas de significación ajustada basadas en las sanciones de la ley del logaritmo iterado son, en efecto, las pruebas estadísticas ajustadas que debe interpretarse en relación con los umbrales de las pruebas de significación de cada estudio. Por lo tanto, la combinación de estas dos estrategias en un gráfico no es significativa. El programa TSA proporciona gráficos separados para las pruebas de significación ajustados basados en las funciones de gasto de alfa y de las sanciones de la ley de logaritmos iterados. Para ver la representación gráfica de los límites de gastos de alfa calculados, seleccione la pestaña *Adjusted Boundaries* ubicada la parte superior la gráfica. Para ver la representación gráfica las sanciones calculadas de la ley de logaritmos iterados, seleccione la pestaña *Penalised Tests* ubicada en la parte superior de la gráfica (figura 53).

Adjusted Boundaries Penalised Tests

Figura 53. Visualización del gráfico de los límites de la prueba y las pruebas penalizadas

4.6 Explorando la diversidad entre los ensayos

El programa TSA también proporciona una opción para explorar la diversidad de los estimados y la comparación de los pesos en los tres modelos de efectos aleatorios: DL, SJ, y BT. Estas opciones están disponibles en la pestaña *Diversity* (Figura 54)



Figura 54. Pulse la pestaña *Diversity* para explorar la diversidad de los estimados y comparar los pesos a través de los modelos de efectos aleatorios.

Después de pulsar en la pestaña *Diversity*, debería aparecer una pantalla similar a la mostrada en la figura 55. En la parte superior de la pantalla, los pesos y sus respectivos porcentajes de cada ensayo (filas), utilizando cada uno de los modelos disponibles (columnas), se muestran en la esquina inferior izquierda. Las siguientes informaciones se muestran para cada uno de los tres modelos de efectos aleatorios: la estimación de la inconsistencia I^2 y su correspondiente corrección de heterogeneidad $1/(1-I^2)$, la estimación de la diversidad D^2 y su correspondiente corrección de heterogeneidad $1/(1-D^2)$, y la estimación de la varianza entre los ensayos, τ^2 . La estimación de inconsistencia solo se muestra para el modelo DL. Tenga en cuenta que la estimación de la varianza entre los ensayos es la misma para los modelos DL y BT (ver sección 2.1.3). En la esquina inferior derecha, hay una opción para elegir el número de decimales con los que se debe mostrar todas las cantidades. Pulse en la ventana desplegable para seleccionar el número de decimales.

TSA - Trial Sequential Analysis Viewer - built 18 may 2010

File Settings Help

Meta-analysis Trials TSA Graphs Diversity

Trial	Fixed Weight	Random(DL) Weight	Random(SJ) Weight	Random(BT) Weight
(1995)Trial 1 (10.0/100.0 15.0/102.0)	4.5 [14.43%]	0.14 [10.33%]	7.43 [14.43%]	0.1 [14.45%]
(1998)Trial 2 (27.0/150.0 22.0/148.0)	6.92 [22.19%]	0.22 [15.4%]	11.07 [22.19%]	0.15 [22.19%]
(1999)Trial 3 (7.0/80.0 14.0/85.0)	3.77 [12.08%]	0.12 [9.44%]	6.79 [12.08%]	0.09 [12.11%]
(2001)Trial 4 (28.0/340.0 40.0/320.0)	7.69 [24.65%]	0.25 [28.66%]	20.61 [24.65%]	0.29 [24.64%]
(2004)Trial 5 (25.0/140.0 52.0/140.0)	8.3 [26.64%]	0.27 [36.16%]	26.0 [26.64%]	0.36 [26.61%]

Diversity Type	Random DL	Random SJ	Random BT
I ²	0.49	0.49	Not applicable
I / (1 - I ²)	1.95	1.95	Not applicable
D ²	0.54	0.54	0.38
I / (1 - D ²)	2.17	2.17	1.61
Tau	0.08	0.08	0.08

Decimals: 2

Figura 55. Pestaña *Diversity*.

5 Ejemplos de aplicaciones del TSA

5.1 Los conjuntos de datos

Para ilustrar las aplicaciones del TSA, utilizamos los datos de varias revisiones sistemáticas publicadas. Algunos de los análisis y las aplicaciones que se presentan en este capítulo son nuestras propias modificaciones y adiciones a las que se pueden encontrar en la publicación original.

5.2 Cómo evitar los falsos positivos

En este ejemplo, hemos utilizado los datos de una revisión que comparó las tasas de abandono del hábito de fumar en pacientes que recibieron asistencia hospitalaria y con un seguimiento de menos de 1 mes contra pacientes que no recibieron la intervención.⁶⁴ En la revisión sistemática, las intervenciones y la duración del seguimiento diferían sustancialmente a través de los ensayos incluidos. Por tanto, los autores utilizaron las siguientes categorías de intensidad de la intervención:⁶⁴

1. Contacto individual en el hospital de duración ≤ 15 minutos, sin apoyo de seguimiento.
2. Uno o más contactos en el hospital de duración > 15 minutos, sin apoyo de seguimiento.
3. Cualquier contacto del hospital y tras un seguimiento ≤ 1 mes.
4. Cualquier contacto del hospital y tras un seguimiento > 1 mes.

El metanálisis de la intervención con intensidad número 3 incluyó seis ensayos, 4476 pacientes y 628 eventos. El modelo de efectos fijos produjo un riesgo relativo combinado de 1,05 (IC 95%: 0,91 a 1,21) (el metanálisis de los odds ratios mostró un resultado similar). La inconsistencia estimada (I^2) fue de 9 %, y la diversidad estimada (D^2) fue de 10 %. Se realizó el análisis secuencial de los ensayos de manera retrospectiva, mediante un metanálisis convencional de datos acumulados cada vez que un nuevo ensayo era publicado. El primer ensayo publicado produjo un riesgo relativo de 1,47 (IC 95 %: 1,05 a 2,05). Después del segundo ensayo el riesgo relativo combinado fue (IC 95 %: 1,02 a 1,75) 1,33. El metanálisis que comparó la intervención de la intensidad de la categoría 3 (véase arriba) con el control fue, por tanto, estadísticamente significativa después de los dos primeros ensayos.

Se realizó el TSA en estos datos. Se calculó el tamaño de la información requerida para demostrar o rechazar un incremento del beneficio relativo de 20 % (siendo el dejar de fumar un resultado benéfico). Asumimos como proporción del evento en el grupo control de 14 %, que era más o menos la mediana y la media de la proporción del evento del grupo control. Se utilizó un error tipo I de 5 % y un error e tipo II de 20 %. No hubo corrección de la heterogeneidad. Con estos parámetros, se determinó que 5.218 pacientes era el tamaño de la información requerida. Como el número de pacientes incluidos en el metanálisis no excedió el tamaño de la información requerida, los límites de futilidad demostraron una firme conclusión «negativa».

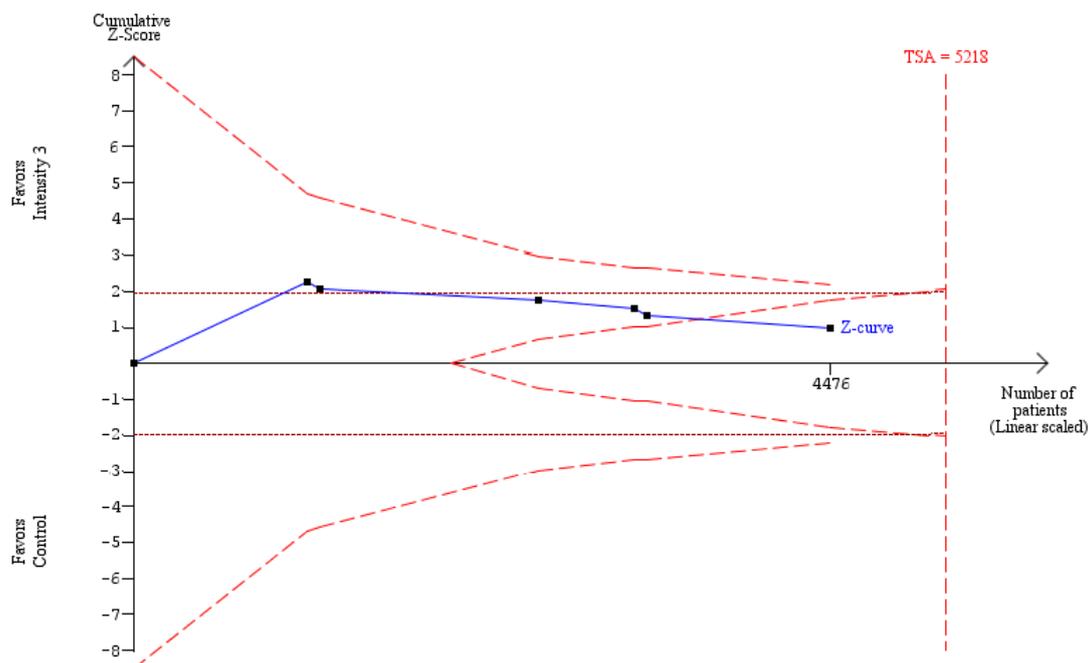


Figura 56. El tamaño de la información requerida para demostrar o rechazar un aumento relativo de 20% en el beneficio de dejar de fumar con una proporción de 14 % en el grupo control, un alfa de 5 % y un beta de 20 % es 5.218 pacientes (línea vertical roja). Las líneas rojas discontinuas representan los límites de monitorización secuencial de los ensayos y los límites de futilidad. La línea azul es la curva Z-acumulada.

El análisis secuencial de los ensayos resultante se muestra en la figura 56. Después del primero y segundo ensayo, la curva Z- acumulada cruzó por encima de 1,96, que corresponde al umbral nominal para la significación estadística, usando técnicas convencionales. Desde el tercer ensayo en adelante, el metanálisis ya no era estadísticamente significativo. Cuando se añadió el último ensayo, el metanálisis cruzó por debajo de los límites de futilidad, demostrando con una potencia de 80 % que el efecto de la intervención con intensidad 3 no es mayor que un aumento relativo del 20 % en dejar de fumar. Es decir, dentro de los supuestos establecidos para la confianza y el tamaño del efecto, esta intervención es ineficaz.

