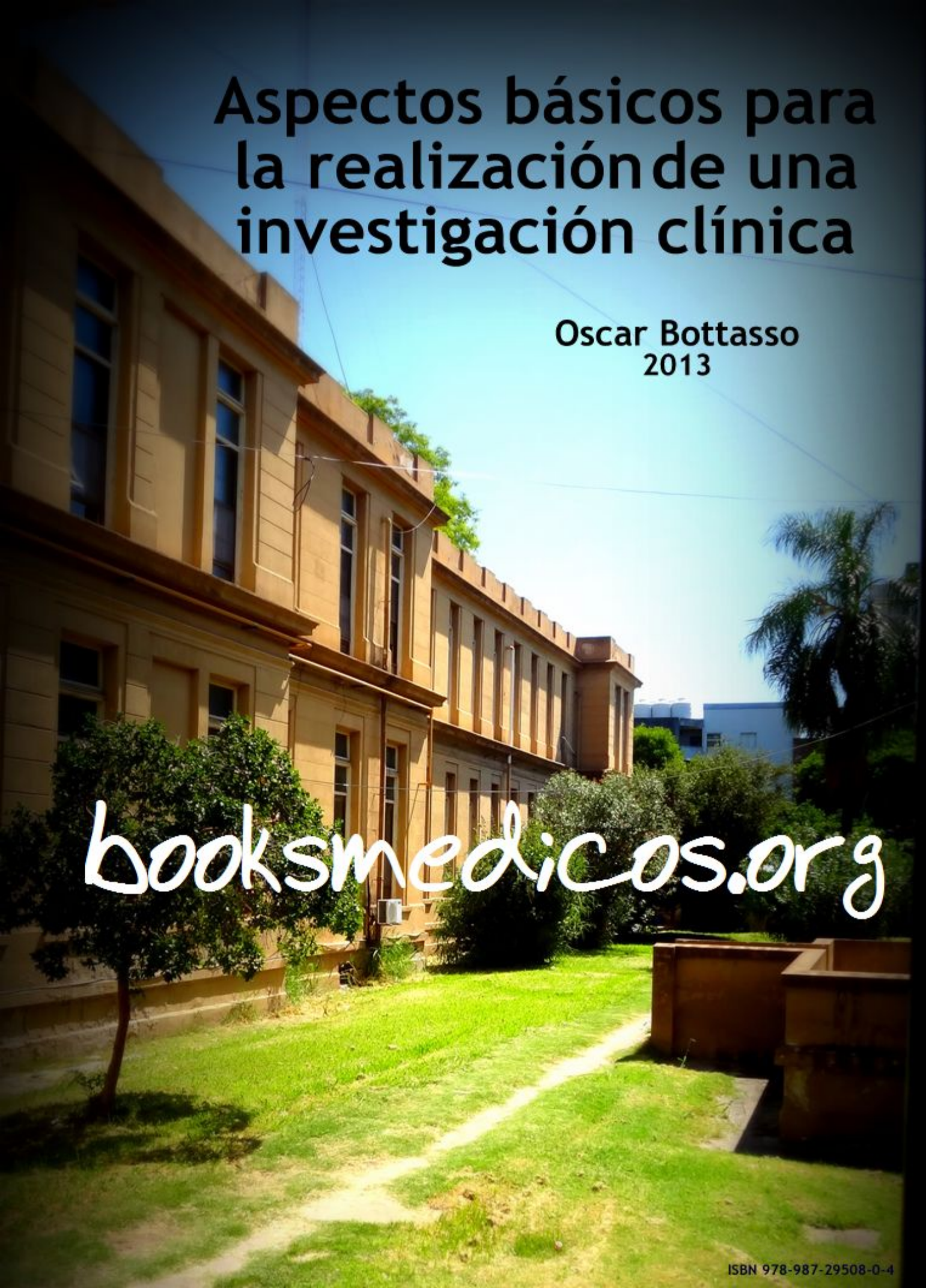
Aspectos básicos para la realización de una investigación clínica

Oscar Bottasso

2013



booksmedicos.org

ISBN 978-987-29508-0-4

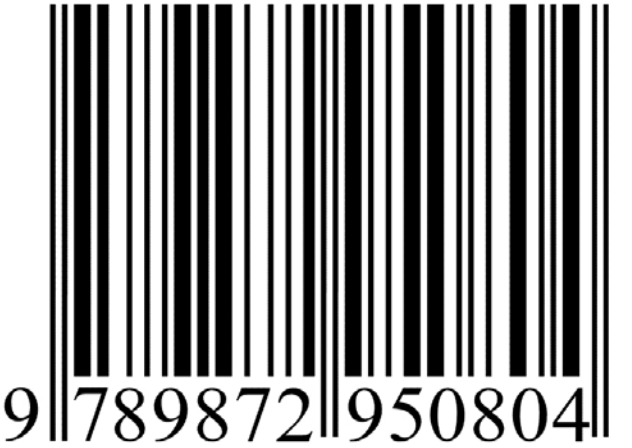
Bottasso, Oscar
Aspectos básicos para la realización de una investigación clínica. - 1a ed. - Buenos Aires: Federación Argentina de Cardiología, 2013.
E-Book.

ISBN 978-987-29508-0-4

1. Medicina. 2. Cardiología. I. Título
CDD 616.12

Fecha de catalogación: 29/05/2013

ISBN 978-987-29508-0-4



A modo de prólogo

Ante la pregunta si es posible conocer mejor, bien podríamos responder que más bien es imperioso. La necesidad de lograr una mayor comprensión de la variada serie de problemas médicos impone una suerte de fuerza directriz capaz de impulsar ese frente de exploración en corrimiento continuo hacia regiones de mayor certeza. Y dentro de este desarrollo el método también hace lo propio para así posibilitarnos una estrategia de aproximación más acabada al objeto de estudio. El *Primus movem* para la preparación del presente texto está vinculado íntimamente con esta problemática.

