



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

TESIS DOCTORAL

Estimación del Tamaño Muestral Requerido en el uso de Variables de Respuesta Combinadas: Nuevas Aportaciones.

Doctorando: Josep Ramon Marsal Mora

Tutora: Dra. Montserrat Martín-Baranera

Director: Dr. Ignacio Ferreira-González

2018

*Departament de Pediatria, d'Obstetrícia i Ginecologia i de Medicina Preventiva.
Programa de Doctorat en Metodologia de la Recerca Biomèdica i Salut Pública.*

Departament de Pediatria, d'Obstetrícia i Ginecologia i de Medicina Preventiva.

Programa de Doctorat en Metodologia de la Recerca Biomèdica i Salut Pública.

Universitat Autònoma de Catalunya (UAB).

**ESTIMACIÓN DEL TAMAÑO MUESTRAL
REQUERIDO EN EL USO DE VARIABLES DE
RESULTADO COMBINADAS: NUEVAS
APORTACIONES**

Memoria presentada por Josep Ramon Marsal para optar al título de Doctor, elaborado a partir del trabajo realizado bajo la dirección del Dr. Ignacio Ferreira González y la tutora Dra. Montse Martín-Baranera. Unitat d'Epidemiologia del Servei de Cardiologia de l'Hospital Universitari Vall d'Hebron y Unitat de Suport a la Recerca de l'institut universitari IDIAP Jordi Gol.

Lleida – Barcelona, 2018

A Inés y Ramón, mis padres.

0. ÍNDICE

0. ÍNDICE.....	6
0.1. Índice de Tablas y Figuras.....	9
0.2. Acrónimos.....	11
1. PRESENTACIÓN.....	13
1.1. Historia.....	14
1.2. Financiación.....	15
1.3. Presentaciones Previas.....	16
1.4. Agradecimientos.....	17
2. INTRODUCCIÓN.....	19
2.1. Definición.....	19
2.2. Bases para la construcción e Interpretación de un CE.....	21
2.3. Ventajas e Inconvenientes en el uso de CE.....	22
2.3.1. Ventajas.....	23
2.3.2. Desventajas.....	26
2.3.3. Diferencia clínicamente relevante.....	27
2.4. Recomendaciones en el uso de CE.....	28
3. JUSTIFICACIÓN.....	31
4. OBJETIVOS.....	35
4.1. Estructura del documento de memoria.....	36
5. MÉTODOS.....	39
5.1. Definiciones, terminología y aspectos matemáticos.....	40
5.1.1. Definición de un CE binario.....	40
5.1.2. Efecto de la intervención sobre el CE.....	42
5.1.3. Grado de asociación entre variables binarias.....	45
5.2. Grado de asociación.....	46
5.2.1. Definición.....	46
5.2.2. Coeficiente de Correlación.....	47
5.2.3. Intervalo de valores posibles.....	49
5.2.4. Otras métricas de interés.....	52
5.3. Tamaño de Muestra Requerido.....	55
6. Cuantificación de la asociación entre eventos binarios.....	61
6.1. Justificación.....	62
6.2. Medidas de similitud.....	64
6.2.1. Coeficiente de correlación ρ_j	65
6.2.2. Probabilidad Conjunta JP_j	66
6.2.3. Índice de Jaccard JJ_j	67

6.2.4.	Índice relativo de Jaccard <i>JJj</i> *	68
6.2.5.	Índice de Ochiai <i>OCj</i>	69
6.2.6.	Coeficiente de Pares Simples <i>SPj</i>	70
6.2.7.	Coeficiente de Sørensen-Dice <i>SDj</i>	71
6.2.8.	Coeficiente de Kulczynski <i>KUj</i>	72
6.3.	Intervalo de valores posibles aplicado a los CS seleccionados.	74
6.4.	Rango y coeficientes estandarizados.....	77
6.5.	Estimación de los CS en los estudios CARE y PARADIGM-HF.....	81
6.6.	Conclusiones del objetivo 1	83
7.	Influencia del grado de asociación en el TMR en un CE binario de dos eventos	87
7.1.	Justificación.....	88
7.2.	Definiciones y Asunciones.....	90
7.3.	Material y Métodos.....	92
7.4.	Resultados	95
7.5.	Conclusiones del proyecto 2.....	104
8.	Bin-CE: Una aplicación web para calcular el TMR en un conjunto de eventos.....	107
8.1.	Justificación.....	108
8.2.	Definiciones y asunciones	110
8.3.	Algoritmo Bin-CE.....	113
8.3.1.	Aspectos Técnicos, Software y Código Abierto	113
8.3.2.	Enfoque general de Bin-CE	114
8.3.3.	Entrada de datos (Pantalla 1).....	115
8.3.4.	Entrada de datos (Pantalla 2).....	118
8.3.5.	Revisión de los datos (pantalla 3)	122
8.4.	Algoritmo iterativo	124
8.5.	Resultados (pantalla 4).....	127
8.6.	Ejemplo Numérico	129
8.6.1.	Resultados.....	130
8.6.2.	Análisis de sensibilidad	132
8.7.	Conclusiones del proyecto 3.....	135
9.	Discusión.....	137
10.	Anexos	151
10.1.	Anexo 1: Estimación de la probabilidad conjunta en el grupo intervención.....	151
10.2.	Anexo 2: El efecto combinado del CE está acotado	155
10.3.	Anexo 3: Listado de coeficientes de similitud	159
10.4.	Anexo 4: Rangos y coeficientes estandarizado	168
10.5.	Anexo 5: Resultados de la simulación.....	170
10.6.	Anexo 6: Código Fuente.....	177
10.7.	Anexo 7: The Use of a Binary Composite Endpoint and Sample Size Requirement: Influence of Endpoints Overlap.....	196

11. BIBLIOGRAFIA..... 209

0.1. Índice de Tablas y Figuras

Figura 1 - Mapa actual del conocimiento sobre CE.....	29
Figura 2 – Ejemplo de posibles escenarios con diferentes probabilidades conjuntas en la intersección de dos eventos en un ensayo clínico.....	50
Tabla 1 - Resumen de la interpretación clínica y dirección de la asociación en los CS seleccionados.	73
Tabla 2 - Aplicación de los límites de Fréchet a los 8 CS estudiados.	76
Figura 3A: Distribución de los límites inferior y superior de la probabilidad conjunta, índice de Jaccard, índice relativo de Jaccard y correlación al incrementarse la probabilidad del ER y del EA..	78
Figura 3B: Distribución de los límites inferior y superior del coeficiente de Ochiai, pares simples, Sørensen-Dice y Kulcysnki al incrementarse la probabilidad del ER y del EA.....	79
Tabla 3 - Resultados del estudio CARE y PARADIGHM-HF.	82
Tabla 4 – Resumen de los escenarios simulados.	93
Figura 4 – Ratio de Tamaños de Muestra (SSR) en las situaciones de la 2 a la 4.	98
Figura 5 – Curvas de potencia para diferentes tamaños de muestra para diferentes grados de asociación cuando $PER = 0,10$ y $PEA = 0,10$	100
Figura 6 – Curvas de potencia en función del tamaño de muestra y el grado de asociación cuando $PER = 0,10$ y $PEA = 0,20$	102
Tabla 5 – Resumen del impacto cualitativo del grado de asociación entre componentes en cuanto al TMR, considerando diferentes probabilidades y efectos de los eventos combinados.....	105
Figura 7 – Diagrama de trabajo de Bin-CE.....	115
Figura 8 – Ejemplo de la pantalla 1. Probabilidad de ocurrencia y efecto de cada evento considerado.....	117
Figura 9 - Ejemplo de la pantalla 2. Probabilidad conjunta (asociación de segundo orden).....	121
Figura 10 - Ejemplo de la pantalla 3. Revisión de los datos, TMR del ER y Gráfico del potencial de cada EA.....	123
Figura 11 – Asociación de orden dos y tres entre eventos hipotéticos (A, B y C).....	126
Figura 12 – Tabla de información presentada en la pantalla 4 de Resultados.....	127
Tabla 6 – Eventos primarios y secundarios del estudio STEMI-RADIAL.....	130
Tabla 7 – Análisis de sensibilidad. Resultados de Bin-CE cuando varía el grado de asociación en todo su rango aplicando los límites de Fréchet. Eventos Hemorrágicos.....	133
Tabla 8 – Análisis de sensibilidad. Resultados de Bin-CE cuando se varía el grado de asociación en todo su rango aplicando los límites de Fréchet. Eventos Isquémicos.....	134
Tabla 9 – Listado de los Coeficientes de Similitud estudiados.....	161
Tabla 10 - Rangos de los 8 CS según $\pi_{j1} + \pi_{j2} < 1$ o $\pi_{j1} + \pi_{j2} > 1$	168
Tabla 11 - Coeficientes estandarizados de los 8 CS según $\pi_{j1} + \pi_{j2} < 1$ o $\pi_{j1} + \pi_{j2} > 1$	169
Tabla 12 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,10, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.....	171

Tabla 13 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,20, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.....	173
Tabla 14 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,03, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.....	175
Figura 13 – Rangos no simétricos entre la asociación del grupo control e intervención	153

0.2. Acrónimos

AVC	Accidente Cerebro-Vascular
CE	<i>Inglés: 'Composite Endpoint' o 'Combined Endpoint'</i>
CS	Coefficiente de Similitud
EA	Evento Aditivo
ECA	Ensayo Clínico Aleatorizado
ER	Evento Relevante
FDA	<i>Inglés: 'Food and Drug Administration of United States'</i>
IAM	Infarto Agudo de Miocardio
JI	<i>Inglés: 'Jaccard's Index'</i>
JI*	<i>Inglés: 'Relative Jaccard's Index'</i>
JP	<i>Inglés: 'Joint Probability'</i>
MACE	<i>Inglés: 'Major Adverse Cardiovascular Event'</i>
NNT	Número Necesario de pacientes a Tratar
OC	<i>Inglés: 'Ochiai's Coefficient'.</i>
OR	<i>Inglés: 'Odds Ratio'</i>
RA	Riesgo Atribuible o Riesgo Absoluto
RR	Riesgo Relativo
RT	Razón de Tamaños
SEB	Sociedad Española de Biometría
SP	<i>Inglés: 'Simple Pairs's Coefficient'</i>
SSR	<i>Inglés: 'Sample Size Ratio'</i>
TM	Tamaño de Muestra
TMR	Tamaño de Muestra Requerido
VRC	Variable/s de Resultado Combinada

1.PRESENTACIÓN

Esta memoria ha sido elaborada para la obtención del título de Doctor, según la normativa interna de la Universitat Autònoma de Barcelona y del Departament de Pediatria, d'Obstetrícia i Ginecologia i de Medicina Preventiva. Se presenta, acompañada de la documentación necesaria, para su evaluación dentro del programa de Doctorat en Metodologia de la Recerca Biomèdica i Salut Pública bajo la modalidad: 'Memoria de Tesis'.

El trabajo incluye un apartado de introducción, justificación y objetivos. En estos tres apartados se contextualizará y establecerán las preguntas y retos a responder y solucionar. A continuación, se adjunta un apartado de Métodos y Resultados donde se desarrollan tres proyectos relacionados entre sí, pero con objetivos distintos. Finalmente se adjunta un apartado de conclusiones y discusión.

Producto de esta Tesis se ha publicado un artículo científico en la revista *American Journal of Epidemiology* y actualmente estamos en la segunda fase de revisión de otro artículo en la

revista *PLoS One*. Otro de los productos de la Tesis ha sido el diseño y elaboración de una aplicación web, gratuita y de acceso libre para el cálculo del tamaño de muestra. Esta aplicación, que se llama Bin-CE es accesible en: <https://uesca-apps.shinyapps.io/bincep/>. El doctorando ha tenido que formarse en el lenguaje de programación específico de la librería Shiny de RStudio que es la que permite realizar este tipo de aplicaciones.

El artículo publicado es:

Marsal JR, Ferreira-González I, Bertran S, Ribera A, Permanyer-Miralda G, García-Dorado D. The use of a binary composite endpoint and simple size requirement: Influence of endpoints overlap. *Am J Epidemiol* 2017; 185:832-41.

1.1. Historia

Nuestro grupo de investigación ya hace 11 años (2007) trabajó intensamente en la descripción del uso de las Variables de Resultado Combinadas en la literatura y en los problemas interpretativos que genera. Como fruto de ese trabajo se publicaron varios manuscritos en revistas internacionales (BMJ, *Journal of Clinical Epidemiology*) y se leyó una tesis doctoral(1-4).

En la línea de investigación mencionada, uno de los problemas que no se abordó es la cuantificación del posible beneficio de las Variables de Resultado Combinadas (VRC) para la disminución del tamaño de muestra requerido (la “eficiencia” de las VRC). Se trata de un problema matemático en el que influyen, esencialmente, la magnitud de la correlación entre los eventos que se combinan y la magnitud del efecto de la intervención sobre cada una de las variables de resultado individuales.

Este aspecto es de gran importancia, pues según dichos parámetros la eficiencia de la VRC para disminuir el tamaño de muestra podría ser muy variable e, incluso, podría ser que el uso de la VRC conllevara un mayor tamaño de muestra que el uso de una única variable de resultado. El impacto potencial es que, teniendo en cuantos dichos parámetros en la estimación del tamaño de muestra se puede determinar con gran precisión el número mínimo de pacientes y evitar, de este modo, la inclusión de más pacientes de los necesarios para demostrar un efecto. La presente tesis aborda dicho objetivo: el estudio de la influencia de ambos parámetros en la eficiencia de la VRC para disminuir el tamaño de muestra. El abordaje que se pretendía desde el principio es eminentemente pragmático, con objetivo final de desarrollar una herramienta de ayuda al investigador que se plantea diseñar un ensayo clínico utilizando VRC.

1.2. Financiación

El proyecto de forma general o cualquier parte de éste, no ha sido financiado o recibido dinero ni de forma parcial ni completa en ningún concurso público. Los gastos de este proyecto han sido cubiertos tanto por la Unidad d'Epidemiologia del Servei de Cardiologia del HUVH como por parte de la Unitat de Suport a la Recercad de Lleida del IDIAP Jordi Gol y por el doctorando. Los gastos financiados han sido los habituales: matriculación de cada uno de los cursos, asistencia a jornadas científicas y congresos, así como los gastos correspondientes a la publicación en revistas *per-review*. El uso del servidor de la herramienta Bin-CE es gratuito.

1.3. Presentaciones Previas

El trabajo ha sido presentado de forma parcial en diferentes jornadas y sesiones. Estas presentaciones le alinean con las actividades obligatorias del programa de Doctorado. Se listan dichos lugares y fechas:

- a) Primeras Jornadas Científicas de Estudiantes de la SEB. Facultad de Matemáticas, campus de Burjassot, Universitat de València. 19 y 20 de enero de 2.015. La Sociedad Española de Biometría (SEB) prepara dichas jornadas anualmente orientadas a los estudiantes de doctorado en diferentes universidades españolas. El objetivo de estas jornadas es fomentar el debate entre estudiantes de doctorado.
- b) Seminario: “*Noves aportacions en l’ús i la interpretació de les variables de resultat combinades*”. Unitat d’Epidemiologia del Servei de Cardiologia del HUVH. 3 febrero de 2.015.
- c) Seminario metodológico “Statistics is cool”: “*The use of binary composite endpoint instead of two primary endpoints could imply a lower or higher sample size requirement*”. Unitat de Bioestadística i Epidemiologia de l’Institut de Recerca Biomèdica de Lleida, IRB Lleida. 21 de abril de 2015.
- d) Segunda Jornadas Científicas de Estudiantes de la SEB. “*Assessment of relationship between two binary variables and the effect over the sample size calculation for a composite endpoint*”. Parc de la Recerca Biomèdica de Barcelona (PRBB), Barcelona. 8 y 9 de septiembre de 2.016.
- e) XVI Conferencia Española de Biometría. “*Bin-CE a Shiny App to compute the Sample Size Requirement with Binary Composite Endpoints*”. Facultad de Matemáticas. IMUS. Sevilla. 13-15 de septiembre de 2.017.

1.4. Agradecimientos

Este proyecto de Tesis es el resultado de la colaboración de un largo listado de personas e instituciones. Todas estas personas e instituciones más o menos han colaborado a que el trabajo final llegue a buen puerto.

Agradecido a **Ignacio Ferreira** por plantearme el reto de continuar su propio trabajo de Tesis y ofrecerse a dirigirlo. Aún a pesar de tener una vida muy ocupada, siempre ha sabido encontrar un momento para discutir y revisar cualquier aspecto referente a la Tesis. Agradecido por sus revisiones y críticas siempre constructivas, que han hecho indudablemente que este proyecto de Tesis sea mejor del que hubiera presentado yo. He tenido la suerte de tener un muy buen maestro y director de Tesis.

Agradecido a **Aida Ribera** tanto por su colaboración activa en todos y cada uno de los puntos desarrollados en esta Tesis, y concretamente por la paciencia demostrada en escuchar mis disquisiciones tanto a nivel laboral como personal.

Agradecido a **Purificació Cascant**, por hacerme entender que lo importante ni se mesura, ni se cuantifica, ni se promedia y que mucho menos se debe al azar. Agradecido a **Gaietà Permanyer** por darme, ya hace muchos años, la oportunidad de trabajar en la Unitat d'Epidemiologia y ser un maestro. Agradecido a **Teresa Fernández** porque siempre tiene una sonrisa y lectura positiva de cualquier cosa.

Agradecido también por la colaboración del Institut Universitari IDIAP Jordi Gol por la ayuda económica en la matrícula del Doctorado y porque siempre ha demostrado que la formación de sus empleados es fundamental. Agradecidos especialmente a las coordinadoras de la USR Lleida: Inés Cruz y Marta Ortega.

Agradecido a Sandra Bertrán, Moisés Gómez, Marta Bofill, David Martínez y otros estadísticos por estar siempre disponibles a discutir diferentes aspectos metodológicos y compartir sus avances en este tema conmigo. Agradecido especialmente a **Lupe Gómez** que siempre ha estado dispuesta a discutir metodológicamente cualquier punto que yo no hubiera podido determinar con esta claridad y rigurosidad.

Agradecido a mi madre, hermana, cuñado, sobrinas y amigos. Agradecido a mi padre Ramón que me retó en su día a emprender este viaje que aquí concluye.

2.INTRODUCCIÓN

2.1. Definición

Se define como Variable de Resultado Combinada (VRC) o en inglés como *Composite Endpoint* o *Combined Endpoint* (CE), la combinación de dos o más sucesos clínicamente relevantes en un único evento que se utilizará como variable de resultado principal. Por ejemplo, una VRC podría ser la combinación de los eventos '*Infarto Agudo de Miocardio (IAM)*' y/o '*Accidente Cerebrovascular (ACV)*' en un único evento. Un individuo que presenta o bien uno de los dos eventos o bien los dos contará para el análisis de la VRC ('*IAM y/o ACV*').

Para la definición de una VRC se consideran una serie requisitos (1,3,5,6). Así, es fundamental que la VRC definida tenga relevancia clínica, tanto de forma global como cada uno de los eventos que la conforman. De esta manera, la combinación de los diferentes eventos debería ser una buena medida de beneficio clínico neto global del tratamiento. En el ejemplo anterior, la combinación de los dos eventos (IAM y/o ACV) debería capturar el efecto global del tratamiento mejor que si hubiera incluido tan solo uno de ellos. La selección de eventos individuales que se combinaran en una VRC dependerá de las características de la enfermedad estudiada y de los objetivos de la investigación, pero también tiene que tener en cuenta el tamaño de muestra que se requiere para contestar a la pregunta de investigación. En este sentido, existen ventajas e inconvenientes inherentes en el uso de VRC (3,7-11). En cualquier caso, los aspectos esenciales son que la VRC debería tratar de capturar los posibles beneficios y riesgos del tratamiento o intervención estudiada, que los eventos combinados deben ser relevantes para el paciente y que deben ocurrir con una incidencia lo suficientemente elevada.

Gómez *et al* (12-14) introdujeron una nomenclatura en la definición de una VRC que consideramos acertada e incorporamos. Se plantea combinar diferentes eventos en una VRC, partiendo de la definición de un '*Evento Relevante*' (**ER**) que constituye al que está fundamentalmente dirigido el efecto de la intervención y, por tanto, define el resultado fundamental. A partir de aquí, el ER se puede combinar con uno o más '*Eventos Aditivos*' (**EA**) que pueden capturar diferentes aspectos de la intervención. Bofill (15) y nosotros mismos (16) hemos utilizado esta nomenclatura en los diferentes trabajos publicados.

Se han utilizado VRC tanto en estudios observacionales(17-19) como experimentales y, especialmente, en ensayos clínicos(20-22). Este último diseño es el que han asumido los principales autores y nosotros mismos para desarrollar los diferentes aspectos de este tipo

de variable de resultado. Imaginémonos un hipotético ensayo clínico donde se pretende evaluar si el tratamiento con doble antiagregación aporta mayor beneficio que el uso de aspirina de forma aislada en pacientes a los que se les ha implantado un stent en una arteria coronaria. El ER podría ser la trombosis del stent, que es lo que pretende evitar la doble antiagregación. EA candidatos a combinarse con el ER serían, por ejemplo, la hemorragia mayor, insuficiencia cardíaca o cualquier otro evento asociado a la patología coronaria en estudio. Nótese que de forma directa o indirecta todos ellos capturan diferentes aspectos asociados a la intervención y, por tanto, están interrelacionados entre sí.

Nota: A partir de este momento utilizaremos el término CE: Composite Endpoint/Combined Endpoint en lugar de VRC: Variable de Resultado Combinada.

2.2. Bases para la construcción e Interpretación de un CE.

En las últimas décadas algunos autores como Montori y Ferreira-González (2,3,23), entre otros (8,24–28), han propuesto las bases sobre cómo construir e interpretar correctamente un CE. Montori y Ferreira-González (2) establecieron un sencillo algoritmo, a partir de tres sencillas preguntas, para establecer la validez de un CE a la hora de extraer conclusiones sobre el efecto de una intervención.

1. ¿Los eventos que forman el CE tienen la misma importancia para el paciente?
2. ¿Los eventos que forman el CE tienen una probabilidad de ocurrencia parecida?

3. ¿La intervención estudiada tiene un efecto parecido sobre cada uno de los eventos que forman el CE?

Idealmente, la variable de resultado utilizada en un estudio debería ser clínicamente relevante, fácilmente medible (sin sesgo asociado) y debería ser sensible a la intervención en estudio (24,28,29). En las tres cuestiones anteriores, se pregunta sobre la frecuencia en la que ocurren eventos y sobre el efecto de la intervención en cada uno de ellos, pero no se hace referencia a la posible asociación entre los eventos combinados.

La combinación de dos o más eventos con frecuencias de ocurrencia distintas provocará que los eventos con mayor probabilidad de ocurrencia dirijan la frecuencia de ocurrencia del CE. De igual forma, si combinamos eventos sobre los que la intervención tiene efectos muy heterogéneos, los eventos en los que la intervención tiene mayor eficacia tendrán una mayor influencia en la eficacia sobre el CE. En todas estas situaciones, será determinante el grado de asociación que existe entre los diferentes eventos que se combinan para formar el CE y, como se demostrará a continuación, éste punto, relativamente inexplorado, es determinante.

2.3. Ventajas e Inconvenientes en el uso de CE

Diferentes autores (2,3,8-10,24,25,28,30,31) sugieren las siguientes ventajas y potenciales inconvenientes en el uso del CE.

2.3.1. Ventajas

Incremento de la potencia estadística

La principal ventaja del uso del CE es la **ganancia de potencia estadística**, ya que el número de eventos esperable será igual o superior al número de eventos que se obtendría utilizando tan solo el ER. Cuantos más eventos combinemos mayor será la potencia estadística, siempre que los efectos de los eventos combinados sean en la misma dirección. En cualquier caso, se debería mantener un equilibrio en cuanto a la complejidad del CE y en cuanto a la ganancia de potencia estadística.

Disponer de un número más elevado de eventos y, por lo tanto, gozar de una potencia estadística superior, permite que el número de individuos en los que investigar la eficacia de cierta intervención sea menor. Por lo tanto, la principal ventaja en el uso de un CE es en la **reducción del Tamaño de Muestra Requerido (TMR)**. Dicha reducción de tamaño de muestra, o bien incremento de la potencia estadística, tiene dos consecuencias posibles (5,6,13,16,25,32-36):

1. Una potencial reducción en el tiempo de seguimiento global para demostrar la hipótesis.
2. Una potencial reducción de los costes asociados a la investigación, lo que podría asegurar la viabilidad de ciertos proyectos.

Algunos autores ya han establecido las bases matemáticas para estimar el tamaño de muestra cuando utilizamos un CE. Por ejemplo Sozu (32), ya en 2010 presentó el enfoque

general sobre como estimar el TMR cuando se combinaban dos eventos binarios en una única variable de respuesta. Cuatro años después Gómez *et al.* (12,14) hicieron lo propio cuando las variables combinadas eran del tipo tiempo hasta evento.

Selección de una variable de resultado cuando existen diferentes alternativas

La utilización del CE como variable de resultado primaria puede solventar el problema de la selección de la variable de resultado principal de un estudio cuando hay diferentes variables candidatas para ser el ER(6).

Beneficio neto

El efecto de la intervención habitualmente se puede reflejar en diferentes dimensiones(37–39). El uso de CE permite combinar eventos de cada una de las dimensiones de interés. Un ejemplo puede ser estimar el efecto de la doble antiagregación, pues en este caso existe un efecto beneficioso asociado a la reducción de eventos isquémicos pero un potencial efecto perjudicial asociado a un aumento en las hemorragias. El efecto de este tipo de fármacos se manifiesta en dos dimensiones: a) ahorro de eventos isquémicos y b) incremento del riesgo de hemorragias. Al combinar los eventos ‘trombosis’ y ‘hemorragia’ en una única variable respuesta se pueden capturar diferentes dimensiones de la intervención, algunas con efecto beneficioso y otras con efecto perjudicial, y de esta manera estimar el beneficio neto de la intervención.

Comparaciones múltiples

El hecho de utilizar diferentes variables de respuesta en vez de utilizar tan solo un CE implica un abordaje diferente desde el punto de vista estadístico en cuanto al contraste de hipótesis. Concretamente, el error tipo I debe repartirse entre el diferente número de contrastes estadísticos (Bonferroni o cualquier otro método de repartición). En el caso de tan solo utilizar una única variable de respuesta, todo el error tipo I se utiliza en la única variable respuesta (4,24,40).

Baja incidencia de eventos

La incidencia de eventos graves en el primer mundo es cada vez más baja(34). La utilización de CE permite incrementar el número de efectivos a los que les sucederá alguno de los eventos y por ende reducir el TMR. El TMR habitualmente es inasumible cuando se utiliza un ER cuya incidencia o efecto son bajos ya que se tiende a determinar eventos muy poco frecuentes o contrastar efectos cada vez más pequeños.

Riesgos Competitivos

La interpretación del posible efecto de un ER es complicada en presencia de riesgos competitivos (3,28,31,41).

Imaginémonos un estudio para determinar el efecto de la doble antiagregación en cuanto a la reducción de eventos isquémicos en pacientes muy añosos en el cual definimos 'evento isquémico' como ER. Si el efecto de la intervención pudiera tener un efecto sobre la 'muerte'

y aquellos pacientes del grupo intervención mueren, el efecto estimado de la intervención sobre el ER quedaría sesgado ya que éstos desaparecerían del estudio sin tener un 'evento isquémico'.

Es habitual y recomendable unir los eventos competitivos con el ER conformando un CE. Más adelante insistiremos en el análisis de este tipo de fenómenos ya que otros autores también lo consideran como una desventaja.

2.3.2. Desventajas

La principal desventaja asociada al uso de CE proviene de potenciales problemas de interpretación de los resultados del estudio, pues una incorrecta interpretación en este contexto suele sobreestimar el efecto real de la intervención en la población(2,23). Que los resultados de un ensayo clínico que utiliza un CE sean interpretables depende de condiciones (2): que los eventos combinados tengan importancia clínica, frecuencia de aparición y efecto parecidos. Por eso, es de suma importancia en la comunicación de los resultados de un estudio, que se reporten la frecuencia de aparición y el efecto sobre cada uno de los componentes del CE (42). Se ha documentado que pocas publicaciones especifican la incidencia, el efecto y el grado de asociación de cada uno los eventos que conforman un CE (4,43).

En cualquier caso, la interpretación de los resultados del estudio será más compleja cuanto mayor es la heterogeneidad de la importancia clínica de los eventos que combinan el CE (2,26,34), cuanto mayor es la heterogeneidad del efecto, incluyendo en esta situación los

casos en los que se combinan eventos sobre los que la intervención tiene efectos de sentido opuesto (42), y cuanto mayor es el número de eventos que se combinan (3).

Uso heterogéneo de CE conocidos: MACE

Uno de los campos de la investigación clínica donde más se han usado los CE ha sido en el contexto de la cardiología. En este contexto es usual utilizar el CE **MACE** (*Major Adverse Cardiovascular Event*), que suele combinar los eventos muerte por motivos cardiovasculares, infarto agudo de miocardio e ictus. En un principio se trata de un CE que tiene sentido, pues un gran número de intervenciones cardiológicas pueden prevenir, de forma directa o indirecta, cada uno de estos eventos. Sin embargo, en una revisión bibliográfica realizada en el 2008 (44), cuyo objetivo fue evaluar la heterogeneidad y validez de los CE utilizados en ensayos clínicos cardiovasculares, se documentó que, a parte de esos tres eventos, se incluían otros muchos eventos incrementándose de forma importante la heterogeneidad del CE. Además, se objetivó que, en la mayor parte de los casos, el efecto de la intervención recaía sobre los eventos de menor relevancia clínica, lo que complicaba todavía más la interpretación.

2.3.3. Diferencia clínicamente relevante

Con el objetivo de estimar el TMR para contestar una hipótesis es usual cuantificar qué diferencia se considera clínicamente relevante (34). El TMR debe ser el mínimo necesario para sostener que la diferencia estimada sea significativa tanto a nivel estadístico como a nivel clínico (45). Una diferencia clínica o efecto muy pequeño requerirá un TMR muy

elevado y de parecida forma una mayor evidencia estadística o mayor confianza en la diferencia o efecto estimado también requerirá un elevado TMR.

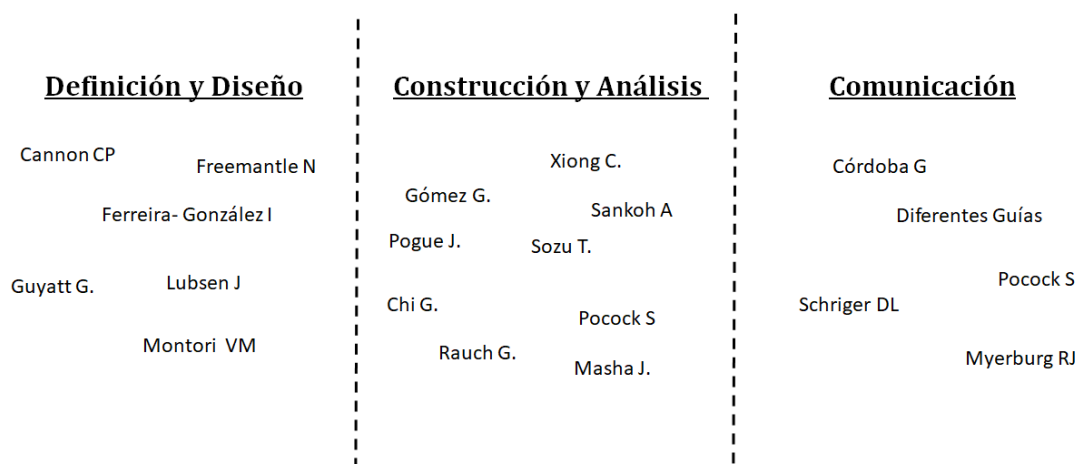
Los clínicos están habituados a determinar qué diferencias o efectos se consideran clínicamente relevantes. No obstante, la combinación de dos o más eventos sobre los que la intervención tiene efectos distintos, no se traduce necesariamente en la suma de efectos(16,32,46,47). De esta manera, el efecto “medio” global de la intervención sobre el CE dependerá del grado de asociación entre los eventos, punto crucial que se tendrá que tener en cuenta a la hora de estimar el TMR. Por lo tanto, cuantificar el efecto que clínicamente se considerará relevante cuando construimos un CE no es trivial, pero sí determinante en la cuantificación del TMR, justificando parte de este trabajo.

La presente Tesis aborda este problema cuando las variables combinadas son categóricas binarias. Otros autores están trabajando en CE de otras naturalezas (48,49) o utilizando destinos abordajes(15).

2.4. Recomendaciones en el uso de CE

Diferentes autores aportan un vasto listado de recomendaciones en la utilización de CE (50) que, según nuestra forma de verlo, se podrían diferenciar en tres etapas: definición y diseño que hacen referencia a aspectos clínicos, una etapa de construcción y análisis de naturaleza metodológica y finalmente una fase de comunicación de los resultados (Figura 1).

Figura 1 - Mapa actual del conocimiento sobre CE



Recomendaciones en su definición y diseño

No se deberían combinar eventos poco importantes desde el punto de vista clínico(3,6), con elevada heterogeneidad en cuanto a su probabilidad de ocurrencia o bien con heterogeneidad del efecto de la intervención sobre ellos(2). Normalmente eso implica no combinar eventos que fisiopatológicamente no se sugiera un efecto de la intervención sobre el evento considerado. En cualquier caso, tanto el CE como todos sus componentes deben ser especificados antes del inicio del estudio, en el momento del diseño. Debemos evitar combinar aquellos componentes que no se asocian o que el efecto de la intervención sea inverosímil, tal y como ya se ha discutido ampliamente.

Recomendaciones en la construcción y análisis

Algunos autores destacan la posibilidad de asignar un peso a cada uno de los eventos que se combinan con el objetivo de determinar la homogeneidad o heterogeneidad en cuanto a

su relevancia y su importancia relativa dentro del constructo (41,50). Diferentes metodologías, recomendaciones y comparación de técnicas han sido realizadas en cuanto al análisis estadístico de los CE (36,38,51-57).

Recomendaciones en la Comunicación de resultados

Como se ha mencionado previamente, deben reportarse los resultados de todos los componentes individuales, como si fueran una variables secundarias para poder valorar la interpretabilidad global del CE (10,39,58). Finalmente, algunos autores (59) recomiendan reportar el grado de asociación entre los componentes del CE.

3.JUSTIFICACIÓN

La toma de decisiones clínicas, eje central de la asistencia sanitaria, se basa en gran medida en el Ensayo Clínico Aleatorizado (ECA), diseño de estudio más utilizado y referenciado en las guías de práctica clínica. En este sentido, en muchos de los ECA relevantes se utiliza como variable de resultado primaria un CE (42,44,68-73,60-67). El uso de CE en el contexto cardiovascular es incluso mayor, ya que el MACE es un CE conocido y habitual (44) que engloba gran parte de las dimensiones en que puede afectar las diferentes intervenciones cardiológicas. Quizás la ventaja más importante en cuanto al uso de CE es la reducción del TMR. Pero esta reducción siempre va acompañada de un incremento en cuanto a la complejidad tanto de análisis estadístico(13,38,78,79,48,51,52,54,74-77) como de interpretación clínica de los resultados(24,42,58).

Si bien la potencial reducción del TMR es una de las ventajas más atribuidas al uso de CE, no existe en la literatura en metodología una aproximación formal al problema del número

óptimo de eventos a combinar para maximizar la eficiencia del CE binario en la reducción del TMR. El problema que se plantea tiene especial relevancia si se considera que la mayoría de las intervenciones que se evalúan en un ensayo clínico pueden tener efectos indeseables en los pacientes y, por tanto, podría conllevar problemas de seguridad. Una aproximación precisa del TMR podría evitar la inclusión innecesaria (fútil) de pacientes para demostrar una hipótesis. La presente Tesis se centrará en el contexto de CE binarios. Se obviarán aquellos CE de cuantifican el tiempo hasta evento porque existen otros investigadores que han establecido los criterios de eficiencia en dicho contexto (12,14).

Aunque la probabilidad de ocurrencia y el efecto de la intervención es bien conocido que afectan al TMR para testar una hipótesis y su influencia ha sido ampliamente estudiada en el contexto de los CE (32), la influencia del grado de asociación entre los eventos que forman el CE en el TMR apenas ha sido explorada. Aunque el concepto 'grado de asociación' o 'grado de relación' o 'interrelación entre eventos'(80,81) es conocido por los investigadores y ha dado lugar a infinidad de bibliografía, su interpretación y cuantificación en el contexto de CE binarios y su influencia en el TMR no se ha estudiado formalmente.

Con la presente Tesis se pretende una aproximación pragmática para la creación de herramientas que objetivamente ayuden a los diferentes profesionales involucrados en el diseño de ECA a crear CE binarios eficientes en cuanto al TMR. Esto ayudaría a mejorar el escenario actual, donde se combinan eventos por diferentes criterios no necesariamente objetivos. Concretamente, se aporta una herramienta que sin duda puede ayudar a la toma de decisiones clínicas basadas en ensayos clínicos. En definitiva, se pretende facilitar a los profesionales una guía que ayudará a discernir entre eventos que deben y no combinarse en función de sus parámetros específicos (probabilidad de ocurrencia, efecto y grado de asociación de los eventos combinados).

4.OBJETIVOS

- Profundizar en el concepto de asociación entre variables binarias. Definir y caracterizar la asociación entre variables binarias, especialmente en el caso de variables binarias que se combinan para formar un CE. Determinar las diferentes formas de estimarla y desarrollar sus características métricas. Comparar las diferentes formas de cuantificar el grado de asociación y determinar si existe algún coeficiente con buenas propiedades para utilizarse en el cálculo del TMR.
- Identificar en qué escenarios el uso de un CE de dos eventos binarios es preferible al uso un único Evento Relevante en cuanto al Tamaño de Muestra Requerido. Determinar en qué sentido y magnitud afectan las variaciones de las incidencias y efectos de los eventos combinados y, en especial, el grado de asociación. Determinar de forma objetiva si el uso de un CE binario es preferible al de un único Evento

Relevante una vez fijados los parámetros que definen los dos eventos y su dependencia (probabilidad de ocurrencia de cada evento, efecto de la intervención sobre cada evento y grado en que se asocian estos eventos.

- Desarrollar una herramienta que determine, a partir de un conjunto de eventos binarios (>2), la combinación óptima para conseguir la máxima reducción del Tamaño de Muestra Requerido

4.1. Estructura del documento de memoria.

Para dar respuesta a los tres objetivos planteados se han realizado tres proyectos independientes entre sí. En el siguiente capítulo de Métodos, se define y desarrollan los conceptos estadísticos y matemáticos comunes a los tres proyectos. No obstante, existen particularidades específicas que deben plantearse en el contexto de cada proyecto. A continuación del capítulo de Métodos, se divide esta memoria en tres apartados de resultados referente a los tres objetivos planteados.

5.MÉTODOS

Hemos dividido este apartado en diferentes secciones. Inicialmente describiremos matemáticamente en la **sección 5.1** los diferentes aspectos que influyen en la definición de un CE, la estimación de eficacia y de la asociación entre variables binarias. En la **sección 5.2** enunciamos en el concepto de asociación entre dos eventos binarios y definimos diferentes métricas. La **sección 5.3** está dedicada a definir los cálculos necesarios para calcular el Tamaño de Muestra Requerido para dos eventos.

5.1. Definiciones, terminología y aspectos matemáticos

5.1.1. Definición de un CE binario

Sea $(X_{ij1}, X_{ij2}, \dots, X_{ijk})$ un vector de k respuestas binarias, donde X_{ijk} es la respuesta del paciente i -ésimo del grupo j para la componente k -ésima. Los posibles valores de las variables son 0 y 1, siendo 0 la ausencia del evento de interés y 1 la presencia de este.

Asumimos que las respuestas son independientemente distribuidas según una distribución bivariada de Bernoulli, con una $E(X_{ijk}) = \pi_{jk}$ y una $V(X_{ijk}) = \pi_{jk} * (1 - \pi_{jk})$, donde j toma valores 1 y 2 (grupo control y grupo intervención) y k toma valores 1 y 2 (ER y EA respectivamente).

Nota: Asumimos para simplificar que solo existen dos grupos de comparación ($j = 1, 2$ o control e intervención) y solo se consideran dos componentes, por lo que el vector de respuestas será (X_{ij1}, X_{ij2}) donde X_{ij1} se define como Evento Relevante (ER) y X_{ij2} como Evento Aditivo (EA). Posteriormente se ampliará la formulación para más de dos componentes.

Definiremos que el CE (X_{ij}^*) ha ocurrido en el paciente al i -ésimo del grupo j -ésimo si cualquier componente del vector de respuestas (X_{ij1}, X_{ij2}) toma valor 1.

$$X_{ij}^* = \begin{cases} 1, & \text{si } X_{ij1} + X_{ij2} \geq 1 \\ 0, & \text{si } X_{ij1} + X_{ij2} = 0 \end{cases}$$

[1]

Fijémonos, que el CE (X_{ij}^*) debe distribuirse también según una distribución de Bernulli que queda completamente definida a partir de la esperanza matemática $E(X_{ij}^*)$:

$$\begin{aligned} E(X_{ij}^*) &= Prob(X_{ij}^* = 1) = Prob((X_{ij1} = 1) \cup (X_{ij2} = 1)) \\ &= Prob(X_{ij1} = 1) + Prob(X_{ij2} = 1) - Prob((X_{ij1} = 1) \cap (X_{ij2} = 1)) \\ &= \pi_{j1} + \pi_{j2} - \pi_{j12} = \pi_j^* \end{aligned}$$

[2]

Donde π_{j12} hace referencia a la probabilidad que ambos sucesos (ER y EA) ocurran durante el seguimiento ($\pi_{j12} = Prob[X_{ij1} + X_{ij2} = 2]$) en el grupo j . Fijémonos que, para estimar la probabilidad del CE en el grupo control, es necesario la probabilidad marginal de cada uno de los eventos y la probabilidad conjunta. Para estimar la probabilidad del CE en el grupo intervención se debe tener en cuenta la eficacia o diferencia clínicamente relevante que se quiere determinar de la intervención para cada uno de los eventos considerados con el objetivo de determinar la probabilidad de ocurrencia de estos eventos en dicho grupo. También es necesario estimar el grado de asociación o probabilidad conjunta en el grupo intervención.

La definición de un CE, cuando se consideran más de dos componentes no es complicada (82,83), aunque la estimación de su función de probabilidad sí lo es. La ecuación [2] presentada para dos componentes se complica rápidamente cuando el número de eventos combinados aumenta(82) al introducir las probabilidades marginales de las K componentes combinadas y de los momentos de orden dos (probabilidad conjunta de cada par de componentes), pero también de los momentos de órdenes superiores. Más adelante se volverá a esta definición.

5.1.2. Efecto de la intervención sobre el CE

La eficacia se puede definir como el efecto que la intervención estudiada tiene en el grupo donde se aplica en comparación con el grupo control. O como la diferencia clínica objetiva entre un grupo al que se ha realizado la intervención en comparación a un grupo control. La eficacia se puede definir para cada una de los componentes considerados ($\theta_{k=1,\dots,K}$) o, una vez construido el CE, se puede definir la eficacia del constructo (θ^*).

Existen diferentes formas para cuantificar el efecto de una intervención sobre un evento binario [3.1 – 3.3]. En la Introducción del trabajo hemos expuesto el concepto de diferencia clínicamente relevante. Dicha diferencia no deja de ser el efecto que tiene el grupo intervención sobre el grupo control, estimado como la diferencia de proporciones. La diferencia de proporciones no es la única forma de cuantificar un efecto. A continuación, se definen las formas más habituales de cuantificar el efecto de una intervención sobre un evento binario y sobre un CE:

- Riesgo Absoluto o Diferencia de proporciones: $\theta_k^{RA} = \pi_{2k} - \pi_{1k}$

$$\theta^{*,RA} = \pi_2^* - \pi_1^* \quad [3.1]$$

- Riesgo Relativo o Razón de proporciones: $\theta_k^{RR} = \frac{\pi_{2k}}{\pi_{1k}}$

$$\theta^{*,RR} = \frac{\pi_2^*}{\pi_1^*} \quad [3.2]$$

- Odds Ratio o Razón de ventajas: $\theta_k^{OR} = \frac{\frac{\pi_{2k}}{1-\pi_{2k}}}{\frac{\pi_{1k}}{1-\pi_{1k}}}$

$$\theta^{*,OR} = \frac{\frac{\pi_2^*}{1-\pi_2^*}}{\frac{\pi_1^*}{1-\pi_1^*}} \quad [3.3]$$

Aunque estos son los parámetros que se utilizan habitualmente en investigación clínica o epidemiología para cuantificar el efecto, existen otros parámetros utilizados en otros campos, como en salud pública, como es el número necesario de pacientes a tratar para evitar un evento (NNT). No obstante, en este trabajo no se utilizarán, puesto que cada parámetro es función de los otros si se conoce la probabilidad en que ocurren los eventos o CE en cada grupo, como se demostrará a continuación.

A partir de las dos probabilidades de ocurrencia del evento k-ésimo (π_{k1}, π_{k2}) o las probabilidades del CE (π_1^*, π_2^*) o también a partir de cualquier coeficiente de eficacia y una de las dos probabilidades de ocurrencia (θ_k, π_k) o (θ^*, π^*), se definen los tres coeficientes de eficacia de forma unívoca. Lo que se comenta a continuación es válido para el efecto de cada uno de los eventos combinados y para el CE, por lo que lo que se trata de forma genérica.

Consideramos que el uso de la diferencia de proporciones o Riesgo Atribuible (RA) no es aconsejable para estimar el efecto de una intervención con un CE, ya que la interpretación de diferencia de proporciones no es válida sin tener en cuenta la probabilidad de ocurrencia del evento o CE. La diferencia de proporciones definida en [3.1] está siempre acotada entre $[-\pi_{k1}, 1 - \pi_{k1}]$, siendo π_{k1} la incidencia marginal o incidencia en el grupo control, $[-\pi_1^*, 1 - \pi_1^*]$ en el caso que la diferencia sea de la proporción de CE en el grupo control e intervención. Por ejemplo, una diferencia de proporciones del 1% no implica ni un efecto grande ni pequeño sin conocer la probabilidad de ocurrencia. Será un efecto grande si la probabilidad marginal es del 3% y será pequeña si la probabilidad es del 20%.

El Riesgo Relativo (RR) o razón de proporciones, consideramos que es mejor parámetro, desde el punto de vista matemático y clínico, para cuantificar el efecto de una intervención sobre un evento o sobre el CE. Retomando el ejemplo anterior, una diferencia de proporciones del 1%, corresponde a un RR = 0.666 cuando la probabilidad marginal es del 3% y un RR = 0.95 cuando la probabilidad marginal es elevada (20%). De esta manera, el efecto de una intervención puede clasificarse como grande o pequeño sin tener en cuenta la proporción de pacientes a los que ocurra el evento o CE. El rango de valores del RR no depende de la probabilidad marginal como en el caso de la diferencia de proporciones y siempre se define entre 0 e infinito.

Finalmente, la Odds Ratio (OR), o razón de ventajas, aun siendo un parámetro de eficacia muy utilizado por sus propiedades matemáticas tiene una interpretación difícil para la mayoría de los clínicos. En el ejemplo que hemos introducido, las OR serían 0.660 en el caso de una elevada eficacia y 0.934 en el otro caso. Como en el caso del RR, su rango siempre está definido entre 0 e infinito.

Por los motivos anteriormente expuestos, hemos decidido utilizar el Riesgo Relativo en el desarrollo de esta Tesis como indicador de eficacia. No obstante, a continuación se presentan las funciones [3.4 – 3.6] donde se comprueba aritméticamente que, a partir de cualquier coeficiente de eficacia y de la probabilidad marginal, se puede estimar cualquier otro de los coeficientes definidos (84):

$$\theta_k^{RA} = \pi_{k2} - \pi_{k1} = \dots = \pi_{k1}(\theta_k^{RR} - 1) = \pi_{k1} \left(\frac{\theta_k^{OR}}{1 - \pi_{k1} + \theta_k^{OR}\pi_{k1}} - 1 \right) \quad [3.4]$$

$$\theta_k^{RR} = \frac{\pi_{k2}}{\pi_{k1}} = \dots = \frac{\theta_k^{RA} + \pi_{k1}}{\pi_{k1}} = \frac{\theta_k^{OR}}{1 - \pi_{k1} + \theta_k^{OR}\pi_{k1}} \quad [3.5]$$

$$\theta_k^{OR} = \frac{\frac{\pi_{k2}}{1 - \pi_{k2}}}{\frac{\pi_{k1}}{1 - \pi_{k1}}} = \dots = \frac{\theta_k^{RR}(1 - \pi_{k1})}{(1 - \theta_k^{RR}\pi_{k1})} = \frac{\theta_k^{RA}}{\pi_{k1}(1 - \theta_k^{RA} - \pi_{k1})} + 1 \quad [3.6]$$

En definitiva, los diferentes coeficientes de eficacia son deducibles una vez fijada la probabilidad de ocurrencia en uno de los grupos [3.3 – 3.6]. Por lo tanto, por comodidad a partir de este momento estimaremos la eficacia utilizando el Riesgo Relativo ($\theta_k^{RR} = RR_k$).

5.1.3. Grado de asociación entre variables binarias

La definición de asociación entre dos variables ha sido discutida ampliamente, incluso en el caso de variables binarias (85–87). Dos variables cualquiera X_1 y X_2 están asociadas cuando la variabilidad de una puede ser explicada por la otra y cuando se sostenga que la

covarianza entre $f(X_1, X_2)$ y $g(X_1, X_2)$ para cualquier función f y g no decreciente sea positiva (88,89). Esta definición puede simplificarse cuando las variables asociadas son binarias, y enunciarse entonces en función de la distribución condicional. Así, dos variables estarán asociadas si la distribución de una condicionada a la otra no coincide con la distribución la marginal de ésta [4].

$$Prob(X_{ij1}|X_{ij2}) \neq Prob(X_{ij1}) \quad [4]$$

5.2. Grado de asociación

5.2.1. Definición

Aunque en la ecuación [2] hemos definido de forma informal que el grado de asociación entre dos variables binarias se puede estimar a partir de la probabilidad conjunta, este parámetro no se ha definido formalmente. Diremos que dos variables binarias están asociadas si su covarianza no es nula [5], y de allí se desprende que la probabilidad conjunta es un indicador de asociación *per se* una vez determinadas las probabilidades de ocurrencia. Como en el anterior punto, se considerará que dos variables binarias están asociadas si su distribución condicionada no coincide con la marginal [4]:

$$COV(X_{ij1}, X_{ij2}) = Prob((X_{ij1} = 1) \cap (X_{ij2} = 1)) - Prob(X_{ij1} = 1)Prob(X_{ij2} = 1) \quad [5]$$

En [4], la $Prob((X_{ij1} = 1) \cap (X_{ij2} = 1))$ corresponde a π_{j12} . Por lo tanto, se considerará un grado de asociación positivo cuando $Prob((X_{ij1} = x) \cap (X_{ij2} = y)) > Prob(X_{ij1} = x)Prob(X_{ij2} = y)$, que es el producto de las probabilidades marginales ($\pi_{j1} * \pi_{j2}$), para cualquier $x, y = 0,1$. Se considerará asociaciones negativas en caso contrario.