5.3 **Confirmación de un resultado positivo**

Para ilustrar la utilización del TSA para confirmar resultados positivos, se utilizaron los datos de una revisión sistemática que comparó la intervención quirúrgica de derivación aorto-coronaria con y sin circulación extracorpórea (CABG, por sus siglas en inglés).⁶⁵

Para este ejemplo, los límites de significación ajustados para la -curva Z acumulada se construyeron bajo el supuesto de que las pruebas de significación pudieron haber realizado cada vez que un nuevo ensayo se añadió al metanálisis. Dada la considerable cantidad de atención que se ha dado en la última década al debate de utilizar o no circulación extracorpórea, esta hipótesis parecía razonable. Se utilizó los límites de supervisión basado en la función gasto de alfa- tipo O'Brien-Fleming, que son relativamente insensibles al número de pruebas de significación repetidas (véase la sección 2.2.3).

En los conjuntos de datos de metanálisis considerado, hubo algunos años en que se publicó más de un ensayo. Para estos años, hemos analizado los ensayos en orden alfabético, según el apellido del primer autor.

5.3.1 Confirmación de la **“answer is in”** Desconozco una traducción adecuada para esta expresión

Para ilustrar la aplicación del TSA para afirmar **“answer is in”**, se utilizó el desenlace fibrilación auricular en este metanálisis. La aparición de la fibrilación auricular se informó en 30 ensayos, el cual incluyó 3.634 pacientes (con dos ensayos que no reportaron eventos).⁶⁵ De acuerdo con las normas convencionales para las pruebas de significación, intervención quirúrgica aorta coronaria sin circulación extracorpórea fue significativamente superior a la misma con bomba para reducir la fibrilación auricular (RR 0,69; IC 95 %: 0,57 a 0,83) (Figura 57). La inconsistencia estimada fue I^2 de 47,3 %, y la diversidad estimada fue D^2 de 49,0 %.

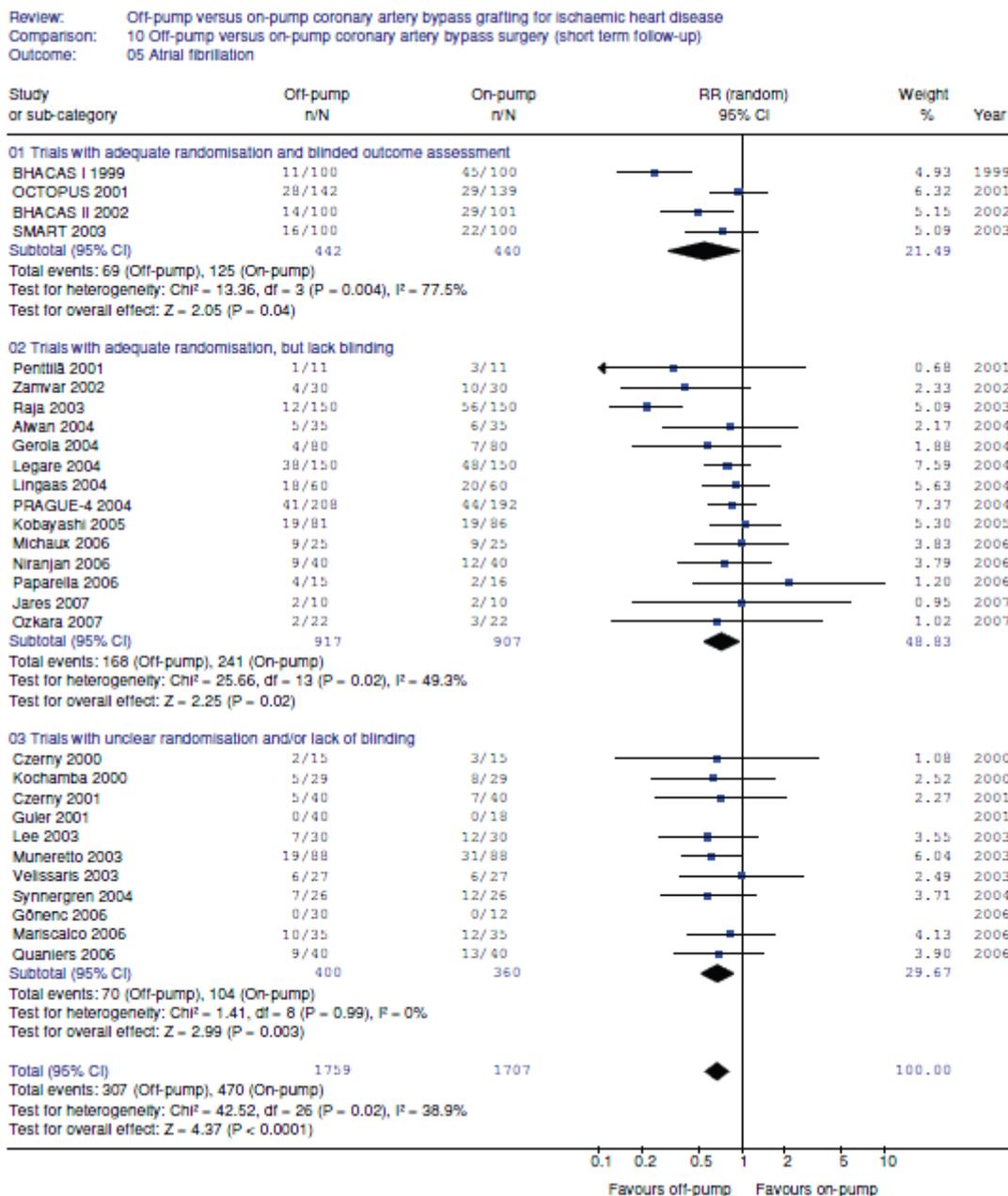


Figura 57. Forest plot del efecto de la derivación aorto coronaria con y sin circulación extracorpórea sobre la frecuencia de fibrilación auricular.

En el metanálisis de ensayos con bajo riesgo de sesgo (1.050 pacientes), el efecto no fue significativo (RR 0,63, IC 95 % 0,35 a 1,13), la heterogeneidad estimada fue I^2 de 77 %, y la diversidad estimada fue D^2 de 79,0 %.

Análisis secuencial de los ensayos de la fibrilación auricular

Se calcularon dos tamaños de información requerida para este ejemplo. En primer lugar, se calculó el tamaño de la información requerida para demostrar o rechazar un efecto de la intervención prevista *a priori* de una reducción relativa del riesgo del 20 %, alfa del 1 % y beta del 10 %, la cual fue de 7.150 pacientes. El valor del efecto de la intervención anticipada de 20 % fue elegido porque se creía que representaba un efecto de la intervención razonable en esta situación clínica. En segundo lugar, se calculó el tamaño de la información para un estimado proveniente de la combinación de la reducción relativa del riesgo a partir de los ensayos de bajo riesgo de sesgo incluidos en la revisión (36,9 %), la cual fue de 1.964pacientes.

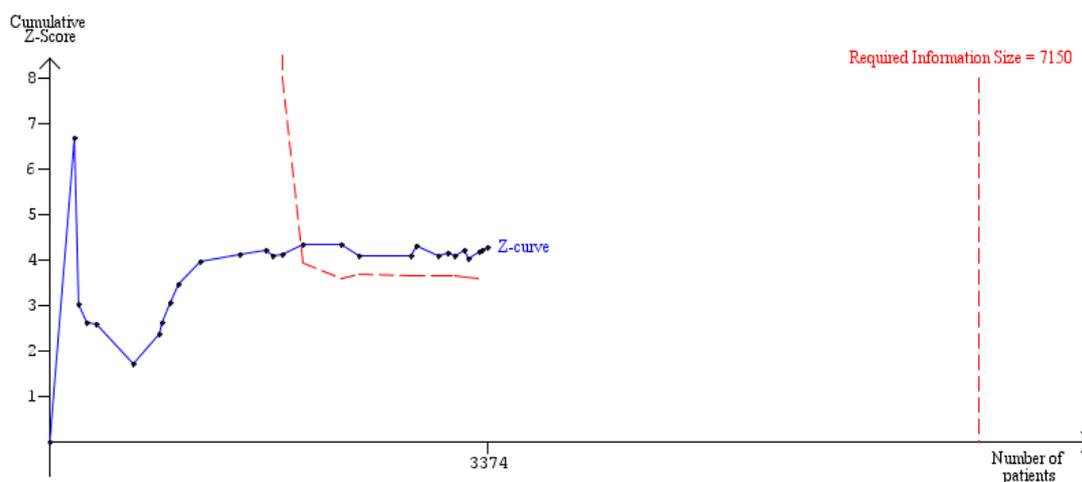


Figura 58. El tamaño de la información requerida ajustada por heterogeneidad para demostrar o rechazar una reducción relativa del riesgo del 20 % (estimación *a priori*) de la fibrilación auricular (con una proporción del grupo control de 27,6 %, un alfa de 1%, y un beta de 10 %) es 7.150 pacientes (línea discontinua vertical roja). La línea discontinua roja en el interior del gráfico- a la izquierda representa los límites de monitorización secuencial de los ensayos que se truncan a partir de los primeros 14 ensayos. La línea azul es la-curve. Z acumulada

Todos los tamaños de la información se derivaron para asegurar un error máximo tipo I de 1%, y un error máximo tipo II de 10 % (es decir, 90 % de potencia). Todos los tamaños de la información se ajustaron por heterogeneidad, mediante la estimación de la diversidad, D^2 . Ambos tamaños de información se obtuvieron asumiendo una proporción del evento de 27,6 % en el grupo con circulación extracorpórea (mediana de la proporción del evento en el control).

La -curva Z acumulada cruzó los límites de vigilancia construidos a partir de los dos cálculos del tamaño de la información (Figura 58 y 59), lo que confirma que la intervención sin circulación extracorpórea es superior a la intervención con circulación extracorpórea para la reducción de la fibrilación auricular.