De uno u otro modo pretendemos que el lector interesado consiga afianzarse cada vez más en la comprensión de las estrategias que rigen la investigación clínica y por ende la producción de conocimiento científico. Conocimiento que a la hora de enmarcarlo debemos recalcar sus rasgos más salientes: posee un método, sustentado en fundamentos lógicos y empíricos, al que debemos adherir en consecuencia; procura ser objetivo, a la vez de tener capacidad descriptiva, explicativa y predictiva. Por último, pero no menos importante, su carácter crítico que lo hace susceptible de ser revisado ante una evidencia refutatoria de los argumentos que se venían sosteniendo. Campo desaconsejable para los dogmáticos, que no obstante pululan, aunque ello no es imputable al método ni mucho menos a la ciencia. Método y conocimiento científico progresan en paralelo, y cada uno realimenta al otro.

Esta falibilidad del método deriva en cierta medida en alguna cuota de escepticismo. Conocimiento y método científicos, deben ser continuamente escrutados para detectar posibles errores. Los resultados científicos son esencialmente provisionales y es casi un acto de puerilidad imaginar que hemos descripto un aspecto del mundo real con total exactitud. Siempre se nos filtrará algún elemento de distorsión. Ni el más precavido y honesto de los científicos está exento de incurrir en un traspiés, que la comunidad científica se lo hará notar.

La investigación científica, al igual que otras actividades humanas, termina instalando una suerte de confiabilidad. Los científicos parten del supuesto por el cual los hallazgos publicados por sus colegas son válidos. La sociedad a su vez confía que los resultados de la investigación reflejan un intento digno de describir al problema en cuestión. El nivel de confianza que caracteriza a la ciencia y su relación con la sociedad ha llevado a una trasposición y aplicabilidad del conocimiento científico como nunca antes vista. A raíz de ello, el investigador debe ser conciente que su trabajo puede tener un gran impacto en la sociedad, puesto que las consecuencias de tal o cual descubrimiento no pueden entreverse en su totalidad. En buena medida porque tal conocimiento se inscribe en un contexto de “no neutralidad”. Así las cosas el debate bioético entre científicos y fuerzas vivas de la comunidad se torna fundamental a fin de reconocer los alcances potenciales de un hallazgo, establecer las debidas precauciones de ser necesario e incluso instaurar un foro de discusión con miras a lograr una pauta de consenso sobre decisiones a futuro.

La ciencia y el acto médico cavan sus cimientos a través de hipótesis, verificaciones experimentales y posteriores interpretaciones. Con las variantes que pueden presentarse en alguna situación en particular, el método es igualmente válido en ambos escenarios. Traspuesto ese umbral, y como una actividad que en definitiva es humana, sobrevendrán las creencias, que inexorablemente están ligadas a la construcción de la realidad, como proyección de nuestros pensamientos y credos más íntimos. Al decir de Einstein “la realidad es una ilusión persistente”.

Con los consabidos altibajos, cierto es que el método ha dado sobradas cuentas de los beneficios que acarrea adherir al mismo. Para los médicos no sólo nos ha facilitado un mejor ejercicio de la profesión, sino que también nos posiciona para comprender con mayor claridad

booksmedicos.org

nuestra base de sustentación. Un buen antídoto para contrarrestar tentaciones de esta milenaria profesión en que las extralimitaciones no fueron precisamente la excepción.

Sumidos en estas cavidades, creemos sin dudarlo, que habitamos lo más elevado de la tierra, que es poco más o menos lo mismo que si uno, teniendo su habitación en las profundidades del Océano, se imaginase que habitaba por cima del mar, y viendo a través del agua el sol y los demás astros, tomase el mar por cielo; y que no habiendo, a causa de su peso y de su debilidad, subido nunca arriba, ni sacado en toda su vida la cabeza fuera del agua, ignorase cuánto más puro y hermoso es este lugar que el que él habita, no habiendo visto, ni tampoco encontrado persona que pudiera enseñárselo.

Diálogos
Platón

INDICE

Introducción. La investigación clínica en la Medicina actual

CAPITULO 1. Cuestiones básicas para la preparación del protocolo[1]

1.1 Tipos de proyectos de investigación (1)
1.2 Secuencia de una investigación (2)
1.3 El origen de la idea a explorar y elementos que hacen a una buena pregunta (4)
1.4 Desarrollo de la cuestión a resolver y estrategias para mejorar los objetivos (5)
1.5 Formulación de los objetivos (6)
1.6 Elección de los participantes (6)
1.7 Tipos de muestreo (7)
1.8 Criterios de evaluación. Tipos de variables, precisión y exactitud. Los datos del archivo (8)
1.9 Cuestionarios y entrevistas. Planificación inicial. Desarrollo y prueba de las preguntas específicas**.** Preguntas abiertas y cerradas. Preguntas directas e indirectas. Asignación de valores. Articulando el cuestionario final. Consideraciones adicionales sobre el entrevistador y la entrevista (11)
1.10 Tipos de hipótesis. Errores tipo I y tipo II. Poder del estudio (16)
Magnitud del efecto. Cálculo del tamaño muestral (19)

CAPITULO 2. Estudios observacionales

2.1 Estudios de cohortes. Estudios prospectivos, retrospectivos, en ambas direcciones, de cohorte doble y con un estudio de casos y controles anidado (25)
Elección de los participantes (27)
Registro de la información (28)
Factores capaces de afectar la validez del estudio (29)
Incidencia y riesgo relativo (30)
Riesgo atribuible (31)

2.2 Estudios transversales (32)
Ejemplo de estudio transversal clásico (35)
Ventajas y desventajas de un estudio transversal (37)

2.3 Estudios de casos y controles (37)
Casos y controles anidado (38)
Consideraciones específicas para el estudio de casos y controles (39)
El coeficiente de desigualdad relativa -Odds Ratio- (40)
Alcances del estudio (42)
Error de muestreo (42)
Error de medición (45)
Errores de clasificación (46)

2.4 Pautas para la publicación de estudios observacionales (46)

2.5 Causalidad (46)

[1] Entre paréntesis se indica el número de página

Los criterios de causalidad de Bradford Hill (50)