Por lo anterior se desprende que la asociación entre dos variables puede ser positiva, negativa, o bien nula en el caso de eventos sin asociación o eventos independientes. Clínicamente dos eventos están asociados positivamente cuando es más probable la ocurrencia de un evento, sabiendo que ha ocurrido el otro. La asociación negativa se define cuando el hecho que ocurra un evento reduce la probabilidad de ocurrencia del otro. Dos eventos son independientes o no están asociados si el riesgo de ocurrencia de un evento no se ve modificado por si el otro ha ocurrido o no.

5.2.2. Coeficiente de Correlación

En general, un buen indicador del grado de asociación lineal es la **correlación de Pearson**(81) (ρ), ya que para variables numéricas se distribuye siempre en un rango entre -1 y +1, correspondiendo valores positivos a asociaciones positivas, correlaciones negativas a asociaciones negativas y una correlación igual a cero indica el escenario en que dos eventos no están relacionados entre sí. Sin embargo, se considera que la correlación de Pearson no es un buen indicador para estimar el grado de asociación entre dos variables binarias ya que, como se comprobará más adelante:

- a) El rango de la correlación no siempre está definido entre -1 y +1 cuando se utiliza en variable categóricas binarias
- b) Una correlación máxima (tanto negativa como positiva) no necesariamente corresponde a correlaciones próximas a -1 o +1.
- c) Una correlación igual a cero no implica una probabilidad conjunta igual a cero, solo indica que los eventos no están asociados.
- d) La correlación puede ser negativa, siendo la probabilidad conjunta siempre definida positiva.

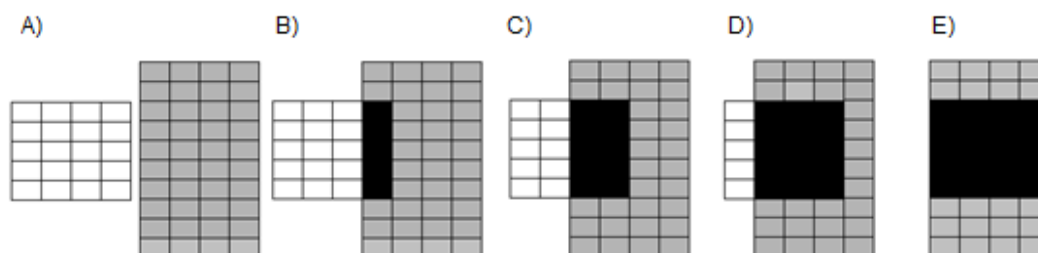
La correlación es tan solo uno de los posibles indicadores para estimar el grado de asociación (φ) entre dos eventos. Otro coeficiente para estimar el grado de asociación que ya se ha presentado es la probabilidad conjunta, pero estos no son los únicos como se verá posteriormente.

El grado de asociación entre dos eventos corresponde a estimar los momentos de orden dos de una distribución bivariada. Cuando se combinan más de dos eventos se puede observar que existen otros indicadores de orden superior. Bahadur y otros (82,83,90) trabajaron en profundidad la distribución de k variables binarias. Cuantificaron que para una distribución k -variada existen $2^k - k - 1$ coeficientes de asociación y que son necesarios $2^k - 1$ parámetros para estimar su distribución de forma precisa.

5.2.3. Intervalo de valores posibles

La probabilidad conjunta $\left(\pi_{j12} = Prob\left((X_{ij1} = 1) \cap (X_{ij2} = 1)\right)\right)$ es un indicador de asociación, ya que cuantifica el porcentaje de pacientes a los que les ocurren los dos eventos y es función de la covarianza cuando las proporciones de cada evento son fijas [4]. Como indicador de asociación si $\pi_{j12} > \pi_{j1}\pi_{j2}$ se considera asociación positiva, y $\pi_{j12} < \pi_{j1}\pi_{j2}$ se considera asociación negativa. En la **Figura 2** podemos observar que cuanto mayor es el grado de asociación mayor es la probabilidad conjunta (área de recuadros negros) y menor es la incidencia del CE construido por el ER (recuadros blancos) y el EA (recuadros grises). Como se puede observar, en el escenario A) de la Figura 2 la probabilidad conjunta es la mínima posible y en el escenario E) la probabilidad conjunta es la máxima posible.

Figura 2 – Ejemplo de posibles escenarios con diferentes probabilidades conjuntas en la intersección de dos eventos en un ensayo clínico.



Ejemplo de 5 posibles escenarios con diferente probabilidad conjunta al construir un CE. Cada recuadro representa un 1% de probabilidad. Los recuadros blancos representan la probabilidad del ER (p.ej.: muerte, $\pi_{ER,1} = 0.20$) y los recuadros grises representan la probabilidad del EA (p.ej.: infarto agudo de miocardio, $\pi_{EA,1} = 0.40$) para el grupo control ($j=1$). La $\pi_{ER-EA,1}$ en cada uno de los escenarios es 0.00 (A), 0.05 (B), 0.10 (C), 0.15 (D) y 0.20 (E) para el grupo control, por lo que la probabilidad del CE según la ecuación [2] es de 0.60, 0.55, 0.50, 0.45 y 0.40, respectivamente.

La Figura 2 presenta que la probabilidad conjunta, cuando se combinan dos eventos con una cierta probabilidad de ocurrencia, queda delimitada. Concretamente se puede observar que, en el caso de la Figura 2, la probabilidad conjunta está acotada entre 0 y 0.20. Fréchet(91) en 1960 indicó que la probabilidad conjunta cuando se combinan dos eventos discretos está acotada dentro de los siguientes límites:

$$[\max\{0; \pi_{j1} + \pi_{j2} - 1\} \leq \pi_{j12} \leq \min\{\pi_{j1}, \pi_{j2}\}] \quad [6]$$

La probabilidad conjunta, al tratarse de una probabilidad debe estar limitada entre 0 y 1. No obstante, cuando ésta se aplica a un CE binario para la cuantificación del grado de asociación está limitada por los límites de Fréchet. Que la probabilidad conjunta esté

acotada implica que la probabilidad de ocurrencia del CE [2], también quede limitada y, por lo tanto, también la de la probabilidad conjunta en el grupo intervención. Conociendo la probabilidad del CE en el grupo control y en el grupo intervención se puede calcular el TMR que estará acotado también.

La probabilidad conjunta se define en cada uno de los grupos ($j=1,2$). Habitualmente ni la probabilidad conjunta ni el rango de Fréchet coincidirán entre el grupo control e intervención incluso asumiendo el mismo grado de asociación (ver posteriores métricas). En el ejemplo planteado en la Figura 2, $\pi_{1,12} \in \{0.00, 0.20\}$, siendo una $\pi_{1,12} > 0.08$ considerada una asociación positiva y viceversa. Si para estos dos eventos considero los hipotéticos efectos (p.ej. $RR_{ER} = 0.75$ y $RR_{EA} = 0.50$) respectivamente, utilizando [3.2] se puede determinar que $\pi_{ER,2} = 0.15$ y $\pi_{EA,2} = 0.20$, por lo que $\pi_{2,12} \in \{0.00, 0.15\}$, considerándose una $\pi_{2,12} > 0.03$ una asociación positiva.

De los límites de Fréchet [6] y la definición de CE a partir de dos eventos [2], se deduce que la probabilidad de ocurrencia del CE estará acotada entre un límite inferior y superior una vez conocida la probabilidad de ocurrencia de cada evento, como se presenta en [7].

$$[\max\{\pi_{j1}, \pi_{j2}\} \leq \text{Prob}(X_{ij}^* = 1) \leq \min\{\pi_{j1} + \pi_{j2}, 1\}] \quad [7]$$

Retomando la Figura 2, el límite superior de [7] corresponde al escenario A) donde los dos eventos son disjuntos (si los representamos los eventos como conjuntos) o $\pi_{j12} = \max\{0, \pi_{j1} + \pi_{j2} - 1\}$. Paralelamente el escenario E) correspondería al límite inferior de

[7], que corresponde al caso donde todos los pacientes con el evento menos probable también experimentan el otro evento o $\pi_{j12} = \min\{\pi_{j1}, \pi_{j2}\}$.

5.2.4. Otras métricas de interés

A continuación, se presentan diferentes características que se utilizarán para comparar y caracterizar las diferentes formas de cuantificar la asociación entre dos variables binarias.

Cuantificación de la asociación para covarianza nula o independencia entre eventos.

Partiendo de la definición de covarianza [5] determinamos qué valor de la probabilidad conjunta (como uno de los parámetros que estiman la asociación) identifica asociaciones positivas o negativas. Este punto de corte en el caso de la probabilidad conjunta es conocido [8].

$$COV(X_{ij1}, X_{ij2}) = 0 \leftrightarrow \pi_{j12} = \pi_{j1}\pi_{j2} \quad [8]$$

En forma general, para otros indicadores de asociación, la cuantificación del punto de corte depende de: a) la propia definición del coeficiente y b) de las probabilidades marginales en algunos casos. Por ejemplo, en el caso de la correlación de Pearson, el punto de corte no depende de las marginales. La correlación de Pearson se define $\rho_{j12} = \frac{COV(X_{ij1}, X_{ij2})}{\sqrt{V(X_{ij1})V(X_{ij2})}} =$

$$\frac{\pi_{j12} - \pi_{j1}\pi_{j2}}{\sqrt{\pi_{j1}(1-\pi_{j1})\pi_{j2}(1-\pi_{j2})}} \quad \text{Se comprueba que: } COV(X_{ij1}, X_{ij2}) = 0 \leftrightarrow \rho_{j12} = 0$$

independientemente el valor de la probabilidad de cada evento.

Para cualquier coeficiente que se puedan definir para cuantificar el grado de asociación entre dos variables binarias, se puede deducir un punto que discrimina entre asociaciones positivas y negativas.

Intervalo de Fréchet y Rango

De la misma forma que se define que para la probabilidad conjunta existe un límite inferior y superior(91), éstos límites se pueden determinar en cualquier coeficiente utilizado en la cuantificación del grado de asociación. Cualquier parámetro definido mediante una función continua, cuando se fije el límite inferior (máxima asociación negativa) y el límite superior (máxima asociación positiva), quedará definido tanto el intervalo como el rango. Definimos el rango entre los límites superior e inferior como:

$$R(\varphi) = Sup(\varphi) - Inf(\varphi) \quad [9]$$

, donde $Sup(\varphi)$ corresponde al límite superior del intervalo definido para el coeficiente φ al aplicar los límites definidos por Fréchet e $Inf(\varphi)$ corresponde al límite inferior del intervalo definido para el coeficiente φ .

Por ejemplo, el límite inferior i superior para la correlación, se deduce en el trabajo de Sozu *et al.* (32):

$$\begin{aligned}
 Inf(\rho_{j12}) &= \max \left\{ -\sqrt{\frac{\pi_{j1}\pi_{j2}}{(1-\pi_{j1})(1-\pi_{j2})}}; -\sqrt{\frac{(1-\pi_{j1})(1-\pi_{j2})}{\pi_{j1}\pi_{j2}}} \right\} \\
 Sup(\rho_{j12}) &= \min \left\{ \sqrt{\frac{\pi_{j1}(1-\pi_{j2})}{\pi_{j2}(1-\pi_{j1})}}, \sqrt{\frac{\pi_{j2}(1-\pi_{j1})}{\pi_{j1}(1-\pi_{j2})}} \right\} \quad [10]
 \end{aligned}$$

Coefficiente estandarizado

Definimos el coeficiente de asociación estandarizado, que servirá para estudiar en profundidad como se distribuyen los diferentes coeficientes utilizados para estimar el grado de asociación, como:

$$Est(\varphi) = \frac{\varphi - Inf(\varphi)}{R(\varphi)} \quad [11]$$

A partir de [11] observamos que el coeficiente estandarizado toma valor 0 cuando el coeficiente toma el valor mínimo o es igual a $Inf(\varphi)$, y 1 cuando este tome el valor máximo $Sup(\varphi)$. Si el rango lo representamos como un segmento con inicio en $Inf(\varphi)$ y final en $Sup(\varphi)$, el coeficiente estandarizado nos localiza el grado de asociación como el recorrido porcentual de dicho segmento.

Si retomamos el ejemplo de la Figura 2 y determinamos el coeficiente estandarizado para la probabilidad conjunta que determina independencia de eventos ($\pi_{1,12} = 0.08$), determinamos a partir de [11] que $Est(\pi_{1,12}) = \frac{0.08-0.00}{0.20} = 0.4$, por lo tanto la asociación está en el 0.4 (40%) del recorrido de un segmento que va entre $Inf(\pi_{1,12})$ y $Sup(\pi_{1,12})$. El escenario de eventos independientes en el caso de la probabilidad conjunta no corresponde al punto medio del segmento.

5.3. Tamaño de Muestra Requerido

En general, el Tamaño de Muestra Requerido depende de diferentes factores, como el efecto que se quiera detectar (RR_k), la probabilidad en que ocurre el evento (π_{jk}), el nivel de significación (error tipo I o α) y la potencia estadística (error tipo II o $1 - \beta$) que se precise. Otros factores pueden influir en el cálculo del TMR, como puede ser el tipo de contraste de hipótesis que se plantea, el tipo de aproximación en la estimación del propio TMR o bien una corrección por comparaciones múltiples.

Imaginemos que planteamos un contraste de hipótesis de igualdad de proporciones para un ER. Otro contraste de hipótesis, no equivalente, puede ser el de igualdad de proporciones para un CE conformado por la combinación del ER y otro EA candidato, tal y como se presenta en [12].

$$\begin{cases} H_0: \pi_{2,RE} = \pi_{1,RE} \\ H_1: \pi_{2,RE} \neq \pi_{1,RE} \end{cases} \quad \begin{cases} H_0^*: \pi_{2,CE} = \pi_{1,CE} \\ H_1^*: \pi_{2,CE} \neq \pi_{1,CE} \end{cases} \quad [12]$$

Una forma de estimar el TMR es utilizando la aproximación a la distribución Normal para un contraste binomial (46,92), y puede ser estimado tanto para el TMR del ER [13] como del CE [14]. Otras aproximaciones al cálculo de TMR pueden ser aplicadas como el método del arcoseno (93) (que proporciona estimaciones más precisas basadas en el cálculo de la probabilidad exacta de Fisher para proporciones pequeñas) o las correcciones por continuidad de Yate's(94) para poblaciones pequeñas. En el trabajo de Sozu *et al.* (32) se pueden consultar dichas aproximaciones.

$$n_{ER} \geq \left(\frac{Z_{\alpha/2} \sqrt{\pi_{ER}(1-\pi_{ER})} + Z_{\beta} \sqrt{\pi_{ER,1}(1-\pi_{ER,1})\pi_{ER,2}(1-\pi_{ER,2})}}{\pi_{ER,1} - \pi_{ER,2}} \right)^2, \pi_{ER} = \frac{\pi_{ER,1} + \pi_{ER,2}}{2} \quad [13]$$

Estimar el TMR en el caso del CE (n^*)[14] implica estimar, a partir de la definición de CE [2], la probabilidad del CE en el grupo control ($\pi_{j=1}^*$) y en el grupo intervención ($\pi_{j=2}^*$). Para la probabilidad en el grupo control ($\pi_{j=1}^*$) es necesario utilizar la probabilidad marginal de cada uno de los eventos y su probabilidad conjunta: $\pi_{j=1,1}, \pi_{j=1,2}, \pi_{j=1,12}$, respectivamente tal y como se define en [2]. La probabilidad del CE en el grupo intervención ($\pi_{j=2}^*$) implica estimar la probabilidad de cada uno de los eventos a partir del efecto esperado de la intervención y estimar la probabilidad marginal de cada uno de ellos según la ecuación [3.2]. En este caso, la probabilidad conjunta de ambos eventos en el grupo intervención deberá estimarse. En el presente trabajo, por simplicidad asumimos que el grado de asociación entre los eventos es el mismo en el grupo intervención y en el grupo control (Anexo 1). Esta asunción, por otra parte, es razonable, pues lo contrario comportaría que la intervención influye en el grado de asociación entre los eventos lo cual, aunque puede ser cierto, no es esperable en la mayoría de los casos. Es importante notar que asumir el mismo 'grado de

asociación' no significa la misma probabilidad conjunta, este punto lo desarrollaremos más adelante.

$$n^* \geq \left(\frac{Z_{\alpha/2} \sqrt{\pi^*(1-\pi^*)} + Z_{\beta} \sqrt{\pi_1^*(1-\pi_1^*)\pi_2^*(1-\pi_2^*)}}{\pi_1^* - \pi_2^*} \right)^2, \pi^* = \frac{\pi_1^* + \pi_2^*}{2} \quad [14]$$

El uso del CE será preferible al uso de tan solo un ER si se determina que el TMR es inferior al requerido al utilizar un ER(14). En este caso estamos asumiendo que los eventos combinados son escogidos de forma adecuada. Recordemos que, en la Introducción de esta Tesis, se listaron las ventajas e inconvenientes y entre ellos existen recomendaciones para seleccionar los componentes. Por ejemplo, se recomienda combinar eventos cuya probabilidad de ocurrencia, efecto de la intervención e importancia para los pacientes sean parecidos. En esta Tesis nos centraremos en la decisión de combinar o no diferentes eventos en cuanto al TMR y asumimos que los eventos combinables son adecuados.

En el caso que la Razón de Tamaños (*Sample Size Ratio, en inglés*) ($SSR = \frac{n_{ER}}{n^*}$) sea superior a 1 será preferible utilizar el CE a utilizar el ER mientras que, en caso contrario, no será eficiente utilizar el CE. Nótese que la eficiencia la estamos definiendo solo desde el punto de vista del número de pacientes a incluir para demostrar un efecto de la intervención, y no tenemos en cuenta otras consideraciones sobre la conveniencia o no conveniencia que implica utilizar un CE en vez de utilizar un único ER.

A partir del cálculo del TMR se puede estimar la potencia estadística (*pw* como acrónimo de *power* en inglés) [15]. La potencia estadística queda determinada una vez especificado el TMR, sea n_{ER} o n^* .

$$pw^* = 1 - \beta = P\left(X \leq \frac{\sqrt{n^*}(\pi_1^* - \pi_2^*) - Z_{\alpha/2} \sqrt{2\pi^*(1-\pi^*)}}{\sqrt{\pi_1^*(1-\pi_1^*)\pi_2^*(1-\pi_2^*)}}\right), \pi^* = \frac{\pi_1^* + \pi_2^*}{2} \quad [15]$$

, donde X es la distribución Normal estandarizada.

La función para estimar la potencia estadística [15] se ha especificado para el CE, pero sin mucha dificultad se puede derivar al uso de un único ER. Dicha función depende directamente de las probabilidades marginales, del efecto y del grado de asociación de cada evento cuando estimamos la potencia estadística del CE. Por lo tanto, matemáticamente se demuestra que, cuando se define un CE a partir de dos eventos y se determinan sus probabilidades marginales y los efectos esperados de la intervención, tanto la potencia estadística como el TMR dependerán del grado de asociación. También que el grado de asociación está acotado entre un límite inferior y uno de superior según los límites descritos por Fréchet. De lo que se deduce que cuando construimos un CE y lo comparamos con un ER, tanto el TMR como el SSR, como la potencia estadística, quedan acotados tanto inferior como superiormente.

Dado un par de probabilidades marginales y un par de efectos esperados, podemos calcular le SSR en todo el rango de asociaciones posibles. Si el intervalo de SSR queda enteramente encima de 1 en ninguna situación será aconsejable utilizar el CE en lugar del ER. En el caso contrario, en el que todo el intervalo del SSR esté debajo del 1, será recomendable su uso independientemente del grado de asociación entre los eventos combinados. El caso intermedio es el caso más interesante desde el punto de vista clínico, pues es entonces

cuando será recomendable el uso del CE solo para algunos grados de asociación y no para otros.

El hecho de que el TMR esté acotado superior e inferiormente es porque también el grado de asociación está acotado y queda delimitado por la probabilidad marginal de cada uno de los eventos combinados. En este caso, también el efecto del CE quedará acotado según el rango determinado por el grado de asociación (Anexo 2).

- ✓ La probabilidad de ocurrencia del CE se puede estimar a partir de la probabilidad de cada evento y la probabilidad conjunta de ambos eventos combinados.
- ✓ El grado de asociación se puede estimar utilizando la probabilidad conjunta.
- ✓ El efecto de la intervención sobre cada evento permite estimar la probabilidad de ocurrencia del CE tanto en el grupo control como en el grupo intervención.
- ✓ Aplicando los límites de Fréchet tanto la probabilidad conjunta como la probabilidad de ocurrencia del CE queda acotada inferior y superiormente.
- ✓ El TMR queda acotado inferior y superiormente a partir de la probabilidad de ocurrencia acotado tanto en el grupo control como intervención.
- ✓ En el caso que el TMR del CE sea inferior al TMR del ER, su uso será preferible al del ER desde el punto de vista del número de individuos a estudiar.

6. Cuantificación de la asociación entre eventos binarios.

El primer objetivo planteado para este proyecto es:

“Profundizar en el concepto de asociación entre variables binarias. Definir y caracterizar la asociación entre variables binarias. Determinar las diferentes formas de estimarla y desarrollar sus características métricas. Comparar las diferentes formas de cuantificar el grado de asociación y determinar si existe alguna con buenas propiedades para utilizarse en el cálculo del TMR. “

6.1. Justificación

La cuantificación de la asociación entre dos eventos es importante para el cálculo del TMR en un CE(32,36). El TMR depende del grado de asociación, siendo mayor o menor en función de si los eventos combinados están más o menos asociados(95). Para conocer en profundidad la potencial reducción del TMR cuando combinamos diferentes eventos es necesario conocer también en profundidad el concepto de asociación. Habitualmente se utiliza la correlación, cuya definición es demasiado robusta para cuantificar el grado de asociación entre dos variables binarias. La cuantificación de la correlación no es trivial en el caso de estimarla para evaluar la interrelación entre eventos binarios. El concepto de asociación entre dos variables binarias no es original de este proyecto de Tesis y ya ha sido discutido en profundidad (89,96), y también su cuantificación, especialmente en el trabajo realizado por Warrens (85,86,97) en su proyecto de Tesis.

El grado de asociación entre dos eventos binarios puede ser cuantificado utilizando decenas de parámetros, que normalmente toman valores entre el -1 y 1 como en el caso de la correlación o bien entre 0 y 1 como en el caso de la probabilidad conjunta. Valores cercanos al límite inferior corresponden a asociaciones negativas intensas y en caso contrario a asociaciones positivas intensas.

A modo de ejemplo, y para ilustrar el desarrollo de este proyecto, utilizaremos los resultados del estudio CARE(98), donde durante 5 años se siguió un total de 4.159 pacientes. Este estudio definió el CE como la unión de '*IAM fatal*' y '*IAM no fatal*'. Al final del seguimiento se registraron 62 pacientes con un '*IAM fatal*' y 308 con '*IAM no fatal*', 6

pacientes registraron ambos eventos. Por lo tanto, el número de pacientes con el CE fue de 364, correspondiendo una correlación de $\rho = 0,0107$.

Otro ejemplo es el estudio PARADIGM-HF(60) en que se incluyeron 8.399 pacientes. El evento relevante '*muerte por causas cardiovasculares*' se objetivó en 1.251 pacientes y el evento adicional '*hospitalización por insuficiencia cardíaca*' en 1.195 y el CE de ambos eventos a 2.031 pacientes. Es fácil cuantificar que 415 pacientes tuvieron ambos eventos, estimándose una correlación de $\rho = 0,227$. Aunque el coeficiente de correlación es superior en el caso del estudio PARADIGM-HF que en el caso del estudio CARE, este parámetro para un clínico es muy poco informativo o intuitivo.

Otra forma de cuantificar la asociación, diferente a la correlación, es la probabilidad conjunta que ya hemos definido anteriormente. La probabilidad condicionada también es un buen indicador del grado de asociación entre dos variables binarias. La probabilidad condicionada se define como la probabilidad que suceda un evento sabiendo que ha ocurrido otro. Si calculamos la probabilidad de que sucedan ambos eventos (probabilidad conjunta) sabiendo que ha ocurrido algún evento (por ejemplo, el ER o el EA o el mismo CE) el resultado será una forma de cuantificar el grado de asociación. En el estudio CARE, por ejemplo, tan solo el 1,65% de los pacientes que tuvieron el CE tuvieron ambos eventos ($6/364 = 0,0165$). Esta probabilidad condicionada parece más interpretable que la correlación. La probabilidad condicionada en el estudio PARADIGM-HF fue del 20,43%. La probabilidad conjunta y la probabilidad condicionada no son el mismo parámetro, no deben confundirse.

En otros contextos, como la zoología, genética y ecología se utilizan diferentes indicadores llamados *medidas de similitud* que pueden ser útiles en la cuantificación del grado de asociación entre componentes binarios de un CE. En la siguiente sección profundizamos en el concepto de asociación entre variables binarias con el objeto de estudiar las diferentes formas de cuantificarlo y sus características. Querremos determinar si, en función de las características métricas de cada coeficiente, alguno de ellos puede ser más apropiado para cuantificar el grado de asociación necesario para estimar el TMR.

6.2. Medidas de similitud

Las medidas de similitud expresan la relación entre dos variables en un número que resume su grado de asociación(99). A partir de este momento, a las medidas de similitud les llamaremos Coeficientes de Similitud (CS). En el siglo pasado se definieron decenas de CS y actualmente aún se definen nuevos indicadores o variaciones de los ya existentes. No se ha realizado una búsqueda exhaustiva de todos los CS, ya que no es el objetivo del proyecto. En el Anexo 3 se presentan 40 CS como resumen de la Tesis del Dr. M.J. Warrens (85–87,97): “*Similarity Coefficients for Binary Data*”, aunque otros CS pueden consultarse en otros trabajos (99).

Antes de ahondar en los diferentes CS se debe comentar que nos vamos a referir a la cuantificación del grado de asociación entre dos únicas variables binarias, por simplicidad. En el caso que el CE estuviera definido por K eventos, deberíamos estimar $\frac{k(k-1)}{2}$ CS que conforman el número de combinaciones de los K eventos cogidos de dos en dos. También se debe indicar que el grado de asociación entre dos eventos puede cuantificarse en cada uno de los grupos de estudio.

Estudiando los CS listados en el Anexo 3, se observa que muchos de estos indicadores se alejan del propósito de cuantificar de forma clínica el grado de asociación entre dos eventos binarios. Poseen propiedades que internamente cuantifican de forma robusta la similitud entre variables(87), pero en contextos específicos que se alejan del propósito de esta Tesis. Se ha seleccionado arbitrariamente 8 CS incluyendo la correlación. Se han seleccionado teniendo en cuenta su interpretación clínica y si se pueden asociar a alguna magnitud conocida. También se ha seleccionado en función de la dificultad en la cuantificación o bien la importancia en los estudios de similitud. Por lo tanto, la decisión de incluir o no un CS concreto se puede considerar que ha sido por convivencia. El Riesgo Relativo [3.2] y la Odds Ratio [3.3], que de alguna manera también se pueden interpretar como SC, se han obviado porque se usan habitualmente para la cuantificación del efecto en nuestro contexto.

6.2.1. Coeficiente de correlación (ρ_j)

La correlación de Pearson se define de forma general como $\rho_j = \frac{COV(X_{ij1}, X_{ij2})}{\sqrt{V(X_{ij1})V(X_{ij2})}}$, estimándose

para variables binarias como $\rho_j = \frac{\pi_{j12} - \pi_{j1}\pi_{j2}}{\sqrt{\pi_{j1}(1-\pi_{j1})\pi_{j2}(1-\pi_{j2})}}$. Examinando la ecuación anterior es

fácil determinar que cuando $\pi_{j12} = \pi_{j1}\pi_{j2}$ el valor de $\rho_j = 0$ y, por lo tanto, las variables X_{ij1} y X_{ij2} no están asociadas ya que su covarianza es nula [5]. Contrariamente, para $\rho_j \neq 0$ dichas variables están asociadas. Cuando $\pi_{j12} > \pi_{j1}\pi_{j2}$ indica una correlación como un grado de asociación positivo y $\pi_{j12} < \pi_{j1}\pi_{j2}$ una correlación y asociación negativa. La correlación de Pearson o el coeficiente Phi de Yule en el contexto de variables binarias fue descrito por Yule(100) en 1912 y un año después por Pearson(81).

Como se ha mencionado previamente, en el estudio CARE (Anexo 3), donde las probabilidades marginales y conjunta son $\pi_{j=1,1} = 0,015$, $\pi_{j=1,2} = 0,074$ y $\pi_{j=1,12} = 0,0014$ respectivamente, se estima una correlación de 0,0107, positiva y por lo tanto los pacientes con un '*IAM no fatal*' tienen una probabilidad superior de tener también un '*IAM fatal*' de los que no han experimentado un '*IAM no fatal*', pero a priori la cuantificación de la correlación no indica si esta asociación es muy intensa o no, ya que como veremos en los próximos puntos su rango en el caso de eventos binarios no es de -1 a +1.

6.2.2. Probabilidad Conjunta (JP_j)

La probabilidad conjunta o en inglés 'Joint Probability' (JP_j) se define como $JP_j = Prob((X_{j1} = 1) \cap (X_{j2} = 1)) = \pi_{j12}$, y corresponde a la probabilidad que ambos eventos sucedan al mismo paciente. Fue definido por Russel and Rao(101) en 1940. En el estudio CARE la probabilidad que un paciente le ocurriera un '*IAM no fatal*' y después un '*IAM fatal*' es de $JP_1 = \pi_{j=1,2} = 0,0014$, lo que significa que el 0,14% de los pacientes tuvieron ambos eventos.

Aunque la interpretación de la probabilidad conjunta es directa, por sí sola no es suficiente sin saber si este valor indica una asociación positiva o negativa. Como $JP_1 = \pi_{j=1,12} = 0,0014 > 0,0011 = \pi_{j=1,1}\pi_{j=1,2}$ podemos concluir que existe una asociación positiva. En el caso del estudio PARADIGM-HF, la probabilidad de tener ambos eventos fue del 4,94%, pero este valor sin saber que a partir de $0,0212 = \pi_{j1}\pi_{j2}$ se definen asociaciones positivas no es suficiente.

6.2.3. Índice de Jaccard ($J I_j$)

El índice de Jaccard ($J I_j$) (102–104), en inglés '*Jaccard's Index*' se define como la probabilidad condicional de que ocurran ambos eventos sabiendo que como mínimo uno de ellos ha ocurrido o lo que es lo mismo, la probabilidad que ocurran ambos eventos sabiendo que ha ocurrido el CE. Por lo tanto:

$$J I_j = \frac{P((X_{j1} = 1) \cap (X_{j2} = 1))}{P((X_{j1} = 1) \cap (X_{j2} = 0)) + P((X_{j1} = 0) \cap (X_{j2} = 1)) - P((X_{j1} = 1) \cap (X_{j2} = 1))}$$

O,

$$J I_j = \frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - \pi_{j12}} = \frac{\pi_{j12}}{\pi_j^*}$$

El punto que discrimina entre asociaciones positivas y negativas es el que corresponde a una covarianza igual a cero ($\pi_{j1}\pi_{j2} = \pi_{j12}$), por lo que $J I_j > \frac{\pi_{j1}\pi_{j2}}{\pi_{j1} + \pi_{j2} - \pi_{j1}\pi_{j2}}$ indican asociaciones positivas y negativas cuando $J I_j < \frac{\pi_{j1}\pi_{j2}}{\pi_{j1} + \pi_{j2} - \pi_{j1}\pi_{j2}}$.

A partir la Figura 2 la interpretación gráfica del indicador es directa. El $J I_j$ se interpreta como el '*porcentaje de solapamiento*' entre eventos . En el escenario A) de la Figura 2 el porcentaje de solapamiento es del 0%, lo que indica que al 0% de los pacientes con alguno de ambos eventos les ha ocurridos los dos. En el escenario E es grado de solapamiento es del 50%, indicando que al 50% de los pacientes a los que les ha sucedido uno de los dos eventos, les ha sucedido los dos.

Este indicador es muy utilizado en estudios de similitud en el contexto de la ecología. En el estudio CARE el $JI_j = 0,0165$, lo que significa que el 1,65% de los pacientes que tuvieron un IAM (ya sea fatal o no), les ocurrió los dos.

6.2.4. Índice relativo de Jaccard (JI_j^*)

El índice relativo de Jaccard (JI_j^*) fue descrito por Szymkiewicz y Simpson(105,106) en 1943. El JI_j^* es una modificación del índice de Jaccard y también es conocido como ‘porcentaje relativo de solapamiento’, se define:

$$JI_j^* = \frac{Prob\left((X_{j1} = 1) \cap (X_{j2} = 1)\right)}{\min\left(Prob\left((X_{j1} = 1)\right), Prob\left((X_{j2} = 1)\right)\right)} = \frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$$

Se interpreta al igual que el índice de Jaccard como una probabilidad conjunta, específicamente como la probabilidad de que ocurran ambos eventos sabiendo que ha ocurrido el menos prevalente de los dos. El punto que discrimina entre grados de asociación positivos y negativos es cuando el $JI_j^* = \max(\pi_{j1}, \pi_{j2})$. Por lo tanto $JI_j^* > \max(\pi_{j1}, \pi_{j2})$ indican asociaciones positivas y cuando $JI_j^* < \max(\pi_{j1}, \pi_{j2})$ asociaciones negativas.

Gráficamente en el ejemplo presentado en la Figura 2 el ‘porcentaje relativo de solapamiento’ es del 0% en el escenario A, ya que el 0% de los pacientes a los que les ha ocurrido el evento

menos prevalente, que en este caso coincide con el ER, han tenido ambos eventos. En cambio, en el escenario E, el 100% de los pacientes a los que les ha ocurrido el ER también les ha ocurrido el otro evento.

En el estudio CARE el $JI_j^* = 0,0968 > \max(\pi_{j1}, \pi_{j2}) = 0,074$, indicando que el 9,68% de los pacientes que tuvieron un '*IAM no fatal*' también tuvieron un '*IAM fatal*' y que esta proporción indica una asociación positiva por ser superior a 7,4%.

6.2.5. Índice de Ochiai (OC_j)

En 1957 Ochiai(107,108) definió un coeficiente que se define como la ratio entre la probabilidad conjunta y la raíz cuadrada del producto de las marginales:

$$OC_j = \frac{\sqrt{\text{Prob}((X_{j1} = 1) \cap (X_{j2} = 1)) \text{Prob}((X_{j1} = 1) \cap (X_{j2} = 1))}}{\sqrt{\text{Prob}((X_{j1} = 1)) \text{Prob}((X_{j2} = 1))}} = \frac{\sqrt{\pi_{j12} \pi_{j12}}}{\sqrt{\pi_{j1} \pi_{j2}}}$$

$$= \frac{\pi_{j12}}{\sqrt{\pi_{j1} \pi_{j2}}}$$

Un $OC_j > \sqrt{\pi_{j1} \pi_{j2}}$ indica un grado de asociación positivo, puesto $\sqrt{\pi_{j1} \pi_{j2}}$ es el punto que identifica independencia de eventos. En el estudio CARE se calcula un $OC_j = 0,0434 > 0,0332 = \sqrt{\pi_{j1} \pi_{j2}}$ y por lo tanto una asociación positiva. Su interpretación no es directa pero el CS de Ochiai es muy utilizado en estudios de similitud, también se conoce como '*cosine similarity*'.

6.2.6. Coeficiente de Pares Simples (SP_j)

En 1958, Sokal presentó el coeficiente de Pares Simples(109), que es un índice de concordancia y corresponde a la probabilidad que los dos eventos tomen el mismo valor, es decir, o que ocurran ambos eventos o que no ocurra ninguno de ellos:

$$SP_j = Prob\left((X_{j1} = 1) \cap (X_{j2} = 1)\right) + Prob\left((X_{j1} = 0) \cap (X_{j2} = 0)\right) = 1 + 2\pi_{j12} - \pi_{j1} - \pi_{j2}.$$

$(1 - SP_j)$ estima la probabilidad presentar tan solo un evento. Clínicamente SP_j se interpreta como la proporción de pacientes sin ningún evento de los considerados más aquellos a los que les ha ocurrido ambos. El punto en que la asociación pasa de positiva a negativa no se ha podido determinar para este indicador.

En el estudio CARE el coeficiente de pares simples es de 0,9139 que significa que el 91,39% de los pacientes tuvieron o bien ambos eventos (*'IAM no fatal'* seguido de un *'IAM fatal'*) o bien ningún IAM. Esto significa que el 8,61% de los pacientes $(1 - SP_j)$ o bien registraron un *'IAM no fatal'* o bien un *'IAM fatal'*.

6.2.7. Coeficiente de Sørensen-Dice (SD_j)

Czekanowsky en 1913(110), Dice en 1945(111) y Sørensen en 1948(112), definieron este CS como extensión del Índice de Jaccard:

$$SD_j = \frac{2\text{Prob}\left((X_{j1} = 1) \cap (X_{j2} = 1)\right)}{\text{Prob}\left((X_{j1} = 1)\right) + \text{Prob}\left((X_{j2} = 1)\right)} = \frac{2\pi_{j12}}{\pi_{j1} + \pi_{j2}}$$

El indicador se estima a partir de la ratio de dos veces la probabilidad conjunta entre la suma de las dos prevalencias marginales. Sustituyendo $\pi_{j12} = \pi_{j1}\pi_{j2}$ en la formula, aritméticamente se deduce el punto que define ausencia de asociación, concretamente:

$$\frac{1}{SD_j} = \frac{1}{2} \left(\frac{1}{\pi_{j1}} + \frac{1}{\pi_{j2}} \right)$$

El valor en el caso del estudio CARE es de 0,0324, que indica una cierta asociación entre el '*IAM no fatal*' y el '*IAM fatal*', pues el valor a partir del cual se considera una asociación positiva es 0,0248. Pero la distancia entre 0,0324 y 0,0248 no puede indicarnos si la asociación positiva es severa o no. El SD_j no tiene una interpretación clínica directa.

6.2.8. Coeficiente de Kulczynski (KU_j)

El último CS escogido fue definido por Kulczynski en 1927(113) y se cuantifica el ratio entre la probabilidad conjunta y la probabilidad de presentar tan solo uno de los dos eventos.

$$KU_j = \frac{Prob\left((X_{j1} = 1) \cap (X_{j2} = 1)\right)}{Prob\left((X_{j1} = 1) \cap (X_{j2} = 0)\right) + Prob\left((X_{j1} = 0) \cap (X_{j2} = 1)\right)}$$
$$= \frac{\pi_{j12}}{\pi_j^* - \pi_{j12}}, \text{ donde } \pi_j^* = E(X_{ij}^*)$$

En el estudio CARE el CS de Kulczynski es de 0,0168. Aritméticamente se puede comprobar que $\frac{1}{KU_j} > \frac{1}{\pi_{j1}} + \frac{1}{\pi_{j2}} - 2$ indica asociaciones positivas.

KU_j se puede interpretar también como la *odds* (ventaja) entre pacientes con dos eventos y pacientes con un único evento. Al poderse expresar como una *odds*, cuando existe una gran asociación entre eventos, el denominador será pequeño y, por lo tanto, el límite superior de KU_j será elevado, pudiendo llegar a infinito en el caso que no existan pacientes con tan solo un evento. KU_j también se puede expresar como: $KU_j = JP_j/(1 - SP_j)$.

En la tabla 1 se resume los 8 CS seleccionados.

Tabla 1 - Resumen de la interpretación clínica y dirección de la asociación en los CS seleccionados.

Medida de Asociación	Interpretación Clínica (considerando 2 componentes)	Expresión	Asociación Positiva
Probabilidad Conjunta (JP_j)	Proporción de pacientes con los dos eventos	$JP_j = Prob((X_{j1} = 1) \cap (X_{j2} = 1)) = \pi_{j12}$	$JP_j > \pi_{j1}\pi_{j2}$
Índice de Jaccard (JI_j)	Proporción de pacientes con dos eventos sabiendo que el CE ha ocurrido	$JI_j = \frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - \pi_{j12}}$	$JI_j > \frac{\pi_{j1}\pi_{j2}}{\pi_{j1} + \pi_{j2} - \pi_{j1}\pi_{j2}}$
Índice Relativo de Jaccard (JI_j^*)	Proporción de pacientes con dos eventos sabiendo que el eventos menos prevalente ha sucedido.	$JI_j^* = \frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$JI_j^* > \max(\pi_{j1}, \pi_{j2})$
Coefficiente de correlación (ρ_j)	No tiene una interpretación clínica directa	$\rho_j = \frac{\pi_{j12} - \pi_{j1}\pi_{j2}}{\sqrt{\pi_{j1}(1 - \pi_{j1})\pi_{j2}(1 - \pi_{j2})}}$	$\rho_j > 0$
Índice de Ochiai (OC_j)	No tiene una interpretación clínica directa	$OC_j = \frac{\pi_{j12}}{\sqrt{\pi_{j1}\pi_{j2}}}$	$OC_j > \sqrt{\pi_{j1}\pi_{j2}}$
Coefficiente de Pares Simples (SP_j)	Proporción de pacientes con o bien dos eventos o con ninguno	$SP_j = 1 + 2\pi_{j12} - \pi_{j1} - \pi_{j2}$	N.A.
Coefficiente Sørensen-Dice (SD_j)	No tiene una interpretación clínica directa	$SD_j = \frac{2\pi_{j12}}{\pi_{j1} + \pi_{j2}}$	$\frac{1}{SD_j} > \frac{1}{2} \left(\frac{1}{\pi_{j1}} + \frac{1}{\pi_{j2}} \right)$
Coefficiente Kulczynski's (KU_j)	Odds de pacientes con dos eventos y con evento	$KU_j = \frac{\pi_{j12}}{\pi_j^* - \pi_{j12}}$	$\frac{1}{KU_j} > \frac{1}{\pi_{j1}} + \frac{1}{\pi_{j2}} - 2$

N.A: No Aplicable. $\pi_j^* = \pi_{j1} + \pi_{j2} - \pi_{j12}$

6.3. Intervalo de valores posibles aplicado a los CS seleccionados.

Como se ha especificado en [6] la probabilidad conjunta está acotada inferior y superiormente por los límites de Fréchet(91) y Bahadur(82). Matemáticamente, la probabilidad conjunta puede tomar valores entre cero y uno, pero cuando las probabilidades marginales se determinan, el rango de valores que puede tomar queda determinado por un límite inferior y un límite superior.

Dado que los 8 CS seleccionados se expresan como una función de las probabilidades marginales, aplicando algunas reglas básicas sobre probabilidades podemos deducir para cada uno de ellos los límites inferiores y superiores (Tabla 2).

En la gran mayoría de las aplicaciones de salud, y especialmente cuando combinamos dos eventos en un CE, raramente las probabilidades marginales son elevadas, y es extraño que la suma de éstas sea superior a 1. No obstante, en la Tabla 2 se han considerado todos los posibles escenarios. Si asumimos que $\pi_{j1} + \pi_{j2} < 1$ observamos que Jl_j^* siempre está acotado por 0 en el límite inferior y en 1 en el límite superior. Volviendo a su interpretación (*porcentaje de pacientes con el evento menos frecuente que también le ha ocurrido el otro evento*) comprobamos que el límite inferior tiene que ser cero y el superior uno. En los otros 7 indicadores no es así.

El límite inferior para todos los SC seleccionados es cero, excepto para la correlación, pues su valor mínimo siempre será negativo. Para el SP_j también puede ser cero, pero este límite solo lo alcanzará cuando $\pi_{j1} + \pi_{j2} = 1$ y ambos sean complementarios. La correlación toma valores negativos hasta el -1 que es el mínimo posible, pero este escenario solo ocurrirá en eventos binarios, tal y como ocurre con SP_j , $\pi_{j1} + \pi_{j2} = 1$, que raramente suceden, por lo que la correlación medida entre variables binarias raramente alcanzará el -1.

El límite superior es 1 cuando $\pi_{j1} = \pi_{j2}$ para todas las medidas de asociación excepto para JP_j y KU_j . En el caso de JP_j será π_{j1} cuando $\pi_{j2} \geq \pi_{j1}$ y π_{j2} cuando $\pi_{j1} \geq \pi_{j2}$. KU_j tiene un límite superior excepcional pues este puede alcanzar el infinito cuando $\pi_{j1} = \pi_{j2}$.

La máxima distancia entre el límite inferior y el límite superior es cuando la prevalencia marginal es la misma en ambos eventos ($\pi_{j1} = \pi_{j2}$) y esto es válido por los 8 CS escogidos.

Cuando se estima el grado de asociación es necesario conocer en qué magnitud se asocian los eventos, tanto positiva como negativamente. Para conocer la intensidad de esta asociación es necesario conocer tanto el límite inferior, como el límite superior, como el punto en que se considera independencia. No es suficiente conocer el límite superior y el límite inferior porque éstos no son equidistantes del punto donde los dos eventos son independientes. Dicho en otras palabras, el punto que marca que ambos eventos son independientes no es equidistante en el rango de posibles valores.

Tabla 2 - Aplicación de los límites de Fréchet a los 8 CS estudiados.

Coeficiente de Similitud	Límite Inferior			Límite Superior		
	General	Si $\pi_{j1} + \pi_{j2} < 1$	Si $\pi_{j1} = \pi_{j2} < 0.5$	General	Si $\pi_{j1} + \pi_{j2} < 1$	Si $\pi_{j1} = \pi_{j2} < 0.5$
Probabilidad Conjunta (JP_j)	$0 \leq \max\{0; \pi_{j1} + \pi_{j2} - 1\}$	0	0	$\min\{\pi_1, \pi_2\} \leq 1$	$\min\{\pi_{j1}, \pi_{j2}\}$	π_{j1}
Índice de Jaccard (JI_j)	$0 \leq \max(0, \pi_{j1} + \pi_{j2} - 1)$	0	0	$\min\left(\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_1}\right) \leq 1$	$\min\left(\frac{\pi_{j1}}{\pi_{j2}}, \frac{\pi_{j2}}{\pi_{j1}}\right)$	1
Índice Relativo de Jaccard (JI_j^*)	$0 \leq \max\left(0, \frac{\pi_{j1} + \pi_{j2} - 1}{\min(\pi_{j1}, \pi_{j2})}\right)$	0	0	$\min\left(\frac{\pi_1}{\min(\pi_1, \pi_2)}, \frac{\pi_2}{\min(\pi_1, \pi_2)}\right) \leq 1$	1	1
Coeficiente de correlación (ρ_j)	$-1 \leq \max\left(-\sqrt{\frac{\pi_{j1}\pi_{j2}}{(1-\pi_{j1})(1-\pi_{j2})}}, -\sqrt{\frac{(1-\pi_{j1})(1-\pi_{j2})}{\pi_{j1}\pi_{j2}}}\right)$	$-\sqrt{\frac{\pi_{j1}\pi_{j2}}{(1-\pi_{j1})(1-\pi_{j2})}}$	$-\frac{\pi_{j1}}{(1-\pi_{j1})}$	$\min\left(\sqrt{\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}}, \sqrt{\frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}}\right) \leq 1$	$\min\left(\sqrt{\frac{\pi_{j1}(1-\pi_{j2})}{\pi_{j2}(1-\pi_{j1})}}, \sqrt{\frac{\pi_{j2}(1-\pi_{j1})}{\pi_{j1}(1-\pi_{j2})}}\right)$	1
Índice de Ochiai (ρ_j)	$0 \leq \max\left(0, \frac{\pi_{j1} + \pi_{j2} - 1}{\sqrt{\pi_{j1}\pi_{j2}}}\right)$	0	0	$\min\left(\sqrt{\frac{\pi_1}{\pi_2}}, \sqrt{\frac{\pi_2}{\pi_1}}\right) \leq 1$	$\min\left(\sqrt{\frac{\pi_{j1}}{\pi_{j2}}}, \sqrt{\frac{\pi_{j2}}{\pi_{j1}}}\right)$	1
Coeficiente de Pares Simples (SP_j)	$0 \leq \max(1 - \pi_{j1} - \pi_{j2}, \pi_{j1} + \pi_{j2} - 1)$	$1 - \pi_{j1} - \pi_{j2}$	$1 - \pi_{j1} - \pi_{j2}$	$\min(1 + \pi_1 - \pi_2, 1 - \pi_1 + \pi_2) \leq 1$	$\min(1 + \pi_{j1} - \pi_{j2}, 1 - \pi_{j1} + \pi_{j2})$	1
Coeficiente Sørensen-Dice (SD_j)	$0 \leq \max\left(0, 2 - \frac{2}{(\pi_{j1} + \pi_{j2})}\right)$	0	0	$\min\left(\frac{2\pi_1}{\pi_1 + \pi_2}, \frac{2\pi_2}{\pi_1 + \pi_2}\right) \leq 1$	$\min\left(\frac{2\pi_{j1}}{\pi_{j1} + \pi_{j2}}, \frac{2\pi_{j2}}{\pi_{j1} + \pi_{j2}}\right)$	1
Coeficiente Kulczynski (KU_j)	$0 \leq \max\left(0, \frac{\pi_{j1} + \pi_{j2} - 1}{2 - \pi_{j1} - \pi_{j2}}\right)$	0	0	$\max\left(\frac{\pi_1}{\pi_2 - \pi_1}, \frac{\pi_2}{\pi_1 - \pi_2}\right) < \infty$	$\max\left(\frac{\pi_{j1}}{\pi_{j2} - \pi_{j1}}, \frac{\pi_{j2}}{\pi_{j1} - \pi_{j2}}\right)$	∞

6.4. Rango y coeficientes estandarizados

En [9] se definió el rango de posibles valores como la diferencia entre el límite superior e inferior aplicando los límites de Fréchet. En [11] se definió el coeficiente estandarizado como la división entre la diferencia entre el valor puntual del CS y el rango definido en [9]. El coeficiente estandarizado se puede interpretar como la distancia a la que se encuentra el verdadero CS en un segmento formado por el rango. Con esta aproximación, dichos puntos pueden compararse entre ellos cuando son estandarizados.