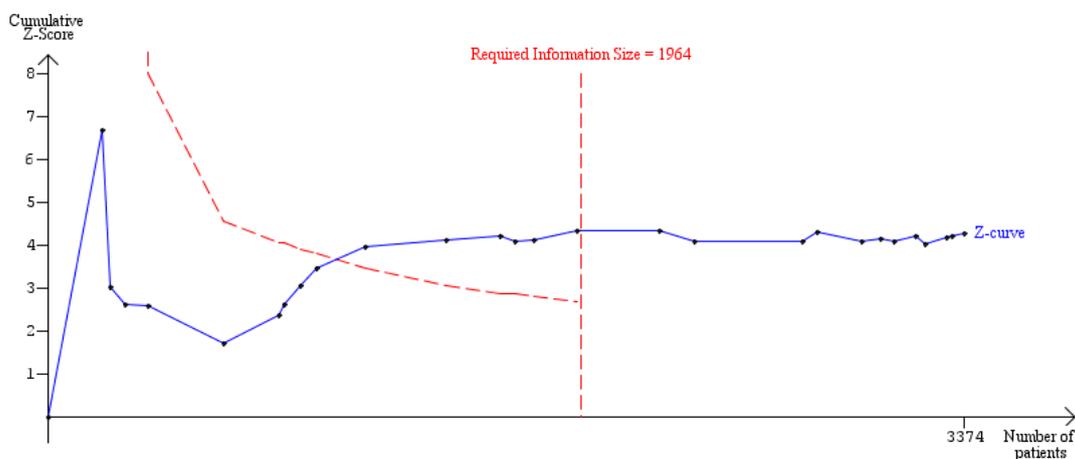


Figura 59. El tamaño de la información requerida ajustada por heterogeneidad para demostrar o rechazar una reducción relativa del riesgo de 36,9 % (estimación con ensayos de bajo riesgo de sesgo) de la fibrilación auricular (con una proporción en grupo control de 27,6 %, un alfa de 1 %, y un beta de 10 %) es 1.964 pacientes (línea discontinua vertical roja). La línea discontinua roja ubicada a la izquierda en el interior del gráfico constituye los límites de seguimiento secuencial de los ensayos que se truncan a partir los primeros 10 ensayos. La línea azul es la -curva Z acumulada.

5.3.2 Evitando sobreestimaciones precoces

Este mismo ejemplo, de fibrilación auricular en CABG, se puede utilizar para ilustrar cómo una sobreestimación del efecto puede ocurrir precozmente en el proceso metanalítico convencional. El metanálisis de la fibrilación auricular fue estadísticamente significativa según el criterio convencional ($p < 0,05$) después de la primer ensayo. Todos excepto uno de los valores de p posteriores en el metanálisis acumulativo fueron también menores que 0,05. De hecho, la mayoría de los subsiguientes valores de p fueron menores que 0,01. La evidencia empírica sugiere que las estimaciones combinadas del efecto, aunque estadísticamente significativa, son inestables cuando existe un número limitado de eventos y pacientes.^{4; 5, 9, 29} Cuando un metanálisis supera el tamaño de información requerida puede garantizar estimaciones combinadas fiables.^{1; 2; 4; 6, 19, 23}

La Tabla 3 muestra la evolución de los efectos del tratamiento en el tiempo, en este ejemplo, al final de cada año. El riesgo relativo combinado fue evidentemente sobrestimado en los dos primeros años y con el apoyo de valores de p menor que 0,01 (los intervalos de confianza convencionales del 99 % excluyeron el 1). En los siguientes tres años, el riesgo relativo combinado fue sobrestimado por un riesgo absoluto de al menos un 10 %. En 2003, el metanálisis cruzó los límites de supervisión del tamaño de la información calculada basada en las estimaciones de bajo riesgo de

sesgo, y en 2004, el metanálisis superó este tamaño de la información requerida. En 2004, el metanálisis también cruzó los límites de supervisión basado en una reducción relativa del riesgo *a priori* de 20 %. Tanto los intervalos de confianza convencionales y ajustados convergieron entre 2002 y 2004.

Year	Total number of			Pooled Effect	99% Confidence Interval	
	Trials	Events	Patients		Conventional	Adjusted
1999	1	55	200	0.24	0.14 to 0.42	0.03 to 7.74
2000	3	74	288	0.39	0.15 to 0.99	0.02 to 7.18
2001	5	143	649	0.57	0.24 to 1.34	0.12 to 2.87
2002	8	204	932	0.52	0.30 to 0.90	0.22 to 1.21
2003 ^a	10	285	1168	0.55	0.37 to 0.81	0.35 to 0.85
2003 ^b	13	391	1722	0.53	0.35 to 0.79	0.34 to 0.83
2004	19	641	2832	0.61	0.46 to 0.82	-
2005	20	679	2999	0.63	0.49 to 0.85	-
2006	25	768	3310	0.67	0.53 to 0.86	-
2007	27	775	3372	0.67	0.53 to 0.86	-

^a First crossing of the boundaries, ^b End of the year

Tabla 3. Muestra la evolución de los efectos combinados (estimaciones de riesgo relativo), los intervalos de confianza de 99 % convencionales y ajustados al final de cada año, con respecto al número acumulado de ensayos, eventos y pacientes. Los intervalos de confianza ajustados 99 % se basan en -el gasto de alfa en relación con el tamaño de la información requerida (1.964 pacientes), utilizando la estimación del riesgo relativo sugerido por los ensayos con bajo riesgo de sesgo.

Este ejemplo ilustra por qué no se debe confiar en estimaciones combinadas basadas en a un número relativamente pequeño de eventos y pacientes (en este caso, a menos de 100 eventos y menos de 300 pacientes). Las estimaciones puntuales de este metanálisis no parecen ser suficientemente fiable hasta cuando al menos se acumulen alrededor de un centenar de eventos y un millar de pacientes. Los intervalos de confianza ajustados sirven para protegerse de inferencias espurias en las primeras etapas de un metanálisis, y acercarse apropiadamente a intervalos de confianza convencionales cuando el número acumulado de pacientes se acerca al tamaño de la información requerida.

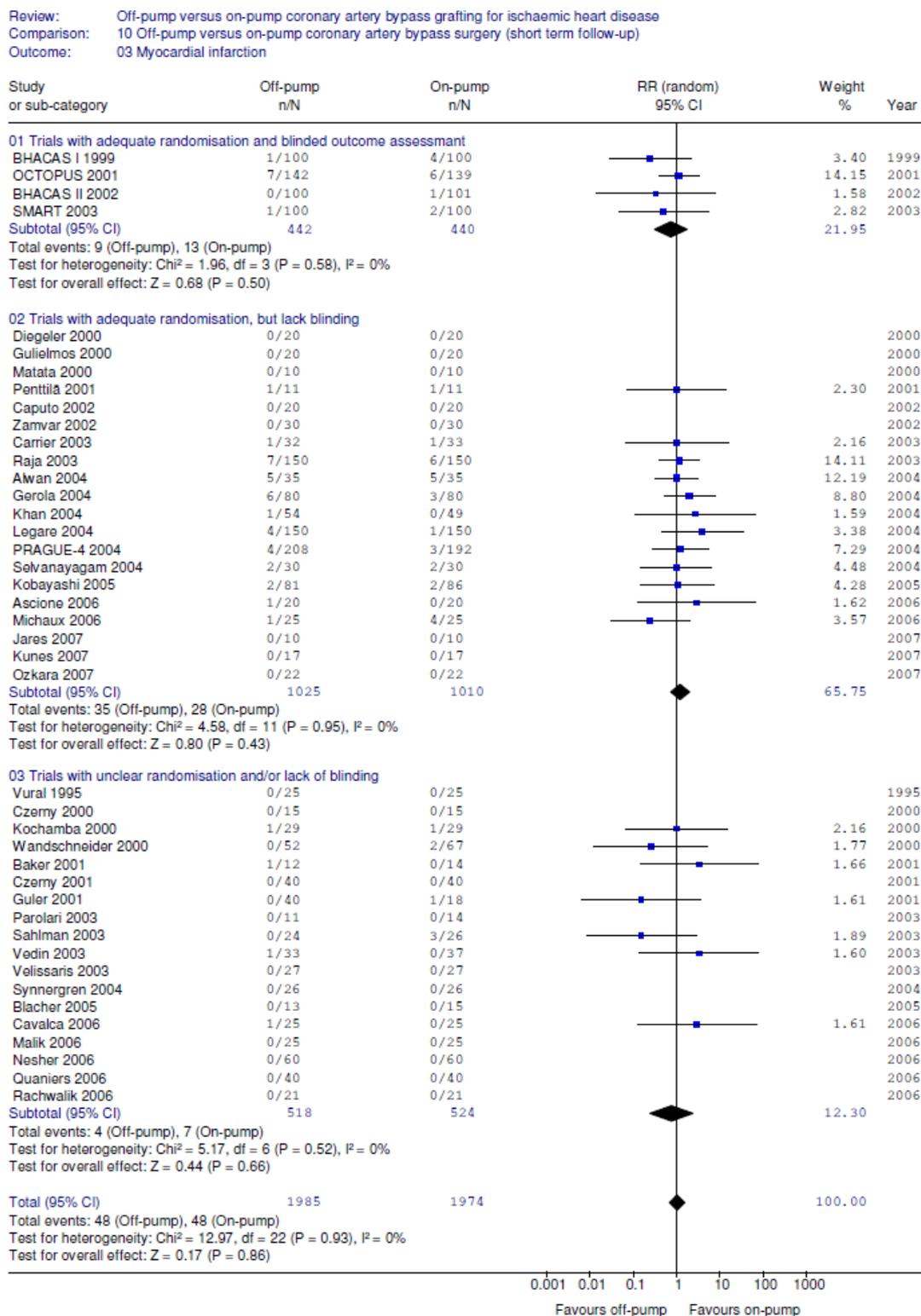


Figura 60. Forest plot del efecto de la CABG con CEC contra CABG sin CEC en el infarto del miocardio.

5.4 Pruebas de inutilidad

El metanálisis, utilizado como ejemplo, que evaluó el impacto de la cirugía de derivación aorto-coronaria con y sin circulación extracorpórea sobre la frecuencia de la fibrilación auricular, también puede utilizarse para ilustrar las pruebas de futilidad, esta vez utilizando el resultado de infarto del miocardio. La frecuencia de este desenlace se reportó en 44 ensayos que incluyeron 4.303 pacientes.⁶⁵ No hubo diferencia significativa entre ambos grupos de comparación (RR 1,06, IC 95 %: 0,72 a 1,54) (Figura 60), este resultado es independiente del riesgo de sesgo. No se detectó heterogeneidad estadística ($I^2 = 0\%$). Diecinueve ensayos (909 pacientes) no reportaron eventos. Cuando estos fueron ajustados por corrección de continuidad, no hubo cambios notables en los resultados (RR 1,05, IC 95 %: 0,74 a 1.48).