CAPITULO 3. Estudios con intervenciones o ensayos clínicos

3.1 Pruebas diagnósticas (56)
Sensibilidad y especificidad (57)
Valor predictivo (58)
Planificación de un estudio diagnóstico (60)
El punto de corte (60)
La curva ROC -receiver operator characteristic- (61)
Limitaciones y estrategias (63)
Consistencia de las pruebas diagnósticas (64)

3.2 Ensayos terapéuticos (67)
Indicaciones para llevar a cabo un ensayo clínico (67)
Las distintas etapas que atraviesa el desarrollo de un medicamento (68)
Consideraciones para el diseño de un ensayo clínico (71)
El ensayo de no inferioridad o equivalencia (73)
Tipos de randomización (74)
El índice de reclutamiento (77)
Efectos adversos (78)
Otros tipos de ensayos clínicos (79)
El ensayo clínico cruzado (80)
Ensayos clínicos en pequeña escala tanto mecanísticos como de eficacia terapéutica (82)
Ensayos clínicos vs. Cohortes para la evaluación del efecto terapéutico (86)

3.4 Lineamientos éticos (88)
Utilización de placebo (91)
El detenimiento de un ensayo clínico (93)
La relación entre el investigador clínico y el sujeto de investigación (94)
Una estrategia global para la bioética (95)
La Comisión de Bioética (96)
Las reglas de la buena práctica clínica (98)
Apéndice. Los contenidos de un protocolo de ensayo clínico (105)

CAPITULO 4. Procesamiento y análisis de los resultados
4.1 Generalidades (108)
4.2 Planificación y análisis (111)
Criterios para la elección de la prueba estadística (115)
Interacciones (118)
4.3 Análisis ajustados para evaluar el efecto potencial de un factor de confusión –FC- (120)
Análisis estratificado y multivariado (123)
4.4 Estrategias particulares para ensayos clínicos (124)
Cálculo de eficacia (124)
Eficacia en vacunas (125)
Análisis de puntos finales cualitativos (126)
Una prueba bien simple (129)

Análisis interinos (130)
Intención de tratar (131)
Análisis por sub-grupos (132)
Puntos finales compuestos (132)
El área bajo la curva (133)
4.5 Recomendaciones estadísticas para las distintas etapas del ensayo clínico (134)
4.6 Los sesgos (135)
4.7 Criterios para valorar un ensayo clínico aleatorizado (141)
4.8 Algunos conceptos sobre meta-análisis (148)
4.9 Indicadores utilizados en la medicina basada en la evidencia (156)
4.10 Pautas para estimar la calidad de la evidencia (159)
4.11 Medicina traslacional (162)

CAPITULO 5. La preparación del manuscrito
5.1 Introducción (172)
5.2 Material y métodos (172)
5.3 Resultados (174)
5.4 Discusión (178)
5.5 Referencias (179)
Apéndice. Declaración de Helsinki (181)

Palabras claves

A mí querida familia

INTRODUCCION

La investigación clínica en la Medicina actual

Como solemos insistir en nuestras clases, los contenidos desarrollados en cualquier texto de Medicina no constituyen en modo alguno el producto de brillantes especulaciones sino de una serie de trabajos de investigación concatenados que en su conjunto determinan el conocimiento científico alcanzado sobre tal aspecto. La lista de referencias bibliográficas que se agrega al final del capítulo habilita al lector para profundizar aún más en torno de lo expuesto.

Como en tantas actividades humanas, el origen de la investigación parte de la antigua Grecia donde Hipócrates y los grandes pensadores de aquella época comienzan a explicar fenómenos naturales por medio de la lógica y la observación proveyendo la base filosófica del método científico (entendiendo como método al modo de decir o hacer algo ordenadamente, o también una guía para sortear obstáculos y evitar errores). A partir del Renacimiento el método científico adquiere una nueva dimensión tras la propuesta de procedimientos cuantitativos por sobre la interpretación escolástica como instancia superadora para entender la realidad. En el campo de la Biomedicina ello promueve el desarrollo de la fisiología y la patología, las que ocasionaron un cambio fundamental en nuestra visión de las enfermedades. Llegamos así al siglo XX donde gracias a la incursión de la estadística el método logra un refinamiento hasta ese momento impensado proporcionando herramientas muy poderosas para el avance de la Medicina, vale decir los estudios observacionales y el ensayo clínico aleatorizado. Por medio de ellos ha sido posible la identificación de causas como así también el diagnóstico, tratamiento y pronóstico de enfermedades de diversa índole, cuya enumeración excede a los propósitos de estas líneas. En el devenir constante, los aportes surgidos tanto de la observación clínica como de investigaciones básicas plantean un nuevo desafío, abordar la multicausalidad. Se habla de causas esenciales, necesarias y contributivas. La psico-neuro-endócrino-inmunología nos propone una red de interacciones que indudablemente influyen sobre el resultado final ante la exposición a un agente nocivo. Para responder a dichas demandas, es cada vez más frecuente el empleo del análisis multivariado que toma en cuenta la participación de factores “extra-biológicos” en la causalidad. Este nuevo enfoque nos llevará a tener una idea cada vez más abarcadora sobre los mecanismos subyacentes de una enfermedad.

Sumado a la importancia natural de la investigación como forma de conocer, no escapa a nuestro análisis que el desarrollo de esta actividad en el claustro Universitario también acarrea ventajas institucionales que van mucho más allá del incremento de la autoestima y engrosamiento del *Curriculum Vitae* de los investigadores. Con el transcurso del tiempo, indefectiblemente se producirá un crecimiento del grupo de investigación con la consecuente formación de recursos humanos tan esenciales para el país. Por otro lado, asistimos a una época donde si bien la información es abundante ello no implica necesariamente conocimiento. Para llegar a esa instancia es necesario jerarquizar, sopesar, relacionar y a veces desechar alguna que otra información. Sin ánimo de adentrarnos en comparaciones odiosas, es dable entrever que el entrenamiento en investigación crea condiciones más favorables para lograr dicho discernimiento, lo que de alguna manera enriquece la actividad docente. La importancia de la investigación también se refleja en la jerarquía que le asignan los organismos gubernamentales involucrados en la acreditación de las carreras médicas de pre y post-grado.