En la Figura 3A y Figura 3B se ha graficado como varía el límite inferior y superior de los 8 CS cuando varían ambas probabilidades marginales desde el 0,01 al 0,40. El objetivo de este ejercicio de simulación es conocer en profundidad como se distribuye el rango de dichos CS y determinar cuáles son sus principales características métricas.

Figura 3A: Distribución de los límites inferior y superior de la probabilidad conjunta, índice de Jaccard, índice relativo de Jaccard y correlación al incrementarse la probabilidad del ER y del EA.

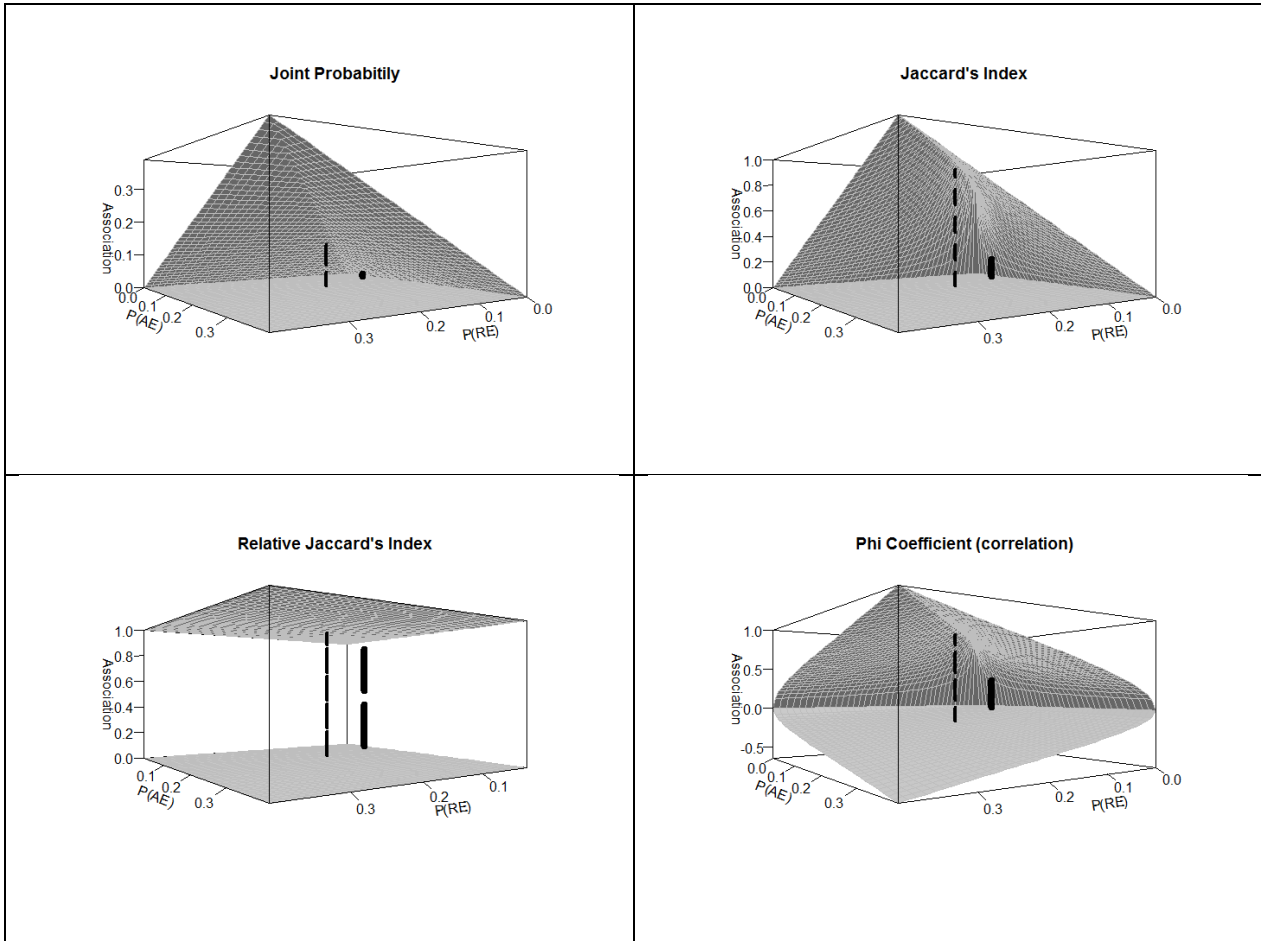
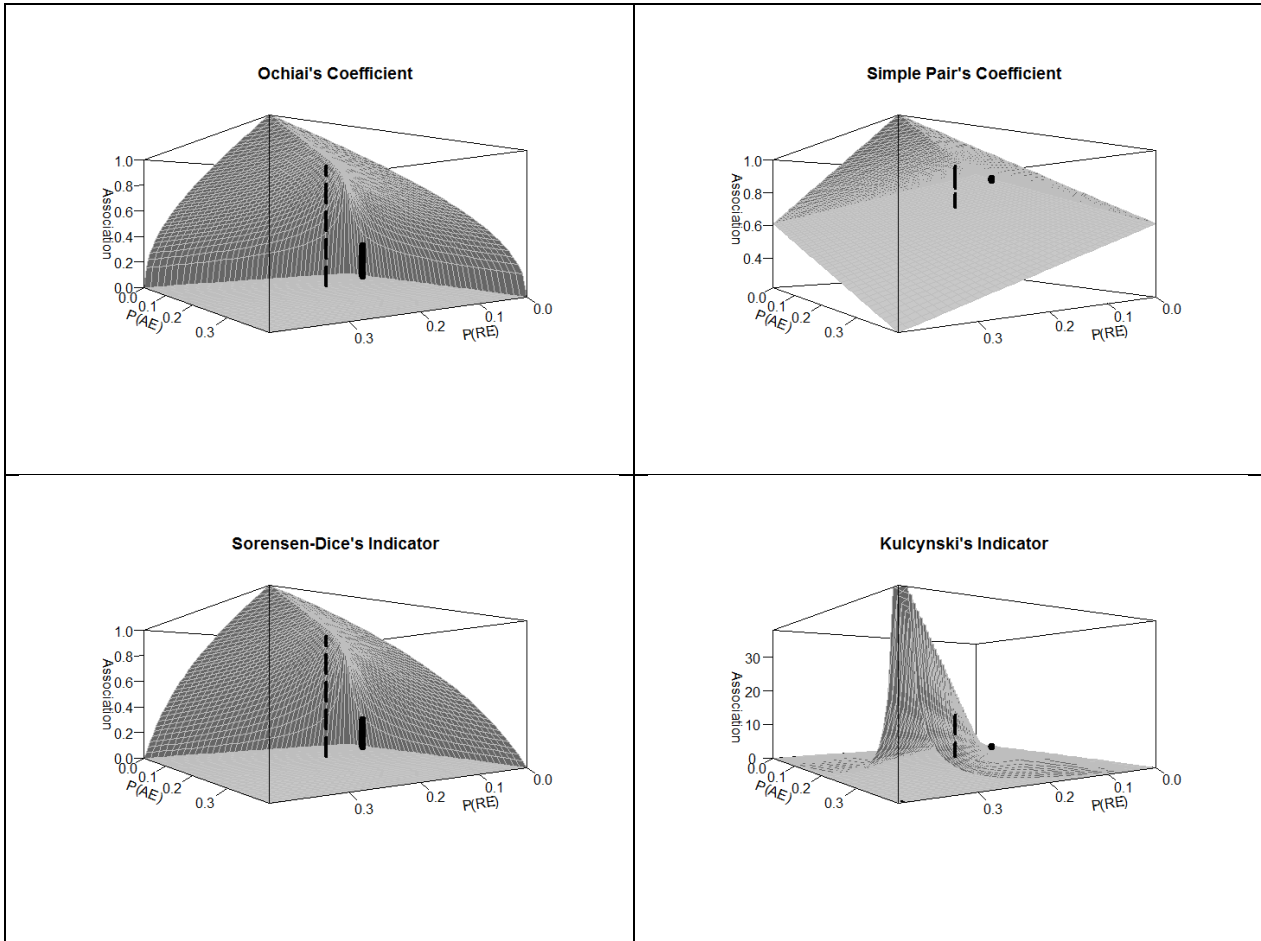


Figura 3B: Distribución de los límites inferior y superior del coeficiente de Ochiai, pares simples, Sørensen-Dice y Kulczynski al incrementarse la probabilidad del ER y del EA.



Cada una de las superficies representa el límite inferior y superior que puede alcanzar cada una de los CS estudiados para cada una de las probabilidades marginales simuladas. La superficie gris oscuro representa el límite superior y la superficie gris claro el límite inferior. La línea vertical negra más gruesa representa el rango de asociaciones del estudio CARE, donde la probabilidad del evento relevante es del 1,2% y la del evento adicional del 6%. También se grafica en cada uno de los casos el rango del estudio PARADIGM-HF (línea vertical negra más fina), en este caso la probabilidad marginal del ER es del 14,9% y del EA del 14,2%.

La primera conclusión que se puede deducir es que el rango varía mucho dependiendo del coeficiente utilizado y de las probabilidades marginales. El rango crece o decrece cuando la probabilidad marginal de cada uno de los eventos es diferente. Por ejemplo, el rango de la correlación, JI_j , OC_j , SD_j y KU_j decrece cuando la probabilidad del evento más prevalente crece y el del menos prevalente está fijo, mientras que JP_j y SP_j permanecen igual.

Paralelamente, cuando el componente más prevalente decrece y el menos prevalente queda fijo, el rango de la correlación, JI_j , OC_j , SD_j y KU_j crecen y el rango de JP_j y SP_j permanecen igual. Es interesante observar que tanto el mínimo como el máximo de JI_j^* siempre son cero y uno respectivamente, independientemente de las prevalencias marginales. Aunque son importantes las variaciones entre los CS estudiados, el máximo rango ocurre cuando las probabilidades marginales son iguales.

Estas conclusiones se alcanzan a partir de la mera observación de las Figuras 3A y Figura 3B, pero es necesario desarrollar las expresiones algebraicas que lo demuestren. En el Anexo 4 se puede consultar, para cada uno de los 8 CS, tanto su rango como su coeficiente estandarizado, tanto para el escenario graficado que es el que habitualmente ocurre ($\pi_{j1} + \pi_{j2} < 1$) como se desarrollan también para escenarios menos frecuentes ($\pi_{j1} + \pi_{j2} > 1$).

En cuanto al coeficiente estandarizado se comprueba que para todos los CS excepto para JI_j y KU_j es $\pi_{j12} / \min(\pi_{j1}, \pi_{j2})$ cuando $\pi_{j1} + \pi_{j2} < 1$ y $\pi_{12} - \pi_1 - \pi_2 + 1 / (1 - \min(\pi_1, \pi_2))$ en aquellos casos que $\pi_{j1} + \pi_{j2} > 1$. El hecho que todos los indicadores excepto el JI_j y KU_j tengan el mismo valor nos indica que el grado de asociación se cuantifica de la misma forma para los diferentes SC independientemente de la métrica.

Se ha de remarcar también que en el caso del Índice Relativo de Jaccard es el único CS que el valor de su coeficiente estandarizado coincide con la misma definición del coeficiente y que su límite inferior y superior están acotados entre 0 y 1 si $\pi_{j1} + \pi_{j2} < 1$ que es el caso más usual.

6.5. Estimación de los CS en los estudios CARE y PARADIGM-HF

Tanto para el estudio CARE como para el estudio PARADIGM-HF se cuantificó el grado de asociación utilizando cada uno de los CS estudiados. Se estimaron también los límites inferior y superior según la aplicación de los límites de Fréchet y se cuantificaron su rango y su valor relativo calculando el coeficiente estandarizado (Tabla 3).

Se ha de notar que si bien los CS expresan estimaciones diferentes del grado de asociación, y sus límites y rangos también son distintos, en todos los CS, excepto para JI_j y KU_j en su forma estandarizada el coeficiente estandarizado fue de 0,097 en el estudio CARE y de 0,347 para el estudio PARADIGM-HF.

Tabla 3 - Resultados del estudio CARE y PARADIGM-HF.

Coeficiente de Similitud	CARE				PARADIGM-HF			
	Valor	Límites	Rango	Estandarizado	Valor	Límites	Rango	Estandarizado
Probabilidad Conjunta (JP_j)	0,0014	(0,000 ; 0,015)	0,015	0,097	0,0494	(0,000 ; 0,142)	0,142	0,347
Índice de Jaccard (JI_j)	0,0165	(0,000 ; 0,201)	0,201	0,082	0,2043	(0,000 ; 0,955)	0,955	0,214
Índice Relativo de Jaccard (JI_j^*)	0,0968	(0,000 ; 1,000)	1,000	0,097	0,3473	(0,000 ; 1,000)	1,000	0,347
Coeficiente de correlación (ρ_j)	0,0107	(-0,035 ; 0,435)	0,470	0,097	0,2269	(-0,170 ; 0,974)	1,144	0,347
Índice de Ochiai (ρ_j)	0,0434	(0,000 ; 0,449)	0,449	0,097	0,3394	(0,000 ; 0,977)	0,977	0,347
Coeficiente de Pares Simples (SP_j)	0,9139	(0,911 ; 0,941)	0,030	0,097	0,8076	(0,709 ; 0,993)	0,285	0,347
Coeficiente Sørensen-Dice (SD_j)	0,0324	(0,000 ; 0,335)	0,335	0,097	0,3393	(0,000 ; 0,977)	0,977	0,347
Coeficiente Kulczynski (KU_j)	0,0168	(0,000 ; 0,252)	0,252	0,066	0,2568	(0,000 ; 21,339)	21,339	0,012

6.6. Conclusiones del objetivo 1

En este punto se listan las diferentes conclusiones del proyecto que aborda el primeo de los objetivos. Éstas se desarrollarán en profundidad y se contextualizarán en el apartado de Discusión de esta Tesis.

1. En general, la magnitud del grado de asociación entre dos eventos binarios depende de la prevalencia marginal en que ocurren dichos eventos.
2. Aunque los Coeficientes de Similitud son útiles para cuantificar el grado de asociación entre dos eventos binarios, muchos de ellos no tienen una interpretación clínica directa, como el índice de Ochiai y el coeficiente de Sørensen-Dice.
3. Los límites y rango de los CS no dependen del grado de asociación, sino que dependen solo de las prevalencias marginales y, por lo tanto, una vez determinadas las prevalencias marginales, los límites y rango quedan definidos independientemente del grado de asociación.
4. El índice relativo de Jaccard es el único CS estudiado que siempre se distribuye entre 0 y 1 (si la suma de probabilidades marginales no supera el 100%). Este es el único indicador estudiado que ni sus límites ni su rango dependen de la probabilidad marginal de cada evento, sino solo dependen exclusivamente del grado (intensidad) de la asociación. Desde el punto de vista clínico, el índice relativo de Jaccard es fácilmente interpretable si lo consideremos como el porcentaje de pacientes de evento menos prevalente a los que también les ha ocurrido el otro evento.

5. La correlación no es un buen indicador del grado de asociación, aunque a partir de su valor sea fácil discriminar asociaciones positivas o negativas. Para una probabilidad conjunta determinada, la correlación puede ser negativa. Una correlación igual a cero no se identifica con un 'solapamiento' nulo o con una probabilidad conjunta nula. Los límites no son -1 y $+1$ para variables binarias y finalmente no necesariamente una correlación intensa negativa o positiva debe estar cerca del -1 o del $+1$.

7. Influencia del grado de asociación en el TMR en un CE binario de dos eventos

El objetivo planteado para este proyecto es:

“Identificar en qué escenarios el uso de un CE de dos eventos binarios es preferible al uso un único Evento Relevante en cuanto al Tamaño de Muestra Requerido. Determinar en qué sentido y magnitud afectan las variaciones de las incidencias y efectos de los eventos combinados y en especial el grado de asociación. Determinar de forma objetiva si el uso de un CE binario es preferible al de un único Evento Relevante una vez fijados los tres conjuntos de parámetros.”

7.1. Justificación

La utilización de CE es una estrategia común en la reducción del TMR en el contexto de los Ensayos Clínicos Aleatorizados (ECA)(23,27,33,35,40). La potencial reducción del TMR se debe al incremento total de pacientes con el evento de interés. Existen ventajas como la reducción del TMR y desventajas como el incremento de la complejidad en la interpretación de los resultados.

Es habitual en ECA en el contexto cardiovascular combinar el evento '*Infarto Agudo de Miocardio*' (IAM) con el evento '*mortalidad*'. La combinación de estos dos eventos es plausible desde el punto de vista biológico, ya que la reducción de IAM podría implicar también una reducción de la '*mortalidad*' y ambos son eventos importantes para los pacientes, por lo es razonable combinarlos en un único CE (1,4,8,44).

Aunque el efecto y la incidencia de los eventos combinados afectan directamente al número de pacientes que tendrán el evento de interés, el grado de asociación también tiene su efecto, aunque habitualmente no se considera. A mayor efecto de los eventos combinados, mayor será la diferencia de ocurrencia de éstos y con un TMR inferior será suficiente. De la misma forma, para eventos que son más frecuentes que otros, el número total de pacientes con el CE será mayor, incrementándose la potencia estadística.

Los beneficios, riesgos e influencia que tiene los eventos combinados ha sido ya ha sido estudiado (1,2,6,8,10,11), pero no ha sido explorado en profundidad la influencia específica que el grado de asociación tiene sobre el TMR (27,40,53,59,83,115). El grado de asociación

entre los eventos combinados puede modificar tanto el error tipo I como tipo II [13-15], por lo que puede modificar el TMR y la potencia estadística(32). La diferencia en cuanto a TMR y la potencia estadística no ha sido estudiada en relación con el grado de asociación de forma exhaustiva(36).

En el contexto del análisis de supervivencia, considerando como variable de resultado el tiempo hasta evento, Gómez y Lagakos(48) han desarrollado un método para determinar la influencia que tendrá la combinación de dos eventos en un CE. Este método, basado en la eficiencia relativa asintótica del contraste de hipótesis y que tiene en cuenta tanto las leyes marginales de probabilidad como la correlación, concluye que la eficiencia se asocia a una variación del TMR(14).

En el segundo proyecto de la presente Tesis exploramos el impacto que tiene el grado de asociación de dos eventos binarios en el TMR de un CE. Se presenta un método sencillo pero objetivo para determinar si la combinación de los dos eventos en un CE es más eficiente, desde el punto de vista del TMR, que la utilización de un único evento.

Como resultado de este proyecto se ha publicado un manuscrito que lleva por título '*The Use of a Binary Composite Endpoint and Sample Size Requirement: Influence of Endpoints Overlap*' en la revista *American Journal of Epidemiology*, 2017; 185(9):832-41 (95). La estructura, tablas y gráficos de este capítulo son de la publicación. Por otro lado, en el anexo 7 se incluye una copia de la publicación.

7.2. Definiciones y Asunciones

Por simplicidad asumimos un ECA con solo dos grupos de comparación (control e intervención). Definimos un Evento Relevante como el que previsiblemente será sensible a la intervención estudiada y cuantifica el efecto de ésta sobre los pacientes. A este ER se plantea combinar también otro Evento Adicional en un único CE con el único objetivo de reducir el TMR.

Interesa determinar si la incidencia de eventos en el grupo control difiere estadísticamente de la del grupo intervención. Para esto se debe realizar un contraste de hipótesis. En este caso, por simplicidad, hemos decidido determinar el contraste de para la diferencia de proporciones [12]. El TMR viene determinado a partir de la aproximación a la distribución Normal, tanto en el caso del ER [13] como en el caso del CE [14]. Para cuantificar el TMR de ER es necesario fijar la incidencia esperada del evento y el efecto esperado de la intervención. En el caso del CE, como se aprecia en la ecuación [2], debemos especificar además el grado de asociación entre los componentes. En el caso de que el TMR del ER sea superior al requerido por el CE, será aconsejable utilizar el CE en lugar del ER de forma aislada. Diremos entonces que la combinación es 'eficiente'. Se ha de hacer notar que este concepto de eficiencia solo abarca el TMR y no se consideran otros factores importantes en la construcción y validación del CE.

Brevemente, porque ya se ha descrito en el Capítulo 4, definimos un vector de respuestas binarias (X_{ij1}, X_{ij2}) , donde X_{ij1} hace referencia al ER y X_{ij2} al EA respectivamente y el subíndice j hace referencia al grupo control e intervención. Definimos que el CE (X_{ij}^*) [1] ocurre si ocurre cualquier de los dos eventos y se estima su probabilidad de ocurrencia

como la suma de la probabilidad de ocurrencia tanto del ER como del EA sustrayéndose la probabilidad de que ocurran ambos eventos [2].

$$E(X_{ij}^*) = \pi_{j1} + \pi_{j2} - \pi_{j12} = \pi_j^* \quad [2]$$

Una vez determinados estos 6 parámetros ($\pi_{j=1,1}, \pi_{j=1,2}, \pi_{j=1,12}, \pi_{j=2,1}, \pi_{j=2,2}, \pi_{j=2,12}$), y teniendo en cuenta las funciones presentadas en [13] y [14], se puede estimar el TMR tanto para el ER como para el EA, fijando los errores tipo I y tipo II. Los parámetros $\pi_{j=1,1}, \pi_{j=1,2}$ hacen referencia a la probabilidad marginal del ER y del EA respectivamente. $\pi_{j=2,1}, \pi_{j=2,2}$ cuantifican la probabilidad de ocurrencia de ambos eventos en el grupo intervención. La estimación de $\pi_{j=2,1}, \pi_{j=2,2}$ es inmediata aplicando las definiciones de efecto presentadas en [3.1 – 3.3].

La Figura 2 presenta gráficamente el concepto de asociación y en el proyecto anterior se realiza un exhaustivo análisis del concepto. El coeficiente $\pi_{j=1,12}$ es la probabilidad conjunta (JP_j), y es una de las formas de cuantificar el grado de asociación. En este proyecto nos hemos centrado en el índice de Jaccard (JI_j) o ‘solapamiento’ por su facilidad de interpretación, lo que lo hace especialmente atractivo para su uso en el mundo clínico. Se podría haber utilizado otros de los coeficientes estudiados en el capítulo anterior.

Para este proyecto asumimos que el grado de asociación en el grupo control ($\pi_{j=1,12}$) es el mismo que el grupo intervención ($\pi_{j=2,12}$), aunque esto no significa que $\pi_{j=1,12} = \pi_{j=2,12}$. Clínicamente no es inverosímil asumir que los eventos se asocian forma parecida y en la misma intensidad tanto en el grupo control como en el grupo intervención. En la Figura 2 y

aplicando los límites de Fréchet determinamos que la probabilidad conjunta puede tomar valores entre el 0,0 y el 0,20. Si, por ejemplo, asumimos que el efecto sobre el ER es de un RR = 0,9 y de un RR = 0,5 para el EA, la probabilidad conjunta del grupo intervención podrá tomar cualquier valor entre 0,0 y 0,18. Si consideramos el escenario E de máximo grado de asociación observamos que $\pi_{j=1,12} = 0,20 \neq 0,18 = \pi_{j=2,12}$.

En el Anexo 1 se determinan los cálculos que se han aplicado a la hora de asumir que el grado de asociación en el grupo control y en el grupo intervención son iguales.

7.3. Material y Métodos

Con el objetivo de estudiar cómo varía el TMR al variar el grado de asociación, se ha realizado un estudio de simulación. Se han simulado diferentes situaciones clínicamente plausibles (Tabla 4), estimando el TMR tanto usando un ER como un CE.

En los diferentes escenarios se ha variado tanto las incidencias de los eventos como el efecto de la intervención. En cada escenario se han simulado 5 distintos grados de asociación, haciendo que coincidan con los escenarios presentados en la Figura 2. Por lo tanto, el primer escenario corresponde a un grado de asociación negativo máximo, que coincide con un $JI_j = 0$. El último escenario corresponde al máximo grado de asociación positiva y coincide con un $JI_j = \min\left(\frac{\pi_{j1}}{\pi_{j2}}, \frac{\pi_{j2}}{\pi_{j1}}\right)$. En otras palabras, los 5 escenarios corresponden a la división equidistante del rango. El coeficiente estandarizado del JI_j es distinto a los otros indicadores estudiados. Si se hubiera utilizado otro indicador las conclusiones serían idénticas, aunque el TMR no coincidiría, excepto en el límite inferior y el límite superior.

Tabla 4 – Resumen de los escenarios simulados.

<i>Prob(ER)</i>	<i>RR(ER)</i>	<i>Prob(EA)</i>	<i>RR(EA)</i>
0,10	De 0,70 a 0,90	0,10	De 0,70 a 1,00
0,10	De 0,70 a 0,90	0,20	De 0,70 a 1,00
0,20	De 0,70 a 0,90	0,10	De 0,70 a 1,00
0,20	De 0,70 a 0,90	0,20	De 0,70 a 1,00
0,03	De 0,70 a 0,90	0,10	De 0,70 a 1,00
0,03	De 0,70 a 0,90	0,01	De 0,70 a 1,00

Los resultados de $Prob(ER) = 0,10$ están en el la Tabla 12 del Anexo 5, Tabla 13 del Anexo 5 para los resultados de $Prob(ER) = 0,20$ y Tabla 14 del Anexo 5 para los resultados de $Prob(ER) = 0,03$.

Específicamente, dentro de cada escenario se simularon 4 diferentes situaciones:

- Situación 1: partiendo de un ER con $Prob(ER) = 0,10$ y efecto incremental $RR(ER)$ de 0,90 a 0,70, se añade un EA con $Prob(EA) = 0,10$ y $Prob(EA) = 0,20$ sin efecto $RR(EA) = 1$.
- Situación 2: $Prob(ER) = 0,10$ con un escaso efecto $RR(ER) = 0,90$ se añade un EA con $Prob(EA) = 0,10$ y $Prob(EA) = 0,20$ con un efecto incremental desde $RR(EA) = 0,90$ a $RR(EA) = 0,70$.
- Situación 3: $Prob(ER) = 0,10$ con un efecto moderado $RR(ER) = 0,80$ se añade un EA con $Prob(EA) = 0,10$ y $Prob(EA) = 0,20$ con un efecto incremental desde $RR(EA) = 0,90$ a $RR(EA) = 0,70$.

- Situación 4: $Prob(ER) = 0,10$ con un efecto importante $RR(ER) = 0,70$ se añade un EA con $Prob(EA) = 0,10$ y $Prob(EA) = 0,20$ con un efecto incremental desde $RR(EA) = 0,90$ a $RR(EA) = 0,70$.

En cada escenario determinado por los parámetros específicos del ER y del EA, se calculó el TMR para el ER y para el CE para cada una de las asociaciones. A partir de ambos TMR se estima la Razón de Tamaños (*Sample Size Ratio, en inglés*) ($SSR = \frac{n_{ER}}{n^*}$). Para $SSR > 1$ el uso del CE requiere un TM inferior al requerido por el EA. Se graficó los SSR en función del grado de asociación para los escenarios más interesantes.

Paralelamente se estimó la potencia estadística fijando diferentes tamaños de muestra. Se estimó tanto para el ER como para el los 5 posibles CE en función de los grados de asociación simulados. Se comparó las curvas de potencia en función del grado de asociación. Se utilizó la ecuación [15] presentada en el capítulo de Métodos.

La simulación se programó en el software estadístico R Studio(116).

7.4. Resultados

Los resultados de la simulación se pueden consultar en el Anexo 5.

En la situación 1 (S1.1 – S1.4), donde añadimos un EA sin efecto alguno, en ningún escenario y para ningún grado de asociación existe un beneficio en cuanto a la reducción del TMR al utilizar el CE ($SSR \leq 1$). En todos ellos existe un incremento del TMR excepto en aquellos casos en los que se añade un EA menos prevalente y con la máxima asociación. En esta última situación el $SSR = 1$. Clínicamente, añadir un evento en el que la intervención no tiene efecto no tiene sentido. El incremento del TMR al utilizar un CE en lugar del ER, es mayor cuanto mayor es la $Prob(EA)$.

En la situación 2 (S2.1 – S2.6), donde a un ER con un efecto pequeño se añade un EA con diferentes efectos y de diferentes prevalencias, el TMR siempre se reduce excepto en el supuesto del mayor grado de asociación. Cuanto mayor es el grado de asociación menor es la reducción del TMR. El máximo potencial en cuanto a la reducción del TMR coincide con la máxima asociación negativa posible, o bien la situación de no ‘solapamiento’ que corresponde con una probabilidad nula igual a cero (conjuntos disjuntos).

La situación 3 (S3.1-S3.6) es idéntica a la situación 2 pero con un incremento del efecto sobre el ER, pasando de un $RR(ER) = 0,90$ a $RR(ER) = 0,80$, por lo que el efecto sobre el ER es moderado. En primer lugar, la simulación indica que a mayor efecto sobre el ER menor es el TMR, por lo que la potencial reducción es inferior (los indicadores SSR son menores entre la situación 2 y 3). Debido a que el TMR al utilizar el ER es inferior, también son

inferiores los cambios en el SSR cuando paso de un JI_j bajo a un JI_j mayor. En esta situación, tanto el grado de asociación como la $Prob(EA)$, determinan si es o no eficiente combinar el EA al ER. El impacto que tiene el grado de asociación en cuanto al TMR es superior cuando el efecto sobre el ER es menor.

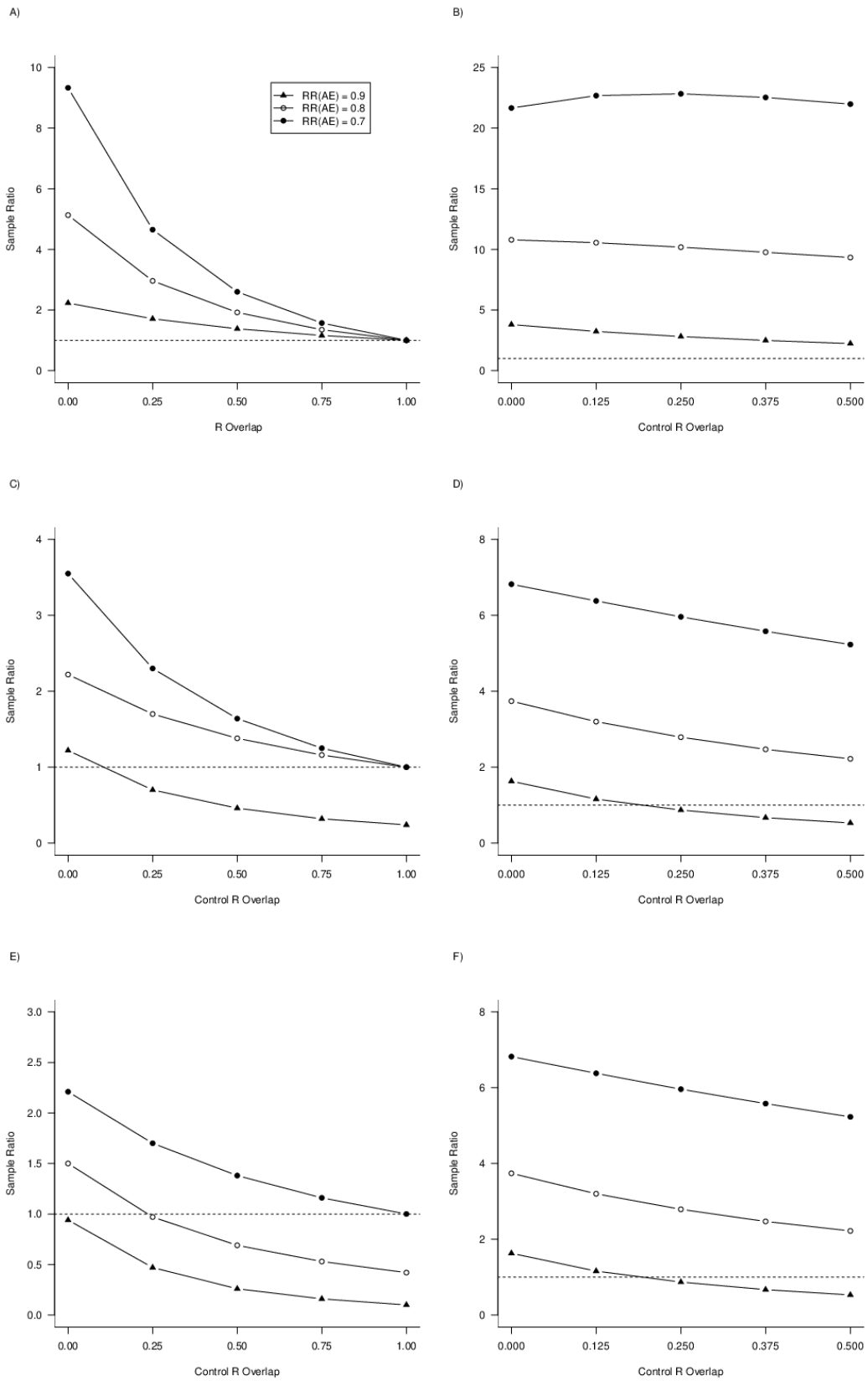
En la situación 4 se combina un EA con un ER sobre el que la intervención tiene un efecto importante (S4.1- S4.6). Como ya se ha detectado en la situación 2, cuanto mayor es el efecto sobre el ER, menor es el TMR y, por lo tanto, menor es la potencial reducción de este (SSR) y menos marcados los cambios entre grados de asociación. En esta situación el grado de asociación también es determinante.

En general, cuanto mayor es el grado de asociación menor es la potencial reducción del TMR, cuando permanecen fijos los otros parámetros. El impacto del grado de asociación es mayor cuando combinamos un EA sobre el que la intervención tiene un efecto inferior al que tiene sobre el ER. Por ejemplo, en las situaciones 3 y 4 cuando añadimos un EA con un efecto inferior al del ER el grado de asociación es determinante en el uso de un CE en vez del uso de un ER de forma aislada. Cuanto mayor es la prevalencia del ER menor es el potencial beneficio de añadir un EA, excepto en la situación en la que el efecto de la intervención sobre el EA sea superior al del ER.

En la Figura 4 se grafica como varía el SSR de cada simulación en función de los diferentes grados de asociación en las situaciones. Se ha obviado la situación donde se añade un EA sin efecto, pues en todos los casos el SSR es inferior a 1. Cada línea representa un efecto del EA diferente. Como mayor efecto tiene el ER, menor es el beneficio de añadir un EA e inferior es la influencia del grado de asociación. El grado de asociación es superior cuanto inferior

es la probabilidad el EA. También cuanto mayor es la probabilidad y el efecto del EA mayor es la propensión de que el $SSR > 1$.

Figura 4 – Ratio de Tamaños de Muestra (SSR) en las situaciones de la 2 a la 4.

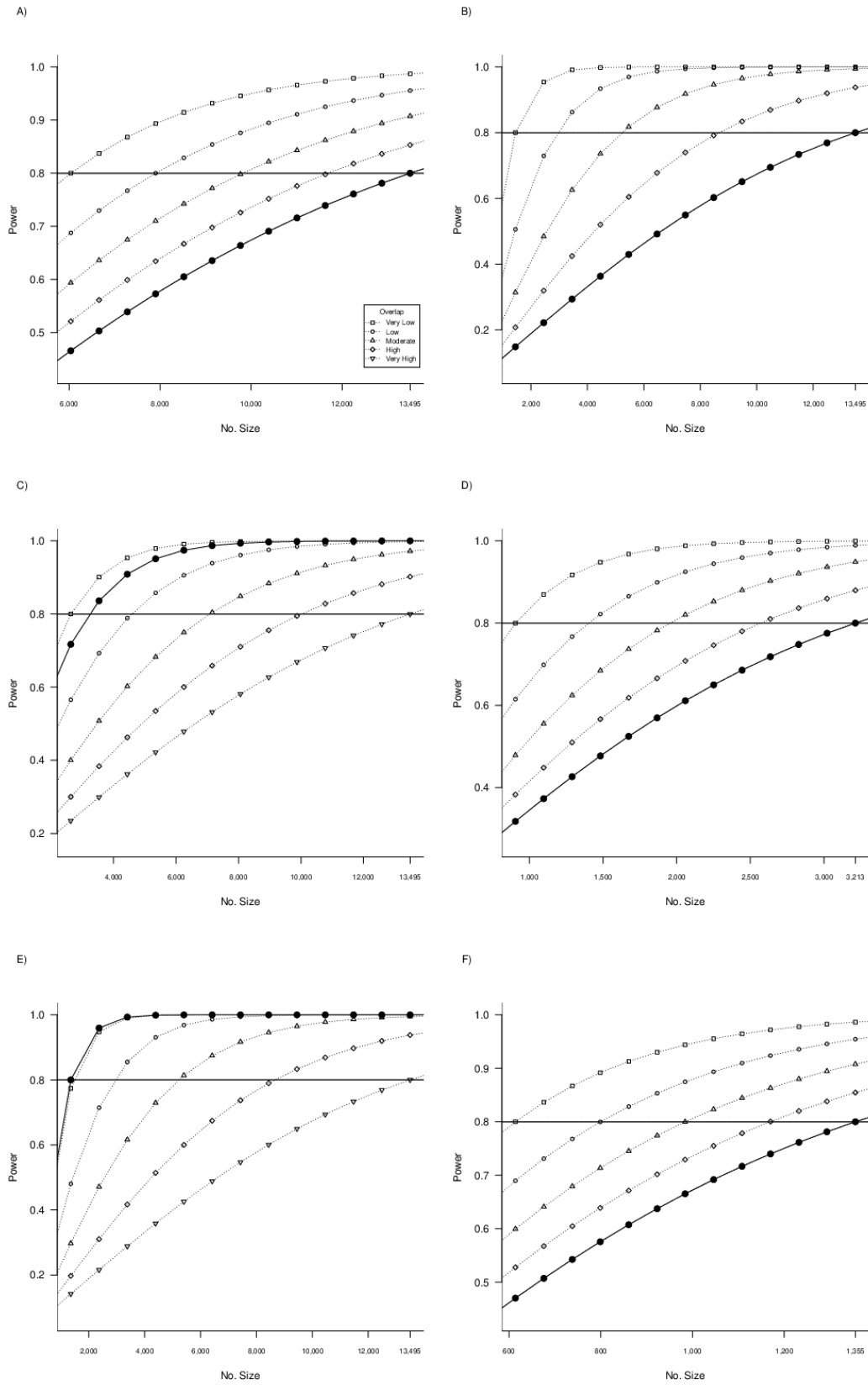


A) $P(EA) = 0,10$ $P(ER) = 0,10$ $RR(ER) = 0,90$. **B)** $P(EA) = 0,20$ $P(ER) = 0,10$ $RR(ER) = 0,90$. **C)** $P(EA) = 0,10$ $P(ER) = 0,10$ $RR(ER) = 0,80$. **D)** $P(EA) = 0,20$ $P(ER) = 0,10$ $RR(ER) = 0,80$. **E)** $P(EA) = 0,10$ $P(ER) = 0,10$ $RR(ER) = 0,70$. **F)** $P(EA) = 0,20$ $P(ER) = 0,10$ $RR(ER) = 0,70$. El rango del grado de asociación depende de la prevalencia de los ER y EA, y está cuantificado utilizando del índice Jaccard. La línea horizontal intermitente corresponde a la situación donde el $SSR = 1$, por lo que escenarios por encima de la línea indican que el TMR al utilizar el CE es inferior al TMR al utilizar tan solo el ER.

La Figura 5 y Figura 6 ilustran la potencia estadística de los diferentes grados de asociación y tamaños de muestra simulados. La línea continua hace referencia a la potencia estadística del ER para cada tamaño de muestra y las líneas discontinuas hacen referencia a la potencia del CE en cada uno de los grados de asociación estudiados.

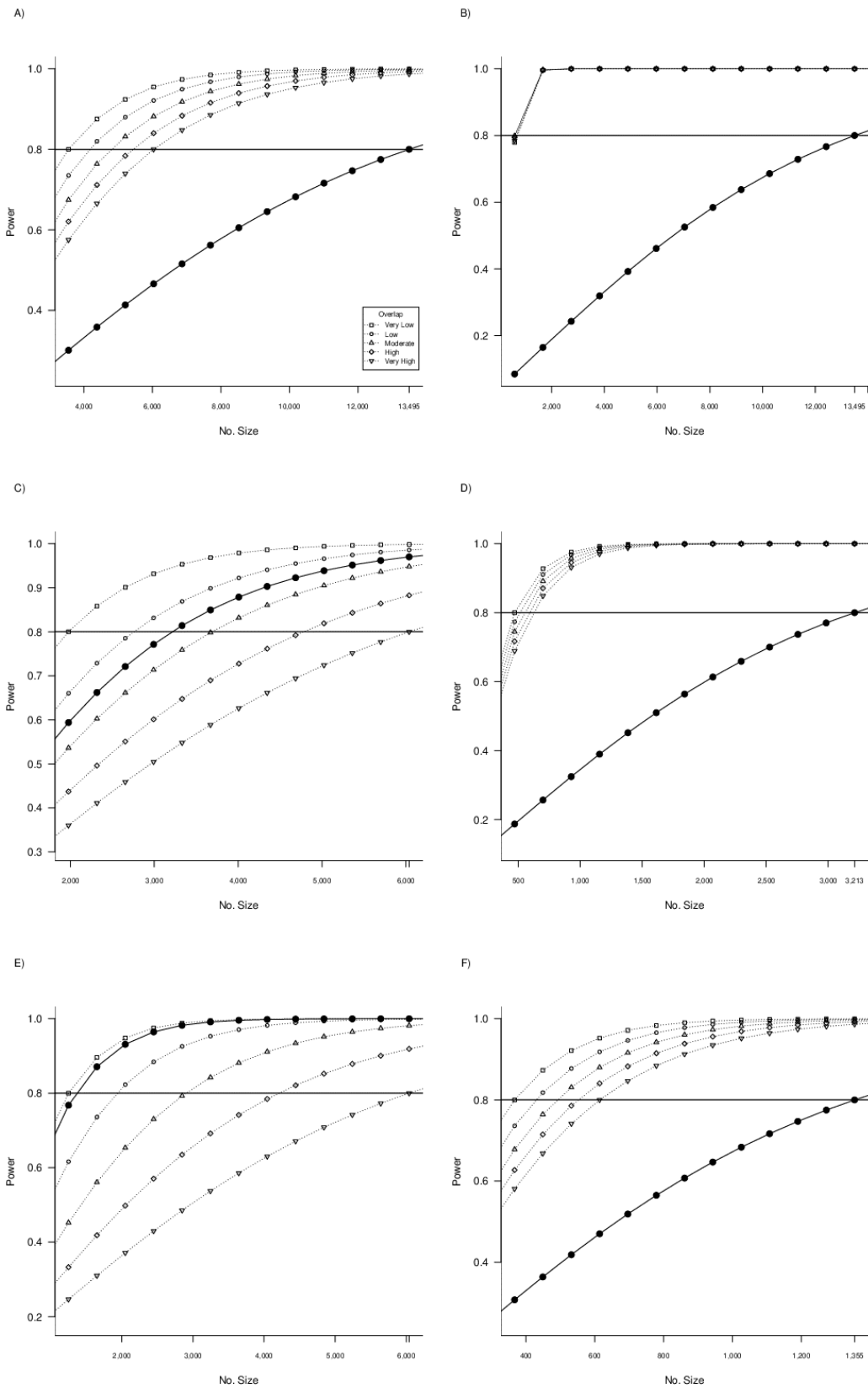
Cuanto mayor es el efecto sobre el EA mayor es la potencia estadística, y este se ve incrementado cuando $J|I_j$ es mínimo (asociación negativa máxima). En algunos escenarios la potencia estadística del CE siempre es superior a la del ER (Figura 6B, 6D y 6F) y en algunas siempre es más potente el ER al CE (Figura 5E, Figura 6E). Aunque existen situaciones (Figura 5C y Figura 6C) que será más o menos potente utilizar el CE según el grado de asociación.

Figura 5 – Curvas de potencia para diferentes tamaños de muestra para diferentes grados de asociación cuando $P(ER) = 0,10$ y $P(EA) = 0,10$.



A) $RR(ER) = 0,90$ y $RR(EA) = 0,90$ **B)** $RR(ER) = 0,90$ y $RR(EA) = 0,70$ **C)** $RR(ER) = 0,80$ y $RR(EA) = 0,90$ **D)** $RR(ER) = 0,80$ y $RR(EA) = 0,70$ **E)** $RR(ER) = 0,70$ y $RR(EA) = 0,90$ **F)** $RR(ER) = 0,70$ y $RR(EA) = 0,70$. La línea negra continua representa la curva de potencia del ER y las líneas grises discontinuas representan las curvas de potencia para el CE en diferentes escenarios del grado de asociación. Una línea horizontal señala la potencia estadística de 0,80.

Figura 6 - Curvas de potencia en función del tamaño de muestra y el grado de asociación cuando $P(ER) = 0,10$ y $P(EA) = 0,20$.



A) $RR(ER) = 0,90$ y $RR(EA) = 0,90$ **B)** $RR(ER) = 0,90$ y $RR(EA) = 0,70$ **C)** $RR(ER) = 0,80$ y $RR(EA) = 0,90$
D) $RR(ER) = 0,80$ y $RR(EA) = 0,70$ **E)** $RR(ER) = 0,70$ y $RR(EA) = 0,90$ **F)** $RR(ER) = 0,70$ y $RR(EA) = 0,70$.

La línea negra continua representa la curva de potencia del ER y las líneas grises discontinuas representan las curvas de potencia para el CE en diferentes escenarios del grado de asociación. Una línea horizontal señala la potencia estadística de 0,80.

7.5. Conclusiones del proyecto 2

En el presente proyecto se han simulado diferentes escenarios y situaciones clínicamente plausibles con el objetivo de estudiar las principales combinaciones de parámetros. En cada una de ellas se ha estimado el TMR en el caso del uso de un ER aislado y de un CE, se ha calculado el SSR y la influencia final que la combinación de los eventos tiene sobre el TMR. Se han simulado 5 grados de asociación en 5 cortes equidistantes del rango estimado a partir del índice de Jaccard (Anexo 5). El resumen de los resultados puede consultarse en la tabla 5. De forma general, se puede concluir:

- Cuando se combina un EA sobre el que la intervención tiene un efecto superior al del ER, independientemente del grado de asociación, siempre se reducirá el TMR. La influencia que tiene el grado de asociación sobre el TMR se reduce cuando aumenta el efecto de la intervención sobre el EA.
- Cuanto mayor es la probabilidad conjunta o el 'solapamiento', menor es el potencial beneficio de combinar un EA con un ER. Una probabilidad conjunta o 'solapamiento' elevado identifican grados de asociación positiva elevados y viceversa. El beneficio de combinar un EA con una asociación negativa con el ER es variable y depende de otros factores, como el efecto de la intervención sobre el EA añadido o su incidencia. En ocasiones, el grado de asociación es determinante en para la decisión sobre el uso del CE.

- La situación más incierta ocurre cuando se combina un EA sobre el que la intervención tiene un efecto inferior al que tiene sobre el ER. En este caso el TMR puede ser superior o inferior, y dependerá del grado de asociación entre ambos eventos, del efecto sobre el EA y de su incidencia.

Tabla 5 - Resumen del impacto cualitativo del grado de asociación entre componentes en cuanto al TMR, considerando diferentes probabilidades y efectos de los eventos combinados.

Escenario		Impacto Cualitativo del grado de asociación ^a	Tamaño de muestra Requerido en el uso del CE ^b		
$P(EA)$	$RR(EA)$		Sin considerar el Grado de Asociación	Considerando el Grado de Asociación	
				Negativa	Positiva
$\leq P(ER)$	1	++	☹	☹	☹☹
$\geq P(ER)$	1	++	☹	☹	☹☹
$\leq P(ER)$	$< RR(ER)$	++	☺	☺☺	☺
$\leq P(ER)$	$> RR(ER)$	+++	☹/☺	☺	☹
$\geq P(ER)$	$< RR(ER)$	+	☺☺	☺☺☺	☺☺
$\geq P(ER)$	$> RR(ER)$	+++	☹/☺	☺	☹

$P(ER)$: Probabilidad del ER. $P(EA)$: Probabilidad del EA. $RR(ER)$: Efecto del ER. $RR(EA)$: Efecto del EA.

a: Impacto Cualitativo del grado de asociación: El impacto del grado de asociación en escenarios (+) es inferior al de los escenarios (++) y este inferior al de los escenarios (+++).

b: Tamaño de Muestra Requerido al utilizar el CE sin tener en cuenta el grado de asociación y tras considerarlo en su rango desde el índice Jaccard mínimo (grado de asociación negativo máximo) a un índice de Jaccard máximo (grado de asociación máxima posible). El TMR al utilizar un CE en los escenarios marcados con '☹' es siempre mayor que la utilización solo del RE. En los escenarios marcados con '☹☹' el TMR es superior a los marcados con '☹'. El TMR al utilizar el CE en los escenarios marcados con '☺' es siempre inferior al TMR en el caso del ER. En los escenarios '☺☺☺' la reducción es mayor a los marcados con '☺☺' y los marcados con '☺'. El TMR al utilizar un CE en los escenarios marcados con '☹/☺' dependerá del grado de asociación.

8. Bin-CE: Una aplicación web para calcular el TMR en un conjunto de eventos.

El objetivo planteado para este proyecto es:

“Desarrollar una herramienta que determine, a partir de un conjunto de eventos binarios, la combinación óptima para conseguir la máxima reducción del Tamaño de Muestra Requerido”.

8.1. Justificación

La estimación del TMR cuando se combinan dos eventos en un CE binario no es trivial, ya que los diferentes parámetros (incidencia, efecto y grado de asociación), pueden interactuar entre ellos de tal forma que el TMR pueda ser mayor o menor. Al combinarse dos o más eventos en un CE binario, éste puede estimar el 'efecto neto' de la intervención en diferentes direcciones. Un ejemplo es la utilización del CE MACE en ECA cardiovasculares. El MACE habitualmente combina diferentes eventos como '*muerte cardiovascular*', '*ictus*' e '*IAM*'. Pero en otras ocasiones se pueden añadir o eliminar eventos: '*hemorragia mayor*', '*necesidad de revascularización*', '*muerte no cardiovascular*'. En estas diferentes combinaciones de eventos se estiman diferentes aspectos de la intervención, pero también precisan de diferentes TMR. Cuanto mayor es el número de eventos combinados más compleja será su interpretación. La combinación de unos eventos u otros en un CE hasta el momento es una decisión más o menos subjetiva del profesional que diseña el ECA. Hasta donde conocemos, no existe ninguna herramienta objetiva desarrollada para la selección de unos eventos frente a otros para construir un CE que sea lo más eficiente posible.