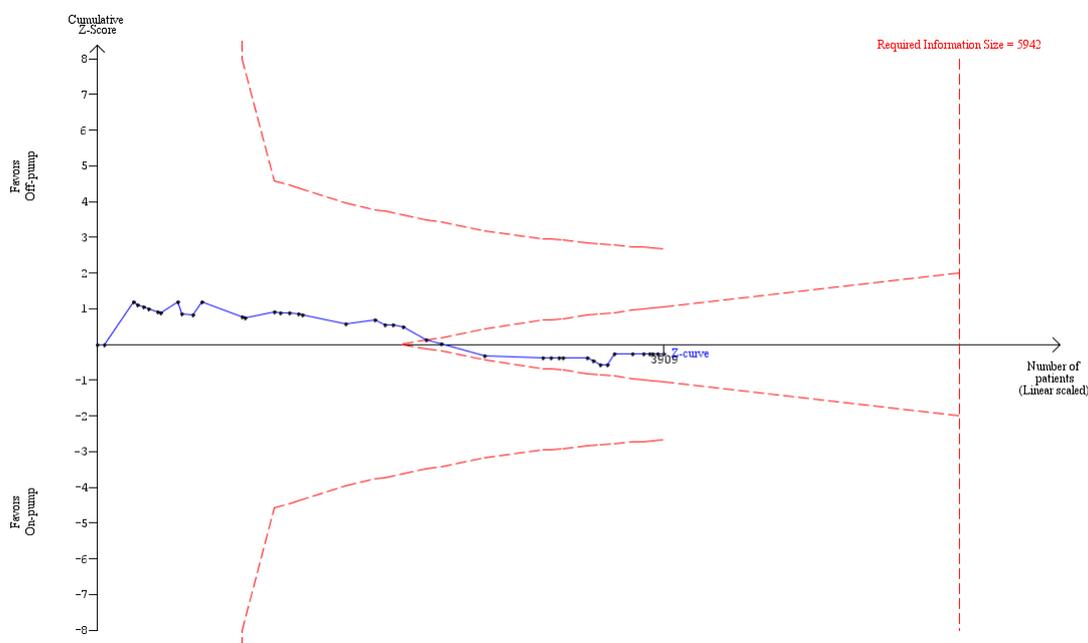


Figura 61. El tamaño de la información requerida ajustada por heterogeneidad para demostrar o rechazar una reducción del riesgo relativo del 33 % (estimación *a priori*) de infarto del miocardio (IM) (con una incidencia de infarto de miocardio en el grupo con bomba de 3,9 %, un alfa de 5 %, y un beta de 20 %) es 5.942 pacientes (línea vertical roja). A la izquierda de esta línea observe unas, líneas rojas discontinuas descendentes que representan los límites de monitorización secuencial de los ensayos; las líneas rojas discontinuas ascendentes representan la región de futilidad. La línea azul es la curva Z acumulada.

Se calculó el tamaño de la información requerida para demostrar o rechazar *a priori* un efecto de la intervención de una reducción del riesgo relativo del 33 %. El valor del 33 % fue elegido porque se cree que representa un efecto razonable de la intervención en esta situación clínica. En contraste con el cálculo del tamaño de la información para

la fibrilación auricular, se utilizó un error máximo tipo I de 5 %, y un error máximo tipo II de 20 % (potencia de 80 %). Se utilizó la mediana como la proporción de infartos de miocardio en los grupos con-bomba (excluyendo los ensayos sin eventos) como la proporción de eventos del grupo control (3,9 %). En conjunto, estos supuestos produjeron un tamaño de la información requerida de 5.942. La curva Z acumulada cruzó los límites de futilidad (Figura 61), y por lo tanto es posible inferir que ninguno de los grupos de comparación es más efectivo de un 33 % con respecto al otro. Este hallazgo, por supuesto, se asocia con un riesgo del 20% de ser un hallazgo con «falso futilidad» (error tipo II de 20 %).

5.5 **La estimación del tamaño de la muestra de un nuevo ensayo clínico**

Cuando un metanálisis no ha cruzado los límites de monitorización ni los límites de futilidad, es posible aproximar el número de pacientes que deben ser asignados al azar en un próximo ensayo para que el metanálisis cruce cualquiera de los dos límites. Un estudio reciente sobre la metodología ilustra esta estrategia mediante un metanálisis de la quimioprofilaxis con isoniazida para prevenir la tuberculosis en pacientes con VIH positivo.²⁵ Este metanálisis incluyó nueve ensayos, 2.911 pacientes, y 131 eventos y produjo un riesgo relativo combinado de 0,74 (IC 95 % 0,53 a 1,04). La heterogeneidad y la diversidad estimadas fueron de 0 %.

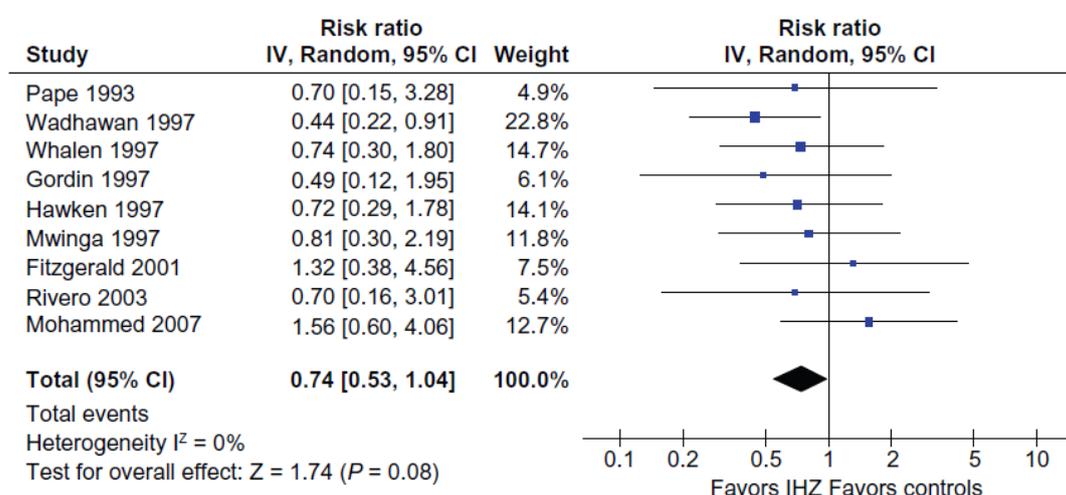


Figura 62. Forest plot de los efectos de los ensayos individuales de la quimioprofilaxis con isoniazida comparadas con el control para la prevención de la tuberculosis en individuos infectados por el VIH y negativos para la prueba con derivados de proteína purificada (PPD)

Se estimó el tamaño de la información necesaria para detectar una reducción relativa de riesgo del 25 % en la tuberculosis con un alfa de 5 % y beta de 20 % (potencia de

80 %). El tamaño de la información requerida se basó en un supuesto de una proporción de eventos en el grupo control de 5 % (aproximadamente la mediana entre los ensayos). También se corrigió la heterogeneidad del tamaño de la información requerida asumiendo una diversidad (D^2) de 20 %. Esto produjo un tamaño de información requerida de 10.508 pacientes. Los límites de monitorización estadística y los de futilidad se construyeron posteriormente en función de los niveles de error y el tamaño de la información requerida.

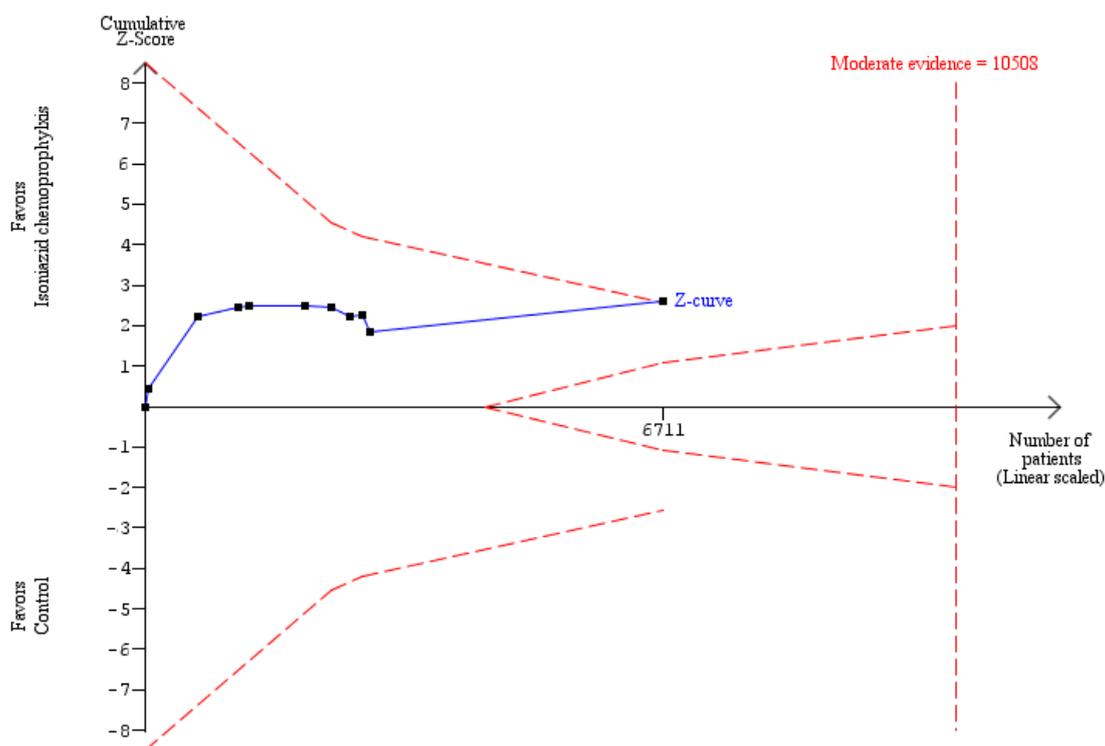


Figura 63. Análisis secuencial prospectivo de ensayos comparando isoniazida con control para la prevención de la tuberculosis. A la izquierda, las líneas discontinuas rojas descendentes representan los límites de monitorización secuencial de los ensayos. A la derecha, las líneas rojas punteadas ascendentes representan la región de futilidad. La línea azul es la curva Z acumulada. El final de la línea de la curva Z acumulada representa un ensayo clínico imaginado que hace que el metanálisis concluya que la isoniazida previene la tuberculosis.