Sobradas son las razones para que nuestras Universidades implementen estrategias que propendan a una sistemática y efectiva formación de profesionales en investigación.

CAPITULO 1

1.1 Tipos de proyectos de investigación

La realización de un estudio clínico no podría llevarse a cabo sin establecer de antemano el protocolo al cual los investigadores deberán adherir. Dicho documento no sólo nos permitirá presentarlo ante una agencia financiadora para obtener el subsidio, sino que también será la herramienta fundamental para que el investigador organice sus actividades en forma lógica y efectiva.

Si tratáramos de identificar el origen de una investigación, resulta claro que la pregunta es la idea seminal que inmediatamente nos llevará a definir los objetivos. No obstante ello, debemos tener en cuenta que la cuestión formulada deberá contar con una fundamentación capaz de establecer con toda claridad la importancia del interrogante a explorar.

Como excepción a toda regla, en los estudios epidemiológicos descriptivos no necesariamente se parte de una hipótesis sino de la necesidad de conocer cómo se distribuye una enfermedad en la población según la edad, género, raza, estilo de vida, etc. Dentro de esta clasificación también se incluyen las comunicaciones sobre algún caso en especial (“case reports”) o series de casos.

La palabra descriptiva en ciencia tiene una cierta connotación peyorativa inmerecida. Tengamos en cuenta, por ejemplo, que el enfoque descriptivo es menos proclive al sesgo. Asimismo, los estudios de laboratorios se basan en recolectar datos de la experimentación por lo que todo sería finalmente descriptivo en alguna medida. Los científicos atribuyen como descriptivo a los casos en que la información se obtiene sin una pregunta en mente, sin una hipótesis orientada a explicar un determinado fenómeno. Descriptivo también puede estar referido a lo observacional e inductivo mientras que nos reservamos la expresión hipotético-deductivo o mecanístico cuando la pregunta es la fuerza conductora. Es bueno recordar sin embargo que muchas disciplinas científicas como la astronomía, arqueología, y paleontología son casi descriptivas en su totalidad. Ciencias descriptivas como la taxonomía, anatomía, botánica, y paleontología han sido fundamentales para el desarrollo de la teoría evolutiva. Muchos procedimientos actuales de la investigación como secuenciación, microarreglos, bioinformática, proteómica son descripciones sofisticadas que posibilitan la caracterización de nuevos fenómenos y la generación de preguntas acordes. En definitiva la transición de una investigación descriptiva a la fundamentada en hipótesis refleja la maduración del campo. Las observaciones descriptivas juegan un papel muy importante en el progreso científico, particularmente durante las exploraciones iniciales que promueven las brechas tecnológicas, pero en sí mismas son raramente conclusivas en términos de explicación. Sí permiten formular las futuras hipótesis de trabajo para los estudios de corte analíticos o explicativos.

Para el caso en que nuestro propósito se enrole en este último tipo de investigación, el paso siguiente será delinear el diseño del estudio. A tal efecto es necesario considerar en primer término si nuestra propuesta plantea modificar los eventos que se pretenden estudiar. Para el caso que no sea así, el estudio será observacional, vale decir que el investigador sólo observará los eventos. Un ejemplo de esto son los estudios de Casos y Controles. Si por el contrario deseamos intervenir sobre el problema en cuestión, estaremos frente a un estudio experimental, al que habitualmente se lo denomina Ensayo Cínico. En el mismo se analizará en qué medida una intervención puede modificar un resultado, por ejemplo la aplicación de un tratamiento para la hipertensión arterial.

A la par de estas consideraciones y especialmente en lo que hace a los estudios observacionales sobreviene inmediatamente una segunda cuestión: la referida a en cuántas ocasiones se efectuarán las mediciones. En caso de que la persona sea examinada en una sola instancia, se tratará de un estudio transversal (*cross-sectional*), mientras que si los participantes son sometidos a un seguimiento el estudio será longitudinal o de cohorte. Vale la pena señalar que los estudios transversales, cuyo propósito es mostrar características de una enfermedad de acuerdo a determinadas variables demográficas, corresponden a la categoría de investigaciones descriptivas, es decir que no son analíticos *sensu stricto*.

1.2 Secuencia de una investigación

Tras haber decidido qué tipo de estudio deseamos realizar, pasaremos a determinar quiénes son las personas que podrían participar de la investigación, las variables a utilizar como instrumentos de medición y los procedimientos estadísticos a los cuales serán sometidos los datos ulteriores. La serie de procedimientos a seguir en la realización de un estudio se halla ilustrada en la Tabla 1.1.

El hecho germinal de un proyecto de investigación es sin lugar a dudas un interrogante sobre el cual aún no se conoce la respuesta. En esta especie de juego de preguntas y respuestas, en primer término habrá que definir con toda claridad qué es lo que deseamos contestar como para intentar resolver el problema bajo análisis. Dentro de ese contexto también debemos estar informados respecto de lo conocido acerca del problema, qué posiciones existen en cuanto al mismo, y estudios previos sobre la temática. En torno de estas cuestiones se comienza a "construir" un documento que finalmente llegará a ser el protocolo de estudio.

Esta pregunta inicial nos lleva inexorablemente a esbozar en sus rasgos generales la población más conveniente para analizar la variable de impacto que está directamente ligada a la cuestión que nos convoca. Con este borrador se comienza a trabajar en una parte esencial del proyecto, referida al tipo de diseño a aplicar y el plan de acción para llevar a cabo la investigación. Es un tiempo de discusión y reflexión entre los investigadores del estudio, para lo cual es necesario poner todas las opiniones sobre la mesa no sólo en lo referente a las cuestiones metodológicas sino también a las posibilidades y limitaciones existentes en el medio en el cual se desarrollarán las actividades. Indudablemente es la etapa donde el documento alcanza su madurez y quedan establecidos los lineamientos fundamentales del estudio, vale decir los objetivos, el tipo de población en la que se llevará a cabo la exploración y las mediciones que se pretenden efectuar.