La probabilidad de ocurrencia o incidencia acumulada al final del seguimiento y el efecto de cada uno de los eventos considerados debe estimarse, pero también el grado de asociación. Cuando se consideran más de dos eventos para ser combinados debemos estimar, a parte de la asociación de cada uno de los pares, la asociación entre tres eventos, cuatro, etcétera(82,115). Los indicadores para estimar la asociación entre dos componentes estudiados en el proyecto anterior (87,97) no son adecuados para estimar la asociación de

tres o más eventos. La probabilidad conjunta de tres o más eventos es un indicador válido, pero en estas situaciones el cálculo del grado de 'solapamiento' no es trivial.

En el proyecto anterior hemos demostrado que el grado de asociación puede ser determinante en cuanto a la decisión de utilizar un único ER o bien un CE(117), ya que no todas las combinaciones de eventos reducen el TMR. Por ejemplo, un hipotético e irracional CE, podría ser el resultante de combinar los eventos '*IAM*' y '*Elevación de los niveles normales de Troponinas*'. Un paciente al que le ocurra cualquiera de los dos eventos será un paciente con el CE, pero esos eventos están tan sumamente asociados que podemos considerar que prácticamente significan lo mismo. Pero si combinamos los eventos '*IAM*' e '*ictus*', aunque se espera una cierta proporción de pacientes que sufran ambos eventos, el número neto de pacientes con alguno de los dos eventos será superior al número de pacientes que les ocurre un único evento(95).

Los hallazgos de los proyectos anteriores se generalizarán a la combinación de K eventos en el presente proyecto. Se presenta una herramienta web que implementa un método para cuantificar el TMR y que debe servir a clínicos, epidemiólogos y bioestadísticos a determinar cuál es la mejor combinación de eventos posible.

En el proceso del diseño y construcción de un CE es necesaria la estimación de los diferentes parámetros necesarios en el cálculo del TMR. Como hemos visto, una estimación incorrecta de estos coeficientes puede variar el cálculo del TMR, y por lo tanto la decisión de si combinar o no un evento. Esta herramienta ha de ser capaz de que los usuarios puedan realizar análisis de sensibilidad para poder determinar si la decisión respecto a la construcción de un CE es robusta.

En este proyecto hemos creado la aplicación Bin-CE la cual, a partir de un conjunto de K eventos candidatos a ser combinados, encuentra la mejor combinación posible de ellos para minimizar el TMR, pudiendo emplear o no todos los eventos introducidos. Bin-CE ha sido programado utilizando el lenguaje de programación libre R, cargado en un servidor también de uso libre y su código fuente se puede consultar en el Anexo 6. Se puede acceder en su versión beta en la dirección: <https://uesca-apps.shinyapps.io/bincep/> y utilizase sin necesidad de instalar ningún software.

8.2. Definiciones y asunciones

En esta ocasión asumimos que existe un ER y un conjunto de $(K-1)$ EA, siendo $K \geq 2$. El objetivo final es contrastar diferencias entre grupos referente a la variable de resultado que es, o bien un ER, o un CE que combine el ER con alguno o con todos los EA posibles. Esta decisión se toma solo desde el punto de vista del TMR.

El cálculo del TMR se realiza asumiendo por defecto la aproximación a la distribución Normal y el contraste de hipótesis para la diferencia de proporciones(32,36,92). Bin-CE incorpora otras aproximaciones, como el método del arcoseno, utilizado cuando el contraste de hipótesis se realiza mediante la distribución exacta de Fisher(93). Ambos métodos se pueden corregir mediante la corrección de Yate's(94). Bin-CE permite el cálculo del TMR para cada tipo de aproximación de forma rápida, porque se ha programado en un entorno reactivo (al cambiar cualquier parámetro de entrada se actualiza automáticamente todo el sistema).

El objetivo de Bin-CE, es calcular el TMR de un posible CE que minimice el TMR y que pueda estar compuesto por un número variable de eventos candidatos a ser combinados. Si el TMR del CE es inferior al del ER el investigador preferirá el uso del CE al del ER(14).

Como anteriormente, se define X_{ijk} como la respuesta binaria del i-ésimo paciente, del j-ésimo grupo en el componente k-ésimo. Como en casos anteriores el índice j toma valores 1 y 2 (grupo control e intervención) pero en este caso el subíndice k puede tomar valores desde 1 hasta K (siendo K el número total de componentes candidatos a ser combinados).

Se asume que el número de pacientes en cada grupo ($j = 1,2$) es idéntico ($n_{j=1} = n_{j=2}$), también que como mínimo habrá dos componentes a ser combinadas ($K > 1$). Se reserva el índice $k=1$ para el Evento Relevante y los otros subíndices ($k > 1$) se utilizan para los Eventos Adicionales.

Como se definió en el apartado 4 de métodos $E(X_{ijk}) = \pi_{ik}$ y $V(X_{ijk}) = \pi_{ik}(1 - \pi_{ik})$.

Imaginemos un hipotético CE que combine dos componentes, el ER ($X_{ijk=1}$) y el EA k'-ésimo ($X_{ijk=k'}$), denotaremos este hipotético CE como $X_{ij}^{1k'}$. A partir de la definición [1]:

$$X_{ij}^{1k'} = \begin{cases} 1, & \text{si } X_{ijk=1} + X_{ijk=k'} \geq 1 \\ 0, & \text{si } X_{ijk=1} + X_{ijk=k'} = 0 \end{cases}$$

Adaptando la ecuación presentada en [2] podemos estimar para cualquier par de componentes y de forma generalizada la probabilidad de ocurrencia de cualquier CE que combine el ER con un EA candidato ($k > 1$):

$$E(X_{ij}^{1k'}) = Prob(X_{ij}^{1k'} = 1) = Prob((X_{ij1} = 1) \cup (X_{ijk'} = 1)) = Prob(X_{ij1} = 1) + Prob(X_{ijk'} = 1) - Prob((X_{ij1} = 1) \cap (X_{ijk'} = 1)) = \pi_{i1} + \pi_{ik'} - \pi_{i,1k'}$$

donde $\pi_{i=1,1k'}$ hace referencia a la probabilidad conjunta entre el ER y el k' -ésimo EA en el grupo control. Asumimos que dicha relación igual en el grupo control ($i=1$) que en el grupo intervención ($i=2$). Esta afirmación implica unos cálculos que se especifican en el Anexo 1 de esta Tesis.

A partir de 6 parámetros ($\pi_{i=1,1}, \pi_{i=1,k'}, \pi_{i=1,1k'}, \pi_{i=2,1}, \pi_{i=2,k'}, \pi_{i=2,1k'}$) se puede determinar el TMR (32,36,95) para el CE ($X_{ij}^{1k'}$). Los parámetros referentes a la probabilidad de ocurrencia, como los de eficacia, habitualmente son conocidos por el investigador o son de fácil estimación/asunción, pero el grado de asociación o como están relacionados los diferentes componentes entre ellos no es trivial (59) y en este caso se debe estimar para los $k^*(k-1)/2$ pares de asociaciones.

El grado de asociación entre dos eventos binarios se puede estimar utilizando diferentes coeficientes (Anexo 3). Todos los coeficientes estudiados en el proyecto 1 de esta Tesis son función de las probabilidades marginales tanto del ER como del k -ésimo EA y de su probabilidad conjunta. Cada uno de los coeficientes tiene unas características determinadas en función de sus límites, rango y valor estandarizado. En el proyecto 1 se seleccionó el

índice relativo de Jaccard (JI_j^*) o ‘*porcentaje relativo de solapamiento*’ como el mejor indicador por sus características, pero se comprobó también que una vez fijado su valor y la probabilidad de cada uno de los eventos, aplicando un poco de aritmética se puede pasar de un indicador a otro.

Bin-CE en su versión beta ha sido programado estimando la asociación a partir de la probabilidad conjunta en cada par de eventos ($\pi_{i=1,kk'}$). En siguientes versiones tenemos previsto programar Bin-CE de tal forma que el usuario pudiendo utilizar otros coeficientes como: correlación, índice Jaccard o índice relativo de Jaccard.

8.3. Algoritmo Bin-CE

8.3.1. Aspectos Técnicos, Software y Código Abierto

Bin-CE es una herramienta web de acceso gratuito. Se puede acceder utilizando cualquier navegador a partir del siguiente enlace: <https://uesca-apps.shinyapps.io/bincep>. Toda su interface está en inglés porque pretende ser de uso global. Bin-CE ha sido programado utilizando la librería Shiny(118) de R Studio(116).

El lenguaje R(119) de programación estadística ha sido utilizado para desarrollar la aplicación Bin-CE en R Studio. Bin-CE puede ser utilizada directamente accediendo con el servidor de Shiny (<https://www.shinyapps.io/>) sin necesidad de que el investigador usuario necesite instalarse ni R Studio ni tampoco la librería Shiny. Para los usuarios más

avanzados, el código fuente es accesible (Anexo 6) contribuyendo a la iniciativa 'Open Source'. En este código fuente se encuentran las funciones necesarias para implementar Bin-CE y poder replicar cualquier resultado o bien modificarlas en función de las necesidades específicas de cada programador.

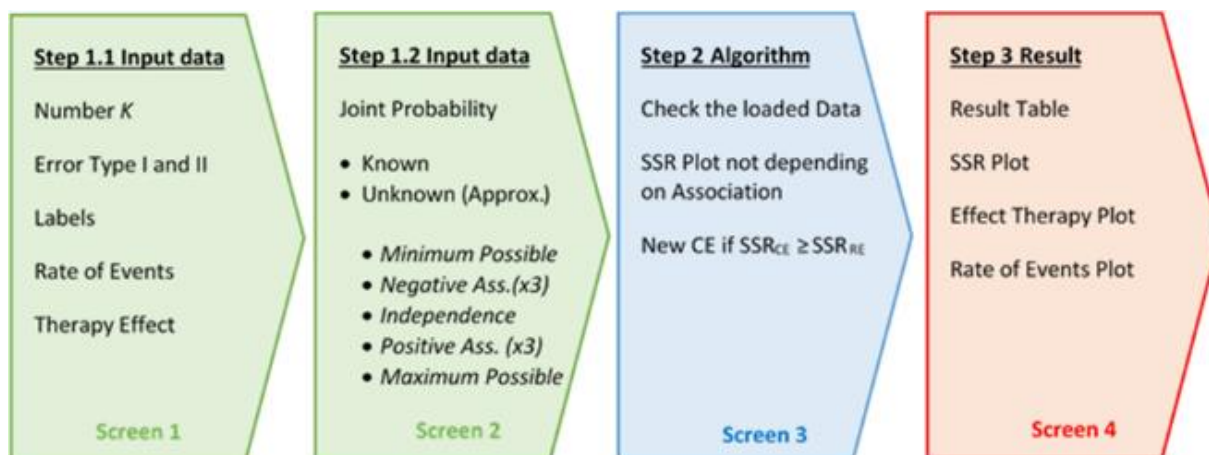
8.3.2. Enfoque general de Bin-CE

Bin-CE ha sido programado para combinar el ER hasta con 9 EA ($K \leq 10$). Es una decisión arbitraria para no incrementar excesivamente el código, también porque en la práctica habitual encontrar un CE de más de 10 eventos es inverosímil.

Bin-CE se ha programado a partir de 4 pantallas reactivas (Figura 7). Las 2 primeras pantallas están reservadas para cargar la información. En la tercera, se presenta la información entrada y da unos primeros resultados. Bin-CE asegura la integridad de los datos cargados. Por ejemplo, no permite registrar probabilidades marginales fuera del rango 0 - 1 o si detecta alguna inconsistencia la corrige. Si existe alguna probabilidad conjunta fuera de los límites de Fréchet(91), Bin-CE automáticamente corrige el dato (estas modificaciones también se presentan en la tercera pantalla). Si se define, por ejemplo, un ER con una probabilidad de ocurrencia del 10% y un EA con una del 6%, no se puede especificar una probabilidad conjunta del 8%. Aplicando los límites de Fréchet la probabilidad conjunta solo está definida entre el 0% y el 6%. Bin-CE detecta el error y lo sustituye por el escenario de no asociación $\pi_{j,1k'} = \pi_{j1}\pi_{jk'} = 0,006$. La última pantalla presenta los resultados referentes a la mejor combinación (si es que existe) y también los resultados de los pasos intermedios, que se han ido alcanzando en cada iteración.

Que las pantallas sean reactivas significa que cuando se varía cualquier dato en cualquiera de ellas, automáticamente todo el sistema actualiza los resultados. Esta característica es importante ya que permite realizar análisis de sensibilidad fácilmente.

Figura 7 – Diagrama de trabajo de Bin-CE.



8.3.3. Entrada de datos (Pantalla 1)

Bin-CE permite combinar hasta 9 eventos al ER que se prefije. En primer lugar, en la primera pantalla se indica el **número de eventos** a combinar y automáticamente aparecen tantas entradas como eventos. Para cada una de estas entradas el usuario debe registrar: la **etiqueta** con que identificará cada evento (por ejemplo 'IAM'), la **probabilidad de ocurrencia** (probabilidad de ocurrencia en el grupo control o incidencia acumulada al final del seguimiento) y el **efecto** que se asume para cada uno de los componentes.

Aunque, como hemos visto en el apartado 4 de Métodos, el efecto de la intervención sobre un evento se puede estimar a partir de la diferencia de proporciones y del Odds Ratio, hemos decidido estimar el efecto de la intervención a partir del Riesgo Relativo. El motivo es

porque consideramos que es de fácil estimación por los clínicos y es el más adecuado en el contexto de los ECA(120). A parte como se comprobó en [3.4 – 3.6] los tres indicadores son intercambiables entre ellos(84).

El **error tipo I**, **error tipo II** y el **tipo de contraste de hipótesis** para estimar el TMR, se prefija en 0,05, 0,20 y ‘contraste asintótico Normal sin corrección’ respectivamente. Estos tres parámetros se pueden cambiar y los resultados son recalculan automáticamente.

Finalmente, en la primera pantalla el usuario puede determinar cuál de los eventos entrados es el Evento Relevante. En caso de que ningún evento se considere como el Evento Relevante, Bin-CE asigna como ER aquel evento que precise un TMR menor (Figura 8). Es interesante asignar la categoría de ER a diferentes eventos como análisis de sensibilidad con el objetivo de estudiar si el hecho de ‘forzar’ que un evento sea seleccionado para el CE varía los resultados finales.

Figura 8 – Ejemplo de la pantalla 1. Probabilidad de ocurrencia y efecto de cada evento considerado.

Input Data (ratio of events, effects) Association Data Results

Please enter:

- a) The total number of candidate outcomes and their labels
- b) The rate of events (%) for each candidate in the control group
- c) The expected treatment effect for each candidate
- d) You may wish to select the Relevant Endpoint

Variables

	Label	Rate (0.00-1.00)	Effect
1	GI Bleeding	0.014	0.19
2	Hb ≥4 without	0.006	0.48
3	Hb ≥3 with	0.037	0.22
4	Hematoma > 15	0.060	0.09
5	Access Complication	0.011	0.24

En la primera pantalla, el usuario debe fijar el número de eventos, el contraste de hipótesis, error tipo I, error tipo II. El usuario debe introducir: a) Etiqueta de cada evento considerado, b) Probabilidad de ocurrencia de cada evento en el grupo control, c) Efecto de la intervención a partir del RR y d) si algún evento es considerado ER (RE – Relevant Event / AE – Additional Event).

8.3.4. Entrada de datos (Pantalla 2)

Bin-CE, en la versión actual beta, utiliza la probabilidad conjunta para el estudio del grado de asociación. El grado de asociación entre cada uno de los pares posibles de eventos puede ser conocido, desconocido o sólo conocido parcialmente (algunas asociaciones son conocidas por el usuario y otras no).

Algunos autores indican que los artículos donde se presentan los resultados de un ECA no se especifica el grado de asociación de forma habitual (42,59). Bin-CE permite, o bien especificar la matriz de probabilidades conjuntas si es conocida o bien, si es desconocida, que será lo más habitual, permite entrar al usuario una aproximación semiparamétrica basada en los límites de Fréchet(91). Para los usuarios clínicos es fácil estimar la relación entre dos eventos a partir de esta aproximación. En caso de que el usuario conozca solo algunas asociaciones, se pueden utilizar ambas opciones.

La aproximación semiparamétrica considera **9 diferentes escenarios** dependiendo de la magnitud y dirección del grado de asociación. La idea es parecida a la presentada en la Figura 2 del proyecto 1, o bien la simulada en el proyecto 2, donde se dividía todo rango de posibles valores de asociación en 5 cortes, aunque en ese caso eran cortes equivalentes. Un escenario de los 9 contemplados, corresponde a la situación donde no existe asociación entre eventos ($\pi_{kk'} = \pi_{ik}\pi_{ik'}$). Se reservan también dos de los 9 escenarios para las situaciones extremas, que corresponden al límite inferior y superior de los límites de Fréchet.

La probabilidad conjunta máxima sucede cuando el evento menos probable es parte del evento más probable. Por ejemplo, si definimos el evento '*Muerte Cardiovascular*' y '*Muerte por cualquier causa*', el grado de asociación será el máximo posible pues todas las '*Muertes Cardiovasculares*' son a su vez '*Muertes por cualquier causa*'. Contrariamente, el grado de asociación mínimo se presenta, por ejemplo, entre los eventos '*ictus isquémicos*' y '*ictus hemorrágico*' ya que sus definiciones son mutuamente excluyentes. La probabilidad conjunta en este caso debe ser 0.

Se reservan otros tres escenarios diferentes para asociaciones positivas y negativas. Cada uno de los tres escenarios cuantifica la magnitud de la asociación positiva/negativa como fuerte, moderada y débil.

A partir de la aproximación semiparamétrica, Bin-CE utiliza: a) la probabilidad marginal de cada uno de los eventos, b) el rango entre el límite inferior y superior [10] de Fréchet y c) la esperanza matemática ($\pi_{kk'} = \pi_{ik}\pi_{ik'}$) de la asociación. En [16] se presenta la expresión matemática para estimar el valor de la probabilidad en situación de no asociación. El límite inferior y superior [10] coinciden con las asociaciones máximas tanto negativas como positivas. A continuación, se presentan las expresiones para estimar los diferentes grados de asociación tanto negativos [17] como positivos [18].

$$\pi_{j,kk'} = \pi_{jk}\pi_{jk'} \quad [16]$$

$$\pi_{j,kk'} = \pi_{j,kk'} - \left(\frac{[\pi_{j,kk'} - \text{Inf}(\pi_{j,kk'})]}{3} \right) \Theta \quad [17]$$

$$\pi_{j,kk'} = \pi_{j,kk'} + \left(\frac{[\text{Sup}(\pi_{j,kk'}) - \pi_{j,kk'}]}{3} \right) \Theta \quad [18]$$

El parámetro θ toma valor 1, 2 o 3 si el usuario ha escogido la opción: débil, moderada o fuerte. Que hayamos dividido en tres grados la magnitud de la asociación es arbitrario. Se ha considerado que tres niveles de asociación son suficientes para discernir las diferentes magnitudes, sin complicar excesivamente su estimación. Ver Figura 9.

Las ecuaciones [17] y [18] son diferentes, puesto que el escenario de independencia entre eventos binarios no corresponde a la mitad del rango descrito por los límites de Fréchet.

Figura 9 - Ejemplo de la pantalla 2. Probabilidad conjunta (asociación de segundo orden).

bin-CE: Sample Size Calculator for Clinical Trials using binary Composite Endpoint

This App computes Sample Size (SS) when are combined two or more endpoints/outcomes. Two outcomes are combined in each step only when the SS of the CE is lower than the previous outcome.
The user can fix an outcome (RE) although if any RE are fixed the App will select the outcome which needs a lower SS. This App ends when any combination can reduce the SS

RE : The outcome that is expected to drive the main effect of the treatment. This outcomes will be fixed. When any RE is defined, this App will select as RE the outcome which need a low SS

AE : The outcome/s that the researcher considers to combine with the RE in a CE.


Contact: Mr. Josep Ramon Marsal (josepmarsal@yahoo.es)

Num. of variables
5

Type I error: 0.05 Type II error: 0.2

[1] "Sig: 0.05 % Pow: 80 %"

Testing Methods
Asymptotic Normal method



Input Data (ratio of events, effects) Association Data Results

Association:
Please indicate the degree of association between candidate outcomes , using the Joint Probability. There are several ways to quantify the degree of association:

Do you know the matrix of association
Some

	HF Hospit	Death for	Atrial F:	Decline f
V1	0.049			
V2				
V3				
V4				

LOW : Mutual Exclusive Outcomes (0), Str.- : Strong Negative Association (1), Mod.- : Moderate Negative Association (2), Low- : Low Negative Association (3), None : Components Not Associated (4), Low+ : Low Positive Association (5), Mod.+ : Moderate Positive Association (6), Str.+ : Strong Positive Association (7) HIGH : One Outcome includes the other (8)

	V2	V3	V4	V5
V1	None	HIGH	Low+	Low+
V2		Low+	Low+	Low+
V3			Low+	Low+
V4				Low+

En la segunda pantalla se debe introducir la probabilidad conjunta entre cada uno de los pares de eventos, en el caso que se conozca. Una aproximación semiparamétrica se puede utilizar para estimar la asociación entre pares cuya probabilidad conjunta se desconozca. En el ejemplo se conoce la probabilidad conjunta del primer par de eventos y se desconocen todas las otras, por la que se asume una asociación positiva débil para todas a excepción entre el evento 1 y el evento 3 que se considera la máxima asociación posible ('Muerte Cardiovascular' y 'Muerte por cualquier motivo').

8.3.5. Revisión de los datos (pantalla 3)

En la tercera pantalla (Figura 10) Bin-CE muestra los datos cargados y corregidos si ha habido alguna asociación fuera de los rangos esperados. Se muestra también el TMR para el ER o para el EA que minimiza el TMR que es considerado ER. Se presenta el gráfico que ilustra el potencial que tiene cada uno de los EA al combinarse con el ER. Este gráfico, para cada posible combinación, ilustra el TMR máximo y mínimo que coincide con las probabilidades conjuntas de los extremos de los límites de Fréchet.

Este gráfico ayuda a los usuarios a identificar la influencia del grado de asociación en el TMR. También permite detectar algunas combinaciones que nunca reducen el TMR o, contrariamente, aquellas que independientemente del grado de asociación siempre lo van a reducir.

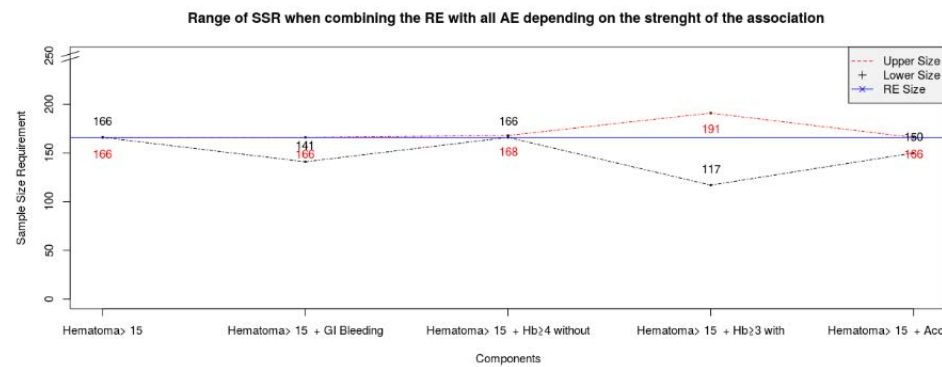
Figura 10 - Ejemplo de la pantalla 3. Revisión de los datos, TMR del ER y Gráfico del potencial de cada EA.

The data when the calculator repair the errors

n	label	re	rate	efect	ass1	ass2	ass3	ass4	ass5
1	GI Bleeding	0	0.014	0.19	NA	8.4e-05	0.000518	0.00084	0.000154
2	Hb>4 without	0	0.006	0.48	NA	NA	0.000222	0.00036	0.000066
3	Hb>3 with	0	0.037	0.22	NA	NA	NA	0.00222	0.000407
4	Hematoma> 15	0	0.060	0.09	NA	NA	NA	NA	0.000660
5	Access Compl	0	0.011	0.24	NA	NA	NA	NA	NA

Sample size requirement (starting with the RE)

Iter	Label	Rate	RR	n
1	1 Hematoma> 15	0.06	0.09	166



En esta pantalla los usuarios pueden chequear que los datos entrados sean los correctos (etiquetas, probabilidades marginales, efectos estimados para cada evento, la matriz triangular con las probabilidades o bien entradas por el usuario o bien estimadas por Bin-CE a partir de la aproximación semiparamétrica del usuario). También se puede consultar el TMR para el ER o el evento que ocupe su lugar si ninguno se ha especificado. Bin-CE ilustra el potencial que tiene la combinación de cada uno de los EA con el ER sin tener en cuenta el grado de asociación, estimando el TMR máximo y mínimo.

8.4. Algoritmo iterativo

En la primera iteración, Bin-CE estima el TMR para el ER (n_{ER}) y para los (K-1) hipotéticos CE contruidos al combinar el ER con los K-1 EA ($n_{k=2..K}^*$). Bin-CE selecciona de aquellos K-1 hipotéticos CE de dos componentes aquel con un SSR inferior, definiendo los K-1 $SSR_{k=2..K} = \frac{n_{ER}}{n_{k=2..K}^*}$. Bin-CE selecciona la combinación entre el ER y el EA que reduce más el TMR del ER(n_{ER}), si es que existe. En caso que existan dos o más EA que reducen exactamente los mismo el TMR, se seleccionará el primer EA de la lista.

En el caso que efectivamente exista una combinación que reduce el TMR del ER, Bin-CE elimina del grupo de eventos candidatos tanto el ER como el EA seleccionado y se añade el nuevo CE conformado por esos dos eventos al listado de candidatos. El CE de dos eventos que se añade al listado de eventos candidatos se considera el nuevo ER. El siguiente paso consiste en analizar si es eficiente agregar un nuevo AE al CE que se ha determinado en el paso anterior (el nuevo ER).

Una vez definido el CE seleccionado en la iteración 1 como nuevo ER y estimados sus parámetros, Bin-CE resuelve el mismo problema que el planteado inicialmente pero ahora con K-1 componentes (se ha eliminado el ER y un EA de la iteración anterior y se ha añadido el CE de dos componentes). Bin-CE finaliza el proceso iterativo en el caso que: a) no existan ningún EA que combinar o b) ningún EA combinable reduzca el TMR del ER (n_{ER}). En la última iteración el ER es un CE de diferentes eventos en el caso que haya encontrado alguna combinación.

Para definir el CE como nuevo ER en cada una de las iteraciones, Bin-CE ha de estimar sus parámetros: proporción marginal, efecto y el grado de asociación. La probabilidad marginal

del CE aplicando [2] y la estimación del efecto se ha estudiado en el proyecto anterior (16,32,36). No obstante, la estimación del grado de asociación entre el nuevo CE y el resto de EA, implica un análisis algo más detallado.

El grado de asociación entre el nuevo ER (CE definido en la iteración anterior) y el resto de AE candidatos (una vez eliminados los dos eventos) debe ser estimado, ya que hasta el momento no se ha definido. Se ilustrará dicha estimación utilizando la Figura 11, donde se presentan tres hipotéticos eventos (A, B y C). En este ejemplo, el evento A y B se combinan en un CE (CE_{AB}) en la iteración 1 ya que es la combinación que reduce más el TMR del ER (por ejemplo, el evento A).

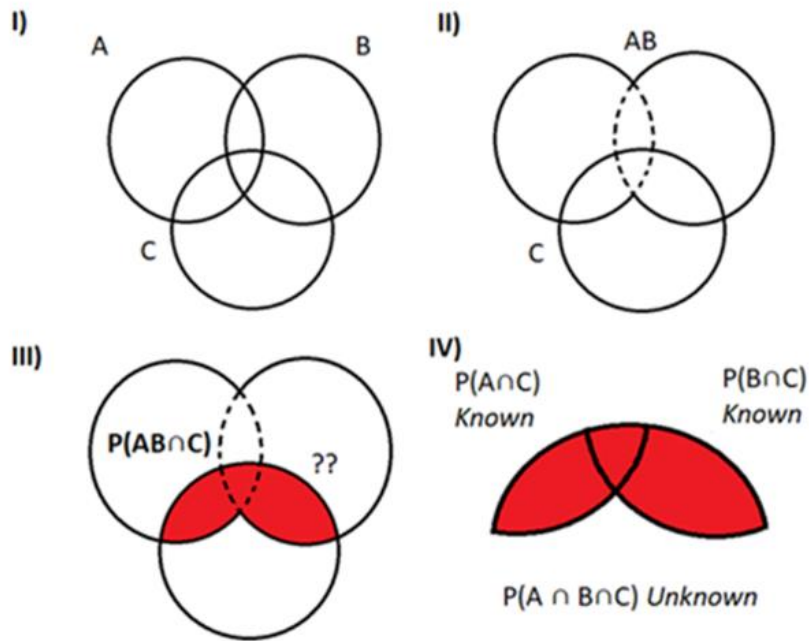
La probabilidad conjunta entre el CE_{AB} y el evento C se ha coloreado de color rojo y es la que se debe estimar. La probabilidad conjunta entre los eventos AC y BC es conocida o estimada a partir de la aproximación semiparamétrica, pero la probabilidad conjunta de los tres eventos (π_{ABC}) es desconocida.

Este problema se ha resuelto en Bin-CE, asumiendo que una buena aproximación al grado de asociación de tercer orden podría ser $\pi_{ABC} = \pi_{AC} * \pi_{AB}$. Esta es una decisión arbitraria. Sin embargo, una buena solución a este problema podría ser que en próximas versiones de Bin-CE fuera el mismo usuario quien pudiera decidir qué tipo de asociación de tercer orden considera.

Para estimar la asociación entre los eventos CE_{AB} y C, que es el parámetro de interés, es necesario estimar π_{ABC} . La probabilidad conjunta de que ocurran los tres eventos (π_{ABC}) debe estar sujeta a los límites de Fréchet. π_{ABC} y por lo tanto también está acotada inferior y superiormente. El producto de las dos probabilidades conjuntas ($\pi_{AC} * \pi_{AB}$) garantiza que

π_{ABC} siempre estará dentro de los límites válidos. Este proceso se debe repetir tantas veces como eventos aún sean candidatos a ser combinados en futuras iteraciones.

Figura 11 – Asociación de orden dos y tres entre eventos hipotéticos (A, B y C).



La asociación entre un nuevo hipotético CE formado por la combinación de A y B y el evento adicional C (probabilidad coloreada de color rojo) es el resultado de combinar la probabilidad conjunta entre los pares de eventos AC y BC. El valor de esta probabilidad según [2] puede estimarse como: $Prob((X_{AC} = 1) \cap (X_{BC} = 1)) = \pi_{AC} + \pi_{BC} - \pi_{ABC} \approx \pi_{AC} + \pi_{BC} - \pi_{AC} * \pi_{BC}$. Bin-CE estima la probabilidad entre los tres eventos a partir de la esperanza matemática ($\pi_{AC} * \pi_{BC}$). Aunque esta es un aproximación al verdadero valor de π_{ABC} , el error potencial tiene que ser bajo, pues el verdadero valor puede estar entre los límites de Fréchet ($\max\{0; \pi_{AC} + \pi_{BC} - 1\} \leq \pi_{ABC} \leq \min\{\pi_{AB}, \pi_{BC}\}$).

8.5. Resultados (pantalla 4)

Bin-CE en la pantalla de resultados, presenta una tabla con la mejor combinación de eventos, la probabilidad de ocurrencia del ER y la de los CE secuencialmente si los hay, sus efectos, el TMR y la reducción del TMR (teniendo como referencia el TMR del ER) (Figura 12).

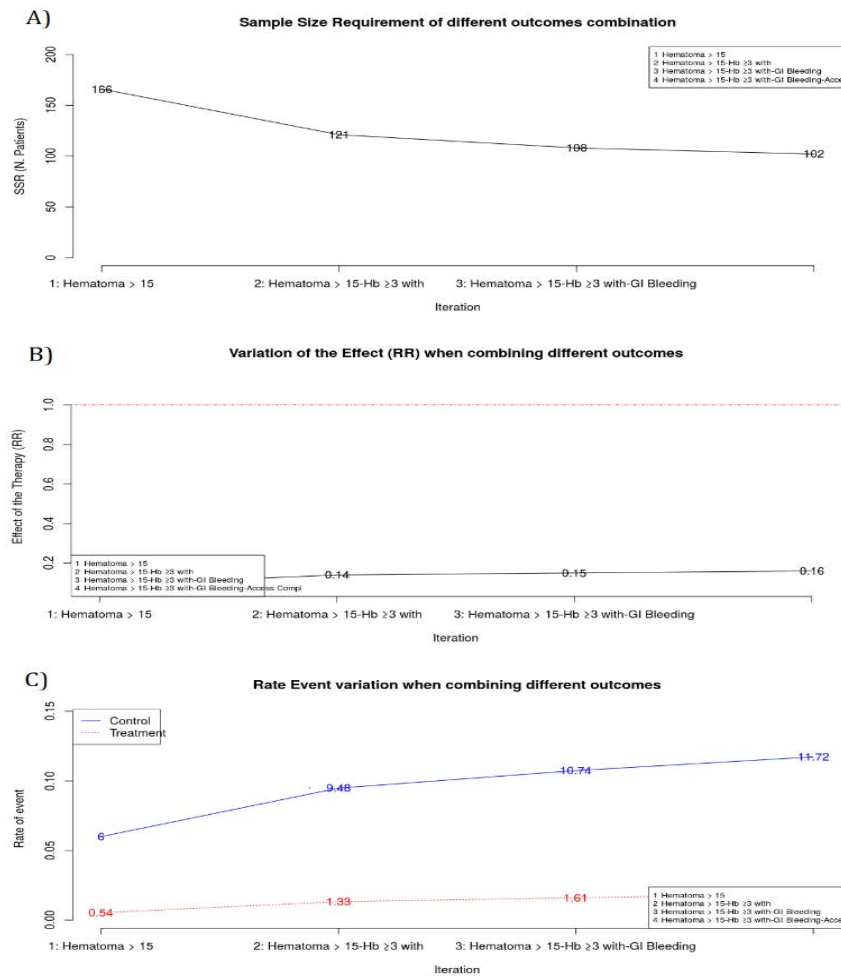
Figura 12 – Tabla de información presentada en la pantalla 4 de Resultados.

Input Data (ratio of events, effects)		Association	Data	Results		
Iter		Label	Rate	RR	n	%SSR
1	1	Hematoma > 15	0.06	0.09	166	100.00
2	2	Hematoma > 15-Hb \geq 3 with	0.0948	0.14	121	72.89
3	3	Hematoma > 15-Hb \geq 3 with-GI Bleeding	0.1074	0.15	108	65.06
4	4	Hematoma > 15-Hb \geq 3 with-GI Bleeding-Access Compl	0.1172	0.16	102	61.45

Cada línea corresponde al CE seleccionado en cada iteración de Bin-CE. Se presenta: la etiqueta de los eventos combinados, la probabilidad de ocurrencia, el efecto del CE (RR), el TMR para cada grupo y la proporción de pacientes comparando con e TMR de la iteración 1.

Bin-CE también presenta gráficamente como varían el TMR, la proporción de ocurrencia y el efecto en cada una de las iteraciones (Figura 13).

Figura 13 – Gráficos presentados en la pantalla 4 de Resultados



A) TMR en cada iteración. El eje Y representa el TMR y el eje X representa cada iteración.

En el ejemplo, en la primera iteración el TMR del ER 'Hematoma > 15' es de 166 individuos por grupo. En la segunda iteración, el CE 'Hematoma > 15' y/o 'Descenso de la Hemoglobina ≥ 3 g/dl en hemorragia abierta' precisa un TMR de 120, inferior al del ER. Bin-CE en la última iteración combina 4 componentes y el TMR es de 101 individuos.

B) Efecto de la intervención en cada iteración de Bin-CE. En el ejemplo, el efecto en la primera iteración es de un RR de 0,09, éste crece (decrece la eficacia) hasta 0,16 en la última iteración.

C) Proporción de ocurrencia en cada iteración. Este gráfico, en cada iteración de Bin-CE, se dibuja la proporción de ocurrencia tanto en el grupo control (línea azul) y en el grupo intervención (línea roja). La probabilidad de ocurrencia inicial en el grupo control es del 6% y de 0,54% en el grupo intervención. En la segunda interacción se pasó a 9,48% y 1,33% respectivamente. En la tercera y cuarta iteración la variación de la probabilidad de ocurrencia varió poco, como se observó en A) donde la reducción del TMR de las últimas iteraciones fue escasa.

8.6. Ejemplo Numérico

El estudio STEMI-RADIAL(61) es ensayo aleatorizado, multicéntrico y con brazos paralelos. Los pacientes incluidos referían un IAM dentro de las primeras 12 primeras horas y recibían tratamiento invasivo. Estos pacientes fueron aleatorizados a que las técnicas invasivas fueran a partir de acceso femoral o radial. La hipótesis de base es que la aproximación radial es más segura que la aproximación femoral.

El evento primario del estudio fue la incidencia acumulada de '*Sangrado Mayor*' y '*Complicaciones del acceso vascular*' a los 30 días. Los componentes del CE fueron: '*Sangrado Gastrointestinal*', '*Descenso de la hemoglobina $\geq 4g/dl$ sin sangrado*', '*Descenso de la hemoglobina $\geq 3g/dl$ con sangrado*', '*Hematoma $> 15cm$* ', '*Transfusión (diferente a bypass arterial)*' y '*Complicaciones del acceso vascular*'. Como evento secundario a los 30 días se incluyó el CE MACE, definido como combinación de '*Muerte*', '*IAM*', '*íctus*' y '*bypass coronario*'. Como en el estudio STEMI-RADIAL incluye eventos isquémicos y hemorrágicos ilustraremos el funcionamiento de Bin-CE separadamente en ambos grupos de componentes (Tabla 6). Las Figuras 8, 10 y 12 utilizadas para ilustrar el uso de Bin-CE, hacen referencia a los eventos hemorrágicos.

Tabla 6 – Eventos primarios y secundarios del estudio STEMI-RADIAL.

	Radial (n = 348)		Femoral (n=359)		Efecto ^b		
	n	%	n	%	RR	OR	Dif.
Eventos Isquémicos							
Muerte	8	2,30%	11	3,06%	1,33	1,34	0,77%
Infarto	4	1,15%	3	0,84%	0,73	0,72	-0,31%
Ictus	1	0,29%	1	0,28%	0,97	0,97	-0,01%
Bypass	32	9,20%	28	7,80%	0,85	0,84	-1,40%
Eventos Hemorrágicos							
Sangrado Gastrointestinal	5	1,44%	1	0,28%	0,19	0,19	-1,16%
Hb descenso ≥ 4g/dl sin sangrado	2	0,57%	1	0,28%	0,48	0,48	-0,30%
Hb descenso ≥ 3g/dl con sangrado	13	3,74%	3	0,84%	0,22	0,22	-2,90%
Hematoma > 15 cm	21	6,03%	2	0,56%	0,09	0,09	-5,48%
Transfusión ^a	3	0,86%	0	0,00%	0,00	0,00	-0,86%
Complicación Acceso Vascular	4	1,15%	1	0,28%	0,24	0,24	-0,87%

^a La transfusión se excluyó de los análisis ya que en el grupo femoral no se presentó ningún caso y el efecto estimado es inverosímil. ^b: El efecto fue estimado utilizando el Riesgo Relativo, Odds Ratio y diferencia de proporciones según [3.1 – 3.3]. Bin-CE utiliza el Riesgo Relativo. Hb: Hemoglobina, RR: Riesgo Relativo, OR: Odds Ratio y Dif: Diferencia de proporciones.

8.6.1.Resultados

En la primera pantalla (Figura 8), el usuario debe introducir la probabilidad de ocurrencia de cada componente en el grupo radial (grupo control), el efecto esperado a partir del RR y el grado de asociación (parecida a Figura 9) para cada par de componentes posibles. Como en el ejemplo el grado de asociación es desconocido se utilizó la aproximación semiparamétrica. Bin-CE retorna la estimación de la probabilidad conjunta utilizando [16 - 18] (Figura 10). Entre los eventos hemorrágicos, por simplicidad se asumió que los componentes no estaban asociados [16].

Bin-CE en la tercera pantalla, grafica el potencial de cada uno de los EA cuando se combinan con el ER, sin tener en cuenta el grado de asociación y por lo tanto como varía el TMR en

todo el rango de asociaciones. Como podemos observar, si se combinan los eventos '*Hematoma > 15cm*' y '*Hb ≥ 3g/dl con sangrado*' el TMR del CE resultante puede variar entre 117 y 191. Se debe tener en cuenta que como el TMR para el ER es de 166 individuos, el grado de asociación entre dichos componentes puede ser determinante. Ningún evento fue declarado ER, pero Bin-CE seleccionó '*Hematoma > 15cm*' como ER porque era el evento que inferior TMR tiene por sí solo.

En la Figura 12 y la Figura 13 se comprueba que el CE óptimo en cuanto al TMR es la combinación '*Hematoma >15cm*', '*Hb ≥ 3g/dl con hemorragia*', '*Sangrado gastrointestinal*' y '*Complicación de acceso*', que requiere un TMR de 102 individuos por grupo. Utilizando esta combinación, reducimos un 38,55% el TMR (61,45 % del TMR inicial), pasando de 166 individuos a 102 por grupo.

Se puede comprobar que la combinación '*Hematoma >15cm*' y '*Hb ≥ 3g/dl con hemorragia*' es la combinación que contribuye más a la reducción del TMR (reducción del 27,11%), pasando de 166 a 121 (reducción de 45 individuos). La combinación del tercer componente implica una reducción menor pasando de 121 a 108 (reducción de 13 individuos y una reducción adicional del 7,83% del TMR inicial). La última combinación reduce el TMR en 6 individuos, una reducción del 3,61%. La contribución de cada iteración en la reducción del TMR inicial es paulatinamente inferior: 27,71%, 7,83% y 3,61% respectivamente. Bin-CE no incluye más eventos ya que ningún evento restante tiene la capacidad de reducir el TMR del CE de la cuarta iteración.

8.6.2. Análisis de sensibilidad

Se ha variado el grado de asociación utilizando la aproximación semiparamétrica pasando por los 9 escenarios posibles, considerando que todos los pares de eventos tenían el mismo grado de asociación (Tabla 7 y Tabla 8). En el ejemplo inicial se ha asumido que todos los eventos no estaban asociados, ahora se repite el proceso asumiendo todo el rango de posibles grados de asociación.

La tabla 7 presenta los resultados del análisis de sensibilidad para los eventos hemorrágicos y la tabla 8 los de los eventos isquémicos. En cuanto a los componentes hemorrágicos, observamos que el grado de asociación es más determinante que en el caso de los eventos isquémicos en cuanto a los CE conformados.

Para los componentes hemorrágicos, se observa que cuando se pasa de asociaciones positivas muy intensa a asociaciones no tan intensas, el número de componentes combinados crece rápidamente. El caso más favorable, desde el punto de vista del TMR es cuando existe la asociación negativa máxima (probabilidad conjunta entre eventos igual a cero), en ese caso teórico el TMR se situaría en 97 individuos por grupo. En los eventos hemorrágicos la asociación es importante tanto en cuanto el número de eventos combinados como en la potencial reducción. En el caso óptimo (eventos disjuntos) se pasa de una talla de 166 a 97, que representa una reducción del 58% del TMR inicial.

Tabla 7 – Análisis de sensibilidad. Resultados de Bin-CE cuando varía el grado de asociación en todo su rango aplicando los límites de Fréchet.

Eventos Hemorrágicos.

<i>Grado de asociación</i>	<i>CE</i>	<i>#Componentes</i>	<i>%¹</i>	<i>RR</i>	<i>TMR</i>	<i>%²</i>
Límite Inferior	<i>Hematoma > 15-Hb ≥ 3 con-Sang GI- Compl Acces-Hb ≥ 4</i>	5	12,80%	0,18	97	58,43%
Negativa Fuerte	<i>Hematoma > 15-Hb ≥ 3 con- Sang GI</i>	3	11,01%	0,15	105	63,25%
Negativa Moderada	<i>Hematoma > 15-Hb ≥ 3 con- Sang GI</i>	3	10,92%	0,15	106	63,86%
Negativa Leve	<i>Hematoma > 15-Hb ≥ 3 con - Sang GI - Compl Acces</i>	4	11,84%	0,16	101	60,84%
Sin asociación	<i>Hematoma > 15-Hb ≥ 3 con - Sang GI - Compl Acces</i>	4	11,72%	0,16	102	61,45%
Positiva Leve	<i>Hematoma > 15-Hb ≥ 3 con - Sang GI - Compl Acces</i>	4	9,40%	0,14	122	73,49%
Positiva Moderada	<i>Hematoma > 15-Hb ≥ 3 con - Sang GI</i>	3	7,74%	0,14	149	89,76%
Positiva Fuerte	<i>Hematoma > 15- Sang GI</i>	2	6,33%	0,10	160	96,39%
Límite Superior	<i>Hematoma > 15</i>	1	6,00%	0,09	166	100%

Hb: Hemoglobina, Sang GI: Sangrado Gastrointestinal, Compl Acces: Complicaciones de acceso. %¹: Incidencia acumulada del CE. %²: Proporción del TMR para el RE en cada una de las combinaciones.

Tabla 8 – Análisis de sensibilidad. Resultados de Bin-CE cuando se varía el grado de asociación en todo su rango aplicando los límites de Fréchet. Eventos Isquémicos.

<i>Grado de asociación</i>	<i>CE</i>	<i>#Componentes</i>	<i>%¹</i>	<i>RR</i>	<i>TMR</i>	<i>%²</i>
Límite Inferior	<i>Bypass - IAM</i>	2	10,35%	0,84	4724	73,62%
Negativa Fuerte	<i>Bypass - IAM</i>	2	10,32%	0,84	4771	74,35%
Negativa Moderada	<i>Bypass - IAM</i>	2	10,30%	0,84	4819	75,10%
Negativa Leve	<i>Bypass - IAM</i>	2	10,27%	0,84	4867	75,85%
Sin asociación	<i>Bypass - IAM</i>	2	10,24%	0,84	4917	76,62%
Positiva Leve	<i>Bypass - IAM</i>	2	9,98%	0,84	5230	81,50%
Positiva Moderada	<i>Bypass - IAM</i>	2	9,72%	0,84	5580	86,96%
Positiva Fuerte	<i>Bypass - IAM</i>	2	9,46%	0,85	5973	93,08%
Límite Superior	<i>Bypass</i>	1	9,20%	0,85	6417	100%

%¹: Incidencia acumulada del CE. %²: Proporción del TMR para el RE en cada una de las combinaciones.

8.7. Conclusiones del proyecto 3

En este proyecto hemos presentado Bin-CE como una herramienta que ayuda a tomar la decisión del número de eventos combinar en un CE binario en función de la minimización del TMR. Bin-CE es una herramienta pública y de acceso libre (<https://uesca-apps.shinyapps.io/bincep/>) para todos los usuarios que accedan. Su utilización ha sido pensada para que los usuarios puedan manejarla sin muchos conocimientos estadísticos. Su entorno reactivo aporta solidez a la herramienta, ya que rápidamente se puede realizar modificaciones en los parámetros de entrada y actualizarse los resultados inmediatamente sin ejecutar ninguna función preestablecida.

Aunque muchos hemos sido los autores que hasta el momento hemos desarrollado las fórmulas y bases matemáticas para estimar el TMR o el grado de asociación de un CE binario(32,36,51,59,79,92,117), en ningún caso que conozcamos se ha generalizado el problema a la combinación de más de 2 eventos. Basándonos en los trabajos de Sozu (32) y Xiong (36), Bin-CE ha sido programado para estimar el TMR de hasta 9 combinaciones de eventos en un CE, y siempre teniendo en cuenta que debía ser una herramienta que en la que el grado de asociación entre eventos fuera un elemento esencial. De esta forma, Bin-CE incorpora una forma sencilla para estimar la asociación real entre dos eventos binarios sin necesidad de conocer en profundidad el tema, a partir del conocimiento clínico.

9. Discusión

En este trabajo de Tesis hemos profundizado en el estudio del posible beneficio cuando utilizamos un CE binario en cuanto a la reducción del Tamaño de Muestra Requerido (TMR). Se trata de un problema básicamente estadístico o matemático, por lo que conociendo en profundidad la influencia de los diferentes parámetros involucrados en el cálculo del TMR es posible cuantificar dicho 'beneficio'. No obstante, esta Tesis pretende ir un paso más allá y sentar las bases matemáticas sobre un problema que no se ha planteado hasta el momento, como es el cálculo del TMR en el caso de un CE binario de tres o más eventos. Se trata de un trabajo esencialmente pragmático cuya motivación es facilitar la tarea del investigador clínico que se plantea el uso de un CE.

El cálculo del TMR depende de tres parámetros: incidencia de cada uno de los eventos combinados, efecto de la intervención sobre cada uno de los eventos considerados y el grado en que se asocian dichos eventos. Nuestro trabajo ha documentado que la cuantificación de la asociación entre dos eventos binarios puede ser determinante en el Tamaño de Muestra Requerido (TMR). La potencial influencia de la asociación en el TMR, depende tanto de la propia magnitud de asociación, como de la incidencia de cada uno de los eventos. Un mismo valor de asociación por ejemplo una probabilidad conjunta del 4%, indicará mayor o menor grado de asociación

en función de la incidencia acumulada de cada uno de los eventos combinados. Cuanto más incidentes menor será la asociación, para esta posición es imprescindible conocer cuál es el valor que indica el escenario de no asociación.