Para estimar cuántos pacientes requieren ser aleatorizados en un ensayo clínico para que el metanálisis sea concluyente, se aproximó el número de pacientes en un futuro ensayo imaginado que haría que la curva Z acumulada cruzara los límites de monitorización o los límites de futilidad. Si un futuro ensayo clínico lograra un metanálisis concluyente con un resultado positivo, se asumiría que los ensayos tendrían la misma proporción de eventos en el grupo control y tamaño del efecto de la intervención tal como fueron postulados en las consideraciones del tamaño de la información. Es decir, se asumió que un ensayo tendría una proporción de eventos

del 5 % en el grupo control y una reducción del riesgo relativo del 25 % (es decir, el ensayo tendría una proporción de eventos en el grupo de intervención de 3,75 %).

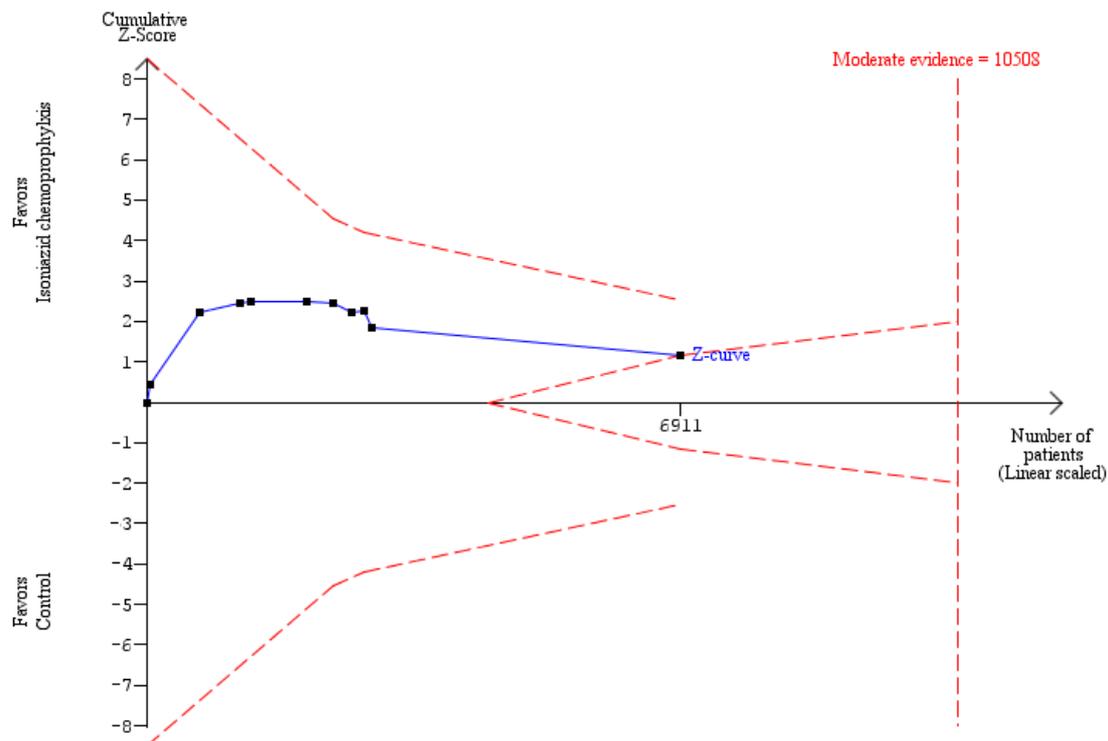


Figura 64. Análisis secuencial prospectivo de ensayos de isoniazida comparados contra el control para la prevención de la tuberculosis. A la izquierda, la línea discontinua roja descendente representa los límites de monitorización secuencial de los ensayos. A la derecha, las líneas discontinuas rojas ascendentes representan la región de futilidad. La línea azul es la curva Z acumulada. El final de la curva Z representa un ensayo imaginado que hace que el metanálisis concluya que la isoniazida no es capaz de prevenir la tuberculosis.

Si un futuro ensayo clínico hiciera el metanálisis concluyente con un resultado futil, se asumiría que la proporción de eventos del grupo de intervención también sería 5 % (es decir, ausencia del efecto). Esto debe ser interpretado que alrededor de 3.800 pacientes (1.900 pacientes en cada grupo de comparación) se necesitarían para producir un metanálisis positivo concluyente (Figura 63). Aproximadamente 4.000 pacientes (2.000 pacientes en cada grupo de comparación) se necesitarían para producir un metanálisis concluyente que muestre futilidad (Figura 64).

5.6 Otras aplicaciones publicadas del Análisis Secuencial de los Ensayos

Los autores de este manual han escrito numerosas revisiones sistemáticas utilizando TSA en, al menos, un metanálisis.^{14; 24; 63; 65-74} La tabla 4 proporciona una breve descripción de estas publicaciones (ordenados por año de publicación).

First author	Journal (year)	Meta-analyses
Bangalore ⁷⁵	<i>BMJ</i> (2011)	Angiotensin receptor blockers (ARB) vs control for i) non-fatal myocardial infarction ii) all-cause mortality iii) cardiovascular mortality iv) angina pectoris v) stroke vi) heart failure vii) new onset diabetes
Bangalore ⁷⁶	<i>Archives of Neurology</i> (2011)	Carotid artery stenting (CAS) vs carotid endarterectomy on i) death, myocardial infarction or stroke ii) periprocedural death or stroke iii) periprocedural stroke
Bangalore ⁷⁷	<i>Lancet Oncology</i> (2011)	i) Angiotensin receptor blockers vs. comparison: effect on cancer risk and on cancer-related death ii) Angiotensin converting enzyme inhibitors vs. comparison: effect on cancer risk and on cancer-related death iii) Beta-blockers vs. comparison: effect on cancer risk and on cancer-related death iv) Calcium channel blockers vs. comparison: effect on cancer risk and on cancer-related death v) Diuretics vs. comparison: effect on cancer risk and on cancer-related death
Afshari A ²⁴	<i>The Cochrane Library</i> (2010)	i) Inhaled nitric oxide vs control for acute respiratory distress syndrome ii) Inhaled nitric oxide vs control for lung injury
Awad T ⁶³	<i>Hepatology</i> (2010)	Peginterferon alfa-2a vs peginterferon alfa-2b for hepatitis C
Brok J ⁶⁶	<i>J Alim Pharm & Ther</i> (2010)	Ribavirin plus interferon vs interferon for hepatitis C
Nielsen N ⁷⁰	<i>Int J Cardiol</i> (2010)	Hypothermia vs control after cardiac arrest

Tarnow-Mordi WO ⁷²	<i>Pediatrics</i> (2010)	i) Probiotics vs control to reduce mortality in newborn ii) Probiotics vs control to reduce necrotizing enterocolitis in newborn
Knorr U ⁶⁹	<i>Psychoneuroendocrinology</i> (2010)	Salivary cortisol in depressed patients vs control persons
Bangalore S ¹⁴	<i>The Lancet</i> (2009)	i) Perioperative beta-blockade vs placebo for mortality ii) Perioperative beta-blockade vs placebo for myocardial infarction
Brok J ⁶⁷	<i>The Cochrane Library</i> (2009)	Ribavirin monotherapy vs placebo for hepatitis C
Whitfield K ⁷³	<i>The Cochrane Library</i> (2009)	Pentoxifylline vs control for alcoholic hepatitis
Moller CH ⁶⁵	<i>Europ Hearj J</i>	i) Off-pump vs on-pump CABG for atrial fibrillation ii) Off-pump vs on-pump CABG for myocardial infarction
Ghandy GY ⁶⁸	<i>Mayo Clin Proc</i> (2008)	i) Perioperative insulin infusion vs control for Mortality ii) Perioperative insulin infusion vs control for Morbidity
Rambaldi A ⁷¹	<i>J Alim Pharm & Ther</i> (2008)	Glucocorticosteroids vs control for alcoholic hepatitis
Whitlock R ⁷⁴	<i>Europ Heart J</i> (2008)	Prophylactic steroid use vs control for patients undergoing cardiopulmonary bypass
Afshari A ²⁴	<i>BMJ</i> (2007)	Antithrombin III vs control for reducing cardiac...

Tabla 4. Metanálisis publicados en los cuales se utilizó Análisis Secuencial de Ensayos Clínicos (TSA).

6 Apéndices

6.1 *Medidas de efecto para metanálisis de los datos dicotómicos y continuos*

Los errores estándar de las medidas de efecto respectivos se calculan de manera similar a los métodos utilizados en la versión.5 del *Review Manager*.²⁷

Para cada ensayo, se denota el número de eventos observados (por ejemplo, muertes) en los dos grupos de intervención, e_A y e_B , y el número total de participantes, n_A y n_B , en los dos grupos de intervención.

Los errores estándar para las diferencias de riesgo, riesgos relativos y odds ratios se calculan mediante las siguientes fórmulas:

$$se(RD) = \sqrt{\frac{e_A(1-e_A)}{n_A^3} + \frac{e_B(1-e_B)}{n_B^3}}$$

$$se(RR) = \sqrt{\frac{1}{e_A} + \frac{1}{e_B} - \frac{1}{n_A} - \frac{1}{n_B}}$$

$$se(OR) = \sqrt{\frac{1}{e_A} + \frac{1}{e_B} + \frac{1}{(1-e_A)} + \frac{1}{(1-e_B)}}$$

Para el odds ratio de Peto, el error estándar está dada por:

$$se(OR) = \sqrt{1/v}$$

donde

$$v = \frac{(n_A n_B (e_A + e_B) ((1-e_A) + (1-e_B)))}{(n_A + n_B)^2 (n_A + n_B - 1)}$$

6.2 Estrategia de efectos aleatorio

6.2.1 Fórmulas del método Biggerstaff-Tweedie

Dada $f_{DL}(t)$ denota la función de densidad de probabilidad de la estimación de DL τ^2 y dada $F_{DL}(t)$ denota la correspondiente $F_{DL}(t)$ función de distribución acumulada.