Luego sobreviene la etapa de la ejecución del estudio, el tiempo de los logros y dificultades, evaluaciones que no se concretan, datos faltantes, tripulantes que desean o quizás terminan abandonando el bote. Como toda empresa humana la investigación no está exenta de estas vicisitudes y le compete al investigador principal ser el piloto de tormenta para que el proyecto arribe a buen puerto. El "oficio" de investigador no sólo requiere de capacitación para la labor sino también de tenacidad y persistencia, más aún en los ámbitos donde la necesidad de investigar no ha sido entendida por quienes deberían ser efectivamente la clase dirigente. Tras un período de meses y a veces años, los investigadores, acuerdan en que se ha analizado un número lo suficientemente adecuado de participantes dándose por concluido el estudio. Es la hora de hacer el balance, cuál ha sido el estudio real, las características de la muestra finalmente conformada y las variables que concretamente se lograron medir. El análisis estadístico mostrará los hallazgos concretos del estudio, permitiéndonos establecer inferencias sobre la validez interna del mismo. Algunos autores utilizan una expresión mucho más

fuerte “la verdad del estudio”, la cual preferimos no emplear puesto que ninguna investigación está exenta de errores. Ello nos lleva a advertir que al momento de extraer conclusiones la validez interna del estudio estará de alguna manera acotada por la presencia del error. Esto adquiere niveles de mayor “peligrosidad” si los investigadores pretenden avanzar en su inferencia adentrándose en el terreno de la validez externa, la que conduce a la enunciación de verdades *quasi* universales. Desde nuestra visión puede ser un salto al vacío.

Tabla 1.1. Componentes y secuencia de una investigación

Momentos	**Planteos y realizaciones**	**Quiénes participan**	**Qué se evalúa**
Punto de partida	Pregunta a elucidar	Población en estudio	Variable de impacto
Preparación del protocolo	Diseño y planificación	Muestra propuesta	Variables que se piensan evaluar
Realización del estudio	Estudio resultante	Muestra definitiva	Variables efectivamente Evaluadas

A fin de que las inferencias sean lo más acertado posible habrá que prestar especial cuidado a los factores de error tanto en el diseño e implementación del estudio, como así también en el análisis de los resultados. Las fuentes de error que pueden restar peso a las conclusiones son lo que se define como error aleatorio y error sistemático.
El error aleatorio está relacionado con la presencia de algún factor desconocido que pueda distorsionar la muestra y los datos sobre lo que estamos trabajando (en cualquier dirección). Una manera de reducir la influencia de este tipo de error es aumentar el tamaño de la muestra. El error sistemático, por el contrario, tiene que ver con la existencia de un sesgo que altera los resultados en una dirección. En líneas generales el error sistemático puede agruparse en dos grandes categorías, error de selección y error en la observación o modo de informar el dato. El sesgo de selección es una eventualidad que puede darse cuando los criterios que debe satisfacer el candidato a ser incluido en el estudio no son uniformes. Supongamos que en un estudio de casos y controles, los requisitos para incluir casos son distintos a los utilizados para los controles y dichos criterios guardan relación a su vez con el factor que suponemos representa un riesgo para el desarrollo de la enfermedad en cuestión.
El sesgo de observación o recolección es el resultado de diferencias sistemáticas en el modo que se recaban los datos referidos a las variables predictivas y de impacto. En estudios de casos y controles donde la variable de impacto ya se conoce (presencia o ausencia de la enfermedad) puede ocurrir un sesgo por la tendencia a buscar con mayor detenimiento el factor de riesgo que potencialmente se encuentra involucrado. En los estudios de cohorte la falta de seguimiento de los participantes es una fuente de error sistemático la cual puede plantear serias dificultades. La falta de criterios estrictos para definir los distintos valores que toman las variables predictivas y de impacto lleva indefectiblemente a una incorrecta clasificación del factor de riesgo y la enfermedad,

respectivamente. Este sesgo se torna aún más serio si la distribución del mismo difiere entre los grupos.
Sea como fuere, los errores aleatorio y sistemático pueden determinar que se produzca un error de muestreo y afecte la inferencia poblacional; o bien un error de medición y en ese caso el problema surgirá cuando tratemos de concluir en base a los datos recogidos respecto del fenómeno de interés.
Para ir adentrándonos en un tema que luego será tratado según el tipo de estudio, la tabla 1.2 brinda una serie de recomendaciones orientadas a reducir el error.
Reducir el error significa que si bien no es posible eliminarlo, debemos propender a que el mismo no sea lo suficientemente significativo como para ocasionar un cambio sustancial en las conclusiones. Al planear un estudio conviene tener en claro que en algún momento deberán efectuarse concesiones entre lo que es la validez científica por un lado y la factibilidad por el otro. Una vez formulado el plan, la recomendación a tener en cuenta es analizar si lo que hemos escrito servirá para contestar la pregunta en el contexto de algo factible de implementar con un nivel de error aceptable. Si la respuesta es negativa habrá que empezar de nuevo. Para no caer en la desazón, es bueno recordar una definición de W. Churchill “el éxito no es más que ir de fracaso en fracaso sin perder el entusiasmo”.

Tabla 1.2. Estrategias para reducir el error

Momento del estudio	Error aleatorio	Error sistemático
Diseño e implementación	Aumentar el tamaño muestral Mejorar la precisión de las evaluaciones.	Mejorar el diseño. Por ejemplo, si en un estudio de casos y controles, los primeros se hallan hospitalizados, tratar de que los controles también estén internados. Ser muy estrictos en el modo en que se recolectarán los datos. Asegurarse de la información pre-existente que se consigna en las historias clínicas.
Análisis de los resultados	Utilizar métodos estadísticos correctos, calcular los intervalos de confianza.	De existir una duda razonable sobre algún sesgo, consultar con expertos en epidemiología analítica para ver si es posible identificarlo.