En este trabajo se ha definido que el uso de un CE es más '*eficiente*' que el uso de un único Evento Relevante (ER), simplemente teniendo en cuenta si el TMR es inferior o no. Hay que tener en cuenta que esta es una aproximación puramente pragmática y algo simplista, pues el concepto de 'eficiencia' en este contexto no es completo sin tener en cuenta que, por ejemplo, cuanto mayor es el número de eventos combinados mayor será la complejidad de interpretación y análisis. Los diferentes pros y contras no se pueden cuantificar de forma objetiva, por lo que hemos restringido el trabajo a la reducción del TMR. Se ha estudiado como afectan los diferentes parámetros en cuanto al 'beneficio' de combinar dos eventos binarios en un CE, dedicando un especial interés al grado de asociación, que es el parámetro más desconocido.

En este sentido hemos concluido que es siempre aconsejable combinar un evento cuyo efecto es superior al efecto del ER, incluso teniendo una incidencia de ocurrencia inferior a éste e independientemente del grado de asociación. Además, hemos documentado que el grado de asociación en determinados escenarios es determinante (Tabla 5), siendo este el escenario donde debe tenerse una especial atención a la cuantificación. En este sentido, el potencial 'beneficio' o reducción del TMR mediante el uso de un CE se minimiza cuando mayor es el grado de asociación (asociación positiva). Por ello interesa combinar eventos o bien no asociados, o con asociaciones grandes, pero en sentido negativo (la ocurrencia de un evento disminuye la probabilidad de ocurrencia del otro). El escenario donde el grado de asociación es más determinante ocurre cuando se combinan un EA sobre el que la intervención tiene una magnitud de eficacia inferior a la que tiene sobre el ER. En este caso,

el TMR puede ser mayor o menor del requerido con el uso aislado del ER, y ello dependerá de la magnitud de la intervención sobre cada uno de los eventos asociados.

Cuando se define un CE de tres o más eventos es necesaria la estimación de un número superior de parámetros. A parte de los parámetros ya estimados en el caso de un CE de dos eventos, en este nuevo escenario también es necesario estimar los momentos de grado 3, 4 o superiores. Abordar el problema de forma precisa es prácticamente imposible puesto que: a) el grado de asociación entre 3 o más eventos es desconocido casi siempre y b) el número de parámetros que se deben estimar crece de forma exponencial con el número de eventos combinados. Con el objetivo de minimizar este problema y aportar una herramienta que ayude a los investigadores a tomar decisiones, decidimos construir Bin-CE (<https://uesca-apps.shinyapps.io/bincep/>) que solo tiene en cuenta el grado de asociación de cada par de eventos.

Bin-CE es una calculadora que estima el TMR cuando se pretende combinar diferentes eventos binarios en un CE. Es de acceso gratuito y su código es libre para que otros investigadores puedan adaptarla a escenarios específicos. Bin-CE minimiza el problema de la dimensión, combinando iterativamente los CE de dos eventos más “eficientes” desde el punto de vista del TMR. Aborda el problema del desconocimiento del grado de asociación entre eventos de una forma intuitiva. Así, en el caso que el usuario no conozca el grado de asociación puede decidir entre 9 escenarios diferentes, definidos por los límites de Fréchet, desde “eventos muy poco asociados” a “eventos muy asociados” e incluso escoger las opciones de máxima y mínima asociación. Esta aproximación semiparamétrica consideramos que es una forma original de estimar de forma aproximada, y en virtud de la experiencia clínica, el grado de asociación real que difícilmente se conoce con precisión. Bin-CE ha sido construido utilizando la librería Shiny de RStudio, que permite programar las

funciones de forma reactiva de tal forma que permite realizar estudios de sensibilidad de forma inmediata.

En trabajos previos algunos autores (32,36) ya habían establecido las bases matemáticas y el enfoque general de cómo calcular el TMR para un CE cuando se combinan dos eventos binarios. En esos trabajos ya se anunciaba que el grado de asociación entre los eventos combinados era un factor a tener en cuenta en el cálculo. Sin embargo, las publicaciones de dichos trabajos se limitan a la estimación del TMR matizando que éste variará en función de la asociación. En los primeros trabajos, no se planteó que el combinar dos eventos puede incrementar el TMR en vez de reducirlo, ni que el grado de asociación pudiera ser uno de los factores claves.

Geraldine Rauch (27,53,121,122), entre otros autores, aportó evidencia en cuanto al cálculo del TMR focalizándose en otros aspectos diferentes al grado de asociación, como es el problema de la multiplicidad de contrastes. Hasta este momento todos los autores consideraron la correlación como estimador de la asociación, seguramente por su enfoque más matemático.

Hasta donde nosotros conocemos, Sozu(32) en 2010 fue el primer autor que simuló diferentes escenarios a partir de diferentes correlaciones. Su objetivo no era determinar cómo afecta la relación entre los eventos combinados, sino ejemplificar las ecuaciones desarrolladas. En ese mismo trabajo, publicado en la revista *Statistics in Medicine*, ya se comunicó que el rango de la correlación cuando se estimaba en dos variables binarias no estaba limitado entre -1 y +1, sino que correspondía a un rango inferior. En el transcurso

del presente trabajo de Tesis, otro trabajo parecido ha sido publicado por Bofill(15) mostrando conclusiones parecidas a nuestro trabajo publicado (117).

Los anteriores autores se focalizaron en el cálculo del TMR desde un punto de vista matemático o estadístico, presentando con mayor o menor detalle y mayor o menor profundidad las diferentes fórmulas y variaciones de éstas, que nosotros hemos utilizado en la Tesis. No obstante, en ninguno de ellos, excepto en los últimos trabajos de Bofill(15) y Gómez(12), se presentan guías con las que los clínicos puedan discriminar entre los diferentes escenarios sin necesidad de cuantificar exactamente el TMR. En ambos trabajos las conclusiones son parecidas, partiendo desde abordajes matemáticos diferentes, lo cual confiere un grado de robustez superior. El cálculo del TMR cuando se utilizan CE en el contexto del análisis de supervivencia ha sido estudiado también en profundidad, desde el punto de vista matemático (13,14,48,49).

Aunque nuestro trabajo se basa y fundamenta en las ecuaciones publicadas por los autores comentados, hemos intentando profundizar más en el proceso de cuantificación de los diferentes parámetros. Además, hemos generalizando el cálculo para CE de más de 2 eventos y hemos aportado herramientas que puedan ayudar a los clínicos a tomar decisiones cuando utilicen CE. Ningún autor, que conozcamos nosotros, ha planteado hasta el momento el reto de cuantificar el TMR cuando se combinan más de dos eventos. La Unitat d'Epidemiologia del Servei de Cardiologia, que es donde este trabajo se ha desarrollado, es una unidad especializada en metodología de la investigación de un servicio clínico-asistencial, por lo que su desarrollo presenta un enfoque clínico, a parte de la necesaria profundidad matemática que le corresponde.

En el trabajo hemos adaptado los métodos teniendo en cuenta que los usuarios de estas técnicas serán fundamentalmente investigadores clínicos. Así, aunque podría haber justificaciones matemáticas para el uso preferente de *odds ratio* para cuantificar el efecto de la intervención y el uso de la *correlación* para estimar el grado de asociación entre dos eventos, estos parámetros son menos intuitivos, desde el punto de vista clínico, que el *riesgo relativo* o la *diferencia de proporciones*, y que la probabilidad conjunta. No obstante, en nuestro trabajo hemos demostrado que estos son intercambiables entre sí y preferibles en el entorno de los ensayos clínicos (84,120).

En la misma línea hemos determinado que la correlación aun teniendo características matemáticas muy interesantes en el contexto de variables continuas, no es el mejor indicador cuando se aplica a variables binarias. De la misma forma, ningún autor ha tenido en cuenta que la correlación está acotada por un límite inferior y otro superior, a excepción de Sozu(32). Y, en el caso de Sozu, aun habiendo definido los límites de la correlación, no los aplicó en sus simulaciones.

Cuando profundizamos en el concepto de asociación descubrimos que existen trabajos muy rigurosos sobre el tema, como los publicados por Warren(85,86,97) y Todeschini (99). No obstante, ningún de estos trabajos fue desarrollado en el contexto de los CE, sino en otros campos de la investigación diferentes a los ensayos clínicos o investigación clínica. El concepto estadístico de asociación (89,96) es conocido desde hace tiempo y está desarrollado de forma exhaustiva por diferentes matemáticos ya a finales del siglo pasado. Por ejemplo, Fréchet(91) ya a mediados del pasado siglo comunicó que el grado de asociación entre dos variables categóricas estaba acotado, y sentó las bases a partir de la cual Sozu calcula los límites de la correlación cuando se aplica a eventos binarios. En este trabajo nosotros hemos descrito diferentes alternativas a la correlación que, bajo nuestro

punto de vista, tienen una mayor interpretabilidad clínica. Consideramos que la probabilidad conjunta o que la “proporción de *solapamiento* de los eventos”, son buenas alternativas a la correlación. Hasta el momento, ningún autor ha abordado otros parámetros de cuantificación de la asociación de eventos, en el contexto de los CE, que puedan resultar más sencillos e intuitivos para los clínicos.

Una aportación adicional del presente trabajo es la generalización de la estimación del TM, cuando se combinan tres o más eventos en un CE. El TMR depende de la distribución conjunta de los K eventos combinados(82,83). Esta distribución k -dimensional queda plenamente definida a partir de $2^k - 1$ parámetros. k parámetros ya son utilizados para cuantificar las k probabilidades marginales y los $2^k - k - 1$ restantes corresponden a asociaciones de dos, tres o más eventos, según el número de eventos considerados. Hemos propuesto un algoritmo que resuelve de forma iterativa el problema k -dimensional convirtiéndolo en un problema más sencillo. De esta manera, Bin-CE determina cual es CE de dos eventos óptimo en cuanto al TMR y lo incorpora a los eventos candidatos a formar parte del CE final, pasando de un problema de dimensión k a un problema de dimensión $k-1$. En el aplicativo Bin-CE se han incorporado los conocimientos de otros autores y los adquiridos. Sin embargo, al renunciar a estimar la distribución estadística de forma precisa se introduce un error de estimación, concretamente la estimación de la asociación entre más de dos eventos.

La reducción del TMR es consecuencia del incremento de la potencia estadística como ya se ha estudiado en profundidad (5,6,13,16,25,32-36). No obstante, esta reducción del TMR no es la misma en todas las combinaciones posibles de eventos que forman parte de un CE, comprobándose que existen uniones más 'eficientes' que otras. Bin-CE, al tratarse de una herramienta que estima todas las posibles combinaciones, garantiza que a cada paso se utiliza la combinación óptima de eventos. Sin embargo, el investigador ha de ser muy cauto con el uso de esta potencial utilidad, pues la principal desventaja asociada al uso de CE proviene de potenciales problemas de interpretación de los resultados del estudio, lo cual no resuelve Bin-CE. Es más, una incorrecta interpretación en este contexto suele sobreestimar el efecto real de la intervención en la población(2,23).

El impacto metodológico de los resultados obtenidos en esta Tesis, en cuanto al cálculo del TMR, hay que ponerlo en contexto. Las bases matemáticas ya han sido definidas ampliamente de forma general y específicamente en el cálculo del TMR. No obstante, la aplicación de esta formulación utilizando estudios de simulación en diferentes escenarios aporta respuestas a preguntas que aún no habían sido respondidas. Por ejemplo, aunque ya se intuía que combinar un EA con una eficacia mayor y una incidencia mayor reduciría el TMR, en nuestro estudio se pudo confirmar y cuantificar en qué grado se reduce. Además, en nuestro trabajo se documenta que existen combinaciones de variables de resultado donde el grado de asociación es determinante. Así, aunque existe la creencia de que la unión de diferentes eventos en un CE reduce siempre el TMR, hemos comprobado que esto no es cierto y que depende tanto del efecto de la intervención, como de la incidencia del evento, pero también, y esto es lo realmente determinante, del grado de asociación de los eventos considerados.

Por todo ello pensamos que la estimación del grado de asociación es importante, sobre todo en aquellos escenarios donde puede ser determinante (*combinar un EA de menor eficacia*). Asimismo recomendamos, como otros autores(9,42,59), que al publicar los resultados de ECA donde la variable de resultado es un CE se publique también la magnitud de relación entre los eventos. La estimación de la relación entre eventos es necesaria para estimar el TMR de forma directa o utilizando una aproximación, como Bin-CE incluye.

El estudio de los diferentes parámetros para estimar el grado de asociación ha sido un ejercicio matemático intenso. El conocimiento adquirido se ha aplicado en el estudio del TMR del CE para dos eventos binarios, donde hemos simulado diferentes escenarios estimando el grado de asociación a partir del índice Jaccard. A partir del conocimiento de su distribución se ha implementado en Bin-CE la aproximación semiparamétrica. En este caso, la aplicación de los límites de Fréchet, el estudio de su distribución y sus métricas, no deja de ser un ejercicio matemático relativamente sencillo, aunque desde el punto de vista clínico no sea evidente. A nivel clínico, hemos recomendado distintos indicadores a la correlación para estimar el grado de asociación.

El impacto real de esta Tesis consideramos que está en el aplicativo Bin-CE. Hemos aportado a la comunidad científica una herramienta gratuita y que permite a los investigadores estimar el TMR cuando combinan diferentes eventos binarios de una forma rápida y sencilla. Los usuarios pueden variar los diferentes parámetros con el objetivo de profundizar en la potencial reducción del TMR de su caso específico. En el caso de CE en el contexto de tiempo hasta evento existe alguna herramienta parecida (CompARE)(12) pero no en el caso de CE binarios.

Bin-CE, permite al usuario cambiar algunos parámetros generales como el error tipo I y tipo II o bien la aproximación asumida en el contraste de hipótesis. En la actual versión, Bin-CE calcula el TMR para el contraste de diferencia de proporciones que, aunque no es el único, sí que es el más utilizado. Por otro lado, el investigador podría estar interesado en comprobar que el efecto de cierta intervención es estrictamente superior al efecto en el grupo control. Este tipo de contraste actualmente no está implementado. Bin-CE pretende ser una herramienta que crezca con las diferentes necesidades y algunas ampliaciones que ya se han detectado.

Imaginémonos un ECA donde la variable respuesta es un CE y tras realizar el estudio se comprueba que, efectivamente, el grupo intervención tiene una incidencia del CE diferente al grupo control. El clínico cuestionará si el efecto estimado depende de algún evento en concreto o no y, por lo tanto, se debería realizar un nuevo contraste de hipótesis el cual requiere un incremento en el TMR. Montori y Ferreira recomendaban combinar eventos con un efecto y una proporción parecida. El problema de multiplicidad de contraste ya ha sido estudiado en profundidad, incluso en el contexto de CE binarios de 2 eventos(11,39,40,53,57,123), pero está incluido en la versión actual de Bin-CE. Sin duda debe incorporarse en futuras versiones de Bin-CE.

Existe una gran heterogeneidad en cuanto al tipo y número de contrastes de hipótesis que uno puede plantear en el contexto de los CE. Por ejemplo, un contraste para determinar si la probabilidad de que ocurran como mínimo x eventos de los y combinados en un grupo sea diferente al otro o cualquier otro tipo de contraste que a uno se le pueda ocurrir. Bin-CE, en su versión actual, no contempla dicha heterogeneidad. Su código fuente está disponible en el Anexo 6 de este trabajo, para que los usuarios expertos en programación puedan adaptarlo a sus necesidades.

En nuevas versiones de la herramienta indudablemente deberemos relajar algunas asunciones que se han prefijado con el objetivo de simplificar el problema. El grado de asociación se estima para el grupo control y se asume que este será equivalente en el grupo intervención. Esta asunción es verosímil, pero podría no ser así, por lo que Bin-CE debería ser modificado para que se pudiera entrar la matriz de asociaciones también para el grupo intervención. En esta misma línea, hasta el momento Bin-CE permite cuantificar el efecto a partir del *Riesgo Relativo*. Esta asunción deberá relajarse en nuevas versiones, y que sea el usuario el que decida si entrar el efecto a partir de otros indicadores: *diferencia de proporciones* o *Odds Ratio*.

Una de las limitaciones más importantes de Bin-CE sin duda es que hasta el momento no se ha validado de forma pragmática. La futura validación deberá contemplar dos aspectos diferentes. En primer lugar, se deberá demostrar que el cálculo del TMR es el mínimo para cuantificar el efecto previsto del CE con la potencia estadística fijada. En segundo lugar, se debe comprobar que el algoritmo iterativo detecta realmente el mejor CE a partir del conjunto de eventos candidatos. Bin-CE incorpora un algoritmo tipo '*greedy step-wise forward*' que resuelve localmente el problema asumiendo que se alcanzará una solución global. Estamos considerando pues que al combinar cada vez el EA al ER que más reduce el TMR se llegará al mínimo TMR global. A partir de simular todas las posibles combinaciones de CE deberemos asegurar que los resultados coincidan.

La validación de que el TMR es el correcto debe realizarse mediante estudios de simulación y técnicas de remuestreo. Se deberán simular diferentes eventos binarios plausibles en el contexto clínico y la combinación de estos teniendo en cuenta que existe un grado de asociación entre ellos y un efecto teórico. Será imprescindible simular diferentes muestras

para determinar si el TMR presentado por Bin-CE alcanza la potencia estadística que se detecta en las simulaciones. Esta simulación debe realizarse aplicando las técnicas de simulación propuestas por Oman (124,125).

Paralelamente, el algoritmo iterativo de Bin-CE encuentra la mejor combinación de dos eventos y ésta se incorpora en la nueva iteración hasta que no se encuentra ninguna combinación mejor o bien ya no existen eventos a ser combinados. Este algoritmo no resuelve el problema de forma específica, ya que para eso sería necesario conocer la distribución *k-variada*, lo cual hemos considerado inviable. Sin embargo, a partir de la simulación diferentes escenarios *k-variados* podríamos cuantificar el error cometido en esta aproximación. Sin haber hecho un estudio formal, consideramos que el error cometido es muy pequeño, ya que la asociación de tres eventos por el hecho de ser una asociación también debe estar dentro de los límites definidos de Fréchet. No obstante, estos límites por definición son muy inferiores a los límites de Fréchet cuando se aplican a la asociación de dos eventos. En definitiva, es plausible que el error asumido implique una pequeña variación en la incidencia del CE y, por lo tanto, también una pequeña variación del TMR sería necesario realizar una validación formal.

El número de ECA donde la variable de respuesta principal es un CE binario es probablemente muy inferior a los ECA donde se utiliza un CE en el contexto del análisis de supervivencia (tiempo hasta evento). Habitualmente se usan cuando el efecto de la intervención es un tiempo de seguimiento corto o bien cuando es difícil precisar el momento en que el evento es incidente. Aún que su uso es menos frecuente, cada vez existen más publicaciones donde la respuesta del ECA es un CE binario, especialmente en el contexto cardiovascular. Creemos que un estudio interesante sería realizar una búsqueda bibliográfica exhaustiva con el objetivo de determinar si el tamaño de muestra con el que se

realiza el estudio tiene o no suficiente potencia estadística una vez conocidos los parámetros reales de los eventos combinados. Este estudio creemos que es necesario, ya que el TMR debe calcularse como si el CE fuera un ER y por lo tanto el investigador debe estimar la incidencia y el efecto del CE. Esta aproximación puede hacer que el número de pacientes recogido realmente es suficiente para garantizar o no la potencia estadística deseable.

En resumen, este trabajo ha abordado el problema del cálculo del TMR cuando se utilizan eventos binarios. Hemos profundizado en el conocimiento sobre los CE de dos eventos binarios respondiendo al objetivo planteado. Se ha estudiado también en profundidad la cuantificación de la interrelación que se establece entre dos eventos. Fruto de este estudio se ha comprobado que la estimación de la asociación no es sencilla y depende de diferentes aspectos, como son el parámetro escogido para estimarla, la incidencia de cada evento o incluso el punto a partir del cual se determinan asociaciones positivas y negativas. Como resultado de este estudio hemos propuesto una forma sencilla de aproximar la estimación que se ha incorporado a la herramienta Bin-CE. Ésta ha sido creada para resolver un problema innovador que nadie hasta el momento había resuelto en el contexto de los CE binarios. Bin-CE ha sido desarrollado teniendo en cuenta las necesidades de los usuarios finales que no son más que los clínicos.

Como producto de esta Tesis se ha publicado un artículo en la *American Journal of Epidemiology* (Anexo 7). Se ha programado la herramienta Bin-CE, gratuita y accesible en : <https://uesca-apps.shinyapps.io/bincep/>. Actualmente se está realizando la segunda revisión de un artículo en la revista *PLoS One* donde se presenta Bin-CE.

10. Anexos

10.1. Anexo 1: Estimación de la probabilidad conjunta en el grupo intervención.

En el desarrollo de la Tesis hemos asumido que el grado de asociación entre dos eventos binarios es el mismo entre el grupo control y el grupo intervención. Esta asunción, que clínicamente tiene sentido, no implica necesariamente que el valor de la asociación coincida en ambos grupos.

Imaginémonos dos eventos con una probabilidad cada uno del 0,2 (20%) y el efecto del primero es bajo ($RR = 0,70$) y el efecto del otro evento es superior ($RR = 0,50$). Si asumimos que en ambos grupos los dos eventos no están asociados (escenario de independencia) y

utilizamos la probabilidad conjunta como indicador de asociación, fácilmente podemos cuantificar la asociación.

En el caso del grupo control, la probabilidad conjunta cuando consideramos eventos independientes es del 0,04 (4%) que corresponde al producto de las dos probabilidades marginales ($0,04=0,20*0,20$). En el caso del grupo intervención y asumiendo el mismo grado de asociación debemos multiplicar cuantificar primeramente las incidencias acumuladas de ambos eventos. La incidencia del primer evento será $0,2*0,7=0,14$ (14%) y de $0,2*0,5=0,10$ (10%) para el segundo. Una vez estimada la incidencia de cada uno de los eventos, el grado de asociación bajo el escenario de independencia es $0,14*0,10=0,014$ (1,4%).

Por lo tanto, queda demostrado que el hecho de asumir el mismo grado de asociación entre el grupo control y el grupo intervención no implica que la asociación tome el mismo valor numérico en ambos grupos. Haciendo un poco más allá, podemos observar que según los límites de Fréchet, la probabilidad conjunta del grupo control ($\pi_{12,j=1}$) y los límites de la probabilidad conjunta del grupo intervención ($\pi_{12,j=2}$) son distintos ya que sus probabilidades marginales también lo son.

En este punto se comprueba también estos los límites y rangos de ambos grupos no son simétricos. Fijémonos que las probabilidades conjuntas del grupo control e intervención pueden tomar cualquier valor comprendido entre los límites $Inf(\pi_{j12})$ y $Sup(\pi_{j12})$ definidos en la ecuación [10] como:

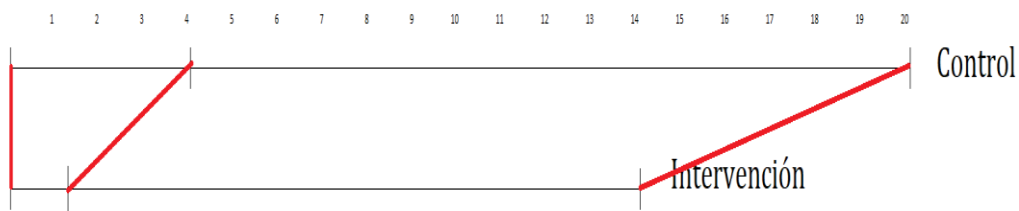
$$Inf(\pi_{j12}) = \max\{0; \pi_{j1} + \pi_{j2} - 1\}$$

$$Sup(\pi_{j12}) = \min\{\pi_{j1}, \pi_{j2}\}$$

De lo que se puede estimar que el rango de posibles valores para el ejemplo anterior en el grupo control va desde el 0 hasta el 0,20 y desde el 0 hasta el 0,14 en el grupo intervención. Por lo tanto, si asumimos la mayor de las asociaciones negativas (que coincide con el límite inferior) en ambos grupos se sitúa en cero. Pero asumiendo la mayor de las asociaciones positivas (límite superior del intervalo) observamos que no coinciden los valores.

La probabilidad conjunta del grupo control puede tomar valores entre 0%-20%, siendo la probabilidad conjunta del 4% el escenario que indica no asociación entre eventos (o el valor por lo que la asociación pasa de negativa a positiva). En el caso del grupo intervención el rango es inferior, del 0% al 14% y el valor que indica independencia es el 1,4% (Figura 13).

Figura 13 – Rangos no simétricos entre la asociación del grupo control e intervención



Teniendo en cuenta que el rango de las asociaciones no es simétrico entre ambos grupos y asumiendo que queremos el mismo grado de asociación entre grupos, deberemos estimar

la probabilidad conjunta del grupo intervención de forma distinta dependiendo si la probabilidad conjunta del grupo control es inferior o superior a la esperanza matemática. Fijémonos que cuando asumimos independencia en ambos grupos deberemos estimar puntos del rango diferentes y cuando se asuman otros grados de asociación estos deberán respetar la magnitud de asociación. Por lo tanto, se estimará de forma diferentes dependiendo de si la asociación es positiva o es negativa.

Para asociaciones negativas

$$\pi_{12,j=2} = \frac{\pi_{12,j=1}}{E(\pi_{12,j=1}) - Inf(\pi_{12,j=1})} * [E(\pi_{12,j=2}) - Inf(\pi_{12,j=2})]$$

donde $E(\pi_{12,j}) = \pi_{1j}\pi_{2j}$, identifica la probabilidad conjunta del escenario donde no se asocian los eventos.

Para asociaciones positivas

$$\pi_{12,j=2} = E(\pi_{12,j=2}) + \left[(Sup(\pi_{12,j=2}) - E(\pi_{12,j=2})) * \frac{\pi_{12,j=1} - E(\pi_{12,j=1})}{Sup(\pi_{12,j=1}) - E(\pi_{12,j=1})} \right]$$

Esta aproximación asigna de forma proporcional valores para el grado de asociación en el grupo intervención, teniendo en cuenta que: a) el rango de posibles valores es distinto y b) que el escenario de no asociación no divide en rango de forma simétrica.

10.2. Anexo 2: El efecto combinado del CE está acotado

Aritméticamente se deduce que el efecto del CE está acotado. El efecto combinado del CE se puede estimar como la combinación de los efectos de los componentes específicos del CE, corregido por una función del grado de asociación. Una vez determinadas las probabilidades y los efectos de cada uno de los eventos combinados (suponemos un CE binario de dos eventos), teniendo en cuenta los posibles grados de asociación, el efecto del CE queda determinado.

Definimos dos eventos binarios X_{ijk} , tal que j toma valores 1 y 2 para el grupo control e intervención respectivamente y k también toma valores 1 y 2 para el ER y el EA respectivamente. Definimos el CE $X_{i,j}^*$ como la combinación de los dos eventos binarios. Dadas sus probabilidades marginales ($\pi_{j=1,k=1}$ y $\pi_{1,k=2}$) y sus efecto estimados usando la diferencia de probabilidades ($\theta_1^{RA} = \pi_{j=2,k=1} - \pi_{1,1}$ y $\theta_2^{RA} = \pi_{2,2} - \pi_{1,2}$). Definimos la probabilidad del CE en cada grupo utilizando [2] de la siguiente forma:

$$Prob(X_{i,j=1}^* = 1) = \pi_{j=1}^* = \pi_{1,1} + \pi_{1,2} - \pi_{1,12}$$

$$Prob(X_{i,j=2}^* = 1) = \pi_{j=2}^* = \pi_{2,1} + \pi_{2,2} - \pi_{2,12}$$

y su efecto como [3.1]

$$\theta^{*,RA} = \pi_{j=2}^* - \pi_{j=1}^*$$

$$\theta^{*,RA} = \pi_{2,1} + \pi_{2,2} - \pi_{2,12} - \pi_{1,1} - \pi_{1,2} + \pi_{1,12}$$

$$\theta^{*,RA} = \theta_1^{RA} + \theta_2^{RA} - \theta_{12}^{RA}$$

Donde defino $\theta_{12}^{RA} = \pi_{2,12} - \pi_{1,12}$, como la diferencia de probabilidades conjuntas entre el grupo intervención y el grupo control. El efecto de la intervención sobre el CE puede escribirse como la suma del efecto sobre cada uno de los eventos individuales, corrigiendo o sustrayendo el efecto sobre las probabilidades conjuntas.

El efecto sobre cada uno de los eventos combinados es fijo, pero el efecto sobre las probabilidades conjuntas no está determinado, pero sí que está acotado porque aplicando los límites de Fréchet [6 y 9] sabemos que tienen una cota inferior y otra superior:

$$Inf(\pi_{j=1,12}) = \max\{0; \pi_{j=1,1} + \pi_{j=1,2} - 1\} \leq \pi_{j=1,12} \leq \min\{\pi_{j=1,1}, \pi_{j=1,2}\} = Sup(\pi_{j=1,12})$$

$$Inf(\pi_{j=2,12}) = \max\{0; \pi_{j=2,1} + \pi_{j=2,2} - 1\} \leq \pi_{j=2,12} \leq \min\{\pi_{j=2,1}, \pi_{j=2,2}\} = Sup(\pi_{j=2,12})$$

La probabilidad conjunta tanto en el grupo control como en el grupo intervención puede tomar cualquier valor entre la cota inferior y superior. Por lo tanto, la diferencia de proporciones conjuntas debe estar también acotada de forma algebraica. A continuación, definimos las 4 situaciones diferentes según la combinación de valores extremos de las probabilidades combinadas.

Escenario 1: Tanto la probabilidad conjunta del grupo control como la probabilidad conjunta del grupo intervención son las mínimas posibles: $\pi_{j=1,12} = Inf(\pi_{j=1,12})$ y

$\pi_{j=2,12} = \text{Inf}(\pi_{j=2,12})$. En este caso defino $\theta_{12}^{RA} = \pi_{2,12} - \pi_{1,12} = \max\{0; \pi_{j=2,1} + \pi_{j=2,2} - 1\} - \max\{0; \pi_{j=1,1} + \pi_{j=1,2} - 1\}$.

Nota: $\theta_{12}^{RA} = 0$ si $\pi_{j=1,1} + \pi_{j=1,2} \leq 1$.

Escenario 2: La probabilidad conjunta del grupo control es la mínima y la probabilidad

conjunta del grupo intervención es la máxima: $\pi_{j=1,12} = \text{Inf}(\pi_{j=1,12})$ y $\pi_{j=2,12} =$

$\text{Sup}(\pi_{j=2,12})$. En este caso defino $\theta_{12}^{RA} = \pi_{2,12} - \pi_{1,12} = \min\{\pi_{j=2,1}, \pi_{j=2,2}\} -$

$\max\{0; \pi_{j=1,1} + \pi_{j=1,2} - 1\}$.

Nota: $\theta_{12}^{RA} = \min\{\pi_{j=2,1}, \pi_{j=2,2}\}$ si $\pi_{j=1,1} + \pi_{j=1,2} \leq 1$

Escenario 3: La probabilidad conjunta del grupo control es la máxima y la probabilidad

conjunta del grupo intervención es la mínima: $\pi_{j=1,12} = \text{Sup}(\pi_{j=1,12})$ y $\pi_{j=2,12} =$

$\text{Inf}(\pi_{j=2,12})$. En este caso defino $\theta_{12}^{RA} = \pi_{2,12} - \pi_{1,12} = \max\{0; \pi_{j=2,1} + \pi_{j=2,2} - 1\} -$

$\min\{\pi_{j=1,1}, \pi_{j=1,2}\}$.

Nota: $\theta_{12}^{RA} = -\min\{\pi_{j=1,1}, \pi_{j=1,2}\}$ si $\pi_{j=2,1} + \pi_{j=2,2} \leq 1$

Escenario4: La probabilidad conjunta tanto del grupo control como del grupo intervención

son las máximas posibles: $\pi_{j=1,12} = \text{Sup}(\pi_{j=1,12})$ y $\pi_{j=2,12} = \text{Sup}(\pi_{j=2,12})$. En este caso

defino $\theta_{12}^{RA} = \pi_{2,12} - \pi_{1,12} = \min\{\pi_{j=2,1}, \pi_{j=2,2}\} - \min\{\pi_{j=1,1}, \pi_{j=1,2}\}$.

Podemos concluir que el efecto combinado $\theta^{*,RA}$ de los efectos, estará acotado ya que se define como la suma de efectos individuales restándole el efecto de la probabilidad conjunta y ésta en los 4 casos extremos está acotada por las distribuciones marginales.

10.3. Anexo 3: Listado de coeficientes de similitud

Cuando definimos un CE a partir de dos eventos binarios (X_{ij1} para el ER y X_{ij2} para el EA) una tabla de contingencia 2x2 puede resumir toda la información.

		<i>Evento Aditivo - X_{ij2}</i>		
<i>Evento Relevante - X_{ij1}</i>	<i>Sí {1}</i>	<i>No {0}</i>	Total	
<i>Sí {1}</i>	<i>a</i>	<i>b</i>	<i>a + b = p₁</i>	
<i>No {0}</i>	<i>c</i>	<i>d</i>	<i>c + d = q₁</i>	
Total	<i>a + c = p₂</i>	<i>b + d = q₂</i>	<i>n</i>	

a: Número o proporción de pacientes con ambos eventos (ER y EA). **b:** Número o proporción de pacientes con el ER y no el EA.

c: Número o proporción de pacientes con el EA y no el ER. **d:** Número o proporción de pacientes sin el ER ni el EA.

p₁ = a + b: Número o proporción de paciente con el ER. **q₁ = c + d:** Número o proporción de pacientes en el ER.

p₂ = a + c: Número o proporción de pacientes con el EA. **q₂ = b + d:** Número o proporción de pacientes sin el EA.

Con los coeficientes definidos en [2] ($\pi_{j1}, \pi_{j2}, \pi_{j12}$) la función bivariada queda completamente definida (82). En la Tabla 9 de este anexo el parámetro p_1 corresponde a π_{j1} ,

p_2 a π_{j2} y a por π_{j12} . Conociendo las probabilidades marginales de los dos eventos y su probabilidad conjunta la tabla anterior puede reescribirse como:

		Evento Aditivo - X_{ij2}		
Evento Relevante - X_{ij1}	<i>Sí {1}</i>	<i>No {0}</i>	Total	
<i>Sí {1}</i>	π_{j12}	$\pi_{j1} - \pi_{j12}$	π_{j1}	
<i>No {0}</i>	$\pi_{j2} - \pi_{j12}$	$1 - \pi_j^*$	$1 - \pi_{j1}$	
Total	π_{j2}	$1 - \pi_{j2}$	1	

π_{j1} = Probabilidad marginal del ER, π_{j2} = Probabilidad marginal del EA. π_{12} = Probabilidad conjunta o probabilidad de la intersección de ambos eventos. π_j^* = Probabilidad del CE $\pi_j^* = P(X_{ij}^*) = \pi_{j1} + \pi_{j2} - \pi_{j12}$.

A partir de este momento obviaremos el sufijo referente al grupo de estudio, pues el grado de asociación se define igualmente en ambos grupos.

Por lo tanto por ejemplo la tabla de contingencia correspondiente al estudio CARE(98) es:

Estudio CARE		Evento Aditivo - IAM no fatal		
Evento Relevante -IAM fatal	<i>Sí {1}</i>	<i>No {0}</i>	Total	
<i>Sí {1}</i>	6 [0,14%]	56 [1,35%]	62 [1,49%]	
<i>No {0}</i>	302 [7,26%]	3.795 [91,25%]	4.097 [98,51%]	
Total	308 [7,41%]	3.851 [92,59%]	4.159 [100%]	

Este anexo presenta un resumen de la Tesis del Dr. Warrens M.J. (85–87,97): “Similarity Coefficients for Binary Data”. Está ordenado cronológicamente por el año de publicación. Se ha listado el nombre del índice o autor que lo publicó, su definición a partir de la fórmula, una referencia bibliográfica y finalmente su cuantificación en el estudio CARE. Aunque esta lista es muy extensa no se puede considerar exhaustiva pues el número de CS va creciendo con el tiempo. Otros CS pueden consultarse en (99).

Tabla 9 – Listado de los Coeficientes de Similitud estudiados.

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
1	Peir1, Peri2	Peirce (1884) (126)	$S_{Peir1} = \frac{ad - bc}{p_1(1 - p_1)} = \frac{\pi_{12} - \pi_1\pi_2}{\pi_1(1 - \pi_1)}$ $S_{Peir2} = \frac{ad - bc}{p_2(1 - p_2)} = \frac{\pi_{12} - \pi_1\pi_2}{\pi_2(1 - \pi_2)}$	$S_{Peir1} = 0,0231$ $S_{Peir2} = 0,0049$
2	Doo	Doolittle (1885) (127), Pearson (1926) (80)	$S_{Doo} = \frac{(ad - bc)^2}{p_1q_1p_2q_2} = \frac{(\pi_{12} - \pi_1\pi_2)^2}{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}$	$S_{Doo} = 0,0001$
3	Yule1	Yule (1900)(128), Montgomery and Crittenden (1977)(129)	$S_{Yule1} = \frac{ad - bc}{ad + bc} = \frac{\pi_{12} - \pi_1\pi_2}{\pi_{12} + \pi_1\pi_2 + 2\pi_{12}(\pi_{12} - \pi_1 - \pi_2)}$	$S_{Yule1} = 0,1476$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
4	Chi-square	Pearson (1905), quoted by Yule and Kendall (1950) (130)	$x^2 = \frac{n(ad - bc)^2}{p_1 q_1 p_2 q_2} = \frac{n(\pi_{12} - \pi_1 \pi_2)^2}{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}$	$x^2 = 0,4737$
5	Forbes	Forbes (1907) (131)	$S_{Forbes} = \frac{na}{p_1 p_2} = \frac{\pi_{12}}{\pi_1 \pi_2}$	$S_{Forbes} = 1,3068$
6	Jl	Jaccard (1912) (103)	$S_{Jl} = \frac{a}{a + b + c} = \frac{\pi_{12}}{\pi^*}$	$S_{Jl} = 0,0165$
7	Phi	Yule (1912) (100), Pearson & Heron (1913) (81)	$S_\rho = \frac{ad - bc}{\sqrt{p_1 q_1 p_2 q_2}} = \frac{\pi_{12} - \pi_1 \pi_2}{\sqrt{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}}$	$S_\rho = 0,0107$
8	Yule2	Yule (1912) (100)	$S_{Yule2} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} = \frac{\sqrt{\pi_{12}(1 - \pi^*)} - \sqrt{(\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}}{\sqrt{\pi_{12}(1 - \pi^*)} + \sqrt{(\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}}$	$S_{Yule2} = 0,0742$
9	SD	Gleason (1920) (132), Dice (1945) (111), Sørensen (1948) (110,112), Nei and Li (1979) (133)	$S_{SD} = \frac{2a}{p_1 + p_2} = \frac{2\pi_{12}}{\pi_1 + \pi_2}$	$S_{Gleas} = 0,0324$
10	Mich	Michael (1920) (134)	$S_{Mich} = \frac{4(ad - bc)}{(a + d)^2 + (b + c)^2} = \frac{4(\pi_{12} - \pi_1 \pi_2)}{(\pi_{12} + (1 - \pi^*))^2 + ((\pi_1 - \pi_{12})(\pi_2 - \pi_{12}))^2}$	$S_{Mich} = 0,0016$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
11	KU1, KU2	Kulczynski (1927) (113), Driver & Kroeber (1932) (135)	$S_{KU1} = \frac{1}{2} \left(\frac{a}{p_1} + \frac{a}{p_2} \right) = \frac{1}{2} \left(\frac{\pi_{12}}{\pi_1} + \frac{\pi_{12}}{\pi_2} \right)$ $S_{KU2} = \frac{a}{b+c} = \frac{\pi_{12}}{\pi^* - \pi_{12}}$	$S_{KU1} = 0,0581$ $S_{KU2} = 0,0168$
12	BB	Braun-Blanquet (1932) (136)	$S_{BB} = \frac{a}{\max(p_1, p_2)} = \frac{\pi_{12}}{\max(\pi_1, \pi_2)}$	$S_{BB} = 0,0195$
13	OC	Driver & Kroeber (1932) (135), Ochiai (1957) (107), Fowlkes and Mallows (1983) (137)	$S_{OC} = \frac{a}{\sqrt{p_1 p_2}} = \frac{\pi_1}{\sqrt{\pi_1 \pi_2}}$	$S_{OC} = 0,0434$
14	KR	Kuder & Richardson (1937) (138), Cronbach (1951) (139)	$S_{KR} = \frac{4(ad - bc)}{p_1(1 - p_1) + p_2(1 - p_2) + 2(ad - bc)} = \frac{4(\pi_{12} - \pi_1 \pi_2)}{\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2) + 2(\pi_{12} - \pi_1 \pi_2)}$	$S_{KR} = 0,0161$
15	JP	Russel & Rao (1940) (101) - Intersection Probability	$S_{JP} = \frac{a}{a + b + c + d} = \pi_{12}$	$S_{RR} = 0,0014$
16	JI*	Simpson (1943) (105,106) - Relative	$S_{JI^*} = \frac{a}{\min(p_1, p_2)} = \frac{\pi_{12}}{\min(\pi_1, \pi_2)}$	$S_{JI^*} = 0,0968$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
		Jaccard Index or Overlap Coefficient		
17	Dice1, Dice2	Dice (1945) (111), Wallace (1983) (140), Post & Snijders (1993) (141)	$S_{Dice1} = \frac{a}{p_1} = \frac{\pi_{12}}{\pi_1}$ $S_{Dice2} = \frac{a}{p_2} = \frac{\pi_{12}}{p_2}$	$S_{Dice1} = 0,0195$ $S_{Dice2} = 0,0968$
18	Loe	Loevinger (1947,1948) (142,143), Mokken (1971) (144), Sijtsma & Molenaar (2002) (145)	$S_{Loe} = \frac{(ad - bc)}{\min(p_1q_2, p_2q_1)} = \frac{(\pi_{12} - \pi_1\pi_2)}{\min(\pi_1(1 - \pi_2); \pi_2(1 - \pi_1))}$	$S_{Loe} = 0,0245$
19	Cole1, Cole2	Cole (1949) (146)	$S_{Cole1} = \frac{(ad - bc)}{p_1q_2} = \frac{(\pi_{12} - \pi_1\pi_2)}{\pi_1(1 - \pi_2)}$ $S_{Cole2} = \frac{(ad - bc)}{p_2q_1} = \frac{(\pi_{12} - \pi_1\pi_2)}{\pi_2(1 - \pi_1)}$	$S_{Cole1} = 0,0049$ $S_{Cole2} = 0,3068$
20	GK	Goodman & Kruskal (1954) (147)	$S_{GK} = \frac{2\min(a, d) - b - c}{2\min(a, d) + b + c} = \frac{2\min(\pi_{12}, 1 - \pi^*) - \pi_1 + \pi_2}{2\min(\pi_{12}, 1 - \pi^*, d) + \pi_1 + \pi_2 + 2\pi_{12}}$	$S_{GK} = 0,9844$
21	Scott	Scott (1955) (148)	$S_{Scott} = \frac{4ad - (b + c)^2}{(p_1 + p_2)(q_1 + q_2)} = \frac{4\pi_{12} - (\pi_1 + \pi_2)^2}{(\pi_1 + \pi_2)((1 - \pi_1) + (1 - \pi_2))}$	$S_{Scott} = -0,0126$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
22	SP	Sokal & Michener (1958) (109), Rand (1971) (149), Brennan & Light (1974) (150) – Simple Pair's	$S_{SP} = \frac{a + d}{a + b + c + d} = \pi_{12} + 1 - \pi^*$	$S_{SP} = 0,914$
23	Sorg	Sorgenfrei (1958) (151), Cheethan & Hazel (1969) (152)	$S_{Sorg} = \frac{a^2}{p_1 p_2} = \frac{\pi_{12}^2}{\pi_1 \pi_2}$	$S_{Sorg} = 0,0019$
24	Cohen	Cohen (1960) (153)	$S_{Cohen} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} = \frac{2(\pi_{12} - \pi_1 \pi_2)}{\pi_1(1 - \pi_2) + \pi_2(1 - \pi_1)}$	$S_{Cohen} = 0,0078$
25	RT	Rogers & Tanimoto (1960)(154), Farkas (1978)(155)	$S_{RT} = \frac{a + d}{a + 2(b + c) + d} = \frac{1 - \pi_1 - \pi_2 + 2\pi_{12}}{1 + \pi_1 + \pi_2 - 2\pi_{12}}$	$S_{RT} = 0,8415$
26	Sti	Stiles (1961) (156)	$S_{Sti} = \log_{10} \frac{n \left(ad - bc - \frac{n}{2} \right)^2}{p_1 q_1 p_2 q_2} = \log_{10} \frac{\left((\pi_{12} - \pi_1 \pi_2) - \frac{1}{2} \right)^2}{\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)}$	$S_{Sti} = -0,7054$
27	Ham	Hamann (1961) (157), Holley & Guilford (1964)	$S_{Sti} = \frac{a - b - c + d}{a + b + c + d} = 1 - 2(\pi_1 + \pi_2 - 2\pi_{12})$	$S_{Sti} = 0,8278$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
		(158), Hubert (1977) (159)		
28	Mount	Mountford (1962) (160)	$S_{Mount} = \frac{2a}{a(b+c) + 2bc} = \frac{2\pi_{12}}{\pi_{12}(\pi_1 + \pi_2 - 2\pi_{12}) + 2\pi_1 + 2\pi_2 - 4\pi_{12}}$	$S_{Mount} = 0,0003$
29	FM	Fager & McGowan (1963) (161)	$S_{FM} = \frac{a}{\sqrt{p_1 p_2}} - \frac{1}{2\sqrt{\max(p_1, p_2)}} = \frac{\pi_{12}}{\sqrt{\pi_1 \pi_2}} - \frac{1}{2\sqrt{\max(\pi_1, \pi_2)}}$	$S_{FM} = 0,0149$
30	SS1, SS2, SS3, SS4	Sokal and Sneath 1963 (162)	$S_{SS1} = \frac{a}{a + 2(b+c)} = \frac{\pi_{12}}{\pi_{12} + 2(\pi_1 + \pi_2 - 2\pi_{12})}$ $S_{SS2} = \frac{2(a+d)}{2a + b + c + 2d} = \frac{2 - 2\pi_1 - 2\pi_2 + 4\pi_{12}}{2 - \pi_1 - \pi_2 + 2\pi_{12}}$ $S_{SS3} = \frac{1}{4} \left(\frac{a}{p_1} + \frac{a}{p_2} + \frac{d}{q_1} + \frac{d}{q_2} \right) = \frac{1}{4} \left(\frac{\pi_{12}}{\pi_1} + \frac{\pi_{12}}{\pi_2} + \frac{(1-\pi^*)}{(1-\pi_1)} + \frac{(1-\pi^*)}{(1-\pi_2)} \right)$ $S_{SS4} = \frac{ad}{\sqrt{p_1 p_2 q_1 q_2}} = \frac{\pi_{12}(1-\pi^*)}{\sqrt{\pi_1(1-\pi_1)\pi_2(1-\pi_2)}}$	$S_{SS1} = 0,0083$ $S_{SS2} = 0,9550$ $S_{SS3} = 0,5070$ $S_{SS4} = 0,0415$
31	McC	McConnaughey (1964) (163)	$S_{McC} = \frac{a^2 - bc}{p_1 p_2} = \frac{\pi_1 \pi_{12} + \pi_2 \pi_{12} - \pi_1 \pi_2}{\pi_1 \pi_2}$	$S_{McC} = -0,8837$
32	RG	Rogot & Goldberg (1966) (164)	$S_{RG} = \frac{a}{p_1 + p_2} + \frac{d}{q_1 + q_2} = \frac{\pi_{12}}{\pi_1 + \pi_2} + \frac{1 - \pi^*}{(1 - \pi_1) + (1 - \pi_2)}$	$S_{RG} = 0,4937$
33	John	Johnson (1967) (165)	$S_{John} = \frac{a}{p_1} + \frac{a}{p_2} = \frac{\pi_{12}}{\pi_1} + \frac{\pi_{12}}{\pi_2}$	$S_{John} = 0,1163$

Id	Indicador	Autor / Año	Coeficiente de Similitud	Valor en el estudio CARE
34	HD	Hawkins & Dotson (1968) (166)	$S_{HD} = \frac{1}{2} \left(\frac{a}{a+b+c} + \frac{d}{b+c+d} \right) = \frac{1}{2} \left(\frac{\pi_{12}}{\pi^*} + \frac{1-\pi^*}{1-\pi_{12}} \right)$	$S_{HD} = 0,4651$
35	Fleiss	Fleiss (1975) (167)	$S_{Fleiss} = \frac{(ad-bc)(p_1q_2+p_2q_1)}{p_1p_2q_1q_2} = \frac{(\pi_{12}-\pi_1\pi_2)(\pi_1(1-\pi_1)+\pi_2(1-\pi_2))}{\pi_1(1-\pi_1)\pi_2(1-\pi_2)}$	$S_{Fleiss} = 0,0146$
36	Clem	Clement (1976) (168)	$S_{Cle} = \frac{aq_1}{p_1} + \frac{dp_1}{q_1} = \frac{\pi_{12}(1-\pi_1)}{\pi_1} + \frac{(1-\pi^*)\pi_1}{(1-\pi_1)}$	$S_{Cle} = 378,54$
37	BUB1, BUB2	Baroni-Urabani & Buser (1976) (169)	$S_{BUB1} = \frac{a + \sqrt{ad}}{a+b+c + \sqrt{ad}} = \frac{\pi_{12} + \sqrt{\pi_{12}(1-\pi^*)}}{\pi^* + \sqrt{\pi_{12}(1-\pi^*)}}$ $S_{BUB2} = \frac{a - b - c + \sqrt{ad}}{a+b+c + \sqrt{ad}} = \frac{2\pi_{12} - \pi^* + \sqrt{\pi_{12}(1-\pi^*)}}{\pi^* + \sqrt{\pi_{12}(1-\pi^*)}}$	$S_{BUB1} = 0,305$ $S_{BUB2} = -0,391$
38	KF1, KF2	Kent & Foster (1977) (170)	$S_{KF1} = \frac{-bc}{bp_1 + cp_2 + bc} = \frac{-(\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}{(\pi_1 - \pi_{12})\pi_1 + (\pi_2 - \pi_{12})\pi_2 + (\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}$ $S_{BUB2} = \frac{-bc}{bq_1 + cq_2 + bc} = \frac{-(\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}{(\pi_1 - \pi_{12})(1-\pi_1) + (\pi_2 - \pi_{12})(1-\pi_2) + (\pi_1 - \pi_{12})(\pi_2 - \pi_{12})}$	$S_{KF1} = -0,1491$ $S_{BUB2} = -0,0120$
39	HL	Harris & Lahey (1978) (171)	$S_{HL} = \frac{a(q_1 + q_2)}{2(a+b+c)} + \frac{d(p_1 + p_2)}{2(b+c+d)} = \frac{\pi_{12}(2 - \pi_1 - \pi_2)}{2(\pi^*)} + \frac{(1-\pi^*)(\pi_1 + \pi_2)}{2(1-\pi_{12})}$	$S_{HL} = 234,56$
40	Digby	Digby (1983) (172)	$S_{Digby} = \frac{(ad)^{3/4} - (bc)^{3/4}}{(ad)^{3/4} + (bc)^{3/4}} = \frac{(\pi_{12}(1-\pi^*))^{3/4} - ((\pi_1 - \pi_{12})(\pi_2 - \pi_{12}))^{3/4}}{(\pi_{12}(1-\pi^*))^{3/4} + ((\pi_1 - \pi_{12})(\pi_2 - \pi_{12}))^{3/4}}$	$S_{Digby} = 0,111$

donde $\pi^* = P(X^*) = \pi_1 + \pi_2 - \pi_{12}$

10.4. Anexo 4: Rangos y coeficientes estandarizado

Tabla 10 - Rangos de los 8 CS según $\pi_{j1} + \pi_{j2} < 1$ o $\pi_{j1} + \pi_{j2} > 1$