Definiendo los pesos de los ensayos como una función de t para $w_i(t) = (\sigma_i^2 + t)^{-1}$ y utilizando la distribución obtenida del estimador de τ_{DL}^2 el llamado estimador frecuentista de Bayes de los pesos de los ensayos se puede obtener:

$$\begin{aligned} w_i^*(\tau^2) &= E[w_i^*(\tau_{DL}^2)] \\ &= F(t \cdot 1_{(0,\infty)}(t)) w_i(0) + \int_0^\infty w_i^*(t) f_{DL}(t) dt \end{aligned}$$

El estimador global del efecto de la intervención de la población general puede ser calculado mediante:

$$\mu_{BT} = \frac{\sum_i w_i^* Y_i}{\sum_i w_i^*}$$

Y la varianza

$$Var(\mu_{BT}) = \frac{1}{\left(\sum_i w_i^*(\tau_{DL}^2)\right)^2} \left[\sum_i \left(w_i^*(\tau_{DL}^2) \right) \left(s_i^2 + \tau_{DL}^2 \right) \right]$$

asegurando así que la varianza del estimador global del efecto se ajusta con respecto a la incertidumbre asociada con la estimación de la varianza entre ensayos.

6.3 Análisis secuencial de los ensayos

6.3.1 Incremento del error tipo I debido a la repetición de pruebas de significación

Por las leyes de la teoría básica de la probabilidad, cuando un dato se examina dos veces en el tiempo, y cuando un α de 5% se usa como un umbral para ambas pruebas (o un valor Z de 1,96), la probabilidad de que las dos intervenciones sean declaradas estadísticamente significativa bajo la hipótesis nula es:

$$\begin{aligned}
\Pr(H_0 \text{ rejected}) &= \Pr(|Z_1| \geq 1.96 \text{ or } |Z_2| \geq 1.96) \\
&= \Pr(|Z_1| \geq 1.96) \cdot \Pr(|Z_2| \geq 1.96 \mid |Z_1| < 1.96) \\
&= (1 - \Pr(|Z_1| < 1.96)) \cdot (1 - \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96)) \\
&= 1 - \Pr(|Z_1| < 1.96) - \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96) + \Pr(|Z_1| < 1.96) \cdot \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96) \\
&= 1 - \Pr(|Z_1| < 1.96) - \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96) + \Pr(|Z_1| < 1.96 \text{ or } |Z_2| < 1.96) \\
&= 0.05 - \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96) + \Pr(|Z_1| < 1.96 \text{ or } |Z_2| < 1.96) \\
&> 0.05
\end{aligned}$$

Cuando la desigualdad es evidente a partir del hecho de que

$$\Pr(|Z_2| < 1.96 \text{ or } |Z_1| < 1.96) > \Pr(|Z_2| < 1.96 \mid |Z_1| < 1.96)$$

Lo anterior es fácilmente generalizable para cualquier valor de α y para cualquier número de repetición de las pruebas de significación

6.3.2 Los métodos alternativos no implementados en el software TSA

Una amplia gama de métodos están disponibles para las pruebas de significación repetidas en los ensayos clínicos aleatorizados, algunos de los cuales también pueden tener aplicación en el metanálisis.³⁰ Las estrategias implementadas en el TSA están construidas alrededor del monitoreo del estadístico Z estandarizada. Otras estrategias secuenciales que han recibido cierta atención en el contexto del metanálisis están construidas para la supervisión de otros estadísticos.

Una estrategia que ha recibido recientemente un poco de atención es el análisis secuencial (monitoreo) de los puntajes de rendimiento o el puntaje estadístico de probabilidad para el efecto metanalizado.⁷⁸⁻⁸¹ En el escenario del metanálisis convencional el puntaje de rendimiento para cada ensayo es simplemente el efecto estimado del tratamiento en el ensayo multiplicado por su varianza, y el puntaje de rendimiento en un metanálisis es la suma de los puntajes de rendimiento de los ensayos. En el análisis secuencial de los puntajes de rendimiento, la información se mide como una información estadística (es decir, la información de Fischer). El puntaje de rendimiento es graficado (eje y) contra la información estadística (eje x), y supervisado con algunos límites. Tal como con los límites basados en los gastos de alfa y de beta, el método secuencial para el seguimiento de los puntajes de rendimiento produce límites de superioridad, inferioridad y futilidad. Ejemplos de tales límites se ilustran en la figura 65 a continuación.

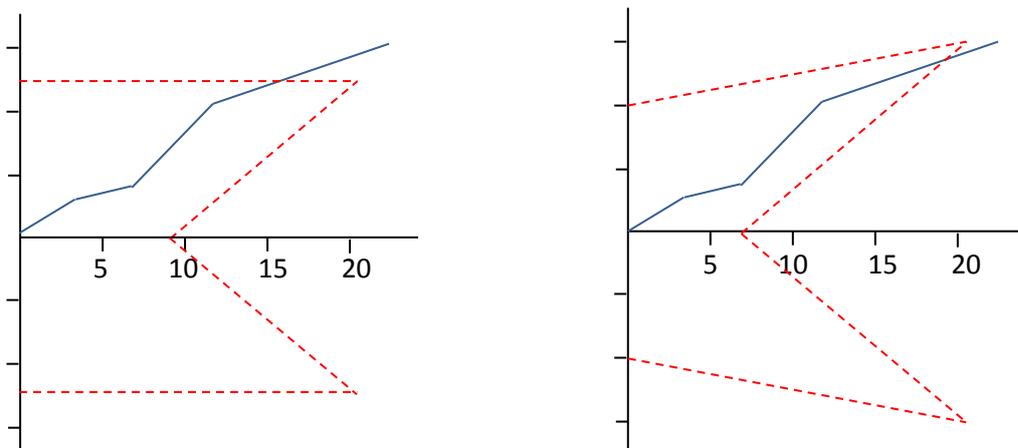


Figura 65. Ejemplo de dos tipos de límites de supervisión del metanálisis secuencial de los puntajes de rendimiento. El gráfico de la izquierda muestra lo que correspondería a límites de significación de gasto de alfa de O'Brien-Fleming y límites futilidad de gasto de beta de O'Brien-Fleming. El gráfico de la derecha ilustra lo que corresponde a lo que normalmente se conoce como límites del triángulo de Whitehead. Este último está diseñado para minimizar el riesgo total de error estadístico (es decir, tanto error tipo I como tipo II).

Al igual que diferentes funciones de gasto de alfa generan diferentes tipos de límites de significación ajustados, la prueba del triángulo se puede utilizar para construir diferentes tipos de límites (y de manera similar para las funciones de gasto de beta- y límites de inutilidad).^{45;80} Por ejemplo, un caso especial de las pruebas del triángulo genera límites equivalentes a los de O'Brien-Fleming cuando se acumula la información estadística (gráfico de la izquierda en la figura 65).

Los límites secuenciales de puntaje de rendimiento de tipo O'Brien-Fleming fueron recientemente explorados empíricamente y mediante *simulation*.⁸⁰ Un estudio realizado por van der Tweel y Bollen comparó los límites de significación O'Brien-Fleming (implementados en el software TSA) con los límites de puntaje de rendimiento secuencial tipo O'Brien-Fleming en seis metanálisis.⁸⁰ Estos seis metanálisis fueron los seleccionados inicialmente (y al azar) como ejemplos ilustrativos de los métodos en el artículo propuesto sobre la corrección de la heterogeneidad del tamaño de la información para el análisis secuencial de los ensayos que se describen en la sección 2.2.1. de este manual.⁶

Tweel y Bollen encontraron que los dos métodos fueron idénticos en la evaluación de la significación. Un estudio de simulación, realizado por Higgins *et al*, investigó el error tipo I y la cobertura del intervalo de confianza ajustado asociado con los límites secuenciales del puntaje de rendimiento tipo O'Brien-Fleming, utilizando diferentes estrategias del modelo de efectos aleatorio.⁷⁹ Esos autores encontraron que con este diseño el modelo de efectos aleatorio convencional de DerSimonian-Laird y el enfoque

Biggerstaff-Tweedie no generó resultados satisfactorios, pero sí con el enfoque semi-bayesiano utilizando una distribución de información Gamma sobre la variación entre ensayos. Otro ejemplo de los límites secuenciales del puntaje de rendimiento es la prueba del triángulo propuesta por Whitehead.^{45; 81} Los límites producidos a partir de este método se ilustran en el gráfico de la derecha de la figura 65. Los límites de las pruebas del triángulo están estadísticamente construidas para generar el menor riesgo posible de error en conjunto (ya sea un error tipo I o tipo II).^{30; 45} Este énfasis, en la minimización de ambos tipos de error, desvía esta técnica a favor del riesgo total de error sobre el riesgo de error tipo I. En el contexto de la investigación médica, la teoría convencional no apoya este equilibrio; la prevención del error alfa siempre se ha considerado más importante.