1.3 El origen de la idea a explorar y elementos que hacen a una buena pregunta

La realización de una investigación parte de una idea que exige la mayor creatividad posible de nuestra parte. Si ella está dotada de gran originalidad desde su concepción primaria mucho mejor, de lo contrario deberemos refinarla y para ello muchas veces es necesario consultar con expertos en el tema. Lamentablemente, los días en que el investigador podía elegir con bastante libertad qué tipo de investigación pretendía realizar han quedado muy lejos. Hoy por hoy se requiere conocer de antemano las prioridades que fijan las Instituciones financiadoras de la investigación. En otros términos: si pretendemos obtener financiación, a la hora de formular nuestro objetivo

tratemos de que el mismo coincida de alguna manera con las pautas establecidas en la convocatoria. Muchas Universidades del mundo cuentan con un Departamento donde se brinda información sobre las distintas entidades financiadoras de proyectos de investigación. Con ello se facilita que el investigador tenga más chance de encontrar una Institución interesada en su línea de investigación.

Si bien no existe una receta respecto al modo de confeccionar la pregunta, conviene tener en cuenta que las rupturas, inconsistencias, combinaciones, analogías y paradojas son ingredientes fundamentales en el "caldero de las ideas". El estar mentalmente preparados para discernir las falencias posibilita el rompimiento con una interpretación alternativa del problema. Es muy recomendable que el investigador sea una persona ilustrada puesto que ello acarrea no pocas ventajas. Por un lado permite sopesar y combinar ideas preexistentes para proveer una nueva interpretación de los fenómenos. A la par de ello, nos ahorraríamos los estériles debates entre **cultivados ignorantes e ilustrados incultos** de las ciencias duras y sociales. En un plano igualitario se sitúan la analogía y la paradoja. La ciencia abunda en ejemplos donde la consideración de ideas ajenas a la disciplina en cuestión o a la misma ciencia provee de formas innovadoras de concepción del problema en cuestión. Nadie desconoce, por otro lado, que un hecho paradojal es un estímulo poderoso para la imaginación científica.

Al formular la pregunta conviene tener en cuenta los factores que contribuyen a que nuestro interrogante sea la simiente de un proyecto de investigación. Un aspecto central en estas cuestiones está referido a la originalidad del planteo, y en qué medida la investigación servirá para arrojar claridad en torno a problemas no resueltos, sea porque provee nuevos hallazgos o bien porque permite confirmar o rechazar postulados previos en un contexto diferente. Más allá de la relevancia puramente científica, hay preguntas que su elucidación tendrá una repercusión inmediata sobre las políticas de salud y la práctica médica. Otros factores que contribuyen tanto en la fundamentación de la pregunta como en el desarrollo de la investigación son los aspectos éticos, el interés que la propuesta suscita entre los investigadores, y la factibilidad de su realización, sea porque el número de pacientes y la experiencia técnica son los adecuados, o bien porque se dispone de suficiente tiempo y dinero.

1.4 Desarrollo de la cuestión a resolver y estrategias para mejorar los objetivos

Inicialmente suele ocurrir que la idea parezca un tanto vaga o inapropiada, para lo cual es aconsejable redactar en 1-2 páginas la fundamentación del aspecto sobre el que se desea indagar, como así también las bases para reclutar la población y efectuar las mediciones. Con el borrador en las manos, los investigadores consiguen visualizar una serie de deficiencias que podrían determinar la no factibilidad del proyecto. Un entusiasmo excesivo suele llevarnos a formular una pregunta demasiado amplia la cual requeriría un emprendimiento *quasi* faraónico para ser contestada. Ello se subsana acotando los interrogantes y reduciendo el número de variables. Al efectuar estas modificaciones se deben tomar todos los recaudos para que la cuestión no termine siendo irrelevante o poco novedosa. Continuando con la lectura del documento puede surgir que el número de participantes sea insuficiente para la propuesta, con lo cual se deberán expandir los criterios de inclusión, reducir los criterios de exclusión o bien prolongar el tiempo de reclutamiento. Si el grupo de investigadores responsables no cuenta con un experto en el tema a abordar, es probable que el diseño presente ciertas deficiencias factibles de ser corregidas mediante la consulta con un especialista en la materia. Dos cuestiones adicionales están referidas al costo que insume la realización del estudio y los aspectos éticos del mismo. Algunos investigadores suelen consultar al

Comité de Bioética para cerciorarse de que el protocolo elaborado no se aparta de las normas que rigen la investigación biomédica.

1.5 Formulación de los objetivos

La enunciación de los propósitos generales del proyecto debe señalar los principales aspectos a investigar brindando una idea clara del tipo de estudio a llevar a cabo, como así también de la potencial relación entre la variable predictiva y la(s) de impacto. En la investigación básica es más frecuente que a las mismas se las denominen variables independientes y dependientes, respectivamente.

Al definir los objetivos específicos se deberá proveer información acerca de cómo se evaluará la dimensión del efecto de la variable predictiva (por ejemplo una intervención) sobre la variable de impacto, tratando de ofrecer en la medida de lo posible la precisión y exactitud con las que se medirá el efecto. Para ello se deberán establecer definiciones concisas de los valores que adoptará la variable de impacto, no sólo por las cuestiones operativas de nuestro estudio sino también para los fines comparativos con los resultados obtenidos por otros autores, lo que finalmente permite establecer pautas de consenso. Al trabajar sobre estos aspectos del protocolo es aconsejable tener presente los estándares internacionales corrientemente utilizados para la definición de muchas enfermedades. Si la investigación que pretendemos llevar a cabo es preventiva la principal variable de impacto es la incidencia de nuevos casos en los grupos que reciben o no la intervención. Si el ensayo fuera terapéutico, obviamente el interés se centrará en la cantidad de fracasos y éxitos según los tratamientos administrados.