<i>Coficiente de Similitud</i>	<i>Rango</i>	
	$\pi_{j1} + \pi_{j2} < 1$	$\pi_{j1} + \pi_{j2} > 1$
Probabilidad Conjunta (JP_j)	$\min\{\pi_{j1}, \pi_{j2}\}$	$\min\{\pi_{j1}, \pi_{j2}\} - (\pi_{j1} + \pi_{j2} - 1)$
Índice de Jaccard (JI_j)	$\min\left(\frac{\pi_{j1}}{\pi_{j2}}, \frac{\pi_{j2}}{\pi_{j1}}\right)$	$\min\left(\frac{\pi_{j1}}{\pi_{j2}}, \frac{\pi_{j2}}{\pi_{j1}}\right) - \pi_{j1} - \pi_{j2} + 1$
Índice Relativo de Jaccard (JI_j^*)	1	$1 - \frac{\pi_{j1} + \pi_{j2} - 1}{\min(\pi_{j1}, \pi_{j2})}$
Coficiente de correlación (ρ_j)	$\min\left(\sqrt{\frac{\pi_{j1}(1-\pi_{j2})}{\pi_{j2}(1-\pi_{j1})}}, \sqrt{\frac{\pi_{j2}(1-\pi_{j1})}{\pi_{j1}(1-\pi_{j2})}}\right) + \sqrt{\frac{\pi_{j1}\pi_{j2}}{(1-\pi_{j1})(1-\pi_{j2})}}$	$\min\left(\sqrt{\frac{\pi_{j1}(1-\pi_{j2})}{\pi_{j2}(1-\pi_{j1})}}, \sqrt{\frac{\pi_{j2}(1-\pi_{j1})}{\pi_{j1}(1-\pi_{j2})}}\right) + \sqrt{\frac{(1-\pi_{j1})(1-\pi_{j2})}{\pi_{j1}\pi_{j2}}}$
Índice de Ochiai (ρ_j)	$\min\left(\sqrt{\frac{\pi_{j1}}{\pi_{j2}}}, \sqrt{\frac{\pi_{j2}}{\pi_{j1}}}\right)$	$\min\left(\sqrt{\frac{\pi_{j1}}{\pi_{j2}}}, \sqrt{\frac{\pi_{j2}}{\pi_{j1}}}\right) - \frac{\pi_{j1} + \pi_{j2} - 1}{\sqrt{\pi_{j1}\pi_{j2}}}$
Coficiente de Pares Simples (SP_j)	$\min(1 + \pi_{j1} - \pi_{j2}, 1 - \pi_{j1} + \pi_{j2}) - 1 + \pi_{j1} + \pi_{j2}$	$\min(1 + \pi_{j1} - \pi_{j2}, 1 - \pi_{j1} + \pi_{j2}) - \pi_{j1} - \pi_{j2} + 1$
Coficiente Sørensen-Dice (SD_j)	$\min\left(\frac{2\pi_{j1}}{\pi_{j1} + \pi_{j2}}, \frac{2\pi_{j2}}{\pi_{j1} + \pi_{j2}}\right)$	$\min\left(\frac{2\pi_{j1}}{\pi_{j1} + \pi_{j2}}, \frac{2\pi_{j2}}{\pi_{j1} + \pi_{j2}}\right) - 2 + \frac{2}{(\pi_{j1} + \pi_{j2})}$
Coficiente Kulczynski (KU_j)	$\max\left(\frac{\pi_{j1}}{\pi_{j2} - \pi_{j1}}, \frac{\pi_{j2}}{\pi_{j1} - \pi_{j2}}\right)$	$\max\left(\frac{\pi_{j1}}{\pi_{j2} - \pi_{j1}}, \frac{\pi_{j2}}{\pi_{j1} - \pi_{j2}}\right) - \frac{\pi_{j1} + \pi_{j2} - 1}{2 - \pi_{j1} - \pi_{j2}}$

Tabla 11 - Coeficientes estandarizados de los 8 CS según $\pi_{j1} + \pi_{j2} < 1$ o $\pi_{j1} + \pi_{j2} > 1$

Coeficiente de Similitud	Coeficiente estandarizado	
	$\pi_{j1} + \pi_{j2} < 1$	$\pi_{j1} + \pi_{j2} > 1$
Probabilidad Conjunta (IP_j)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Índice de Jaccard (II_j)	$\frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - \pi_{j12}} \min\left(\frac{\pi_{j2}}{\pi_{j1}}, \frac{\pi_{j1}}{\pi_{j2}}\right)^{-1}$	$\frac{\frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - \pi_{j12}} - \pi_{j1} - \pi_{j2} + 1}{\min\left(\frac{\pi_{j2}}{\pi_{j1}}, \frac{\pi_{j1}}{\pi_{j2}}\right) - \pi_{j1} - \pi_{j2} + 1}$
Índice Relativo de Jaccard (II_j^*)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Coeficiente de correlación (ρ_j)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Índice de Ochiai (ρ_j)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Coeficiente de Pares Simples (SP_j)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Coeficiente Sørensen-Dice (SD_j)	$\frac{\pi_{j12}}{\min(\pi_{j1}, \pi_{j2})}$	$\frac{\pi_{j12} - \pi_{j1} - \pi_{j2} + 1}{1 - \min(\pi_{j1}, \pi_{j2})}$
Coeficiente Kulczynski (KU_j)	$\frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - 2\pi_{j12}} \max\left(\frac{\pi_{j2}}{\pi_{j2} - \pi_{j1}}, \frac{\pi_{j1}}{\pi_{j2} - \pi_{j1}}\right)^{-1}$	$\frac{\frac{\pi_{j12}}{\pi_{j1} + \pi_{j2} - 2\pi_{j12}} - \frac{\pi_{j1} + \pi_{j2} - 1}{2 - \pi_{j1} - \pi_{j2}}}{\min\left(\frac{\pi_{j2}}{\pi_{j1}}, \frac{\pi_{j1}}{\pi_{j2}}\right) - \frac{\pi_{j1} + \pi_{j2} - 1}{2 - \pi_{j1} - \pi_{j2}}}$

10.5. Anexo 5: Resultados de la simulación

- Tabla 12 presenta los resultados cuando la probabilidad del ER es del 0,10.
- Tabla 13 presenta los resultados cuando la probabilidad del ER es del 0,20.
- Tabla 14 presenta los resultados cuando la probabilidad del ER es del 0,03

Tabla 12 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,10, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.

	Evento Relevante			Evento Adicional		Grado de Asociación									
						A=Muy Bajo		B=Bajo		C=Moderado		D=Alto		E=Muy Alto	
	P (ER)	RR (ER)	n (ER)	P (EA)	RR (EA)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)
S1.1	0,10	0,90	13.495	0,10	1,00	24.640 (0,55)	0,95	86.849 (0,16)	0,97	340.738 (0,04)	0,98	2.177.280 (0,01)	0,99	∞ (0,00)	1,00
S1.2	0,10	0,90	13.495	0,20	1,00	32.646 (0,41)	0,97	84.869 (0,16)	0,98	267.288 (0,05)	0,99	1.445.762 (0,01)	0,99	∞ (0,00)	1,00
S1.3	0,10	0,70	1.355	0,10	1,00	2.629 (0,52)	0,85	8.637 (0,16)	0,90	32.277 (0,04)	0,94	199.035 (0,01)	0,98	∞ (0,00)	1,00
S1.4	0,10	0,70	1.355	0,20	1,00	3.553 (0,38)	0,90	8.872 (0,15)	0,93	27.047 (0,05)	0,96	142.423 (0,01)	0,98	∞ (0,00)	1,00
S2.1	0,10	0,90	13.495	0,10	0,90	6.039 (2,23)	0,90	7.903 (1,71)	0,90	9.767 (1,38)	0,90	11.631 (1,16)	0,90	13.495 (1,00)	0,90
S2.2	0,10	0,90	13.495	0,20	0,90	3.553 (3,80)	0,90	4.174 (3,23)	0,90	4.796 (2,81)	0,90	5.417 (2,49)	0,90	6.039 (2,23)	0,90
S2.3	0,10	0,90	13.495	0,10	0,80	2.629 (5,13)	0,85	4.565 (2,96)	0,87	7.042 (1,92)	0,88	10.029 (1,35)	0,89	13.495 (1,00)	0,90
S2.4	0,10	0,90	13.495	0,20	0,80	1.251 (10,8)	0,83	1.279 (10,6)	0,82	1.326 (10,2)	0,81	1.383 (9,76)	0,81	1.447 (9,33)	0,80
S2.5	0,10	0,90	13.495	0,10	0,70	1.447 (9,33)	0,80	2.900 (4,65)	0,84	5.194 (2,60)	0,86	8.611 (1,57)	0,88	13.495 (1,00)	0,90
S2.6	0,10	0,90	13.495	0,20	0,70	623 (21,7)	0,77	595 (22,7)	0,74	591 (22,8)	0,73	599 (22,5)	0,71	614 (22,0)	0,70
S3.1	0,10	0,80	3.213	0,10	0,90	2.629 (1,22)	0,85	4.565 (0,70)	0,87	7.042 (0,46)	0,88	10.029 (0,32)	0,89	13.495 (0,24)	0,90
S3.2	0,10	0,80	3.213	0,20	0,90	1.977 (1,63)	0,87	2.759 (1,16)	0,88	3.695 (0,87)	0,89	4.787 (0,67)	0,89	6.039 (0,53)	0,90
S3.3	0,10	0,80	3.213	0,10	0,80	1.447 (2,22)	0,80	1.888 (1,70)	0,80	2.330 (1,38)	0,80	2.771 (1,16)	0,80	3.213 (1,00)	0,80
S3.4	0,10	0,80	3.213	0,20	0,80	858 (3,74)	0,80	1.005 (3,20)	0,80	1.153 (2,79)	0,80	1.300 (2,47)	0,80	1.447 (2,22)	0,80
S3.5	0,10	0,80	3.213	0,10	0,70	905 (3,55)	0,75	1.398 (2,30)	0,77	1.955 (1,64)	0,78	2.564 (1,25)	0,79	3.213 (1,00)	0,80
S3.6	0,10	0,80	3.213	0,20	0,70	471 (6,82)	0,73	504 (6,38)	0,72	539 (5,96)	0,71	576 (5,58)	0,71	614 (5,23)	0,70

	Evento Relevante			Evento Adicional		Grado de Asociación									
						A=Muy Bajo		B=Bajo		C=Moderado		D=Alto		E=Muy Alto	
	P (ER)	RR (ER)	n (ER)	P (EA)	RR (EA)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)
S4.1	0,10	0,70	1.355	0,10	0,90	1.447 (0,94)	0,80	2.900 (0,47)	0,84	5.194 (0,26)	0,86	8.611 (0,16)	0,88	13.495 (0,10)	0,90
S4.2	0,10	0,70	1.355	0,20	0,90	1.251 (1,08)	0,83	1.938 (0,70)	0,85	2.900 (0,47)	0,87	4.228 (0,32)	0,89	6.039 (0,22)	0,90
S4.3	0,10	0,70	1.355	0,10	0,80	905 (1,50)	0,75	1.398 (0,97)	0,77	1.955 (0,69)	0,78	2.564 (0,53)	0,79	3.213 (0,42)	0,80
S4.4	0,10	0,70	1.355	0,20	0,80	623 (2,17)	0,77	805 (1,68)	0,78	1.004 (1,35)	0,79	1.219 (1,11)	0,79	1.447 (0,94)	0,80
S4.5	0,10	0,70	1.355	0,10	0,70	614 (2,21)	0,70	799 (1,70)	0,70	985 (1,38)	0,70	1.170 (1,16)	0,70	1.355 (1,00)	0,70
S4.6	0,10	0,70	1.355	0,20	0,70	367 (3,69)	0,70	429 (3,16)	0,70	491 (2,76)	0,70	552 (2,45)	0,70	614 (2,21)	0,70

Las celdas sombreadas identifican aquellos escenarios donde es preferible el uso de un CE ($SSR > 1$). Escenario 1 (S1): EA sin efecto. Escenario 2 (S2): Efecto pequeño del ER y en el EA un incremento progresivo del efecto. Escenario 3 (S3): Efecto moderado del ER y en el EA un incremento progresivo del efecto. Escenario 4 (S4): Gran efecto del ER y en el EA un incremento progresivo del efecto.

P(): Probabilidad. RR: Riesgo Relativo. n: TMR. La tabla presenta el TMR utilizando un ER y un CE considerando diferentes grados de asociación. Para cada asociación se ha estimado el SSR ($SSR > 1$ si $n^* = n_{ER}$). La utilización del CE es apropiada en aquellos escenarios con un $SSR > 1$ (celdas sombreadas). El efecto final del CE también ha sido estimado como el RR. Se han fijado el error tipo 1 a 0,05 y el error tipo 2 a 0,20 en los cálculos.

Tabla 13 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,20, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.

Evento Relevante			Evento Adicional		Grado de Asociación										
					A=Muy Bajo		B=Bajo		C=Moderado		D=Alto		E=Muy Alto		
P (ER)	RR (ER)	n (ER)	P (EA)	RR (EA)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	
S1.1	0,20	0,90	6.039	0,10	1,00	8.079 (0,75)	0,93	6.906 (0,87)	0,92	6.373 (0,95)	0,91	6.373 (0,95)	0,91	6.039 (1,00)	0,90
S1.2	0,20	0,90	6.039	0,20	1,00	9.335 (0,65)	0,95	35.303 (0,17)	0,97	144.418 (0,04)	0,98	948.805 (0,01)	0,99	∞ (0,00)	1,00
S1.3	0,20	0,70	614	0,10	1,00	858 (0,72)	0,80	709 (0,87)	0,76	648 (0,95)	0,74	622 (0,99)	0,72	614 (1,00)	0,70
S1.4	0,20	0,70	614	0,20	1,00	1.015 (0,60)	0,85	3.542 (0,17)	0,90	13.734 (0,04)	0,94	86.854 (0,01)	0,98	∞ (0,00)	1,00
S2.1	0,20	0,90	6.039	0,10	0,90	3.553 (1,70)	0,90	4.174 (1,45)	0,90	4.796 (1,26)	0,90	5.417 (1,11)	0,90	6.039 (1,00)	0,90
S2.2	0,20	0,90	6.039	0,20	0,90	2.310 (2,61)	0,90	3.242 (1,86)	0,90	4.174 (1,45)	0,90	5.106 (1,18)	0,90	6.039 (1,00)	0,90
S2.3	0,20	0,90	6.039	0,10	0,80	1.977 (3,05)	0,87	2.759 (2,19)	0,88	3.695 (1,63)	0,89	4.787 (1,26)	0,89	6.039 (1,00)	0,90
S2.4	0,20	0,90	6.039	0,20	0,80	1.015 (5,95)	0,85	1.881 (3,21)	0,87	3.015 (2,00)	0,88	4.406 (1,37)	0,89	6.039 (1,00)	0,90
S2.5	0,20	0,90	6.039	0,10	0,70	1.251 (4,83)	0,83	1.938 (3,12)	0,85	2.900 (2,08)	0,87	4.228 (1,43)	0,89	6.039 (1,00)	0,90
S2.6	0,20	0,90	6.039	0,20	0,70	564 (10,7)	0,80	1.199 (5,04)	0,84	2.228 (2,71)	0,86	3.786 (1,60)	0,88	6.039 (1,00)	0,90
S3.1	0,20	0,80	1.447	0,10	0,90	1.251 (1,16)	0,83	1.279 (1,13)	0,82	1.326 (1,09)	0,81	1.383 (1,05)	0,81	1.447 (1,00)	0,80
S3.2	0,20	0,80	1.447	0,20	0,90	1.015 (1,43)	0,85	1.881 (0,77)	0,87	3.015 (0,48)	0,88	4.406 (0,33)	0,89	6.039 (0,24)	0,90
S3.3	0,20	0,80	1.447	0,10	0,80	858 (1,69)	0,80	1.005 (1,44)	0,80	1.153 (1,25)	0,80	1.300 (1,11)	0,80	1.447 (1,00)	0,80
S3.4	0,20	0,80	1.447	0,20	0,80	564 (2,57)	0,80	785 (1,84)	0,80	1.005 (1,44)	0,80	1.226 (1,18)	0,80	1.447 (1,00)	0,80
S3.5	0,20	0,80	1.447	0,10	0,70	623 (2,32)	0,77	805 (1,80)	0,78	1.004 (1,44)	0,79	1.219 (1,19)	0,79	1.447 (1,00)	0,80
S3.6	0,20	0,80	1.447	0,20	0,70	356 (4,06)	0,75	583 (2,48)	0,77	845 (1,71)	0,78	1.135 (1,27)	0,79	1.447 (1,00)	0,80

Evento Relevante			Evento Adicional		Grado de Asociación										
					A=Muy Bajo		B=Bajo		C=Moderado		D=Alto		E=Muy Alto		
P (ER)	RR (ER)	n (ER)	P (EA)	RR (EA)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	
S4.1	0,20	0,70	614	0,10	0,90	623 (0,99)	0,77	595 (1,03)	0,74	591 (1,04)	0,73	599 (1,03)	0,71	614 (1,00)	0,70
S4.2	0,20	0,70	614	0,20	0,90	614 (1,09)	0,80	1.199 (0,51)	0,84	2.228 (0,28)	0,86	3.786 (0,16)	0,88	6.039 (0,10)	0,90
S4.3	0,20	0,70	614	0,10	0,80	471 (1,30)	0,73	504 (1,22)	0,72	539 (1,14)	0,71	576 (1,07)	0,71	614 (1,00)	0,70
S4.4	0,20	0,70	614	0,20	0,80	356 (1,72)	0,75	583 (1,05)	0,77	845 (0,73)	0,78	1.135 (0,54)	0,79	1.447 (0,42)	0,80
S4.5	0,20	0,70	614	0,10	0,70	367 (1,67)	0,70	429 (1,43)	0,70	491 (1,25)	0,70	552 (1,11)	0,70	614 (1,00)	0,70
S4.6	0,20	0,70	614	0,20	0,70	243 (2,53)	0,70	336 (1,83)	0,70	429 (1,43)	0,70	521 (1,18)	0,70	614 (1,00)	0,70

Las celdas sombreadas identifican aquellos escenarios donde es preferible el uso de un CE ($SSR > 1$). Escenario 1 (S1): EA sin efecto. Escenario 2 (S2): Efecto pequeño del ER y en el EA un incremento progresivo del efecto. Escenario 3 (S3): Efecto moderado del ER y en el EA un incremento progresivo del efecto. Escenario 4 (S4): Gran efecto del ER y en el EA un incremento progresivo del efecto.

P(): Probabilidad. RR: Riesgo Relativo. n: TMR. La tabla presenta el TMR utilizando un ER y un CE considerando diferentes grados de asociación. Para cada asociación se ha estimado el SSR ($SSR > 1$ si $n^* = n_{ER}$). La utilización del CE es apropiada en aquellos escenarios con un $SSR > 1$ (celdas sombreadas). El efecto final del CE también ha sido estimado como el RR. Se han fijado el error tipo 1 a 0,05 y el error tipo 2 a 0,20 en los cálculos.

Tabla 14 - TMR en el uso de un CE cuando la probabilidad de observar el ER es 0,03, en diferentes escenarios en función de la probabilidad del EA, el efecto de ambos eventos y el grado de asociación.

<i>Evento Relevante</i>			<i>Evento Adicional</i>		<i>Grado de Asociación</i>										
					<i>A=Muy Bajo</i>		<i>B=Bajo</i>		<i>C=Moderado</i>		<i>D=Alto</i>		<i>E=Muy Alto</i>		
<i>P (ER)</i>	<i>RR (ER)</i>	<i>n (ER)</i>	<i>P (EA)</i>	<i>RR (EA)</i>	<i>n (SSR)</i>	<i>RR (CE)</i>	<i>n (SSR)</i>	<i>RR (CE)</i>	<i>n (SSR)</i>	<i>RR (CE)</i>	<i>n (SSR)</i>	<i>RR (CE)</i>	<i>n (SSR)</i>	<i>RR (CE)</i>	
<i>S1.1</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,10</i>	<i>1,00</i>	195.327 (0,25)	0,98	431.054 (0,11)	0,98	1.186.474 (0,04)	0,99	5.732.226 (0,01)	1,00	∞ (0,00)	1,00
<i>S1.2</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,01</i>	<i>1,00</i>	64.565 (0,75)	0,92	57.285 (0,84)	0,92	52.868 (0,91)	0,91	50.077 (0,96)	0,90	48.292 (1,00)	0,90
<i>S1.3</i>	<i>0,03</i>	<i>0,70</i>	4.815	<i>0,10</i>	<i>1,00</i>	21.268 (0,23)	0,93	45.884 (0,10)	0,95	123.790 (0,04)	0,97	587.486 (0,01)	0,99	∞ (0,00)	1,00
<i>S1.4</i>	<i>0,03</i>	<i>0,70</i>	4.815	<i>0,01</i>	<i>1,00</i>	6.634 (0,73)	0,78	5.772 (0,83)	0,75	5.283 (0,91)	0,73	4.991 (0,96)	0,71	4.815 (1,00)	0,70
<i>S2.1</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,10</i>	<i>0,90</i>	10.054 (4,80)	0,90	10.914 (4,42)	0,90	11.774 (4,10)	0,90	12.635 (3,82)	0,90	13.495 (3,58)	0,90
<i>S2.2</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,01</i>	<i>0,90</i>	35.864 (1,35)	0,90	38.971 (1,24)	0,90	42.078 (1,15)	0,90	45.185 (1,07)	0,90	48.292 (1,00)	0,90
<i>S2.3</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,10</i>	<i>0,80</i>	3.099 (15,58)	0,82	3.097 (15,59)	0,82	3.120 (15,48)	0,81	3.160 (15,28)	0,80	3.213 (15,03)	0,80
<i>S2.4</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,01</i>	<i>0,80</i>	22.662 (2,13)	0,88	27.997 (1,72)	0,88	34.031 (1,42)	0,89	40.789 (1,18)	0,89	48.292 (1,00)	0,90
<i>S2.5</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,10</i>	<i>0,70</i>	1.449 (33,33)	0,75	1.395 (34,62)	0,73	1.367 (35,33)	0,72	1.356 (35,33)	0,71	1.355 (35,64)	0,70
<i>S2.6</i>	<i>0,03</i>	<i>0,90</i>	48.292	<i>0,01</i>	<i>0,70</i>	15.536 (3,11)	0,85	20.933 (2,31)	0,86	27.892 (1,73)	0,88	36.831 (1,31)	0,89	48.292 (1,00)	0,90
<i>S3.1</i>	<i>0,03</i>	<i>0,80</i>	11.454	<i>0,10</i>	<i>0,90</i>	6.567 (1,74)	0,88	8.019 (1,43)	0,88	9.653 (1,19)	0,89	11.476 (1,00)	0,90	13.495 (0,85)	0,90
<i>S3.2</i>	<i>0,03</i>	<i>0,80</i>	11.454	<i>0,01</i>	<i>0,90</i>	11.265 (1,02)	0,82	11.164 (1,03)	0,82	11.186 (1,02)	0,81	11.291 (1,01)	0,80	11.454 (1,00)	0,80
<i>S3.3</i>	<i>0,03</i>	<i>0,80</i>	11.454	<i>0,10</i>	<i>0,80</i>	2.398 (4,78)	0,80	2.602 (4,40)	0,80	2:805 (4,08)	0,80	3.009 (3,81)	0,80	3.213 (3,56)	0,80

Evento Relevante			Evento Adicional			Grado de Asociación									
						A=Muy Bajo		B=Bajo		C=Moderado		D=Alto		E=Muy Alto	
P (ER)	RR (ER)	n (ER)	P (EA)	RR (EA)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	n (SSR)	RR (CE)	
S3.4	0,03	0,80	11.454	0,01	0,80	8.511 (1,35)	0,80	9.247 (1,24)	0,80	9.983 (1,15)	0,80	10.718 (1,07)	0,80	11.454 (1,00)	0,80
S3.5	0,03	0,80	11.454	0,10	0,70	1.203 (9,52)	0,72	1.236 (9,27)	0,72	1.273 (9,00)	0,71	1.313 (8,72)	0,70	1.355 (8,45)	0,70
S3.6	0,03	0,80	11.454	0,01	0,70	6.634 (1,73)	0,78	7.747 (1,48)	0,78	8.924 (1,28)	0,79	10.161 (1,13)	0,79	11.454 (1,00)	0,80
S4.1	0,03	0,70	4.815	0,10	0,90	4.607 (1,05)	0,85	6.104 (0,79)	0,87	8.009 (0,60)	0,88	10.428 (0,46)	0,89	13.495 (0,36)	0,90
S4.2	0,03	0,70	4.815	0,01	0,90	5.301 (0,91)	0,75	5.036 (0,96)	0,73	4.895 (0,98)	0,72	4.830 (1,00)	0,71	4.815 (1,00)	0,70
S4.3	0,03	0,70	4.815	0,10	0,80	1.906 (2,53)	0,78	2.208 (2,18)	0,78	2.527 (1,91)	0,79	2.862 (1,68)	0,80	3.213 (1,50)	0,80
S4.4	0,03	0,70	4.815	0,01	0,80	4.320 (1,11)	0,72	4.416 (1,09)	0,72	4.534 (1,06)	0,71	4.668 (1,03)	0,70	4.815 (1,00)	0,70
S4.5	0,03	0,70	4.815	0,10	0,70	1.013 (4,75)	0,70	1.099 (4,38)	0,70	1.184 (4,07)	0,70	1.270 (3,79)	0,70	1.355 (3,55)	0,70
S4.6	0,03	0,70	4.815	0,01	0,70	3.579 (1,35)	0,70	3.888 (1,24)	0,70	4.197 (1,15)	0,70	4.506 (1,07)	0,70	4.815 (1,00)	0,70

Las celdas sombreadas identifican aquellos escenarios donde es preferible el uso de un CE ($SSR > 1$). Escenario 1 (S1): EA sin efecto. Escenario 2 (S2):

Efecto pequeño del ER y en el EA un incremento progresivo del efecto. Escenario 3 (S3): Efecto moderado del ER y en el EA un incremento progresivo del efecto. Escenario 4 (S4): Gran efecto del ER y en el EA un incremento progresivo del efecto.

P(): Probabilidad. RR: Riesgo Relativo. n: TMR. La tabla presenta el TMR utilizando un ER y un CE considerando diferentes grados de asociación. Para cada asociación se ha estimado el SSR ($SSR > 1$ si $n^* = n_{ER}$). La utilización del CE es apropiada en aquellos escenarios con un $SSR > 1$ (celdas sombreadas). El efecto final del CE también ha sido estimado como el RR. Se han fijado el error tipo 1 a 0,05 y el error tipo 2 a 0,20 en los cálculos.

10.6. Anexo 6: Código Fuente

```
# Bin-CE, version 1.0
# 14th June, 2018
# Marsal Mora, Josep Ramon - Unidad de Epidemiologia del Hospital Universitario Vall d'Hebron.
#           - Univeritat Autònoma de Barcelona
#           - USR LLeida. IDIAP Jordi Gol
# joseprmarsal@yahoo.es
#
#
# Main Function:
# shinyServer(input, output)
#
# Create, Debug and Check the Data Set:
# f_bd()
# reset_ass()
# reset_ds()
# depura()
#
# Numeric Functions:
# ssize()
# First_resul()
# all1()
# engine()
# SR1()
# select_best()
#
# Plot Functions:
# potential_graph()

library(shiny)
library(plotrix)

# ShinyServer Function drives all the analysis

shinyServer(function(input, output) {

  original<-reactive({f_bd(input)})
  alfa<-reactive({input$e1})
  beta<-reactive({input$e2})
  test<-reactive({as.numeric(input$test)})

  output$label1 <- renderText({ input$label1 })
  output$label2 <- renderText({ input$label2 })
  output$label3 <- renderText({ input$label3 })
  output$label4 <- renderText({ input$label4 })
  output$label5 <- renderText({ input$label5 })
  output$label6 <- renderText({ input$label6 })
  output$label7 <- renderText({ input$label7 })
  output$label8 <- renderText({ input$label8 })
  output$label9 <- renderText({ input$label9 })
```

```

output$label10 <- renderText({input$label10 })

output$dataTable1<-renderPrint({original()})
original2<-reactive({
  original_depurada<-depura(original())
  return(original_depurada)
})

bd<-reactive({
  x<-reset_ass(original2())
  x1<-reset_ds(x,as.numeric(input$ass_1))
  re<-99

  if (max(x1$re)==1) {re<-subset(x1,x1$re==1)$n}
  if (re==99) {
    nre<-data.frame()
    for (i in 1:nrow(x)){
      nre[i,1]<-x1[i,1]
      nre[i,2]<-x1[i,2]
      nre[i,3]<-ssize(x1[i,4],x[i,4]*x[i,5],alfa=alfa(),beta=beta(),test=test())
    }
    re<-as.numeric(subset(nre,nre$V3==min(nre$V3))$V1)

    for (i in 1:nrow(x)){
      if (x1[i,1]==re) {x1[i,3]<-1}
    }
    return(x1)
  })

output$dataTable2<-renderPrint({original2()})
results1<-reactive({first_resul(bd(),alfa=alfa(),beta=beta(),test=test())})

```

```

output$dataTable3<-renderPrint({results1()})

output$graph1<-renderPlot({potential_graph(bd(),results1(),beta=beta(),alfa=alfa(),test=test())})

all_results<-reactive({
  bd<-bd()
  alfa<-alfa()
  beta<-beta()
  test<-test()
  all1(base=bd,alfa=alfa,beta=beta,test=test())
})

output$dataTable4<-renderPrint({all_results()})

output$alfa<-renderPrint(paste("Sig: ",alfa(),"%", "Pow: ",100-beta()*100,"%"))

output$graph2<-renderPlot({
  results<-all_results()
  maxi<-as.numeric(max(results[,5]))

  maxc<-ceiling(ceiling(maxi)/(10^trunc(log10(maxi))))*(10^trunc(log10(maxi)))
  plot(results[,1],results[,5],type="l",ylab="SSR (N. Patients)",xlab="Iteration",axes=F,ylim=c(0,maxc))
  text(x=results[,5],labels=results[,5], cex=1)
  axis(2)

  axis(1,at=results[,1],labels=paste(results[,1],results[,2],sep=': '))
  legend("topright",pch=as.character(1:nrow(results)),results[,2],cex=0.7)
  title("Sample Size Requirement of different outcomes combination")
})

output$graph3<-renderPlot({

```

```

results<-all_results()

min<-as.numeric(min(results[,4]))

max<-as.numeric(max(results[,4]))

if (min>1){mylim<-c(0.9,max*1.2)}

if (max<1){mylim<-c(0.8*min,1.1)}

plot(results[,1],results[,4],type="l",ylab="Effect of the Therapy (RR)",xlab="Iteration",axes=F,ylim=mylim)

abline(h=1,col="red",lty=4)

text(x=results[,4],labels=results[,4], cex=1)

axis(2)

axis(1,at=results[,1],labels=paste(results[,1],results[,2],sep=": "))

legend("bottomleft",pch=as.character(1:nrow(results)),results[,2],cex=0.7)

title("Variation of the Effect (RR) when combining different outcomes")

})

output$graph4<-renderPlot({

results<-all_results()

max<-as.numeric(max(results[,3]))

dif<-max-as.numeric(min(results[,3]))

max<-max+dif/2

plot(results[,1],results[,3],type="l",ylab="Rate of event",xlab="Iteration",axes=F,ylim=c(0,max),col="blue",lty=1)

lines(results[,1],as.numeric(results[,3])*as.numeric(results[,4]),col="red",lty=3)

text(x=results[,3],labels=round(as.numeric(results[,3])*100,2), cex=1,col="blue")

text(x=as.numeric(results[,3])*as.numeric(results[,4]),labels=round(as.numeric(results[,3])*as.numeric(results[,4])*100,2), cex=1,col="red")

axis(2)

axis(1,at=results[,1],labels=paste(results[,1],results[,2],sep=": "))

legend("bottomright",pch=as.character(1:nrow(results)),results[,2],cex=0.7)

legend("topleft",c("Control","Treatment"),lty =c(1,3),col=c("blue","red"))

title("Rate Event variation when combining different outcomes")

})

})

## f_bd () function creates the data frame with all the input coefficients. This data frame
## enters in the iterative algorithm.

f_bd<-function(input){

bd<-data.frame(matrix(NA,nrow=10,ncol=25))

for (i in 1:10) {bd[i,1]<-i}

names(bd)[1]<- "n"

names(bd)[2]<- "label"

names(bd)[3]<- "re"

names(bd)[4]<- "rate"

names(bd)[5]<- "efect"

names(bd)[6]<- "ass1"

names(bd)[7]<- "ass2"

names(bd)[8]<- "ass3"

names(bd)[9]<- "ass4"

names(bd)[10]<- "ass5"

names(bd)[11]<- "ass6"

names(bd)[12]<- "ass7"

names(bd)[13]<- "ass8"

names(bd)[14]<- "ass9"

names(bd)[15]<- "ass10"

names(bd)[16]<- "ass1c"

names(bd)[17]<- "ass2c"

names(bd)[18]<- "ass3c"

```

```
names(bd)[19]<-"ass4c"  
names(bd)[20]<-"ass5c"  
names(bd)[21]<-"ass6c"  
names(bd)[22]<-"ass7c"  
names(bd)[23]<-"ass8c"  
names(bd)[24]<-"ass9c"  
names(bd)[25]<-"ass10c"
```

```
bd[1,2]<-input$label1  
bd[2,2]<-input$label2  
bd[3,2]<-input$label3  
bd[4,2]<-input$label4  
bd[5,2]<-input$label5  
bd[6,2]<-input$label6  
bd[7,2]<-input$label7  
bd[8,2]<-input$label8  
bd[9,2]<-input$label9  
bd[10,2]<-input$label10
```

```
bd[1,3]<-input$re1  
bd[2,3]<-input$re2  
bd[3,3]<-input$re3  
bd[4,3]<-input$re4  
bd[5,3]<-input$re5  
bd[6,3]<-input$re6  
bd[7,3]<-input$re7  
bd[8,3]<-input$re8  
bd[9,3]<-input$re9  
bd[10,3]<-input$re10
```

```
bd[1,4]<-input$prev1
```

```
bd[2,4]<-input$prev2  
bd[3,4]<-input$prev3  
bd[4,4]<-input$prev4  
bd[5,4]<-input$prev5  
bd[6,4]<-input$prev6  
bd[7,4]<-input$prev7  
bd[8,4]<-input$prev8  
bd[9,4]<-input$prev9  
bd[10,4]<-input$prev10
```

```
bd[1,5]<-input$rr1  
bd[2,5]<-input$rr2  
bd[3,5]<-input$rr3  
bd[4,5]<-input$rr4  
bd[5,5]<-input$rr5  
bd[6,5]<-input$rr6  
bd[7,5]<-input$rr7  
bd[8,5]<-input$rr8  
bd[9,5]<-input$rr9  
bd[10,5]<-input$rr10
```

```
if (input$know == 0){  
  bd[,"ass1"]<-NA  
  bd[,"ass2"]<-NA  
  bd[,"ass3"]<-NA  
  bd[,"ass4"]<-NA  
  bd[,"ass5"]<-NA  
  bd[,"ass6"]<-NA  
  bd[,"ass7"]<-NA  
  bd[,"ass8"]<-NA
```

bd[,"ass9"]<-NA

bd[,"ass10"]<-NA

bd[1,"ass1c"]<-99

bd[1,"ass2c"]<-as.numeric(input\$ass12)

bd[1,"ass3c"]<-as.numeric(input\$ass13)

bd[1,"ass4c"]<-as.numeric(input\$ass14)

bd[1,"ass5c"]<-as.numeric(input\$ass15)

bd[1,"ass6c"]<-as.numeric(input\$ass16)

bd[1,"ass7c"]<-as.numeric(input\$ass17)

bd[1,"ass8c"]<-as.numeric(input\$ass18)

bd[1,"ass9c"]<-as.numeric(input\$ass19)

bd[1,"ass10c"]<-as.numeric(input\$ass110)

bd[2,"ass2c"]<-99

bd[2,"ass3c"]<-as.numeric(input\$ass23)

bd[2,"ass4c"]<-as.numeric(input\$ass24)

bd[2,"ass5c"]<-as.numeric(input\$ass25)

bd[2,"ass6c"]<-as.numeric(input\$ass26)

bd[2,"ass7c"]<-as.numeric(input\$ass27)

bd[2,"ass8c"]<-as.numeric(input\$ass28)

bd[2,"ass9c"]<-as.numeric(input\$ass29)

bd[2,"ass10c"]<-as.numeric(input\$ass210)

bd[3,"ass3c"]<-99

bd[3,"ass4c"]<-input\$ass34

bd[3,"ass5c"]<-input\$ass35

bd[3,"ass6c"]<-input\$ass36

bd[3,"ass7c"]<-input\$ass37

bd[3,"ass8c"]<-input\$ass38

bd[3,"ass9c"]<-input\$ass39

bd[3,"ass10c"]<-input\$ass310

bd[4,"ass4c"]<-99

bd[4,"ass5c"]<-input\$ass45

bd[4,"ass6c"]<-input\$ass46

bd[4,"ass7c"]<-input\$ass47

bd[4,"ass8c"]<-input\$ass48

bd[4,"ass9c"]<-input\$ass49

bd[4,"ass10c"]<-input\$ass410

bd[5,"ass5c"]<-99

bd[5,"ass6c"]<-input\$ass56

bd[5,"ass7c"]<-input\$ass57

bd[5,"ass8c"]<-input\$ass58

bd[5,"ass9c"]<-input\$ass59

bd[5,"ass10c"]<-input\$ass510

bd[6,"ass6c"]<-99

bd[6,"ass7c"]<-input\$ass67

bd[6,"ass8c"]<-input\$ass68

bd[6,"ass9c"]<-input\$ass69

bd[6,"ass10c"]<-input\$ass610

bd[7,"ass7c"]<-99

bd[7,"ass8c"]<-input\$ass78

bd[7,"ass9c"]<-input\$ass79

bd[7,"ass10c"]<-input\$ass710

bd[8,"ass8c"]<-99

bd[8,"ass9c"]<-input\$ass89

bd[8,"ass10c"]<-input\$ass810

```

bd[9,"ass9c"]<-99
bd[9,"ass10c"]<-input$ass910

bd[10,"ass10c"]<-99
}

if (input$know == 1){

bd[,"ass1c"]<-NA
bd[,"ass2c"]<-NA
bd[,"ass3c"]<-NA
bd[,"ass4c"]<-NA
bd[,"ass5c"]<-NA
bd[,"ass6c"]<-NA
bd[,"ass7c"]<-NA
bd[,"ass8c"]<-NA
bd[,"ass9c"]<-NA
bd[,"ass10c"]<-NA

bd[1,"ass1"]<-1
bd[1,"ass2"]<-input$a12
bd[1,"ass3"]<-input$a13
bd[1,"ass4"]<-input$a14
bd[1,"ass5"]<-input$a15
bd[1,"ass6"]<-input$a16
bd[1,"ass7"]<-input$a17
bd[1,"ass8"]<-input$a18
bd[1,"ass9"]<-input$a19
bd[1,"ass10"]<-input$a110

```

```

bd[2,"ass2"]<-1
bd[2,"ass3"]<-input$a23
bd[2,"ass4"]<-input$a24
bd[2,"ass5"]<-input$a25
bd[2,"ass6"]<-input$a26
bd[2,"ass7"]<-input$a27
bd[2,"ass8"]<-input$a28
bd[2,"ass9"]<-input$a29
bd[2,"ass10"]<-input$a210

bd[3,"ass3"]<-1
bd[3,"ass4"]<-input$a34
bd[3,"ass5"]<-input$a35
bd[3,"ass6"]<-input$a36
bd[3,"ass7"]<-input$a37
bd[3,"ass8"]<-input$a38
bd[3,"ass9"]<-input$a39
bd[3,"ass10"]<-input$a310

bd[4,"ass4"]<-1
bd[4,"ass5"]<-input$a45
bd[4,"ass6"]<-input$a46
bd[4,"ass7"]<-input$a47
bd[4,"ass8"]<-input$a48
bd[4,"ass9"]<-input$a49
bd[4,"ass10"]<-input$a410

bd[5,"ass5"]<-1
bd[5,"ass6"]<-input$a56
bd[5,"ass7"]<-input$a57
bd[5,"ass8"]<-input$a58

```

```
bd[5,"ass9"]<-input$a59
bd[5,"ass10"]<-input$a510

bd[6,"ass6"]<-1
bd[6,"ass7"]<-input$a67
bd[6,"ass8"]<-input$a68
bd[6,"ass9"]<-input$a69
bd[6,"ass10"]<-input$a610

bd[7,"ass7"]<-1
bd[7,"ass8"]<-input$a78
bd[7,"ass9"]<-input$a79
bd[7,"ass10"]<-input$a710

bd[8,"ass8"]<-1
bd[8,"ass9"]<-input$a89
bd[8,"ass10"]<-input$a810

bd[9,"ass9"]<-1
bd[9,"ass10"]<-input$a910

bd[10,"ass10"]<-1

}

if (input$know == 2){

bd[1,"ass1"]<-1
bd[1,"ass2"]<-input$a12
bd[1,"ass3"]<-input$a13
```

```
bd[1,"ass4"]<-input$a14
bd[1,"ass5"]<-input$a15
bd[1,"ass6"]<-input$a16
bd[1,"ass7"]<-input$a17
bd[1,"ass8"]<-input$a18
bd[1,"ass9"]<-input$a19
bd[1,"ass10"]<-input$a110

bd[2,"ass2"]<-1
bd[2,"ass3"]<-input$a23
bd[2,"ass4"]<-input$a24
bd[2,"ass5"]<-input$a25
bd[2,"ass6"]<-input$a26
bd[2,"ass7"]<-input$a27
bd[2,"ass8"]<-input$a28
bd[2,"ass9"]<-input$a29
bd[2,"ass10"]<-input$a210

bd[3,"ass3"]<-1
bd[3,"ass4"]<-input$a34
bd[3,"ass5"]<-input$a35
bd[3,"ass6"]<-input$a36
bd[3,"ass7"]<-input$a37
bd[3,"ass8"]<-input$a38
bd[3,"ass9"]<-input$a39
bd[3,"ass10"]<-input$a310

bd[4,"ass4"]<-1
bd[4,"ass5"]<-input$a45
bd[4,"ass6"]<-input$a46
bd[4,"ass7"]<-input$a47
```


bd[4,"ass8"]<-input\$a48
bd[4,"ass9"]<-input\$a49
bd[4,"ass10"]<-input\$a410

bd[5,"ass5"]<-1
bd[5,"ass6"]<-input\$a56
bd[5,"ass7"]<-input\$a57
bd[5,"ass8"]<-input\$a58
bd[5,"ass9"]<-input\$a59
bd[5,"ass10"]<-input\$a510

bd[6,"ass6"]<-1
bd[6,"ass7"]<-input\$a67
bd[6,"ass8"]<-input\$a68
bd[6,"ass9"]<-input\$a69
bd[6,"ass10"]<-input\$a610

bd[7,"ass7"]<-1
bd[7,"ass8"]<-input\$a78
bd[7,"ass9"]<-input\$a79
bd[7,"ass10"]<-input\$a710

bd[8,"ass8"]<-1
bd[8,"ass9"]<-input\$a89
bd[8,"ass10"]<-input\$a810

bd[9,"ass9"]<-1
bd[9,"ass10"]<-input\$a910

bd[10,"ass10"]<-1

bd[1,"ass1c"]<-99
bd[1,"ass2c"]<-input\$ass12
bd[1,"ass3c"]<-input\$ass13
bd[1,"ass4c"]<-input\$ass14
bd[1,"ass5c"]<-input\$ass15
bd[1,"ass6c"]<-input\$ass16
bd[1,"ass7c"]<-input\$ass17
bd[1,"ass8c"]<-input\$ass18
bd[1,"ass9c"]<-input\$ass19
bd[1,"ass10c"]<-input\$ass110

bd[2,"ass2c"]<-99
bd[2,"ass3c"]<-input\$ass23
bd[2,"ass4c"]<-input\$ass24
bd[2,"ass5c"]<-input\$ass25
bd[2,"ass6c"]<-input\$ass26
bd[2,"ass7c"]<-input\$ass27
bd[2,"ass8c"]<-input\$ass28
bd[2,"ass9c"]<-input\$ass29
bd[2,"ass10c"]<-input\$ass210

bd[3,"ass3c"]<-99
bd[3,"ass4c"]<-input\$ass34
bd[3,"ass5c"]<-input\$ass35
bd[3,"ass6c"]<-input\$ass36
bd[3,"ass7c"]<-input\$ass37
bd[3,"ass8c"]<-input\$ass38
bd[3,"ass9c"]<-input\$ass39
bd[3,"ass10c"]<-input\$ass310

```

bd[4,"ass4c"]<-99
bd[4,"ass5c"]<-input$ass45
bd[4,"ass6c"]<-input$ass46
bd[4,"ass7c"]<-input$ass47
bd[4,"ass8c"]<-input$ass48
bd[4,"ass9c"]<-input$ass49
bd[4,"ass10c"]<-input$ass410

bd[5,"ass5c"]<-99
bd[5,"ass6c"]<-input$ass56
bd[5,"ass7c"]<-input$ass57
bd[5,"ass8c"]<-input$ass58
bd[5,"ass9c"]<-input$ass59
bd[5,"ass10c"]<-input$ass510

bd[6,"ass6c"]<-99
bd[6,"ass7c"]<-input$ass67
bd[6,"ass8c"]<-input$ass68
bd[6,"ass9c"]<-input$ass69
bd[6,"ass10c"]<-input$ass610

bd[7,"ass7c"]<-99
bd[7,"ass8c"]<-input$ass78
bd[7,"ass9c"]<-input$ass79
bd[7,"ass10c"]<-input$ass710

bd[8,"ass8c"]<-99
bd[8,"ass9c"]<-input$ass89
bd[8,"ass10c"]<-input$ass810

bd[9,"ass9c"]<-99

```

```

bd[9,"ass10c"]<-input$ass910

bd[10,"ass10c"]<-99
}

if (input$know==9999){
for (i in 1:(input$n)) {
for (j in 1:(input$n)) {
min<-max(0,bd[i,"rate"]+bd[j,"rate"]-1)
max<-min(bd[i,"rate"],bd[j,"rate"])
med<-bd[i,"rate"]*bd[j,"rate"]

if (i!=j){
if (bd[i,j+15]<4) {
delta=(med-min)/4
if (bd[i,j+15]==0) {bd[i,j+5]<-min}
if (bd[i,j+15]==1) {bd[i,j+5]<-med-(3*delta)}
if (bd[i,j+15]==2) {bd[i,j+5]<-med-(2*delta)}
if (bd[i,j+15]==3) {bd[i,j+5]<-med-(1*delta)}
}
if (bd[i,j+15]==4) {bd[i,j+5]<-med}
if (bd[i,j+15]>4) {
delta=(max-med)/4
if (bd[i,j+15]==5) {bd[i,j+5]<-med+(1*delta)}
if (bd[i,j+15]==6) {bd[i,j+5]<-med+(2*delta)}
if (bd[i,j+15]==7) {bd[i,j+5]<-med+(3*delta)}
if (bd[i,j+15]==8) {bd[i,j+5]<-max}
}}
}
}
}

```

```

if (i==j) {bd[i,5+j]<-1}
}}}

if (input$know==992){

for (i in 1:(input$n)) {
for (j in i:(input$n)) {

min<-max(0,bd[i,"rate"]+bd[j,"rate"]-1)
max<-min(bd[i,"rate"],bd[j,"rate"])
med<-bd[i,"rate"]*bd[j,"rate"]
actual_ass<-bd[i,j+5]

if (!is.na(actual_ass)){

if (i!=j){

if (input$bd[i,j+15]<4) {
delta=(med-min)/4
if (input$bd[i,j+15]==0) {bd[i,j+5]<-min}
if (input$bd[i,j+15]==1) {bd[i,j+5]<-med-(3*delta)}
if (input$bd[i,j+15]==2) {bd[i,j+5]<-med-(2*delta)}
if (input$bd[i,j+15]==3) {bd[i,j+5]<-med-(1*delta)}}

if (input$bd[i,j+15]==4) {bd[i,j+5]<-med}

if (input$bd[i,j+15]>4) {
delta=(max-med)/4
if (input$bd[i,j+15]==5) {bd[i,j+5]<-med+(1*delta)}
if (input$bd[i,j+15]==6) {bd[i,j+5]<-med+(2*delta)}

```

```

if (input$bd[i,j+15]==7) {bd[i,j+5]<-med+(3*delta)}
if (input$bd[i,j+15]==8) {bd[i,j+5]<-max}}
}
}}
}

bd1<-bd[1:input$n,1:(input$n+5)]
bd2<-bd[1:input$n,16:(input$n+15)]
bd<-cbind(bd1,bd2)
return(bd)
}

## depura() checks the association matrix entered by the user and replace the values under or over the
## limits

depura<-function(bd){
nvar<-nrow(bd)
for (i in 1:(nrow(bd)-1)){
p1<-as.numeric(bd[i,4])
for (j in (i+1):nrow(bd)){
p2<-as.numeric(bd[j,4])

min<-max(0,p1+p2-1)
max<-min(p1,p2)

if (bd[i,j+5]<min & !is.na(bd[i,j+5])) {bd[i,j+5]<-88}
if (bd[i,j+5]>max & !is.na(bd[i,j+5])) {bd[i,j+5]<-88}
if (is.na(bd[i,j+5])) {bd[i,j+5]<-88}

```

```

ass<-bd[i,j+nvar+5]

if (bd[i,j+5]==88){
  med=p1*p2
  if (ass==4) {bd[i,j+5]=med}

  if (ass<4) {
    delta=(med-min)/4
    if (ass==0) {bd[i,j+5]<-min}
    if (ass==1) {bd[i,j+5]<-med-(3*delta)}
    if (ass==2) {bd[i,j+5]<-med-(2*delta)}
    if (ass==3) {bd[i,j+5]<-med-(1*delta)}

    if (ass>4) {
      delta=(max-med)/4
      if (ass==5) {bd[i,j+5]<-med+(1*delta)}
      if (ass==6) {bd[i,j+5]<-med+(2*delta)}
      if (ass==7) {bd[i,j+5]<-med+(3*delta)}
      if (ass==8) {bd[i,j+5]<-max}
    }
  }

}

return(bd[,1:(nrow(bd)+5)])
}

```

reset_ass() function checks the assotions matrix and correct them if it is necessary.

```

reset_ass<-function(bd){
  for (i in 1:(nrow(bd)-1)){
    p1<-as.numeric(bd[i,4])
    for (j in (i+1):nrow(bd)){
      p2<-as.numeric(bd[j,4])

      min<-max(0,p1+p2-1)
      max<-min(p1,p2)
      if (bd[i,j+5]<min & !is.na(bd[i,j+5])) {bd[i,j+5]<-min}
      if (bd[i,j+5]>max & !is.na(bd[i,j+5])) {bd[i,j+5]<-max}
      if (is.na(bd[i,j+5])) {bd[i,j+5]<-bd[j,i+5]}
      if (bd[i,j+5]<min & !is.na(bd[i,j+5])) {bd[i,j+5]<-min}
      if (bd[i,j+5]>max & !is.na(bd[i,j+5])) {bd[i,j+5]<-max}
    }
  }

  for (i in 1:(nrow(bd))){bd[,5+i]<-1}

  vars<-nrow(bd)-sum(is.na(bd[,2]))
  for (i in 1:vars){
    for (j in i:vars){
      bd[j+1,i+5]<-NA
    }
  }
  bd<-subset(bd,is.na(bd[1]))
  return(bd)
}

```

reset_ds() resets the data set before to start the recursive algorithm

```

reset_ds<-function(bd,tas=0){

vars<-nrow(bd)-sum(is.na(bd[,2]))

for (i in 1:(nrow(bd)-1)){
  p1<-bd[i,4]
  for (j in (i+1):nrow(bd)){
    p2<-bd[j,4]

    if (is.na(bd[i,j+5])){
      pmin<-max(p1+p2-1,0)
      pmax<-min(p1,p2)
      range<-pmax-pmin
      exp<-p1*p2
      delta_u<-(pmax-exp)/4
      delta_d<-(-exp-pmin)/4
      if (tas==4){bd[i,j+5]<-exp}
      if (tas>4){bd[i,j+5]<-exp+(delta_u*(tas-4))}
      if (tas<4){bd[i,j+5]<-exp-(delta_d*(4-tas))}
    }
  }
}

bd<-bd[1:vars,1:(5+vars)]
return(bd)
}

```

ssize() function compute the Sample Size using a Normal Aproximation

```
ssize<-function (p1,p2,alfa=0.05,beta=0.2,test=1){
```

```

p<-(p1+p2)/2
denom<-p1-p2
n1<-qnorm(1-(alfa/2))*sqrt(2*p*(1-p))
n2<-qnorm(1-beta)*sqrt((p1*(1-p1)+(p2*(1-p2)))
n<-trunc(((n1+n2)/denom)^2)+1

if (test==1) {n_final<-n}
if (test==2){
  cc=1/(0.5*abs(p1-p2))
  n_final<-n+trunc(cc)+1

  if (test==3){
    zeta<-(qnorm(1-(alfa/2))-qnorm(beta))
    delta<-(asin(sqrt(p1))-asin(sqrt(p2)))
    n_final<-trunc((zeta/delta)^2*5)+1
  }

  if (test==4){
    zeta<-(qnorm(1-(alfa/2))-qnorm(beta))
    delta<-abs(asin(sqrt(p1))-asin(sqrt(p2)))
    ce<-(p1*(1-p1))^(0.5)+(p2*(1-p2))^(0.5)
    n_final<-trunc((zeta+sqrt(zeta^2+(2*ce*delta)))^2/(8*delta^2))+1
  }
}

return(n_final)
}

```

First_resul() function computes the SSR for the first RE. The SSR if any component is combined.

```
first_resul<-function(bd,alfa,beta,test){
```

```

x<-reset_ass(bd)

re<-99

if (max(x$re)==1) {re<-subset(x,x$re==1)$n}

results<-data.frame(as.numeric(1),as.array(x[re,2]),x[re,4],x[re,5],ssize(x[re,4],x[re,4]*x[re,5],alfa,beta,test))

names(results)<-c("Iter","Label","Rate","RR","n")

return(results)

}

```

```

## potential_graph() function computes and graphs the potential SSR using the different combinations between
## the RE and each AE. Note that the exact value of the association is not implemented in this plot.
## Rather the plot shows the potential SSR variation for each combination of outcomes considering
## the whole range of joint probabilities according to the Fréchet bounds.