El desempeño de la prueba del triángulo Whitehead aplicado en el metanálisis se ha explorado en un estudio de simulación, donde se encontró que el método exhibió un inadecuado control del error máximo tipo I en los metanálisis heterogéneos.⁸¹ Los resultados de este estudio sugieren que a más heterogeneidad en un metanálisis se encontrará un peor control del error tipo I en la prueba del triángulo.⁸¹ Hasta la fecha, la literatura contiene un ejemplo de la prueba del triángulo de Whitehead aplicado a un metanálisis que comparó la muerte o enfermedad pulmonar crónica después de ventilación de alta frecuencia con la ventilación mecánica convencional en el tratamiento de recién nacidos prematuros.⁷⁸ En este ejemplo, el metanálisis no demostró diferencia entre las dos intervenciones porque el estadístico de puntaje acumulado cruzó los límites de futilidad.⁷⁸

Restricción estocástica (*Stochastic curtailment*) es otro método para controlar el riesgo de falsos positivos y negativos.^{1; 2} Cuando se aplica a un metanálisis, este método se centra en predecir cuál será el desenlace cuando un metanálisis supera su tamaño de información necesaria.^{1; 2} Más específicamente, la restricción estocástica es un método para calcular la probabilidad de que la tendencia actual de los datos se revertirá antes de superar el tamaño de la información requerida. Cuando la probabilidad de tal reversión es suficientemente pequeña, el metanálisis puede ser considerado concluyente. Dos probabilidades condicionales pueden ser calculadas. En primer lugar, si la tendencia actual de los datos sugiere que la intervención experimental es eficaz, la restricción estocástica se puede utilizar para calcular la probabilidad de rechazar la hipótesis nula cuando el metanálisis sobrepasó el tamaño de la información requerida. Si esta probabilidad condicional es suficientemente alta, el metanálisis se puede considerar concluyente. Del mismo modo, si los datos actuales están sugiriendo una ausencia de tendencia, la restricción estocástica puede ser utilizada para calcular la probabilidad de no rechazar la hipótesis nula una vez que el metanálisis supera el tamaño de la información requerida. Una vez más, si esta probabilidad condicional es suficientemente alta, el metanálisis se puede considerar

concluyente. La restricción estocástica puede ser una herramienta valiosa para ayudar a la toma de decisión, a partir de métodos formales de pruebas de significación. Sin embargo, debido a que la mayoría de los metanálisis están sujetos a sesgos de tendencias temporales, la probabilidad condicional de un cambio de tendencia es muy probable que también esté sesgado.

7 Lista de abreviaturas y notaciones estadísticas

El siguiente capítulo proporciona una guía para abreviaturas, notación y la terminología utilizada en este manual. En algunos casos, estas definiciones pueden variar de otras fuentes. Nuestra intención es ofrecer al lector una guía de cómo se usaron estos términos en este manual.

7.1 *Abreviaturas generales*

AF	- Adjustment Factor (Factor de ajuste)
BT	- Biggersaff-Tweedie
CI	- Confidence Interval (Intervalo de confianza)
D ²	- Diversity (Diversidad)
DL	- DerSimonian-Laird
I ²	- Inconsistency (Inconsistencia)
IF	- Information Fraction (Fracción de información)
IS	- Information Size (Tamaño de la información)
JRE	- Java Runtime Environment (Ambiente de funcionamiento del Java)
MD	- Mean Difference (Diferencia de medias)
OIS	- Optimal Information Size (Tamaño óptimo de la información)
OR	- Odds Ratio
RCT	- Randomised Controlled Trial (Ensayo clínico aleatorizado)
RD	- Risk Difference (Diferencia de riesgo)
RR	- Relative Risk (Riesgo relativo)
RRR	- Relative Risk Reduction (Reducción relativa del riesgo)
SJ	- Sidik-Jonkman
SMD	- Standardised Mean Difference (Diferencia estandarizada de media)
TSA	- Trial Sequential Analysis (Análisis secuencial de ensayos)

7.2 *Notaciones estadísticas*

7.2.1 **Símbolos con letras minúsculas**

<i>c</i>	– The statistical significance threshold with respect $ Z $ (El umbral de significación estadística con respecto $ Z $)
----------	--

c_i	– The adjusted threshold for Z_i under repeated testing (El umbral ajustado para Z_i en pruebas repetidas)
e_x	– The number of events in intervention group X (Número de eventos en el grupo de intervención)
$f_{DL}(t)$	– The probability distribution for the DerSimonian-Laird estimator (La distribución de probabilidad para el estimador de DerSimonian-Laird)
k	– The number of trials in a meta-analysis (Número de ensayos en el metanálisis)
m_x	– The mean response in intervention group X (Respuesta promedio en el grupo de intervención)
n_x	– The number of patients in intervention group X (Número de pacientes en el grupo de intervención)
sd_x	– The standard deviation in intervention group X (Desviación estándar en el grupo de intervención)
v	– Variance estimate (Varianza del estimador)
v_F	– The variance in a fixed-effect model (Varianza en el modelo de efecto fijo)
v_R	– The variance in a random-effects model (Varianza en el modelo de efectos aleatorio)
w_i	– The weight assigned to the i -th trial in a fixed-effect model (Peso asignado al ensayo en el modelo de efecto fijo)
w_i^*	– The weight assigned to the i -th trial in a random-effects model (Peso asignado al modelo de efectos aleatorio)
$w_i(t)$	– The i -th trial weight as a function of the between-trial variance (Peso del ensayo como function de la varianza entre ensayos)

7.2.2 Símbolos con letras mayúsculas

AF	– The heterogeneity adjustment factor (Factor de ajuste de la heterogeneidad)
C	– The sum of the continuity corrections for two groups (Suma de las correcciones de continuidad para dos grupos)
CF_x	– The continuity correction for intervention group X Corrección continuada para el grupo de intervención)
D^2	– The diversity measure used to quantify heterogeneity (Medida de la diversidad para cuantificar heterogeneidad)
$E(X)$	– The expectation of X
H	– A conceptual measure of D^2 (Medida conceptual de diversidad)
H_0	– The null hypothesis (Hipótesis nula)
I^2	– The inconsistency measure used to quantify heterogeneity (Medida de inconsistencia para cuantificar heterogeneidad)

I_j	– The cumulative statistical information after the j -th (La información estadística acumulada después de)
IF_i	– The cumulative information fraction after the i -th trial (Fracción de información acumulada después de un ensayo dado)
$IS_{Patients}$	– The required number of patients in a meta-analysis (Número de pacientes requerido en un metanálisis)
IS_{Events}	– The required number of events in a meta-analysis (Número de eventos requeridos en el metanálisis)
$IS_{Statistical}$	– The required statistical information in a meta-analysis (Información estadística requerida en el metanálisis)
IS_{Fixed}	– The required information size for a fixed-effect model (Tamaño de la información requerida para un modelo de efecto fijo)
IS_{Random}	– The required information size for a random-effects model (Tamaño de la información requerida para un modelo de efectos aleatorios)
OR_i	– The odds ratio estimate of the i -th trial (OR estimado para un ensayo clínico)
P	– The test P-value derived from Z (Prueba del valor de P derivada de la Z)
P_X	– The event rate in intervention group X (Frecuencia del evento en el grupo de intervención)
P^*	– The average event rates of the two treatment groups (Promedio de las frecuencias del evento en los grupos de tratamiento)
$Pr(X)$	– The probability that some event X occurs (Probabilidad de que ocurra un evento X)
$Pr(X Y)$	– The probability that some event X given the event Y occurred (Probabilidad de que ocurra el evento X dado que ocurre el evento Y)
Q	– The Cochran homogeneity test statistic (Prueba estadística de homogeneidad Cochran)
R	– The randomisation ratio (Razón de aleatorización)
RD_i	– The risk difference estimate of the i -th trial (Estimador de la diferencia de riesgo de un ensayo)
RR_i	– The relative risk estimate of the i -th trial (Estimador del riesgo relativo de un ensayo)
S_r	– The sum of trial weights to the r -th power (Suma de los pesos de los ensayos para la potencia r -th)
$SE(X)$	– The standard error of X (Error estándar de X)
$Var(X)$	– The variance of X (Varianza de X)
Z	– The test statistic for whether there exists an intervention effect (Prueba estadística para determinar si existe efecto de una intervención)

Z_i	– The Z -value from the meta-analysis including the first i trials (Valor de Z de un metanálisis incluyendo el primero de los ensayos)
$Z_{1-\alpha/2}$	– The $(1-\alpha/2)$ -th percentile of the standard normal distribution (Percentil de $(1-\alpha/2)$ -th de la distribución normal estándar)
$Z_{1-\beta}$	– The $(1-\beta)$ -th percentile of the standard normal distribution (Percentil de $(1-\beta)$ -th de la distribución normal estándar)
Y_i	– The observed intervention effect in the i -th trial (Efecto de la intervención observada en un ensayo)

7.2.3. Símbolos en letras griegas

α	– The maximum risk of type I error (Riesgo máximo de error tipo I)
$\alpha(t)$	– The cumulative type I error risk as a function of time (Riesgo del error tipo I acumulado como función de tiempo)
β	– The maximum risk of type II error (Riesgo máximo del error tipo II)
$\beta(t)$	– The cumulative type II error risk as a function of time (Riesgo del error tipo II acumulado como función de tiempo)
δ	– The <i>a priori</i> estimate of an anticipated intervention effect (Estimador <i>a priori</i> del efecto anticipado de una intervención)
δ_F	– The anticipated intervention effect in a fixed-effect model (Efecto anticipado de una intervención en el modelo de efecto fijo)
δ_R	– The anticipated intervention effect in a random-effects model (Efecto anticipado de una intervención en el modelo de efectos aleatorio)
λ	– A constant to ensure control of α when penalising Z (Constante para asegurar el control de alfa cuando se penaliza a Z)
μ_i	– The underlying ‘true’ intervention effect of the i -th trial (El efecto verdadero de la intervención de un ensayo)
μ	– The overall ‘true’ intervention effect (El efecto global verdadero de una intervención)
$\hat{\mu}$	– The pooled intervention effect (Efecto combinado de una intervención)
σ^2	– The variance of δ (Varianza de δ)
σ_i^2	– The variance of Y_i (Varianza de Y_i)
τ^2	– The between-trial variance (Varianza entre los ensayos)
τ_{DL}^2	– The DerSimonian-Laird estimate for the between-trial variance (Estimador DerSimonian-Laird para la varianza entre ensayos)
τ_{SJ}^2	– The Sidik-Jonkman estimate for the between-trial variance (Estimador Sidik-Jonkman para la varianza entre ensayos)
$\hat{\theta}$	– The pooled odds ratio of excluding zero-event trials (Odds Ratio combinados excluyendo ensayos sin desenlaces)

- Φ – The cumulative standard normal probability distribution function
(Función de la distribución de probabilidad normal estándar acumulada)

8 Referencias bibliográficas

- (1) Pogue J, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Controlled Clinical Trials* 1997;18:580-593.
- (2) Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;351:47-52.
- (3) Sterne JA, Davey SG. Sifting the evidence - what's wrong with significance tests? *British Medical Journal* 2001;322:226-231.
- (4) Thorlund K, Devereaux PJ, Wetterslev J et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *International Journal of Epidemiology* 2009;38:276-286.
- (5) Trikalinos TA, Churchill R, Ferri M et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology* 2004;57:1124-1130.
- (6) Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *Journal of Clinical Epidemiology* 2008;61:64-75.
- (7) Hu M, Cappelleri J, Lan KK. Applying the law of the iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clinical Trials* 2007;4:329-340.
- (8) Lan KK, Hu M, Cappelleri J. Applying the law of the iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Statistica Sinica* 2003;13:1135-1145.
- (9) Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci U S A* 2001;98:831-836.
- (10) Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. *J Clin Epidemiol* 2009;62:825-830.
- (11) Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *Journal of Clinical Epidemiology* 2008;61:763-769.
- (12) Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive - Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *International Journal of Epidemiology* 2009;38:287-298.
- (13) Higgins JPT, Green S. *Cochrane Handbook for systematic reviews of interventions, version 5.0.0*. John Wiley & Sons, 2009.
- (14) Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet* 2008;372:1962-1976.
- (15) Devereaux PJ, Beattie WS, Choi PT et al. How strong is the evidence for the use of perioperative beta blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. *BMJ* 2005;331:313-321.
- (16) DeMets D, Lan KK. Interim analysis: the alpha spending function approach. *Statistics in Medicine* 1994;12:1341-1352.