En la descripción de los objetivos específicos es aconsejable mencionar el tipo de individuos sobre los que se trabajará. Pueden darse circunstancias en que los investigadores quieran analizar la influencia de la variable predictiva sobre mediciones que si bien no están relacionadas con los propósitos fundamentales del estudio agregan información al mismo constituyendo lo que se denominan objetivos secundarios. Tomemos como ejemplo el caso de un ensayo clínico terapéutico donde además de la eficacia de un antibiótico sobre una enfermedad infecciosa los investigadores también desean saber si se modifican los marcadores inmunológicos asociados a la misma.

1.6 Elección de los participantes

En lo referente a este aspecto vale la pena repasar las definiciones de población y muestra. Una población es un grupo completo de personas con algunas características en común, mientras que la muestra es un subgrupo extraído de esa población. Esta última se conforma en base a los criterios de inclusión y exclusión. Los primeros establecen los perfiles que deberán poseer las personas a efectos de responder a los objetivos del estudio. Para ello deberán definirse las características, demográficas, clínicas, geográficas y temporales. Los criterios de exclusión se preparan teniendo en cuenta si los pacientes presentan patologías concomitantes capaces de favorecer un mayor riesgo (para el caso de un tratamiento) o bien alterar el resultado de lo que se pretende evaluar. Es frecuente excluir de los ensayos clínicos a las personas demasiado jóvenes o viejas como así también a las mujeres embarazadas a menos que sean precisamente ellas las destinatarias de los objetivos del estudio. Entre los criterios de exclusión también se suelen mencionar situaciones vinculadas directamente a las características y posibilidades del paciente tales como, baja probabilidad de seguimiento, incapacidad de proveer datos fidedignos y no adhesión a los requerimientos del estudio, entre otras.

1.7 Tipos de muestreo

Llegado al punto de decidir el modo de reclutar los participantes de la investigación y dependiendo del tipo de diseño se debe optar entre dos alternativas: muestreo probabilístico o no probabilístico.

Muestro Probabilístico

Utiliza un proceso aleatorio por el cual cada integrante de la población tiene chance de ser elegido. De esa manera habrá bastante seguridad respecto a que los fenómenos observados en la muestra son representativos de lo que ocurre en la población.

Los diseños de muestreo probabilístico incluyen las siguientes técnicas: muestreo aleatorio simple, muestreo sistemático, muestreo estratificado, muestreo por cluster o muestreo multi-fases.

- Muestreo aleatorio simple: se asigna un número a cada integrante y luego se confecciona la muestra mediante la utilización de una tabla de números al azar.
- Muestreo sistemático: es una técnica simple, conveniente y muy utilizada. Si bien es atractiva por su simplicidad hay que tomar algunos recaudos. Según la estructura de la población, la información brindada puedo no ir más allá de lo que implica una única observación aleatoria. A raíz de ello, conviene que la población esté secuencialmente ordenada. Por ejemplo las casas a lo largo de una calle o una lista alfabética de nombres extraídos de una guía. O bien puede ser una lista en orden aleatorio de unidades a muestrear. A partir de esto, se toma aleatoriamente una primera unidad y luego se van seleccionado posteriores unidades a intervalos iguales.
- Muestreo aleatorio estratificado: tiene que ver con la división de la población en partes (estratos). Idealmente cada estrato contiene unidades que son homogéneas respecto a las variables de relevancia para el estudio. Se selecciona una muestra aleatoria de cada estrato. Los ejemplos pueden ser áreas geográficas (zonas urbanas o rurales) o niveles socioeconómicos de los habitantes de una ciudad. Cada estrato se muestrea independientemente en general por el método aleatorio simple. El procedimiento sirve para aumentar la eficiencia del muestreo respecto de costos y precisión del estimador.
- Muestreo por "clusters": El agrupamiento *(clustering)* también tiene que ver con la división de la población en grupos pero a diferencia de los estratos, los “clusters” los grupos de unidades poblacionales pueden ser más heterogéneos. Idealmente, cada cluster debería ser un microcosmos de la población (provincias, determinado cursado de un gran establecimiento educativo). Digamos que estratificación sería más aceptable que “*clustering”* en el sentido que el primero tiene a incrementar la precisión y asegurar la representatividad, mientras que el segundo la desciende. En un diseño a una sola fase se analiza la unidad muestreada. En un diseño multi-fase, primero se efectúan las selecciones de unidades grandes, y luego en las fases subsiguientes se realizan las subselecciones dentro de las unidades conformadas en la primera etapa.
- Una variedad aceptada y que posibilita la selección igualitaria es el muestreo con probabilidad proporcional al tamaño; aquí la primera fase del muestreo se efectúa en proporción al número poblacional y al pasar al la segunda fase se aplica un número constante.

Muestreo no probabilístico.
Para los estudios no observacionales sino más bien intervencionistas esta alternativa es más práctica. Se distinguen 3 tipos:

- Muestreo consecutivo: vale decir incluir todo paciente que reúna los criterios de selección para un determinado período de tiempo y número de enfermos. Es el modo más práctico de los procedimientos no probabilísticos.
- Muestreo por conveniencia: consiste en incorporar a los integrantes más disponibles de la población.
- Muestreo discriminado: en este caso el investigador elige de la población a su alcance a los individuos más adecuados para los propósitos del estudio, por ejemplo aquellos que cumplirán con todas las visitas requeridas si el protocolo exige un estricto seguimiento de los pacientes.