```

```

potential_graph<-function(bd,res,alfa,beta,test){

```

```

mult<-as.numeric(res$N[1])

re<-subset(bd,bd$re==1)$n

p1_b<-bd[re,4]

RR1<-bd[re,5]

p1_f<-p1_b*RR1

n<-ssize(p1_b,p1_f,alfa,beta,test)

results<-data.frame()

results[1,1]<-1

results[1,2]<-bd$label[re]

results[1,3]<-n

results[1,4]<-n

pot<-bd[-re,]

for (i in 1:(nrow(pot))){

p2_b<-pot[i,4]

```

```

min_b<-max(0,p1_b+p2_b-1)

max_b<-min(p1_b,p2_b)

ce_min_b<-p1_b+p2_b-min_b

ce_max_b<-p1_b+p2_b-max_b

p1_b,p2_b,min_b,ce_min_b

p1_b,p2_b,max_b,ce_max_b

RR2<-pot[i,5]

p2_f<-p2_b*RR2

min_f<-max(0,p1_f+p2_f-1)

max_f<-min(p1_f,p2_f)

ce_min_f<-p1_f+p2_f-min_f

ce_max_f<-p1_f+p2_f-max_f

n_min<-ssize(ce_min_b,ce_min_f,alfa,beta,test)

n_max<-ssize(ce_max_b,ce_max_f,alfa,beta,test)

if (n_min>n_max) {x<-n_min

n_min<-n_max

n_max<-x}

results[i+1,1]<-i+1

results[i+1,2]<-paste(results[1,2]," +",pot$label[i])

results[i+1,3]<-n_min

results[i+1,4]<-n_max

n_min<-1

n_max<-1}

results$diff<-results[,4]-results[,3]

mult<-1.5*mult/results[1,3]

matplot(c(1,nrow(results)),c(0,results[1,3]*mult),ylab="Sample Size Requirement",type="n",xlab="Components",main="Range of SSR when combining the
RE with all AE depending on the strenght of the association ",axes=F)

matpoints(results[,1], results[,3], pch = 20, cex = 0.5)

matpoints(results[,1], results[,4], pch = 20, cex = 0.5)

matlines(results[,1], results[,3:4],type="l",lty=4)

```

```

axis(1,at=results[,1],labels=results[,2])

axis(2)

axis.break(2,results[1,3]*mult,style = "slash")

legend("topright", c("Upper Size", "Lower Size", "RE Size"), col = c("red", "black", "blue"),
      text.col = "black", lty = c(2, -1, 1), pch = c(NA, 3, 4),
      merge = TRUE, bg = "gray95")

box()

abline(h=results[1,3],lty=1,col="blue")

if (nrow(res)>1) {
  for (i in 1:nrow(res)-1){abline(h=res[i,5],lty=i+1,col="blue")}
}

dis<-0.1*results[1,3]

pmax<-results[,c(1,3)]

pmin<-results[,c(1,4)]

pmax[,3]<-pmax[,2]+dis

pmin[,3]<-pmin[,2]-dis

text(pmax[,c(1,3)],as.character(pmax[,2]),cex=1)

text(pmin[,c(1,3)],as.character(pmin[,2]),cex=1,col="red")

}

```

```

## all1() function compute the SSR for each possible combination of two outcomes. all1() function selects
## the combination which lower SSR needs if any there are and aggregates the outcome to the CE. If any combination
## is done, this function resets the data frame.

```

```

all1<-function(base,alfa,beta,test){
  bd<-base

```

```

  re<-subset(bd,bd$re==1)$n
  results<-data.frame(1,as.array(bd[re,2]),bd[re,4],bd[re,5],ssize(bd[re,4],bd[re,4]*bd[re,5],alfa=alfa,beta=beta,test=test))
  names(results)<-c("Iter","Label","Rate","RR","n")
  stop<-0

  itera<-2
  bd<-reset_ass(bd)
  r<-engine(bd,alfa,beta,test)
  x<-select_best(r)
  bd2<-add_cep(bd,x)
  if (nrow(bd)==nrow(bd2)){stop<-1}
  if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x$lab_cep),as.numeric(x$pcep_b),as.numeric(x$RR_cep),as.numeric(x$n_cep)))}
  if (nrow(bd2)==1){stop<-1}

  itera<-3
  if (stop==0){
    bd2<-reset_ass(bd2)
    r2<-engine(as.data.frame(bd2),alfa,beta,test)
    x2<-select_best(r2)
    bd3<-add_cep(bd2,x2)
    if (nrow(bd2)==nrow(bd3)){stop<-1}
    if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x2$lab_cep),as.numeric(x2$pcep_b),as.numeric(x2$RR_cep),as.numeric(x2$n_cep)))}
    if (nrow(bd3)==1){stop<-1}
  }

  itera<-4
  if (stop==0){
    bd3<-reset_ass(bd3)
    r3<-engine(bd3,alfa,beta,test)
    x3<-select_best(r3)
    bd4<-add_cep(bd3,x3)

```

```

if (nrow(bd3)==nrow(bd4)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x3$lab_cep),as.numeric(x3$pcep_b),as.numeric(x3$RR_cep),as.numeric(x3$n_cep)))}

if (nrow(bd4)==1){stop<-1}

}

itera<-5

if (stop==0){

bd4<-reset_ass(bd4)

r4<-engine(bd4,alfa,beta,test)

x4<-select_best(r4)

bd5<-add_cep(bd4,x4)

if (nrow(bd4)==nrow(bd5)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x4$lab_cep),as.numeric(x4$pcep_b),as.numeric(x4$RR_cep),as.numeric(x4$n_cep)))}

if (nrow(bd5)==1){stop<-1}

}

itera<-6

if (stop==0){

bd5<-reset_ass(bd5)

r5<-engine(bd5,alfa,beta,test)

x5<-select_best(r5)

bd6<-add_cep(bd5,x5)

if (nrow(bd5)==nrow(bd6)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x5$lab_cep),as.numeric(x5$pcep_b),as.numeric(x5$RR_cep),as.numeric(x5$n_cep)))}

if (nrow(bd6)==1){stop<-1}

}

itera<-7

if (stop==0){

bd6<-reset_ass(bd6)

r6<-engine(bd6,alfa,beta,test)

```

```

x6<-select_best(r6)

bd7<-add_cep(bd5,x6)

if (nrow(bd6)==nrow(bd7)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x6$lab_cep),as.numeric(x6$pcep_b),as.numeric(x6$RR_cep),as.numeric(x6$n_cep)))}

if (nrow(bd7)==1){stop<-1}

}

itera<-8

if (stop==0){

bd7<-reset_ass(bd7)

r7<-engine(bd7,alfa,beta,test)

x7<-select_best(r7)

bd8<-add_cep(bd7,x7)

if (nrow(bd7)==nrow(bd8)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x7$lab_cep),as.numeric(x7$pcep_b),as.numeric(x7$RR_cep),as.numeric(x7$n_cep)))}

if (nrow(bd8)==1){stop<-1}

}

itera<-9

if (stop==0){

bd8<-reset_ass(bd8)

r8<-engine(bd8,alfa,beta,test)

x8<-select_best(r8)

bd9<-add_cep(bd8,x8)

if (nrow(bd8)==nrow(bd9)){stop<-1}

if (stop==0){results<-rbind(results,c(as.numeric(itera),as.character(x8$lab_cep),as.numeric(x8$pcep_b),as.numeric(x8$RR_cep),as.numeric(x8$n_cep)))}

if (nrow(bd9)==1){stop<-1}

}

for (i in 1 : nrow(results)){results[i,"%SSR"]<-round(as.numeric(results[i,5])/as.numeric(results[1,5])*100,2)}

```



```

return(results)
}

## engine() computes the SSR for each pair of outcomes.

engine<-function(bd,alfa,beta,test){
for (i in 1:nrow(bd)){if (i!=2){bd[,i]<-as.numeric(bd[,i])}}
re<-as.numeric(subset(bd,bd$re==1)$n)
results<-NULL
out<-NULL

for (i in 1:nrow(bd)){

if (re>i) {ass_b<-as.numeric(bd[i,re+5])}
if (re<i) {ass_b<-as.numeric(bd[i,+5])}

p1_b<-as.numeric(bd[re,4])
p2_b<-as.numeric(bd[i,4])
RR1<-as.numeric(bd[re,5])
RR2<-as.numeric(bd[i,5])
lab1<-as.character(bd[re,2])
lab2<-as.character(bd[i,2])
tass=0
if (i!=re){

# In Bin-CE always is.na(ass_b) is FALSE.

if (is.na(ass_b)) {out<-SR(p1_b,p2_b,RR1,RR2,tass,lab_1=lab1,lab_2=lab2)}

```

```

if (!is.na(ass_b)) {
a_min<-max(0,p1_b+p2_b-1)
a_max<-min(p1_b,p2_b)
dif<-as.numeric(a_max-a_min)
tick<-ass_b/dif
a2_min<-max(0,(p1_b*RR1)+(p2_b*RR2)-1)
a2_max<-min((p1_b*RR1),(p2_b*RR2))
dif2<-a2_max-a2_min
ass_f<-tick*dif2

out<-SR1(p1_b,p2_b,RR1,RR2,over_b=ass_b,over_f=ass_f,tass,lab_1=lab1,lab_2=lab2,alfa,beta,test)
}
}

results<-data.frame(rbind(results,out))

}

return(results)

## SR1() functions compare the SSR of the RE with the combination of the RE+AEI

SR1<-function(p1_b,p2_b,RR1,RR2,over_b,over_f,tass=1,alfa=0.05,beta=0.2,lab_1,lab_2,test){
p1_f<-RR1*p1_b
p2_f<-RR2*p2_b

if (tass==0) {
p12_b<-over_b
p12_f<-over_f
}
if (tass==1){
p12_b<-over_b*(p1_b+p2_b)/(over_b+1)
p12_f<-over_f*(p1_f+p2_f)/(over_f+1)
}
}

```

```

if (tass==2){
  p12_b<-corr_to_p12(p1_b,p2_b,over_b)
  p12_f<-corr_to_p12(p1_f,p2_f,over_f)
}
if (tass==3){
  p12_b<-over_b*min(p1_b,p2_b)
  p12_f<-over_f*min(p1_f,p2_f)
}

pcep_b<-p1_b+p2_b-p12_b
pcep_f<-p1_f+p2_f-p12_f
RR_cep<-pcep_f/pcep_b
n_pe<-ssize(p1_b,p1_f,alfa,beta,test)
n_cep<-ssize(pcep_b,pcep_f,alfa,beta,test)
SR<-round(n_pe/n_cep,2)
pcep_b<-round(pcep_b,4)
RR_cep<-round(RR_cep,2)
pcep_f<-round(pcep_f,4)
over_b<-round(over_b,4)
over_f<-round(over_f,4)

USE<-ifelse(round(n_pe/n_cep,2)<1,"Use Relevant Endpoint","Use CEP")

lab_cep<-paste(lab_1,lab_2,sep="-")

out<-data.frame(cbind(lab_1=as.character(lab_1),
  p1_b=as.numeric(paste(p1_b)),
  RR1=as.numeric(paste(RR1)),
  p1_f=as.numeric(paste(p1_f)),
  n_pe=as.numeric(paste(n_pe)),
  lab_2=as.character(lab_2),
  p2_b=as.numeric(paste(p2_b)),
  RR2=as.numeric(paste(RR2)),
  p2_f=as.numeric(paste(p2_f)),
  lab_cep=as.character(lab_cep),
  pcep_b=as.numeric(paste(pcep_b)),
  RR_cep=as.numeric(paste(RR_cep)),
  pcep_f=as.numeric(paste(pcep_f)),
  n_cep=as.numeric(paste(n_cep)),
  over_b=as.numeric(paste(over_b)),
  over_f=as.numeric(paste(over_f)),
  tass=as.numeric(paste(tass)),
  SR=as.numeric(paste(SR)))

#names(out)<-c("p(RE)_b","RR(RE)","P(RE)_f","n(RE)","p(CEP)_b","RR(CEP)","p(CEP)_f","n(CEP)","Overlap Basal","Overlap Final","SR","Ussage")
return(data.frame(out))
}

## select_best() function selects the best combination (low SSR) between all combinatio candidates.

select_best<-function(results){

for (x in 1:ncol(results)){
  if (x==1 || x==6 || x==10) {results[,x]<-results[,x]}
  else {results[,x]<-as.numeric(paste(results[,x]))}
  x<-subset(results,results$SR==max(results$SR))
  x<-x[!duplicated(x), ]
  if (nrow(x)>1) {x<-subset(x,x$n_cep==min(x$n_cep))}
  return(x)}

## add_cep() function adds the new combination of RE+AE as a new RE and removes the old RE and AE.
## add_cep() function prepares the new data frame for the next iteration.

add_cep<-function(df,best){

```

```

df2<-df

if (best$SR > 1 && nrow(df2)>2){
  df2<-subset(df2,df2$label != best$lab_1)
  df2<-subset(df2,df2$label != best$lab_2)

  if (nrow(df2)>1){

    for (p in 1:(nrow(df2)-1)){
      var<-as.numeric(df2[p,1])
      z<-as.numeric(df2[-p,1])
      for (t in 1:length(z)){
        df2[p,6+t]<-as.numeric(df[var,z[t]+5])
      }}

    if (nrow(df2)>0){
      df2<-df2[,1:(nrow(df2)+6)]
      df2[,ncol(df2)]<-NA
      for (i in 1:nrow(df2)){df2[i,5+i]<-1}
    }

    df2<-
    rbind(df2,c(as.numeric(99999),as.character(best$lab_cep),as.numeric(1),as.numeric(best$pcp_b),as.numeric(best$RR_cep),rep(NA,nrow(df2)+1)))
  }
}

```

```

if (best$SR > 1){
  if (nrow(df2)>0) {df2<-reset_ass(df2)}
  if (nrow(df2)>0) {for (i in 1:nrow(df2)){df2[i,1]<-i}}
  for (i in 1:nrow(df2)){df2[i,5+i]<-1}
}

if (best$SR>1){
  for (i in 1:(nrow(df2)-1)){

    ce_1<-as.numeric(df[df$lab==best$lab_1,]$n)
    ce_2<-as.numeric(df[df$lab==best$lab_2,]$n)
    ae<-as.numeric(subset(df,df$lab==df2[i,2])$n)

    if (ce_1>ae) {ass1<-as.numeric(df[ae,ce_1+5])}
    if (ce_1<ae) {ass1<-as.numeric(df[ce_1,ae+5])}

    if (ce_2>ae) {ass2<-as.numeric(df[ae,ce_2+5])}
    if (ce_2<ae) {ass2<-as.numeric(df[ce_2,ae+5])}

    p<-as.numeric(df$prev[ae])
    ass3<-(min(ass1,ass2)-max(0,ass1+ass2-p))/2+max(0,ass1+ass2-p)
    df2[i,ncol(df2)]<-ass1+ass2-ass3

    rm(ass1,ass2,ass3,p)
  }}
}

if (best$SR > 1 && nrow(df)==2){

```

```
df2<-rbind(df2,c(as.numeric(1),as.character(best$lab_cep),as.numeric(1),as.numeric(best$spcep_b),as.numeric(best$RR_cep),rep(NA,nrow(df2)+1)))  
df2<-df2[nrow(df2),1:5]  
}  
return(df2)  
}
```

**10.7. Anexo 7: The Use of a Binary Composite Endpoint
and Sample Size Requirement: Influence of
Endpoints Overlap.**



Practice of Epidemiology

The Use of a Binary Composite Endpoint and Sample Size Requirement: Influence of Endpoints Overlap

Josep-Ramon Marsal, Ignacio Ferreira-González*, Sandra Bertran, Aida Ribera, Galetà Permanyer-Miralda, David García-Dorado, and Guadalupe Gómez

* Correspondence to Dr. Ignacio Ferreira-Gonzalez, Epidemiology Unit, Department of Cardiology, Vall d'Hebron Hospital, Passeig de la Vall d'Hebron, 119-129, 08035 Barcelona, Spain (e-mail: nachoferreira@secardiologia.es).

Initially submitted November 13, 2015; accepted for publication April 4, 2016.

Although composite endpoints (CE) are common in clinical trials, the impact of the relationship between the components of a binary CE on the sample size requirement (SSR) has not been addressed. We performed a computational study considering 2 treatments and a CE with 2 components: the relevant endpoint (RE) and the additional endpoint (AE). We assessed the strength of the components' interrelation by the degree of relative overlap between them, which was stratified into 5 groups. Within each stratum, SSR was computed for multiple scenarios by varying the events proportion and the effect of the therapy. A lower SSR using CE was defined as the best scenario for using the CE. In 25 of 66 scenarios the degree of relative overlap determined the benefit of using CE instead of the RE. Adding an AE with greater effect than the RE leads to lower SSR using the CE regardless of the AE proportion and the relative overlap. The influence of overlapping decreases when the effect on RE increases. Adding an AE with lower effect than the RE constitutes the most uncertain situation. In summary, the interrelationship between CE components, assessed by the relative overlap, can help to define the SSR in specific situations and it should be considered for SSR computation.

association measures; binary endpoints; composite endpoint; correlated endpoints; outcome assessment; sample size

Abbreviations: AE, additional endpoint; AMI, acute myocardial infarction; CE, composite endpoint; RCT, randomized clinical trial; RE, relevant endpoint; RR, relative risk; SR, sample ratio; SSR, sample size requirement.

Use of a composite endpoint (CE) is a common strategy to reduce sample size in randomized clinical trials (RCT) (1–5). In addition to other potential benefits (6–8), combining 2 or more endpoints in a single CE might increase the number of events, thus reducing sample size requirements (SSRs). This strategy has been widely employed in cardiovascular clinical trials. For instance, the endpoint acute myocardial infarction (AMI) is usually combined with the endpoint “mortality.” It is biologically plausible that reducing AMI could also lead to mortality reductions, and both are important outcomes for patients. The fact that both components are important for clinicians and patients provides a rationale for using CEs (4, 5, 9–12). Although the effect of a new treatment on the CE clearly depends on the effect on each component, the impact of the relationship between components on the statistical power and sample size is not

usually considered. Thus, the increased probability of mortality in patients with AMI, which determines the correlation between both components, could be an important issue in sample size computation. In other words, the presence of a strong or a weak relationship between components of a CE may modify sample size computation in a RCT.

The benefits, risks, and influence on sample size computation of using CEs have been broadly studied (1, 2, 6, 7, 9, 13, 14). However, to our knowledge, the impact of the strength of the relationship between CE components on such computation has not been addressed in depth (3, 4, 15–18). In this sense, it is well known that the degree of relationship between CE components can affect both type I and type II errors when computing the sample size of the CE. However, a numerical approach to quantifying their impact on the power of the CE has not, to our knowledge, been explored. Although in the

setting of 2 or more coprimary binary endpoints, the correlation between endpoints and impact on power and sample size has been explored (16, 17, 19, 20), this is not the case regarding CE components. In the setting of survival analysis, Gómez and Lagakos (21) have developed a method to quantify the efficiency of adding an endpoint to an outcome expected to capture the main effect of the treatment. Their method is valid for time-to-event primary endpoints and is based on the asymptotic relative efficiency between 2 log-rank tests. The method is built from the marginal laws of the times to each of these 2 outcomes and uses Spearman's rank correlation between these 2 times. Other papers discuss the relationship between asymptotic relative efficiency and needed sample sizes (22) and provide recommendations in cardiovascular studies (23).

We propose a method to quantify the strength of the relationship between components of a binary CE. It is based on the quantification of the overlap between them (i.e., the probability of both events happening together) and their relative contribution to the probability of the CE. We explore, based on this method, the impact of the strength of the interrelation between the components on the SSR using the CE, which, ultimately, can be useful when choosing between a binary CE and one of its components.

METHODS

Definitions and assumptions

For simplicity we consider an RCT with only 2 treatment arms. We define a relevant endpoint (RE) as the outcome that is expected to drive the main effect of the treatment. For example, if the treatment is expected to reduce the rate of AMI, then AMI is the RE. The additional endpoint (AE) is the outcome that the researcher considers combining with the RE in a CE to reduce the SSR. For instance, mortality and AMI could be combined in a CE attempting to decrease sample size. The statistical test is based on the difference in events proportion between groups. Sample size estimates are computed, assuming a normal distribution, for the CE and the RE. If using the CE requires a smaller sample than the single RE then researchers could prefer to use the CE (22) instead of the RE alone.

Let (X_{ij1}, X_{ij2}) be the vector of responses where X_{ij1} and X_{ij2} denote the responses of the RE and the AE endpoints

respectively for the j th subject in the i th treatment group ($i = 1, 2$). We assume that the responses are independently distributed as a bivariate Bernoulli distribution with $E(X_{ijk}) = \pi_{ik}$ and $V(X_{ijk}) = \pi_{ik}(1 - \pi_{ik})$, where k only takes the values 1 or 2. Define the response of the CE by X_{ij}^* ; note that it is equal to 1 whenever X_{ij1} or X_{ij2} are equal 1, and it is 0 otherwise. X_{ij}^* is a Bernoulli random variable with expectation:

$$E(X_{ij}^*) = \text{Prob}(X_{ij}^* = 1) = \text{Prob}((X_{ij1} = 1) \cup (X_{ij2} = 1)) \\ = \text{Prob}(X_{ij1} = 1) + \text{Prob}(X_{ij2} = 1) - \text{Prob}((X_{ij1} = 1) \cap (X_{ij2} = 1)) = \pi_{i1} + \pi_{i2} - \pi_{i12} \tag{1}$$

where π_{i12} corresponds to the probability of both outcomes (RE and AE) happening together. Note that with these 6 parameters ($\pi_{i=1,1}, \pi_{i=1,2}, \pi_{i=1,12}, \pi_{i=2,1}, \pi_{i=2,2}, \pi_{i=2,12}$), the sample size for RE and CE can be computed. Usually π_{i1}, π_{i2} are known, but this is not the case for π_{i12} .

The overall possible scenarios concerning the probability of the intersection of both outcomes in an RCT can be represented as in Figure 1.

Figure 1A depicts a scenario without overlapping between endpoints (i.e., no patient who had AMI died). Figure 1B, Figure 1C, and Figure 1D represent weak, moderate, and strong overlap, respectively. Figure 1E represents the maximum possible overlap (i.e., all patients who died had had AMI). Each scenario has different impact on the computation of the sample size.

We define relative overlap as the conditional probability of experiencing the 2 outcomes given that the patient has experienced one of them, which is evaluated as the ratio between the probability of the intersection and the probability of the CE as computed in equation (1). This statistic is well known in ecology and genetics as Jaccard's index of similarity (24–26). It is used to estimate the similarity of sample sets. Using our notation, the relative overlap in group i (for an arbitrary subject j from treatment group i):

$$RO_i = \text{Prob}(X_{i1} = X_{i2} = 1 | X_{i1} + X_{i2} \geq 1) = \frac{\pi_{i12}}{\pi_{i1} + \pi_{i2} - \pi_{i12}} \tag{2}$$

Note that equation (2) corresponds to $P(A \cap B) \div P(A \cup B)$ when A and B stand, respectively, for the events $[X_{ij1} = 1]$

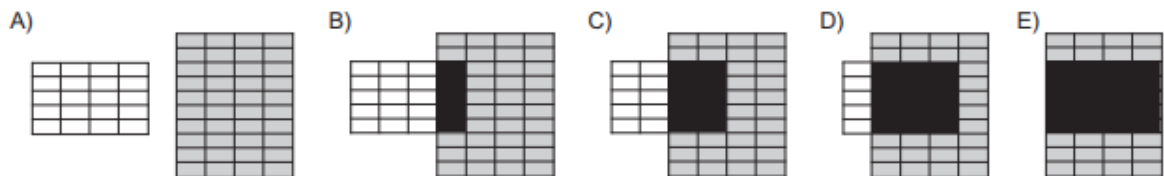


Figure 1. Example of possible scenarios concerning the probability of the intersection of both outcomes in a randomized clinical trial. Each square represents 1% probability. The white square represent the probability of death ($\pi_{i1} = 0.20$), and gray square represents the probability of acute myocardial infarction ($\pi_{i2} = 0.40$) for the i th treatment. The π_{i12} in each scenario are: 0.00 (A), 0.05 (B), 0.10 (C), 0.15 (D), and 0.20 (E), and the probability of the composite endpoint in each scenario was assessed as in equation (1), at 0.60, 0.55, 0.50, 0.45, and 0.40, respectively.

and $[X_{ij2} = 1]$. Following basic relationships between the probability of the union of 2 events and the probability of their intersection it follows that the relative overlap in group i , RO_i , is bounded by $\max\{0, \pi_{i1} + \pi_{i2} - 1\} \leq RO_i \leq \min\left\{\frac{\pi_{i1}}{\pi_{i2}}, \frac{\pi_{i2}}{\pi_{i1}}\right\}$.

The π_{i12} value ranges from 0, in the case of 2 disjoint sets (Figure 1A), to the value of the less prevalent endpoint probability, in the case of completely overlapping sets (Figure 1E); then:

$$\max\{0; \pi_{1j} + \pi_{2j} - 1\} \leq \pi_{i12} \leq \min\{\pi_{i1}, \pi_{i2}\}. \quad (3)$$

From equations (1) and (3), the range of possible values of $E(X_{ij}^*)$ is given by

$$[\{\max\{\pi_{i1}, \pi_{i2}\} \leq \text{Prob}(X_{ij}^* = 1) \leq \min\{\pi_{i1} + \pi_{i2}, 1\}\}]. \quad (4)$$

And from equations (1) to (4), the range of values of RO_i is bounded by:

$$\left[\max\{0, \pi_{i1} + \pi_{i2} - 1\}, \min\left\{\frac{\pi_{i1}}{\pi_{i2}}, \frac{\pi_{i2}}{\pi_{i1}}\right\} \right]. \quad (5)$$

We additionally assume that the magnitude of the relationship between endpoints is the same in treatment ($i = 1$) and control groups ($i = 2$); the conditional probability of having the AE as well as having had the RE (and vice versa) is not affected by the treatment. Note that this does not imply $\pi_{1,12} = \pi_{2,12}$. For example, consider that the proportion of the RE and the AE in the control group ($i = 2$) are 0.20 and 0.40 respectively. From equation (3) it easily follows that the range of values of $\pi_{2,12}$ can fall from 0 to 0.2. Assuming that the effect of the treatment is quantified by the relative risk (RR) and is equal to 0.9 for the RE and 0.5 for the AE, the proportion for the treatment group ($i = 1$) will be 0.18 and 0.20 for the RE and the AE, respectively. Again, from equation (3), $\pi_{1,12}$ can take values from 0 to 0.18. Considering, for instance, the maximum possible overlap (Figure 1E), then $\pi_{2,12} = 0.20 \neq \pi_{1,12} = 0.18$.

Sample size and statistical power calculation

The analysis aimed at quantifying to what extent using a binary CE instead of a single RE can decrease the SSR for a given significance level and for a given power to detect a given proportion difference, considering different overlapping scenarios.

SSRs using the RE or using the CE are calculated to detect differences in the proportion of events between both groups, considering a type I error of 0.05 (note that other hypothesis tests such as superiority could be used). We address 2 different, not equivalent, hypothesis tests, depending on whether we base our primary endpoint on RE or CE. Specifically:

$$\begin{cases} H_0 : \pi_{2,RE} = \pi_{1,RE} \\ H_1 : \pi_{2,RE} \neq \pi_{1,RE} \end{cases} \quad \begin{cases} H_0^* : \pi_{2,CE} = \pi_{1,CE} \\ H_1^* : \pi_{2,CE} \neq \pi_{1,CE} \end{cases}$$

Sample size is calculated using the normal approximation to the binomial test. It can be computed for both the RE and CE as it is presented in equation (6) (27, 28), where π represents the probability of the RE or the probability of the CE as calculated in equation (1), $\pi_{i=1}$ the probability in one group, and $\pi_{i=2}$ the probability in the other group (estimated assuming certain effect). The α and β are the type I and II errors respectively:

$$n \geq \left(\frac{Z_{\alpha/2} \sqrt{2\pi_1(1-\pi_1)} + Z_{\beta} \sqrt{\pi_{i=1}(1-\pi_{i=1}) + \pi_{i=2}(1-\pi_{i=2})}}{\pi_{i=1} - \pi_{i=2}} \right)^2, \quad (6)$$

$$\pi_1 = \frac{\pi_{i=1} + \pi_{i=2}}{2}$$

A CE will be preferred if the number of subjects required (n^*) to detect an effect using the CE is lower than using the single RE (22) (n). We define the sample ratio (SR) as the ratio between sample sizes: $(SR = n/n^*)$. Thus the CE will be preferred in all situations where $SR > 1$.

Finally, given a fixed number of patients, the statistical power (pw) using the CE and the RE can be calculated as follows:

$$pw = 1 - \beta = P\left(X \leq \frac{\sqrt{n}(\pi_{i=1} - \pi_{i=2}) - Z_{\alpha/2} \sqrt{2\pi_1(1-\pi_1)}}{\sqrt{\pi_{i=1}(1-\pi_{i=1}) + \pi_{i=2}(1-\pi_{i=2})}} \right), \quad (7)$$

where X is a standardized normal distribution.

Numerical examples

We simulated several scenarios (Table 1) assuming different values of the event proportion for both the RE and the

Table 1. Prevalence and Range of Effects Simulated for the Relevant and Additional Endpoints in Different Scenarios

P(RE) ^a	RR(RE)	P(AE)	RR(AE)
0.10	0.70 to 0.90	0.10	0.70 to 1.00
0.10	0.70 to 0.90	0.20	0.70 to 1.00
0.20	0.70 to 0.90	0.10	0.70 to 1.00
0.20	0.70 to 0.90	0.20	0.70 to 1.00
0.03	0.70 to 0.90	0.10	0.70 to 1.00
0.03	0.70 to 0.90	0.01	0.70 to 1.00

Abbreviations: P(AE), proportion of additional endpoint; P(RE), proportion of relevant endpoint; RR(AE), relative risk of additional endpoint; RR(RE), relative risk of relevant endpoint.

^a The results of the scenarios with P(RE) = 0.10 are presented in Web Table 1, for P(RE) = 0.20 in Web Table 2, and for P(RE) = 0.03 in Web Table 3.

AE, the effect of therapy on each endpoint, and the degree of overlapping:

Scenario 1: RE with a fixed proportion $\pi_{i=1,1} = 0.10$ and increasing treatment effect (RR_1 from 0.9 to 0.7). An AE with 2 proportions $\pi_{i=1,2} = 0.10$ and $\pi_{i=1,2} = 0.20$ was added, without effect of the therapy on the AE ($RR_2 = 1.0$).

Scenario 2: RE with a fixed proportion $\pi_{i=1,1} = 0.10$ and fixed and low treatment effect ($RR_1 = 0.9$). An AE with 2 proportions $\pi_{i=1,2} = 0.10$ and $\pi_{i=1,2} = 0.20$ was added. Progressively higher therapy effect on AE ($RR_2 = 0.9$ to 0.7).

Scenario 3: RE with a fixed proportion $\pi_{i=1,1} = 0.10$ and fixed and moderate treatment effect ($RR_1 = 0.8$). An AE with 2 proportions $\pi_{i=1,2} = 0.10$ and $\pi_{i=1,2} = 0.20$ was added. Progressively higher treatment effect on AE ($RR_2 = 0.9$ to 0.7).

Scenario 4: RE with a fixed proportion $\pi_{i=1,1} = 0.10$ and fixed and high treatment effect ($RR_1 = 0.7$). An AE with 2 proportions $\pi_{i=1,2} = 0.1$ and $\pi_{i=1,2} = 0.2$ was added. Progressively higher treatment effect ($RR_2 = 0.9$ to 0.7) on AE.

Then, for each scenario, 5 situations of overlapping were considered, from absence (Figure 1A) to maximum overlapping (Figure 1E). The whole range of possible values of relative overlap in each scenario was divided into 5 strata. The Web Material (available at <http://aje.oxfordjournals.org/>) includes all the scenarios showed in Table 1.

For each scenario we calculated sample sizes of the RE and the CE, the SR, and the RR on the CE. The SRs generated in each scenario were plotted to highlight the situations in which the use of CE is more "efficient" than the use of a single RE in terms of sample size. Additionally, we compared the power of using a single RE with that using the CE across different sample sizes and considered different situations regarding the effect of the treatment and the degree of overlapping.

Calculations for the 4 scenarios were replicated, increasing the proportion of the RE to 0.20 (see Web Table 1), decreasing the proportion of the RE to 0.03 (see Web Table 2), and for 2 different scenarios of the proportion of the AE (0.01 and 0.1). The scenarios have been chosen considering a general framework of credible clinical situations in research. Other scenarios could be easily reproduced using the formulas (1–7).

RESULTS

Web Table 3 shows the sample size calculated for the RE and the CE, the SR, and the RR of the CE across all scenarios considered. In scenario 1 (S1.1 – S1.4), adding an AE without effect ($RR = 1$) results in a higher SSR, with all SRs being lower than 1; thus the CE requires a higher sample size than the RE. In addition, the higher the overlap, the lower the effect on the CE and the greater the additional sample size required. In general, an increase in the proportion of the AE leads to either an increase in SSRs when the overlap is very low or low or a reduction when overlap is

moderate or high. Then adding components without effect always results in an increase of the sample size, even in the most favorable situation when the components are disjoint.

In scenario 2 (S2.1 – S2.6), adding an AE with effect results in a lower SSR except for maximum overlap, when for sample size computation a single RE has equivalent effects. Again, the higher the overlap, the lower the SR. However, the impact of overlapping is greater when the proportion of the AE is low (i.e., 0.1). The most favorable situation corresponds to the absence of overlapping, which results in an exponential increase of SRs as long as the RR on the AE increases. In all cases, increasing the AE proportion implies a lower SSR with the CE. The impact of the increase in AE proportion is greater as the effect on the AE increases and the overlap grows.

In scenario 3 (S3.1–S3.6) the effect of the RE was incremented with respect to scenario 2 ($RR = 0.8$ for scenario 3 against $RR = 0.9$ in scenario 2) while maintaining all the other parameters. The comparison of these 2 scenarios suggests that a greater effect on the RE implies lower SSRs, lower SRs, and less marked changes in SRs when increasing the overlap. However, it is to be noted that, when the effect on the AE is lower than the effect on the RE (in both cases the probabilities of events are restricted to $\pi_{i=c,2} = 0.10$ and $\pi_{i=c,2} = 0.20$), overlap can determine a lower (SRs > 1; overlap very low or low) or higher (SRs < 1; overlap moderate to very high) SSR when using the CE. This leads to a higher number of instances—9 instances in scenario 3 and 3 in scenario 2—in which the use of the CE implies the same or greater SSRs than the use of the isolated RE. Thus, the relative overlapping net impact on efficiency may be greater when $RR = 0.9$ (scenario 2) than when $RR = 0.8$ (scenario 3). Also, the impact of the increase of the proportion in AE is greater as the effect on the AE and the overlap increases. In scenario 4, when $RR = 0.7$ (S4.1–S4.6), the effect on the RE is greater than in scenario 3 and scenario 2 ($RR = 0.7$ vs. 0.8 vs. 0.9), and the other conditions are equal.

Similarly, compared with scenario 2 and scenario 3, scenario 4 leads to a lower SSRs, lower SRs, and less marked changes in SRs when increasing overlap. The number of instances in which the use of the CE implies the same or greater SSRs than the use of the isolated RE is greater than in scenario 3 and scenario 2 (15 vs. 9 vs. 3). Therefore, the net impact of overlapping on efficiency may be greater as the effect on the RE increases (scenarios 1 through 4).

In general, in all situations a greater overlap implies higher SSR of the CE (all other conditions being equal). However, the qualitative impact of overlapping is higher when the effect on the AE is lower than the effect on the RE. This can be observed in scenario 3 and scenario 4. In these situations, adding an AE with lower effect than the RE may result in a lower or higher SSR depending on the relative overlap. Globally in 11 out of 22 scenarios explored, the relative overlap magnitude could be determinant for CE use.

In general, the relative benefit expected of adding an AE to an RE, assessed by the SRs, decreases when the proportion of the RE increases (Web Table 2). The exception is when both the proportion and the effect on the AE are greater and the effect on the RE is greater. In cases of low proportion of the RE (Web Table 3), the main impact of the

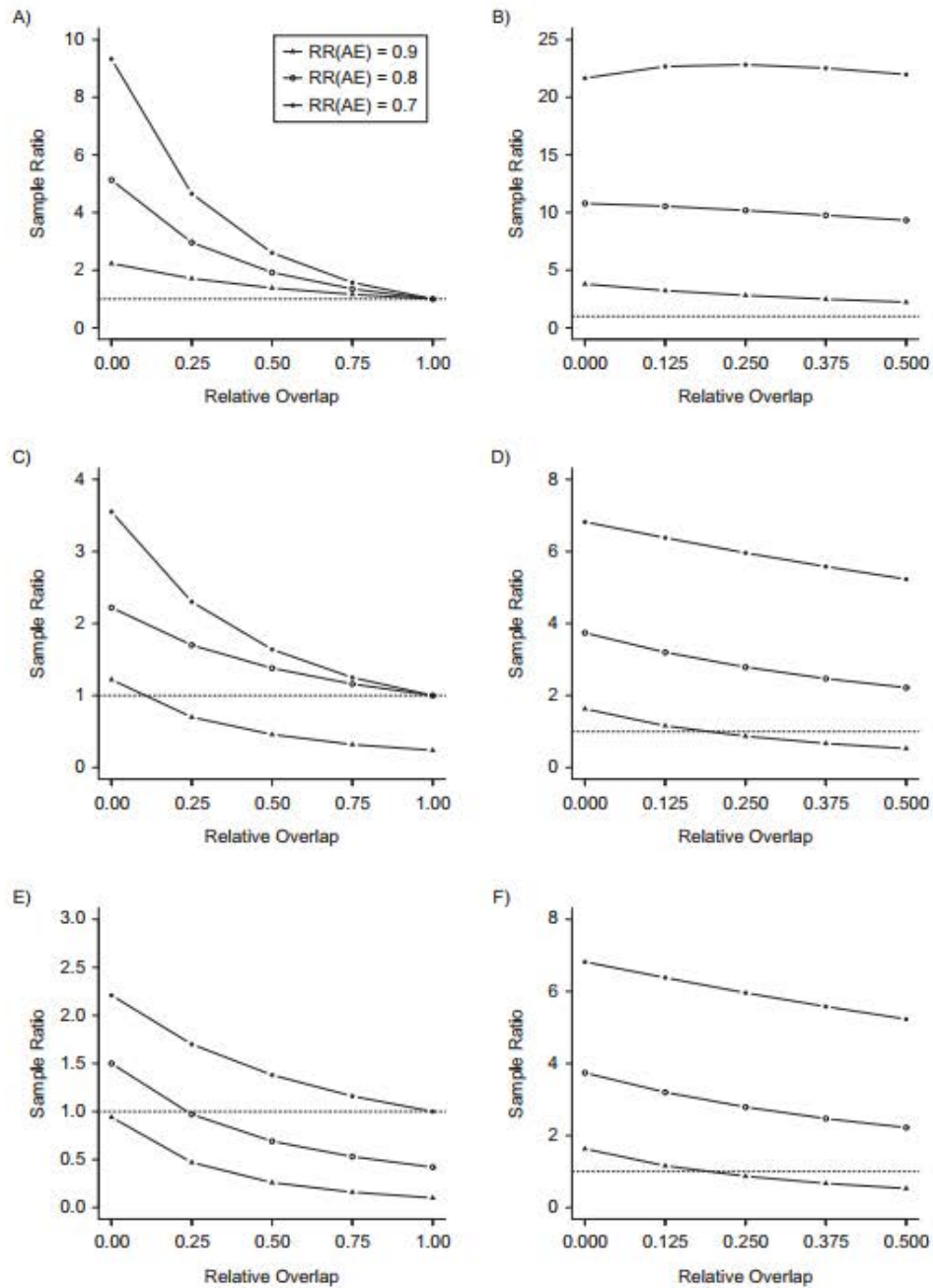


Figure 2. Sample ratio trends in scenarios 2–4. A) $P(AE) = 0.10$, $P(RE) = 0.10$, and $RR(RE) = 0.90$. B) $P(AE) = 0.20$, $P(RE) = 0.10$, and $RR(RE) = 0.90$. C) $P(AE) = 0.10$, $P(RE) = 0.10$, and $RR(RE) = 0.80$. D) $P(AE) = 0.20$, $P(RE) = 0.10$, and $RR(RE) = 0.80$. E) $P(AE) = 0.10$, $P(RE) = 0.10$, and $RR(RE) = 0.70$. F) $P(AE) = 0.20$, $P(RE) = 0.10$, and $RR(RE) = 0.70$. Note that the possible range of relative overlap values (x -axis) depends on the proportions of both the AE and the RE (i.e., lower bound: $\{0, \pi_{11} + \pi_{22} - 1\}$; upper bound: $\min\{\pi_{11}/\pi_{22}, \pi_{22}/\pi_{11}\}$). Dashed black lines denote equal sample size requirement (SSR) using the RE and the composite endpoint (CE); above the dashed black line implies lower SSR using the CE, and below the dashed black line implies higher SSR using the CE. $P(AE)$, proportion of additional endpoint; $P(RE)$, proportion of relevant endpoint; $RR(AE)$, relative risk of additional endpoint; $RR(RE)$, relative risk of relevant endpoint.

CE SSR rests on both the proportion and the effect on the AE. In 25 of the total of 66 scenarios explored, the degree of relative overlap could determine the benefit for SSR of using a CE.

Figure 2 summarizes the behavior of the SR across the spectrum of overlapping in scenarios 2–4 (scenario 1 was not included because all SRs were lower than 1 in all situations). In general, the greater the effect on the RE, the lower the benefit of adding an AE, and the smaller the influence of overlapping. By contrast, the impact of overlapping on SSR is greater with lower proportions of the AE. In general, the more the added AE is prevalent, the greater the benefit of the CE on SSR (only in cases with a $SR > 1$). It must be noted that, as shown above (in Definitions and Assumptions), the possible range of relative overlap values depends on the proportion of both the AE and the RE.

Figures 3 and 4 show the power function, for both the RE (solid line) and the CE (different dotted lines), given different sample size intervals and considering different scenarios concerning the effect on endpoints and the degree of overlapping. As expected, the greater the effect on the AE, the greater the statistical power of the CE, especially when the overlap is small. In some scenarios the statistical power is clearly greater using the CE regardless of the overlap (Figure 4B, 4D, and 4F), whereas in other scenarios it is more efficient to use a single RE (Figures 3E and 4E). However there are situations in which the overlap is determinant to the selection of the single RE or the CE (Figures 3C and 4C).

DISCUSSION

We have addressed the issue of the statistical “efficiency” in sample size and statistical power terms using a binary CE in a conceptual and visual manner. We have employed the SR between the RE and the binary CE that results from adding an AE to evaluate to what extent the proportion of events, the effect on endpoints, and especially the overlap between them determines the SSR using a CE. Examination of different scenarios can be summarized as follows (Table 2):

- A. Should the effect on the AE be superior to the effect on the RE, the use of CE will always be convenient to decrease sample size regardless of the event’s proportion and the relative overlap. Even when adding a low-prevalence AE with similar effect to that on the RE, the CE will be a better solution than the RE. Influence of overlapping will decrease as long as the effect on RE increases.
- B. The greater the degree of overlapping, the lower the potential benefit of the CE to reduce sample size. However, the benefit of combining endpoints with small overlap is variable, depending on other parameters, such as the treatment effect on each endpoint and the event proportion. Occasionally the degree of overlapping will be sufficient to recommend the use of CE.
- C. Adding an AE with lower effect than that on the RE constitutes the most uncertain situation. It could result in either a greater or lower SSR when using the CE, depending on the degree of overlap (i.e., the more, the lower), the effect on the AE (i.e., the greater, the higher),

and the proportion of the AE. In general, increased AE proportion will decrease the SSR using the CE.

The potential benefits of using CE in clinical trials have been widely assessed (1–8). It is well recognized that, among other benefits, CEs can be useful to decrease sample size. Moreover, many authors agree that this could be the main reason for using CEs. However, it is less recognized that, besides other problems, the use of CEs may lead to a statistical power reduction (21, 22). This is obvious when the treatment effect on the AE is in the opposite direction from the effect on the RE. Now we have described other scenarios in which endpoint combination could also lead to increased SSRs even when components have the same direction. Moreover, we have shown situations where, even with substantial treatment effect on the AE, the use of CE could imply a higher SSR to detect the true treatment differences. This is closely related to the degree of overlapping between components.