- (17) Lan KK, DeMets D. Discrete sequential monitoring boundaries for clinical trials. *Biometrika* 1983;659-663.
- (18) Pocock S. Intermittent analyses for randomized clinical trials: the group sequential approach. *Biometrics* 1982;38:153-162.
- (19) GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy* 2009;64:669-677.
- (20) Guyatt G, Mills E. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Medicine* 2008;5:e4.
- (21) Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. *BMC Medical Research Methodology* 2009;9.
- (22) Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-1194.
- (23) Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in a random-effects meta-analysis. *BMC Medical Research Methodology* 2009;9.
- (24) Afshari A, Wetterslev J, Brok J, Moller A. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. *BMJ* 2007;335:1248-1251.
- (25) Thorlund K, Anema A, Mills E. Interpreting meta-analysis according to the adequacy of sample size. An example using isoniazid chemoprophylaxis for tuberculosis in purified protein derivative negative HIV-infected individuals. *Clinical Epidemiology* 2010;2:57-66.
- (26) Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359:519.
- (27) Review Manager (RevMan) [Computer program]. Version 5.1. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2011.
- (28) Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine* 2010;4:e78.
- (29) Gehr B, Weiss C, Porzolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Medical Research Methodology* 2006;6:25.
- (30) Jennison C, Turnbull B. *Group sequential methods with applications to clinical trials*. Chapman&Hall/CRC Press, 2000.
- (31) Berkey C, Mosteller F, Lau J. Uncertainty of the time of first significance in random-effects cumulative meta-analysis. *Controlled Clinical Trials* 1996;17:357-371.
- (32) DerSimonian L, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;7:177-188.
- (33) Sidik K, Jonkman J. Simple heterogeneity variance estimation for meta-analysis. *Journal of Royal Statistical Society(C)* 2005;54:367-384.
- (34) Sidik K, Jonkman J. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 2007;26:1964-1981.
- (35) Brockwell S, Gordon I. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001;20:825-840.
- (36) Brockwell S, Gordon IR. A simple method for inference on an overall effect in meta-analysis. *Stat Med* 2007;26:4531-4543.
- (37) Sanchez-Meca J, Martin-Martinez, F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* 2008;13:31-48.

- (38) Sidik K, Jonkman J. Robust variance estimation for random-effects meta-analysis. *Comp Stat Data An* 2006;50:3681-3701.
- (39) Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J Educa Behav Stat* 2005;30:261-293.
- (40) Makambi KH. The effect of the heterogeneity variance estimator on some tests of efficacy. *J Biopharm Stat* 2004;14:439-449.
- (41) Biggerstaff B, Tweedie R. Incorporating variability in estimates of heterogeneity in the random-effects model in meta-analysis. *Statistics in Medicine* 2009;16:753-768.
- (42) Sweeting M, Sutton A, Lambert P. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis. *Statistics in Medicine* 2004;23:1351-1375.
- (43) Armitage P. Sequential analysis in therapeutic trials. *Annual Review of Medicine* 1969;20:425-430.
- (44) Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977;191-199.
- (45) Whitehead J. *The design and analysis of sequential clinical trials*. John Wiley & Sons, 2000.
- (46) Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 1991;10:1665-1677.
- (47) Higgins JPT, Thompson S. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;21:1539-1558.
- (48) Federov V, Jones B. The design of multicentre trials. *Statistical Methods in Medical Research* 2005;14:205-248.
- (49) Ioannidis J, Patsopoulos N, Evangelou E. Uncertainty in heterogeneity estimates in meta-analysis. *British Medical Journal* 2007;335:914-916.
- (50) Jackson D. The implications of publication bias for meta-analysis' other parameter. *Statistics in Medicine* 2006;25:2911-2921.
- (51) Chan AW, Hrobjartsson A, Haahr M, Gotzsche P, Altman D. Empirical evidence for selective reporting of outcomes in randomized trials. comparison of protocols to published articles. *Journal of American Medical Association* 2004;291:2457-2465.
- (52) Chan AW, Altman D. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal* 2004;171:735-740.
- (53) Dwan K, Altman D, Arnaiz J et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS Medicine* 2008;3:e3081.
- (54) Hrobjartsson A, Chan AW, Haahr M, Gotzsche P, Altman D. Selective reporting of positive outcomes in randomised trials - secondary publication. A comparison of protocols with published reports. *Ugeskr Laeger* 2005;167:3189-3191.
- (55) Kjaergaard L, Villumsen J, Gluud C. Reported methodological quality and discrepancies between small and large randomized trials in meta-analyses. *Annals of Internal Medicine* 2001;135:982-989.
- (56) Moher D, Pham B, Jones A et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-613.
- (57) Schulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of American Medical Association* 1995;273:408-412.
- (58) Wood L, Egger M, Gluud LL et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *British Medical Journal* 2008;336:601-605.

- (59) Guyatt GH, Oxman AD, Vist GE et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.
- (60) Reboussin DM, DeMets DL, Kyungmann K, Lan KKG. Programs for Computing Group Sequential Boundaries Using the Lan-DeMets Method, v.2. 2009.
- (61) O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979;35:549-556.
- (62) Pocock SJ. When to stop a clinical trial. *British Medical Journal* 1992;305:235-240.
- (63) Awad T, Thorlund K, Hauser G, Stimac D, Mabrouk M, Gluud C. Peginterferon alpha-2a is associated with higher sustained virological response than peginterferon alfa-2b in chronic hepatitis C: Systematic review of randomized trials. *Hepatology* 2010;51:1176-1184.
- (64) Rigotti N, Munafo MR, Stead LF. Interventions for smoking cessation in hospitalised patients. *Cochrane Database of Systematic Reviews* 2007.
- (65) Moller CH, Penninga L, Wetterslev J, Steinbruchel DA, Gluud C. Clinical outcomes in randomized trials of off- vs. on-pump coronary artery bypass surgery: systematic review with meta-analyses and trial sequential analyses. *Eur Heart J* 2008;29:2601-2616.
- (66) Brok J, Gluud LL, Gluud C. Meta-analysis: ribavirin plus interferon versus interferon monotherapy for chronic hepatitis C - an updated Cochrane review. *Alimentary Pharmacology and Therapeutics* 2010;32:840-850.
- (67) Brok J, Gluud C. Ribavirin monotherapy for hepatitis C. *Cochrane Database of Systematic Reviews* 2010; Issue 1.
- (68) Ghandi GYMM, Flynn DN, et al. Effect of perioperative insulin infusion on surgical morbidity and mortality: systematic review and meta-analysis of randomized trials. *Mayo Clinique Proceedings* 2008;83:418-430.
- (69) Knorr U, Vinberg M, Kessing LV, Wetterslev J. Salivary cortisol in depressed patients versus control persons: a systematic review and meta-analysis. *Psychoneuroendocrinology* 2010; 35(9):1275-86.
- (70) Nielsen N, Friberg H, Gluud C, Herlitz J, Wetterslev J. Hypothermia after cardiac arrest should be further evaluated - a systematic review of randomised trials with meta-analysis and trial sequential analysis. *International Journal of Cardiology* 2010.
- (71) Rambaldi A, Saconato HH, Christensen E, Thorlund K, Wetterslev J, Gluud C. Systematic review of glucocorticosteroids for alcoholic hepatitis - a Cochrane Hepato-Biliary Group systematic review with meta-analysis and trial sequential analysis of randomised clinical trials. *Alimentary Pharmacology and Therapeutics* 2008;27:1167-1178.
- (72) Tarnow-Mordi WO, Wilkinson D, Trivedi D, Brok J. Probiotics reduce all-cause mortality and necrotizing enterocolitis: it is time to change practice. *Pediatrics* 2010;125:1068-1070.
- (73) Whitfield K, Rambaldi A, Wetterslev J, Gluud C. Pentoxifylline for alcoholic hepatitis C. *Cochrane Database of Systematic Reviews* 2009.
- (74) Whitlock R, Chan S, Devereaux PJ. Clinical benefit of steroid use in patients undergoing cardiopulmonary bypass: a meta-analysis of randomized trials. *European Heart Journal* 2008;29:2592-2600.
- (75) Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. *BMJ* 2011;342:d2234.

- (76) Bangalore S, Kumar S, Wetterslev J et al. Carotid artery stenting vs carotid endarterectomy: meta-analysis and diversity-adjusted trial sequential analysis of randomized trials. *Arch Neurol* 2011;68:172-184.
- (77) Bangalore S, Kumar S, Kjeldsen SE et al. Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials. *Lancet Oncol* 2011;12:65-82.
- (78) Bollen C, Uiterwaal C, Vught A, van der Tweel I. Sequential meta-analysis of past clinical trials to determine the use of a new trial. *Epidemiology* 2006;17:644-649.
- (79) Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med* 2011;30:903-921.
- (80) van der Tweel I, Bollen C. Sequential meta-analysis: an efficient decision making tool. *Clinical Trials* 2010;7:136-146.
- (81) Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Statistics in Medicine* 1997;16:2901-2913.