We have introduced “overlap between endpoints” as a conceptual approach to visualizing how relationships between correlated components of a binary CE may influence statistical power for the CE. However, it is to be noted that overlap between components is neither equivalent to correlation nor the intersection probability. Correlation between endpoints is more related to the conditional probability of having the AE after having had the RE and vice versa, which is in relationship with the overlap but not equivalent. In this case, we used the relative overlap as an approach to quantify to what extent the relationship between components of the CE can affect the power. This ratio is also known as Jaccard’s statistic in ecology and genetics (24, 25), where it is used to assess the similarity between different samples or communities. In the context of CE, we believe that the concept of relative overlap can be more intuitive for clinicians than the concept of correlation and the concept of the intersection probability.

Elements that have an impact on the statistical efficiency of the CE have been well described within time-to-event analysis. Along with the probabilities of observing the event components and the magnitude of treatment effect on each component via the hazard ratio, Spearman’s rank correlation between the times-to-each-component is needed (23). Describing the relationship between the overlap of components of a binary CE and the correlation between them in the context of the time-to-event analysis is beyond the scope of this work and should be addressed in future work.

Schriger et al. (18) have identified that the relationship between endpoints is not usually reported in publications of RCTs. Moreover, sample computations of RCTs do not usually include this parameter. Given its implications for CE construction, an anticipated estimation of the plausible relationship between components would be useful to increase the precision of sample size estimates. As a future line of outcomes research, it would be desirable to quantify the relationship between outcomes in specific populations from real RCT results. For example, knowing the proportion of stroke in patients who had an AMI could be of interest for sample size computations using a CE including these 2 outcomes.

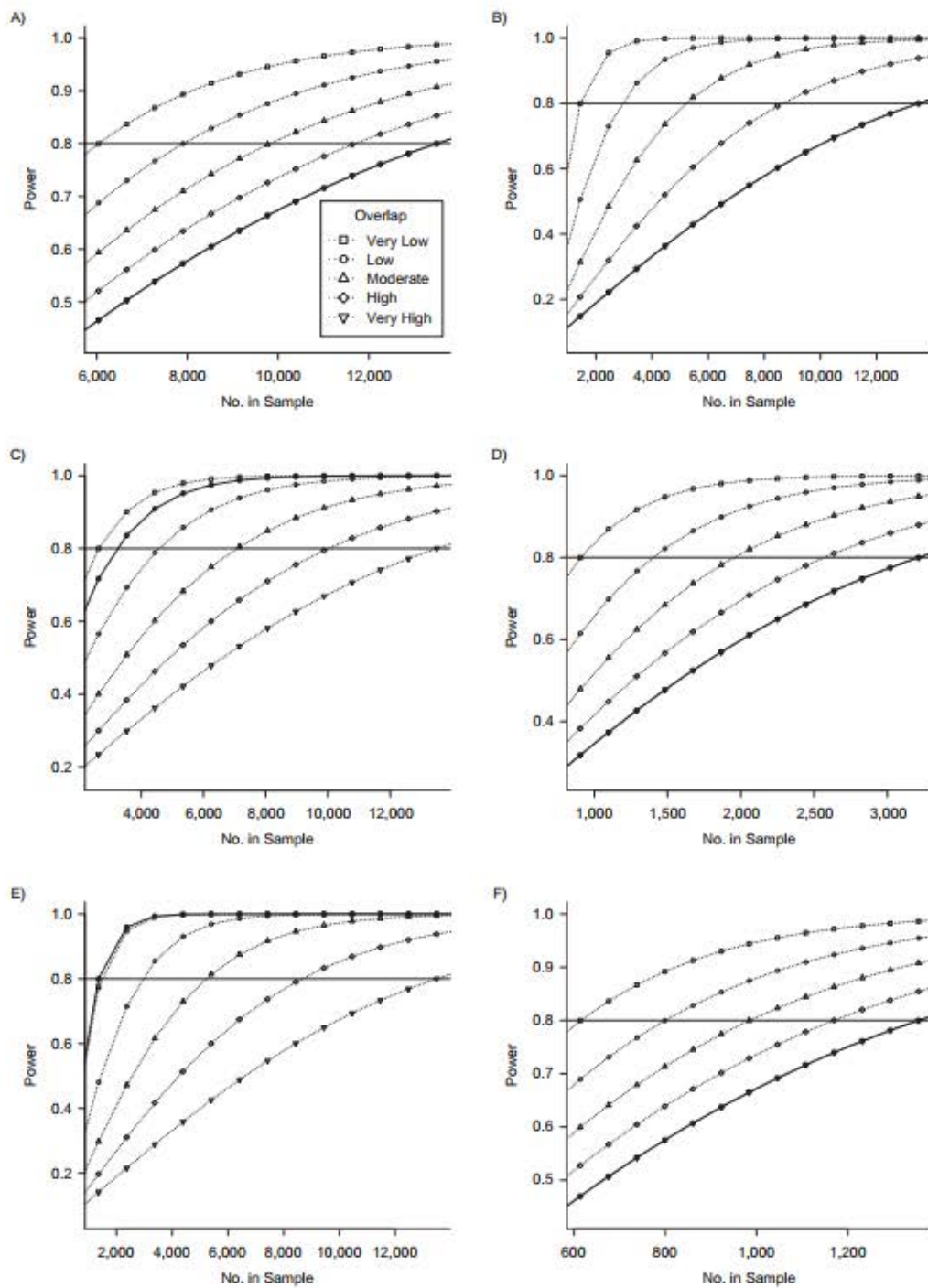


Figure 3. Power curves considering different sample sizes in several scenarios for a $P(\text{RE}) = 0.10$ and a $P(\text{AE}) = 0.10$. A) $\text{RR}(\text{RE}) = 0.90$ and $\text{RR}(\text{AE}) = 0.90$. B) $\text{RR}(\text{RE}) = 0.90$ and $\text{RR}(\text{AE}) = 0.70$. C) $\text{RR}(\text{RE}) = 0.80$ and $\text{RR}(\text{AE}) = 0.90$. D) $\text{RR}(\text{RE}) = 0.80$ and $\text{RR}(\text{AE}) = 0.70$. E) $\text{RR}(\text{RE}) = 0.70$ and $\text{RR}(\text{AE}) = 0.90$. F) $\text{RR}(\text{RE}) = 0.70$ and $\text{RR}(\text{AE}) = 0.70$. Solid black line with closed circles denotes the statistical power curves for the single RE. Dashed lines denote the statistical power curves for the composite endpoint for different relative overlap. Black horizontal lines denote the threshold for a statistical power of 0.80. $P(\text{AE})$, proportion of additional endpoint; $P(\text{RE})$, proportion of relevant endpoint; $\text{RR}(\text{AE})$, relative risk of additional endpoint; $\text{RR}(\text{RE})$, relative risk of relevant endpoint.

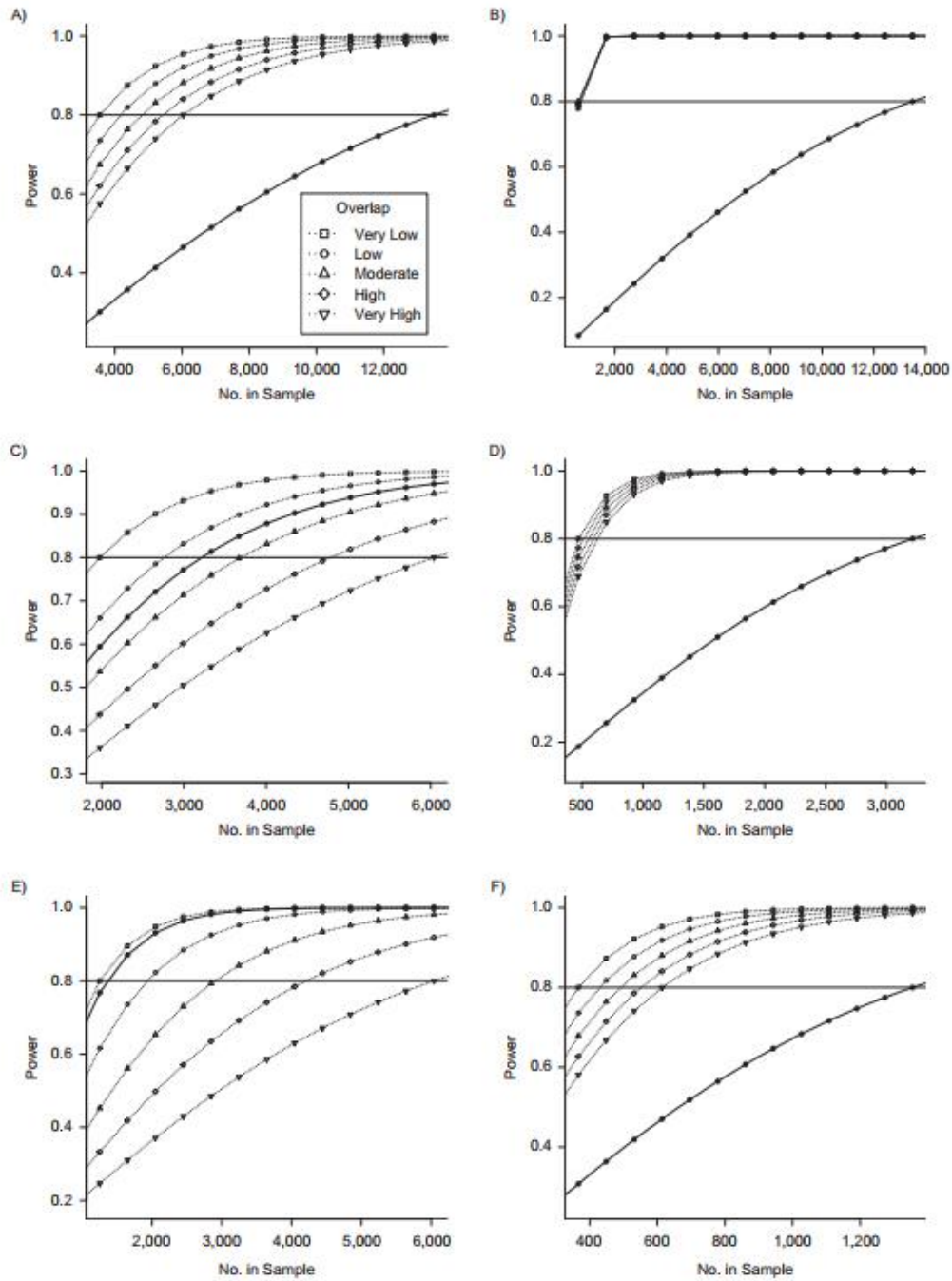


Figure 4. Power curves considering different sample sizes in several scenarios for a $P(RE) = 0.10$ and a $P(AE) = 0.20$. A) $RR(RE) = 0.90$ and $RR(AE) = 0.90$. B) $RR(RE) = 0.90$ and $RR(AE) = 0.70$. C) $RR(RE) = 0.80$ and $RR(AE) = 0.90$. D) $RR(RE) = 0.80$ and $RR(AE) = 0.70$. E) $RR(RE) = 0.70$ and $RR(AE) = 0.90$. F) $RR(RE) = 0.70$ and $RR(AE) = 0.70$. Solid black line with closed circles denotes the statistical power curves for the single RE. Dashed lines denote the statistical power curves for the composite endpoint for different relative overlap. Black horizontal lines denote the threshold for a statistical power of 0.80. $P(AE)$, proportion of additional endpoint; $P(RE)$, proportion of relevant endpoint; $RR(AE)$, relative risk of additional endpoint; $RR(RE)$, relative risk of relevant endpoint.

Table 2. Summary of the Qualitative Impact of Relative Overlapping Between Components on the Sample Size Requirement of the Composite Endpoint Considering Different Scenarios for the Relative Risks and Probabilities of the Components

Scenarios		Qualitative Impact of Overlapping ^a	Sample Size Requirement Using the CE ^b		
P(AE)	RR(AE)		Before Considering Overlap	After Considering Overlap	
				Low Overlap	High Overlap
≤P(RE)	1	++	i	i	ii
≥P(RE)	1	++	i	i	ii
≤P(RE)	<RR(RE)	++	d	dd	d
≤P(RE)	>RR(RE)	+++	i/d	d	i
≥P(RE)	<RR(RE)	+	dd	ddd	dd
≥P(RE)	>RR(RE)	+++	i/d	d	i

Abbreviations: CE, composite endpoint; P(AE), proportion of additional endpoint; P(RE), proportion of relevant endpoint; RR(AE), relative risk of additional endpoint; RR(RE), relative risk of relevant endpoint.

^a The impact of components overlap in scenarios labeled + is lower than in scenarios labeled ++. Impact is highest in scenarios labeled +++.

^b Sample size requirement using the CE in scenarios labeled i is always higher than using only the RE regardless the degree of the overlap. In scenarios labeled ii the increase is higher than scenarios labeled i. Sample size requirement using the CE in scenarios labeled d is always lower than using only the RE regardless the degree of the overlap. In scenarios labeled ddd the decrease is lower than the scenarios labeled dd and lower than scenarios labeled d. Sample size requirement using the CE in scenarios labeled i/d depends on the degree of the overlap. Note: In all scenarios, probability for the CE was assumed to be lower than 0.5.

We have limited our exploration to binary CEs with only 2 components. Quantifying the associations between 3 or more components of a CE and the resulting impact on the SSRs should be addressed in future studies. However, although computationally much more challenging, the essential concepts described here would be valid. In any case, their extension to more complex CEs will be necessary to simulate real-life examples.

We have focused exclusively on the study of the SSR using CEs. CEs may have other additional benefits that could be apparent even if the CE is not efficient. For example, the study of the net benefit of a therapy by a CE including efficacy and safety endpoints could be completely “inefficient,” because the CE may include components with treatment effects in opposite directions. Thus a treatment could reduce the rate of AMI, increasing at the same time the rate of severe hemorrhage events. In this case, computation of sample size based on the net benefit of the therapy (i.e., the rate of the CE AMI or severe hemorrhage) could be inefficient in both sample size and statistical power terms. However, sometimes this approach will be pragmatically essential to depict the net usefulness of a treatment. Even in these extreme cases, the study of the overlap between endpoints will be useful to estimate the SSR.

ACKNOWLEDGMENTS

Author Affiliations: Epidemiology Unit, Department of Cardiology, Vall d'Hebron Hospital, Barcelona, Spain (Josep-Ramon Marsal, Ignacio Ferreira-González, Sandra Bertran, Aida Ribera, Gaietà Permanyer-Miralda, David García-Dorado); Consorcio de Investigación Biomédica en

Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain (Josep-Ramon Marsal, Ignacio Ferreira-González, Sandra Bertran, Aida Ribera, Gaietà Permanyer-Miralda, David García-Dorado); Department of Pediatrics, Obstetrics and Gynecology and Public Health of Universitat Autònoma de Barcelona, Cerdanyola del Vallés, Spain (Josep-Ramon Marsal); Unit of Research Support, Primary Care Research Institute (IDIAP) Jordi Gol, Catalan Institute of Health, Lleida, Spain (Josep-Ramon Marsal); and Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain (Sandra Bertran, Guadalupe Gómez).

G.G. is partially supported by the Ministerio de Economía y Competitividad del Gobierno de España (grants MTM2012-38067-C02-01 and MTM2015-64465_C02_1R) and by the Departament d'Economia i Coneixement de la Generalitat de Catalunya (grant 2014 SGR 464).

Conflict of interest: none declared.

REFERENCES

1. Ferreira-González I, Alonso-Coello P, Solà I, et al. Composite endpoints in clinical trials. *Rev Esp Cardiol*. 2008;61(3):283–290.
2. Chi GY. Some issues with composite endpoints in clinical trials. *Fundam Clin Pharmacol*. 2005;19(6):609–619.
3. Sankoh AJ, D'Agostino RB Sr, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med*. 2003;22(20):3133–3150.
4. Rauch G, Rauch B, Schüler S, et al. Opportunities and challenges of clinical trials in cardiology using composite primary endpoints. *World J Cardiol*. 2015;7(1):1–5.

5. Cannon CP. Clinical perspectives on the use of composite endpoints. *Control Clin Trials*. 1997;18(6):517–529.
6. Cordoba G, Schwartz L, Woloshin S, et al. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ*. 2010;341:c3920.
7. Ferreira-González I, Permyer-Miralda G, Busse JW, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007;60(7):651–657.
8. Holmes DR Jr, Brennan JM, Rumsfeld JS, et al. Clinical outcomes at 1 year following transcatheter aortic valve replacement. *JAMA*. 2015;313(10):1019–1028.
9. Prieto-Merino D, Smeeth L, Staa TP, et al. Dangers of non-specific composite outcome measures in clinical trials. *BMJ*. 2013;347:f6782.
10. European Medicines Agency. *ICH Topic E 9: Statistical Principles for Clinical Trials*. London, UK: European Medicines Agency; 2006: 1–37.
11. Committee For Proprietary Medicinal Products, European Medicines Agency. *Points to Consider on Multiplicity Issues in Clinical Trials*; London, UK: EMEA (European Medicines Agency); 2006: 1–11.
12. Kip KE, Hollabaugh K, Marroquin OC, et al. The problem with composite end points in cardiovascular studies: the story of major adverse cardiac events and percutaneous coronary intervention. *J Am Coll Cardiol*. 2008;51(7):701–707.
13. Ferreira-González I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334(7597):786.
14. Cannon CP, Sharis PJ, Schweiger MJ, et al. Prospective validation of a composite end point in thrombolytic trials of acute myocardial infarction (TIMI 4 and 5). Thrombosis In Myocardial Infarction. *Am J Cardiol*. 1997;80(6):696–699.
15. Lefkopoulou M, Ryan L. Global tests for multiple binary outcomes. *Biometrics*. 1993;49(4):975–988.
16. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984;40(4):1079–1087.
17. Rauch G, Kieser M. Multiplicity adjustment for composite binary endpoints. *Methods Inf Med*. 2012; 51(4):309–317.
18. Schriger DL, Cooper RJ, Lopez-O'Sullivan A, et al. Is the relationship among outcome variables shown in randomized trials? *Trials*. 2015;16(1):57.
19. Sozu T, Sugimoto T, Hamasaki T. Sample size determination in clinical trials with multiple co-primary binary endpoints. *Stat Med*. 2010;29(21):2169–2179.
20. Xiong C, Yu K, Gao F, et al. Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. *Clin Trials*. 2005;2(5):387–393.
21. Gómez G, Lagakos SW. Statistical considerations when using a composite endpoint for comparing treatment groups. *Stat Med*. 2013;32(5):719–738.
22. Gómez G, Gómez-Mateu M. The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses. *SORT*. 2014;38(1):73–88.
23. Gómez G, Gómez-Mateu M, Dafni U. Informed choice of composite end points in cardiovascular trials. *Circ Cardiovasc Qual Outcomes*. 2014;7(1):170–178.
24. Miller JA, Miller JH, Pham DS, et al. Cyberdiversity: improving the informatic value of diverse tropical arthropod inventories. *PLoS One*. 2014;9(12):e115750.
25. Sousa FM, Mendes RM, Mendes RF, et al. Evaluation of genetic similarity between accessions of *Pityrocarpa moniliformis* (angico-de-bezerra) using RAPD markers. *Genet Mol Res*. 2014;13(3):5815–5821.
26. McCormick WP, Lyons NI, Hutcheson K. Distributional properties of Jaccard's index of similarity. *Commun Stat Theory Methods*. 1992;21(1):51–68.
27. Fleiss JL. *Statistical Methods for Rates and Proportions*. Hoboken, NJ: John Wiley & Sons; 1981:33–49.
28. Casagrande JT, Pike MC. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*. 1978;34(3): 483–486.

11. BIBLIOGRAFIA

1. Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Composite endpoints in clinical trials: the trees and the forest. *J Clin Epidemiol*. 2007 Jul;60(7):660-1.
2. Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ*. 2005 Mar 12;330(7491):594-6.
3. Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007 Jul;60(7):651-7; discussion 658-62.
4. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl E a, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334(7597):786.
5. cki. Endpoints used for Relative Effectiveness Assessment Composite endpnts

- [Internet]. [cited 2018 Mar 16]. Available from: https://www.eunethta.eu/wp-content/uploads/2018/03/composite_endpoints.pdf
6. Lauer MS, Topol EJ. Clinical trials - multiple treatments, multiple endpoints, and multiple lessons. *Jama*. 2003;289(19):2575–7.
 7. Ich. ICH HARMONISED TRIPARTITE GUIDELINE STATISTICAL PRINCIPLES FOR CLINICAL TRIALS E9 STATISTICAL PRINCIPLES FOR CLINICAL TRIALS. 1998 [cited 2018 Mar 21]; Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf
 8. Prieto-Merino D, Smeeth L, Staa TP Van, Roberts I. Dangers of non-specific composite outcome measures in clinical trials. *BMJ*. 2013;347(November):f6782.
 9. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* [Internet]. 2010 Mar 23 [cited 2018 Mar 21];340:c869. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20332511>
 10. Hills I. Multiple Endpoints in Clinical Trials Guidance for Industry [Internet]. [cited 2018 Jun 7]. Available from: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>
 11. Medicines Agency E. Guideline on multiplicity issues in clinical trials - for publication [Internet]. 2016 [cited 2018 Jun 7]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224998.pdf
 12. Gómez G, Gómez-Mateu M, Dafni U. Informed choice of composite end points in cardiovascular trials. *Circ Cardiovasc Qual Outcomes*. 2014;7(1):170–8.

13. G G. Statistical Considerations in the Use of a Composite Time-to-Event Endpoint for Comparing Treatment Groups. 2010.
14. Gómez G, Gómez-Mateu M. The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses. *Sort*. 2014;38(1):73–88.
15. Bofill Roig M, Gómez Melis G. Selection of composite binary endpoints in clinical trials. *Biom J* [Internet]. 2017;(July):1–16. Available from:
<http://doi.wiley.com/10.1002/bimj.201600229><http://www.ncbi.nlm.nih.gov/pubmed/29023990>
16. Marsal JR, Ferreira-González I, Bertran S, Ribera A, Permanyer-Miralda G, García-Dorado D, et al. The use of a binary composite endpoint and sample size requirement: Influence of endpoints overlap. *Am J Epidemiol*. 2017;185(9):832–41.
17. Andreasen C, Jørgensen ME, Gislason GH, Martinsson A, Sanders RD, Abdulla J, et al. Association of Timing of Aortic Valve Replacement Surgery After Stroke With Risk of Recurrent Stroke and Mortality. *JAMA Cardiol* [Internet]. 2018 Jun 1 [cited 2018 Sep 19];3(6):506. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/29710128>
18. Andersson C, Mérie C, Jørgensen M, Gislason GH, Torp-Pedersen C, Overgaard C, et al. Association of β -Blocker Therapy With Risks of Adverse Cardiovascular Events and Deaths in Patients With Ischemic Heart Disease Undergoing Noncardiac Surgery. *JAMA Intern Med* [Internet]. 2014 Mar 1 [cited 2018 Sep 19];174(3):336. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24247428>
19. Komoto A, Maiguma T, Teshima D, Sugiyama T, Haruki Y. Effects of pharmacist intervention in Vancomycin treatment for patients with bacteremia due to Methicillin-resistant *Staphylococcus aureus*. Smith TC, editor. *PLoS One* [Internet]. 2018 Sep 6 [cited 2018 Sep 19];13(9):e0203453. Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/30188918>

20. Parshuram CS, Dryden-Palmer K, Farrell C, Gottesman R, Gray M, Hutchison JS, et al. Effect of a Pediatric Early Warning System on All-Cause Mortality in Hospitalized Pediatric Patients. *JAMA* [Internet]. 2018 Mar 13 [cited 2018 Sep 19];319(10):1002. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29486493>
21. Schoeps A, Nebié E, Fisker AB, Sié A, Zakane A, Müller O, et al. No effect of an additional early dose of measles vaccine on hospitalization or mortality in children: A randomized controlled trial. *Vaccine* [Internet]. 2018 Apr 5 [cited 2018 Sep 19];36(15):1965–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29523450>
22. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, et al. Balanced Crystalloids versus Saline in Critically Ill Adults. *N Engl J Med* [Internet]. 2018 Mar 1 [cited 2018 Sep 19];378(9):829–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29485925>
23. Ferreira-González I, Alonso-Coello P, Solà I, Pacheco-Huergo V, Domingo-Salvany A, Alonso J, et al. [Composite endpoints in clinical trials]. *Rev española Cardiol*. 2008 Mar;61(3):283–90.
24. Freemantle N, Calvert MJ. Interpreting composite outcomes in trials. *BMJ* [Internet]. 2010 Aug 18 [cited 2018 Mar 19];341(aug18 3):c3529–c3529. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.c3529>
25. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* [Internet]. 2003 May 21 [cited 2012 Jul 16];289(19):2554–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12759327>
26. Caverly TJ, Matlock DD, Prochazka A V, Lucas BP, Hayward RA. Interpreting Clinical

- Trial Outcomes for Optimal Patient Care: A Survey of Clinicians and Trainees. *J Grad Med Educ* [Internet]. 2016 Feb [cited 2016 Feb 29];8(1):57–62. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26913104>
27. Rauch G, Rauch B, Schüler S, Kieser M. Opportunities and challenges of clinical trials in cardiology using composite primary endpoints. *World J Cardiol* [Internet]. 2015 Jan 26 [cited 2015 Aug 20];7(1):1–5. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4306200&tool=pmcentrez&rendertype=abstract>
 28. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key Issues in End Point Selection for Heart Failure Trials: Composite End Points. [cited 2018 Mar 21]; Available from: [http://www.onlinejcf.com/article/S1071-9164\(05\)01180-2/pdf](http://www.onlinejcf.com/article/S1071-9164(05)01180-2/pdf)
 29. Lubsen J, Kirwan B-A. Combined endpoints: can we use them? *Stat Med* [Internet]. 2002 Oct 15 [cited 2018 Aug 3];21(19):2959–70. Available from: <http://doi.wiley.com/10.1002/sim.1300>
 30. European Medicines Agency Committee For Proprietary Medicinal Products (CPMP). Points to consider on multiplicity issues in clinical trials. 2002. p. 0–10.
 31. Lubsen J, Kirwan B-A. Combined endpoints: can we use them? *Stat Med* [Internet]. 2002 Oct 15 [cited 2018 Mar 21];21(19):2959–70. Available from: <http://doi.wiley.com/10.1002/sim.1300>
 32. Sozu T, Sugimoto T, Hamasaki T. Sample size determination in clinical trials with multiple co-primary binary endpoints. *Stat Med*. 2010;29(21):2169–79.
 33. Chi GYH. Some issues with composite endpoints in clinical trials. *Fundam Clin Pharmacol*. 2005 Dec;19(6):609–19.
 34. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol* [Internet]. 2007 Feb [cited 2018 Mar

- 19];196(2):119.e1-119.e6. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/S0002937806022186>
35. Cannon CP. Clinical perspectives on the use of composite endpoints. *Control Clin Trials*. 1997 Dec;18(6):517-29; discussion 546-9.
 36. Xiong C, Yu K, Gao F, Yan Y, Zhang Z. Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. *Clin Trials*. 2005 Jan;2(5):387-93.
 37. Wittkop L, Smith C, Fox Z, Sabin C, Richert L, Aboulker J-P, et al. Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clin Trials J Soc Clin Trials* [Internet]. 2010 Feb 15 [cited 2018 Mar 19];7(1):19-35. Available from:
<http://journals.sagepub.com/doi/10.1177/1740774509356117>
 38. Pocock SJ, Ariti C a., Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* [Internet]. 2012 Jan [cited 2012 Dec 6];33(2):176-82. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/21900289>
 39. Pocock SJ. Clinical trials with multiple outcomes: A statistical perspective on their design, analysis, and interpretation. *Control Clin Trials* [Internet]. 1997 Dec [cited 2018 Mar 21];18(6):530-45. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/S0197245697000081>
 40. Sankoh AJ, D'Agostino RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. *Stat Med*. 2003;22(20):3133-50.
 41. Hallstrom AP, Litwin PE, Weaver WD. A method of assigning scores to the components of a composite outcome: an example from the MITI trial. *Control Clin*

- Trials [Internet]. 1992 Apr [cited 2018 Mar 21];13(2):148–55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1316829>
42. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ*. 2010;341:c3920.
 43. Ewp. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP) GUIDELINE ON THE EVALUATION OF MEDICINAL PRODUCTS FOR CARDIOVASCULAR DISEASE PREVENTION DRAFT AGREED BY EFFICACY WORKING PARTY ADOPTION BY CHMP FOR RELEASE FOR CONSULTATION GUIDELINE ON THE EVALUATION OF MEDICINAL PRODUCTS FOR CARDIOVASCULAR DISEASE PREVENTION. 2007 [cited 2018 Mar 20]; Available from: <http://www.emea.europa.eu>
 44. Kip KE, Hollabaugh K, Marroquin OC, Williams DO. The problem with composite end points in cardiovascular studies: the story of major adverse cardiac events and percutaneous coronary intervention. *J Am Coll Cardiol*. 2008 Feb 19;51(7):701–7.
 45. Bhardwaj SS, Camacho F, Derrow A, Fleischer AB, Feldman SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. *Arch Dermatol* [Internet]. 2004 Dec 1 [cited 2018 Aug 3];140(12):1520–3. Available from: <http://archderm.jamanetwork.com/article.aspx?doi=10.1001/archderm.140.12.1520>
 46. Casagrande JT, Pike MC. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*. 1978;34(3):483–6.
 47. Collins GS, de Groot J a, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of

- methodological conduct and reporting. *BMC Med Res Methodol.* 2014;14(1).
48. Gómez G, Lagakos SW. Statistical considerations when using a composite endpoint for comparing treatment groups. *Stat Med.* 2012 Jul 1;32(5):719–38.
 49. Gómez G, Plana-Ripoll O, Dafni U. Selection of the primary end point in an observational cohort study. *J Epidemiol Community Health [Internet].* 2016 Oct [cited 2017 Apr 25];70(10):950–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27072869>
 50. DeMets DL, Califf RM. Lessons learned from recent cardiovascular clinical trials: Part I. *Circulation [Internet].* 2002 Aug 6 [cited 2018 Aug 3];106(6):746–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12163438>
 51. Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. *Clin Cancer Res [Internet].* 2012 Apr 15 [cited 2012 Nov 26];18(8):2301–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22282466>
 52. Prentice RL, Kalbfleisch JD, Peterson a V, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics.* 1978 Dec;34(4):541–54.
 53. Rauch G, Kieser M. Multiplicity adjustment for composite binary endpoints. *Methods Inf Med.* 2012;51(4):309–17.
 54. Pogue J, Thabane L, Devereaux PJ, Yusuf S. Testing for heterogeneity among the components of a binary composite outcome in a clinical trial. *BMC Med Res Methodol.* 2010 Jan;10:49.
 55. Wu L, Cook R. *Statistical Issues in the Use of Composite Endpoints in Clinical Trials A Global Analysis : A Multivariate Approach.* 2010.

56. Bakal J a., Roe MT, Ohman EM, Goodman SG, Fox K a. a., Zheng Y, et al. Applying novel methods to assess clinical outcomes: insights from the TRILOGY ACS trial. *Eur Heart J* [Internet]. 2014;36(6):385–92. Available from:
<http://eurheartj.oxfordjournals.org/cgi/doi/10.1093/eurheartj/ehu262>
57. Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. *Control Clin Trials* [Internet]. 1997 Jun [cited 2012 Sep 21];18(3):204–21. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/9204221>
58. Myerburg RJ, Castellanos A. Language and interpretation of clinical trial outcomes: alternates, surrogates, and composites. *Hear Rhythm* [Internet]. 2004 Nov [cited 2018 Aug 3];1(5):538–9. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/S1547527104004151>
59. Schriger DL, Cooper RJ, Lopez-O’Sullivan A, Wystrach C, Altman DG. Is the relationship among outcome variables shown in randomized trials? *Trials*. 2015;16(1):57.
60. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure. *N Engl J Med* [Internet]. 2014;371(11):993–1004. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/25176015>
<http://www.nejm.org/doi/10.1056/NEJMoa1409077>
61. Bernat I, Horak D, Stasek J, Mates M, Pesek J, Ostadal P, et al. ST-segment elevation myocardial infarction treated by radial or femoral approach in a multicenter randomized clinical trial: The STEMI-RADIAL trial. *J Am Coll Cardiol*. 2014;63(10):964–72.
62. Bueno H, Armstrong PW, Buxton MJ, Danchin N, Lubsen J, Roland E, et al. The future

- of clinical trials in secondary prevention after acute coronary syndromes. *Eur Heart J* [Internet]. 2011 Jul [cited 2013 Jan 9];32(13):1583–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21037255>
63. Finn RS, Martin M, Rugo HS, Jones S, Im S-A, Gelmon K, et al. Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* [Internet]. 2016;375(20):1925–36. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa1607303>
64. Foukakis T, Von Minckwitz G, Bengtsson NO, Brandberg Y, Wallberg B, Fornander T, et al. Effect of tailored dose-dense chemotherapy vs standard 3-weekly adjuvant chemotherapy on recurrence-free survival among women with high-risk early breast cancer: A randomized clinical trial. *JAMA - J Am Med Assoc*. 2016;316(18):1888–96.
65. Gillinov AM, Bagiella E, Moskowitz AJ, Raiten JM, Groh MA, Bowdish ME, et al. Rate Control versus Rhythm Control for Atrial Fibrillation after Cardiac Surgery. *N Engl J Med* [Internet]. 2016;374(20):1911–21. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa1602002>
66. Talan DA, Mower WR, Krishnadasan A, Abrahamian FM, Lovecchio F, Karras DJ, et al. Trimethoprim–Sulfamethoxazole versus Placebo for Uncomplicated Skin Abscess. *N Engl J Med* [Internet]. 2016;374(9):823–32. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa1507476>
67. Kitzman DW, Brubaker P, Morgan T, Haykowsky M, Hundley G, Kraus WE, et al. Effect of caloric restriction or aerobic exercise training on peak oxygen consumption and quality of life in obese older patients with heart failure with preserved ejection fraction: A randomized clinical trial. *JAMA - J Am Med Assoc*. 2016;315(1):36–46.
68. Opiyo N, Shepperd S, Musila N, Allen E, Nyamai R, Fretheim A, et al. Comparison of

- alternative evidence summary and presentation formats in clinical guideline development: a mixed-method study. *PLoS One* [Internet]. 2013 Jan [cited 2013 Feb 5];8(1):e55067. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3555827&tool=pmcentrez&rendertype=abstract>
69. Laitio R, Hynninen M, Arola O, Virtanen S, Parkkola R, Saunavaara J, et al. Effect of Inhaled Xenon on Cerebral White Matter Damage in Comatose Survivors of Out-of-Hospital Cardiac Arrest. *Jama* [Internet]. 2016;315(11):1120. Available from:
<http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2016.1933>
70. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): Effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* [Internet]. 2016;387(10013):53–60. Available from:
[http://dx.doi.org/10.1016/S0140-6736\(15\)00056-2](http://dx.doi.org/10.1016/S0140-6736(15)00056-2)
71. Hammel P, Huguet F, Van Laethem JL, Goldstein D, Glimelius B, Artru P, et al. Effect of chemoradiotherapy vs chemotherapy on survival in patients with locally advanced pancreatic cancer controlled after 4 months of gemcitabine with or without erlotinib the LAP07 randomized clinical trial. *JAMA - J Am Med Assoc*. 2016;315(17):1844–53.
72. Ribrag V, Koscielny S, Bosq J, Leguay T, Casasnovas O, Fornecker LM, et al. Rituximab and dose-dense chemotherapy for adults with Burkitt's lymphoma: a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2016;387(10036):2402–11.
73. Dake MD, Ansel GM, Jaff MR, Ohki T, Saxon RR, Smouse HB, et al. Durable Clinical Effectiveness with Paclitaxel-Eluting Stents in the Femoropopliteal Artery: 5-Year Results of the Zilver PTX Randomized Trial. *Circulation*. 2016;133(15):1472–83.

74. Moineddin R, Meaney C, Grunfeld E. On the analysis of composite measures of quality in medical research. *Stat Methods Med Res* [Internet]. 2017;26(2):633–60. Available from: <http://journals.sagepub.com/doi/10.1177/0962280214553330>
75. Pogue J, Devereaux PJ, Thabane L, Yusuf S. Designing and analyzing clinical trials with composite outcomes: consideration of possible treatment differences between the individual outcomes. *PLoS One* [Internet]. 2012 [cited 2018 Aug 3];7(4):e34785. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22529934>
76. Ich C. ICH Topic E 9: Statistical Principles for Clinical Trials. EMEA (European Medicines Agency), London (UK); 2006.
77. Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. *Control Clin Trials* [Internet]. 1997 Jun [cited 2012 Sep 21];18(3):204–21. Available from: [http://dx.doi.org/10.1016/S0197-2456\(96\)00129-8](http://dx.doi.org/10.1016/S0197-2456(96)00129-8)
78. Prentice RL. Discussion of Correlation Induced by Covariate Measurement Errors. 2015;81(394):321–7.
79. Mascha EJ, Imrey PB. Factors affecting power of tests for multiple binary outcomes. *Stat Med*. 2010;29(28):2890–904.
80. Pearson K. On the coefficient of racial likeness. *Biometrika*. 1926;9:105–7.
81. Pearson K, Heron D. On theories of association. *Biometrika*. 1913;(9):159–315.
82. Bahadur RR. A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction, Vol. VI, Stanford Mathematical Studies in the Social Sciences*. Solomon H, editor. *Studies in Item Analysis and Prediction*. Stanford University Press: Stanford, CA; 1961. 158-168.
83. Lefkopoulou M, Ryan L. Global tests for multiple binary outcomes. *Biometrics*.

- 1993;49(4):975–88.
84. George 's S. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. [cited 2018 Jun 15]; Available from:
<https://www.bmj.com/content/bmj/348/bmj.f7450.full.pdf>
 85. Warrens MJ. On Association Coefficients for 2x2 Tables and Properties That Do Not Depend on the Marginal Distributions. *Psychometrika*. 2008 Dec;73(4):777–89.
 86. Warrens MJ. Chance-corrected measures for 2×2 tables that coincide with weighted kappa. *Br J Math Stat Psychol*. 2011 May;64(Pt 2):355–65.
 87. Warrens MJ. Similarity Coefficients for Binary Data [Internet]. 2008. 253 p.
Available from:
<https://openaccess.leidenuniv.nl/bitstream/handle/1887/12987/Full?sequence=2>
 88. Joag-Dev K, Proschan F. Negative Association of Random Variables with Applications. *Ann Stat* [Internet]. 1983 Mar [cited 2016 Nov 16];11(1):286–95.
Available from: <http://projecteuclid.org/euclid.aos/1176346079>
 89. Esary JD, Proschan F, Walkup DW. Association of Random Variables, with Applications. *Ann Math Stat*. 1967;38(5):1466–74.
 90. Legler JM, Lefkopoulou M, Ryan LM. Efficiency and Power of Tests For Multiple Binary Outcomes. *J Am Stat Assoc*. 1995;90(430):680–93.
 91. Fréchet M. Sur les tableaux de corrélation dont les marges et des bornes sont données. *Rev Inst Stat*. 1960;28:10–32.
 92. Fleiss JL. *Statistical Methods for Rates and Proportions*. Sons JW& S, editor. John Wiley & Sons; 1981. 33-49 p.
 93. Marrugat J, Vila J, Pavesi M, Sanz F. [Estimation of the sample size in clinical and epidemiological investigations]. *Med Clin (Barc)*. 1998 Sep 12;111(7):267–76.

94. Yates F. Contingency Tables Involving Small Numbers and the χ^2 Test. *Suppl to J R Stat Soc.* 1934;1:217–35.
95. Marsal J-R, Ferreira-González I, Bertran S, Ribera A, Permanyer-Miralda G, García-Dorado D, et al. The Use of a Binary Composite Endpoint and Sample Size Requirement: Influence of Endpoints Overlap. *Am J Epidemiol.* 2017 May 1;185(9):832–41.
96. Joag-Dev K, Proschan F. Negative Association of Random Variables with Applications. *Ann Stat.* 1983 Mar;11(1):286–95.
97. Warrens MJ. On Similarity Coefficients for 2x2 Tables and Correction for Chance. *Psychometrika.* 2008 Sep;73(3):487–502.
98. Sacks FM, Pfeffer MA, Moye LA, Rouleau JL, Rutherford JD, Cole TG, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. *N Engl J Med.* 1996 Oct 3;335(14):1001–9.
99. Todeschini R, Consonni V, Xiang H, Holliday J, Buscema M, Willett P. Similarity Coefficients for Binary Chemoinformatics Data: Overview and Extended Comparison Using Simulated and Real Data Sets. *J Chem Inf Model [Internet].* 2012 Nov 26 [cited 2016 Nov 16];52(11):2884–901. Available from: <http://pubs.acs.org/doi/abs/10.1021/ci300261r>
100. Yule G. On the methods of measuring the association between two attributes. *J Roy Stat Soc.* 1912;75:579–652.
101. Russell P, Rao T. On habitat and association of species of anophelinae larvae in south-eastern Madras. *J Malar Inst India.* 1940;3:153–78.
102. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull la Société Vaudoise des Sci Nat.* 1901;37:547–79.

103. Jaccard P. The distribution of the flora in the alpine zone. *New Phytol.* 1912;11:37–50.
104. McCormick WP, Lyons NI, Hutcheson K. Distributional properties of jaccard's index of similarity. *Commun Stat - Theory Methods.* 1992;21(1):51–68.
105. Simpson GG. Mammals and the nature of continents. *Am J Sci.* 1943;241:1–31.
106. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJM. Using networks to measure similarity between genes: association index selection. *Nat Methods.* 2013 Dec;10(12):1169–76.
107. Ochiai A. Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions. *Bull Japan Soc Sci Fish.* 1957;22(9):526–30.
108. Barkman J. *Phytosociology and ecology of cryptogamic epiphytes : including a taxonomic survey and description of their vegetation units in Europe.* Assen. Van Gorcum; 1958. 628 p.
109. Sokal R, Michener C. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull.* 1958;28:1409–38.
110. Czekanowski J. *Zarys metod statystycznych w zastosowaniu do antropologii.* Warszawskie TN, editor. Warszawa; 1913.
111. Dice L. Measures of the Amount of Ecologic Association Between Species. *Ecology.* 1945;26(3):297–302.
112. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *K Danske Vidensk Selsk.* 1948;5(4):1–34.
113. Kulczynski S. Die Pflanzenassoziationen der Pieninen. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres C des SM et N,* editor. 1927. 57-203

p.

114. Cannon CP, Sharis PJ, Schweiger MJ, McCabe CH, Diver DJ, Shah PK, et al. Prospective validation of a composite end point in thrombolytic trials of acute myocardial infarction (TIMI 4 and 5). *Thrombosis In Myocardial Infarction. Am J Cardiol.* 1997 Sep 15;80(6):696–9.
115. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics.* 1984;40(4):1079–87.
116. Home - RStudio [Internet]. [cited 2018 Sep 26]. Available from: <https://www.rstudio.com/>
117. Marsal J-R, Ferreira-González I, Bertran S, Ribera A, Permanyer-Miralda G, García-Dorado D, et al. The use of a binary composite endpoint and sample size requirement: Influence of endpoints overlap. *Am J Epidemiol.* 2017;185(9).
118. Shiny [Internet]. [cited 2018 Sep 26]. Available from: <https://shiny.rstudio.com/>
119. R: The R Project for Statistical Computing [Internet]. [cited 2018 Jun 7]. Available from: <https://www.r-project.org/>
120. Balasubramanian H, Ananthan A, Rao S, Patole S. Odds ratio vs risk ratio in randomized controlled trials. *Postgrad Med [Internet].* 2015 Jul 4 [cited 2018 Mar 6];127(4):359–67. Available from: <http://www.tandfonline.com/doi/full/10.1080/00325481.2015.1022494>
121. Rauch G, Jahn-Eimermacher A, Brannath W, Kieser M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat Med [Internet].* 2014 Mar 30 [cited 2015 Aug 20];33(7):1104–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24122841>
122. Rauch G, Kieser M. An expected power approach for the assessment of composite

- endpoints and their components. *Comput Stat Data Anal* [Internet]. 2013 Apr [cited 2018 Oct 10];60:111–22. Available from:
<http://linkinghub.elsevier.com/retrieve/pii/S0167947312003921>
123. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. *Lancet* [Internet]. 2005 [cited 2012 Sep 21];365(9470):1591–5. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/15866314>
124. Oman SD. Easily simulated multivariate binary distributions with given positive and negative correlations. *Comput Stat Data Anal*. 2009;53(4):999–1005.
125. Oman SD. Modelling and generating correlated binary variables. *Biometrika* [Internet]. 2001;88(1):287–90. Available from:
<http://biomet.oupjournals.org/cgi/doi/10.1093/biomet/88.1.287%5Cnpapers3://publication/doi/10.1093/biomet/88.1.287>
126. Peirce CS. The numerical measure of the success of predictions. *Science*. 1884 Nov 14;4(93):453–4.
127. Doolittle MH. The verification of predictions. *Bull Philos Soc Washingt*. 1885;7:122–7.
128. Yule GU. On the association of attributes in statistics. *Philos Trans R Soc A*. 1900;75:257–319.
129. Montgomery AC, Crittenden KS. Improving coding reliability for open-ended questions. *Public Opin Q*. 1977;41:235–43.
130. Yule GU. *An introduction to the theory of statistics*. C. Griffin and company, limited; 1932. 434 p.
131. Forbes SA. On the local distribution of certain Illinois fishes: An essay in statistical ecology. *Bull Illinois State Lab Nat Hist*. 1907;(7):273–303.

132. Gleason HA. Some applications of the quadrat method. Bull Torrey Bot Club. 1920;47:21–33.
133. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 1979 Oct;76(10):5269–73.
134. Michael EL. Marine ecology and the coefficient of association. J Anim Ecol. 1920;8:54–9.
135. Driver HE, Kroeber AL. Quantitative expression of cultural relationship. Univ Calif Publ Am Archaeol Ethnol. 1932;31:211–56.
136. Braun-Blanquet J. Plant Sociology: The Study of Plant Communities. Authorized English translation of Pflanzensoziologie. McGraw-Hil. New York; 1932.
137. Fowlkes EB, Mallow CL. A method for comparing two hierarchical clusterings. J Am Stat Assoc. 1983;78(553–569).
138. Kuder GF, Richardson MW. The theory of estimation of test reliability. Psychometrika. 1937;2:151–60.
139. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
140. Wallace DL. A method for comparing two hierarchical clusterings: Comment. J Am Stat Assoc. 1983;78:569–76.
141. Post WJ, Snijders TAB. Nonparametric unfolding models for dichotomous data. Methodika,. 1993;7:130–56.
142. Loevinger JA. A systematic approach to the construction and evaluation of tests of ability. Psychometrika. 1947;Monograph.
143. Loevinger JA. The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. Psychol Bull. 1948;45:507–30.

144. Mokken RJ. A Theory and Procedure of Scale Analysis. The Netherlands: Mouton: The Hague; 1971.
145. Sijtsma K, Molenaar IJ. Introduction to Nonparametric Item Response Theory. Thousand Oaks;; 2002.
146. Cole LC. The Measurement of Interspecific Associaton. Ecology. 1949;30(4):411-24.
147. Goodman LA, Kruscal WH. Measures of association for cross classifications. J Am Stat Assoc. 1954;49:732-64.
148. Scott W. Reliability of content analysis: The case of nominal scale coding. Public Opin Q. 1955;
149. Rand W. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc. 1971;66:846-50.
150. Brennan R, Light R. Measuring agreement when two observers classify people into categories not defined in advance. Br J Math 1974;27:154-63.
151. Sorgenfrei T. Molluscan Assemblages From the Marine Middle Miocene of South Jutland and Their Environments. Copenhagen: Reitzel; 1958.
152. Cheetham A, Hazel J. Binary (presence-absence) similarity coefficients. J Paleontol. 1969;(43):1130-6.
153. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46.
154. Rogers DJ, Tanimoto TT. A Computer Program for Classifying Plants. Science. 1960;132(3434):1115-8.
155. Farkas GM. Correction for bias present in a method of calculating interobserver agreement. J Appl Behav Anal. 1978;11:188.

156. Stiles HE. The association factor in information retrieval. *J Assoc Comput Mach.* 1961;8:271-9.
157. Hamann U. Merkmalsbestand und Verwandtschaftsbeziehungen der Farinose. Ein Betrag zum System der Monokotyledonen. *Willdenowia.* 1961;2:639-768.
158. Holley JW, Guilford JP. A note on the G-index of agreement. *Educ Psychol Meas.* 1964;24:749-53.
159. Hubert LJ. Nominal scale response agreement as a generalized correlation. *Br J Math Stat Psychol.* 1977;30:98-103.
160. Mountford MD. An index of similarity and its applications to classificatory problems. In P. W. M. London: Butterworths: *Progress in Soil Zoology*; 1962. 43-50 p.
161. Fager EW, McGowan JA. Zooplankton Species Groups in the North Pacific: Co-occurrences of species can be used to derive groups whose members react similarly to water-mass types. *Science.* 1963;140(3566):453-60.
162. Sokal RR, Sneath PHA. *Principles of Numerical Taxonomy.* W. H. Freeman and Company, editor. San Francisco; 1963. 129 p.
163. McConnaughey BH. The determination and analysis of plankton communities. *Mar Res.* 1964;Special No:1-40.
164. Rogot E, Goldbert ID. A proposed index for measuring agreement in test-retest studies. *J Chronic Dis.* 1966;19:991-1006.
165. Johnson SC. Hierarchical clustering schemes. *Psychometrika.* 1967;32(3):241-54.
166. Hawkins RP, Dotson VA. Reliability scores that delude: An Alice in Wonderland trip through the misleading characteristics of interobserver agreement scores in interval coding. In: *Behavior Analysis: Areas of Research and Application.* In E.

- Ramp. Englewood Cliffs: Prentice-Hall; 1968.
167. Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*. 1975 Sep;31(3):651-9.
 168. Clement PW. A formula for computing inter-observer agreement. *Psychol Rep*. 1976;39:257-8.
 169. Baroni-Urbani C, Buser MW. Similarity of binary data. *Syst Zool*. 1976;25:251-9.
 170. Kent RN, Fosterl SL. Direct observational procedures: Methodological issues in naturalistic settings. In: *Handbook of Behavioral Assessment*. New York, NY: John Wiley & Sons; 1977. p. 279-328.
 171. Harris FC, Lahey BB. A method for combining occurrence and nonoccurrence interobserver agreement scores. *J Appl Behav Anal*. 1978 Jan;11(4):523-7.
 172. Digby PG. Approximating the tetrachoric correlation coefficient. *Biometrics*. 1983;39:753-7.