La ventaja de estos 3 últimos tipos de muestreos reside en su eficiencia, es decir le permite al investigador extraer conclusiones a un menor costo en cuanto a tiempo y esfuerzo. La desventaja es que el error (tanto en el diseño como la implementación) puede llevar a inferencias un tanto apartadas en cuanto a lo que acontece en la población.
Para minimizar estas fuentes de error y al decidir sobre el diseño del estudio preguntémonos si la población blanco es la más adecuada para el interrogante sobre el cual se indagará. De no ser así tendremos que redefinir los criterios de inclusión especificando mejor las características de los pacientes, o bien discutir otras alternativas con expertos. Puede darse que la población a la que tenemos acceso no es lo suficientemente representativa de la población blanco, lo que impone modificar los criterios de inclusión teniendo en cuenta las características étnicas, de género, geográficas y rango de edades de esta última.
Superados estos inconvenientes no estaremos exentos de dificultades que podrían surgir durante la implementación. Por ejemplo, sería muy desafortunado que la muestra obtenida no fuese representativa de lo que se pretendía inicialmente. Los factores capaces de llevar a una situación como la descripta son por un lado un error aleatorio en el muestreo, subsanable en parte mediante el incremento en el número de personas o realización de un muestreo estratificado para aumentar algunos subgrupos. Por otro lado también se puede producir un error sistemático muestral (sesgo), como sería una incapacidad de contactar con algún tipo de pacientes, o personas que por alguna razón no desean participar. Como se dice vulgarmente aquí hay que “ponerse las pilas” y brindar una mejor información a los potenciales participantes en torno a los objetivos y necesidad del estudio. Menos frecuente, pero no improbable, es que se hayan cometido errores de cálculo los cuales afectaron la composición y número de la muestra. Por lo anteriormente señalado y por cuestiones a tratar más adelante es bueno tener en cuenta que el error sistemático es una especie de “taladro horadador” de los aspectos esenciales de una investigación actuando en desmedro del éxito de la misma. Recuperarse del ridículo es un trámite doloroso y a veces muy difícil.

1.8 Criterios de evaluación

Un punto central en cualquier tipo de investigación clínica son las mediciones que se deberán efectuar para poder contestar los interrogantes planteados. Si el estudio es de tipo intervencionista, vale decir un ensayo clínico, las variables de impacto (también denominadas de salida o resultantes) tienen que ver con los cambios en cuanto a la incidencia, prevalencia, o mejoría y hasta incluso cura de la enfermedad. Los estudios observacionales, por el contrario, tratarán de identificar factores que puedan determinar

una menor o mayor chance de desarrollar la enfermedad, mientras que las investigaciones básicas o "mecanísticas" utilizan en su mayoría datos del laboratorio. En cualquiera de los ejemplos mencionados es necesario llegar a una definición clara de las variables que se pretenden evaluar no sólo por la consistencia misma del estudio sino por la posibilidad de utilizar tales definiciones en otros estudios y poder comparar los resultados de las distintas investigaciones.

Buena parte de los resultados que se recogen en una investigación está directamente relacionada con los datos que brinda el diagnóstico clínico de la enfermedad, tanto en lo referido al examen físico y el interrogatorio, como así también a la serie de estudios complementarios a que son sometidos los pacientes. No debemos dejar de lado que a la par de los éxitos terapéuticos factibles de ser conseguidos con la utilización de un nuevo tratamiento, también es necesario identificar los potenciales efectos adversos que el mismo puede ocasionar. Para un correcto registro de este tipo de eventualidades los investigadores pueden utilizar sistemas internacionalmente aceptados de gradación de los efectos tóxicos, que en la mayoría de los casos son provistos por el laboratorio patrocinante del estudio.

Así como se debe prestar especial atención a las definiciones, no es menos cierto que las variables a evaluar en un estudio deben reunir algunos requisitos adicionales. Un aspecto fundamental es la relevancia de la medición. Si se tratara de un estudio donde se evalúa la eficacia de un interferón recombinante en enfermos con hepatitis, es aceptable que sumado a la evaluación clínica se incluyan determinaciones de laboratorio específico (por ej. transaminasas) como pautas de seguimiento de la enfermedad. Ubiquémonos en una situación diferente, en este caso los investigadores están interesados en determinar si el tratamiento con un fármaco que tiene acción contra el *Trypanosoma cruzi* en personas infectadas con este parásito puede ser efectivo para reducir la morbilidad y transmisibilidad de la infección. Para el primero de los objetivos es claro que la evaluación se centrará en la realización de electrocardiogramas y algún otro estudio adicional como el ecocardiograma. Para el caso de la transmisibilidad la elección es un tanto más complicada. Si bien se ha observado que los niveles de anticuerpos específicos contra el parásito son más bajos en aquellos que han recibido drogas antiparasitarias dicha estimación no es lo suficientemente sólida como para decidir si el tratamiento es realmente efectivo. Obviamente será necesario contar con pruebas que nos puedan estimar el nivel de la carga parasitaria del individuo, por ejemplo hemocultivos o técnicas moleculares de mayor sensibilidad. Si estos procedimientos no están a nuestro alcance, sea por razones económicas o técnicas, la evaluación no es factible y deberemos resignar el segundo de los objetivos.

La importancia de los criterios de evaluación aparece con toda claridad al momento de efectuar las inferencias, puesto que la validez interna de un estudio depende de cuán representativas son las mediciones que efectuamos de la variable en cuestión. Por su parte, la validez externa del estudio se apoya en parte en la representatividad de nuestra variable sobre el fenómeno de interés.

Con el propósito de brindar una idea general sobre el tipo de variables que se utilizan en investigación clínica, la Tabla 1.3 presenta los tipos de mediciones como así también el tipo de análisis a efectuar.

Tabla 1.3. Tipos de variables

Particularidades	**Tipos de variables**		
	Categóricas o nominales	**Ordinales**	**Cuantitativas (contínuas o discretas)**
Ejemplos	Sexo grupo étnico	Grado de severidad	Peso Corporal, glicemia
Presentación y análisis de los datos	Porcentajes Pruebas de $\chi 2$ y Fisher	Medianas y rangos, pruebas de $\chi 2$, U de Mann-Whitney	Media ± DS Prueba de la *t* de Student

DS, desvío estándar
Las variables continuas toman un número infinito de valores (por ejemplo, el peso), mientras que las discretas tienen una escala finita (número de cigarrillos/día)

Precisión y Exactitud
Antes de concluir sobre el tema de las mediciones es necesario efectuar algunas consideraciones adicionales sobre dos cuestiones vinculadas al modo de cuantificar, como son la precisión y exactitud.
Una medición precisa es aquella que arroja un valor muy parecido cada vez que se lleva a cabo la misma, por ejemplo medir la altura con la misma vestimenta y calzado. La precisión se halla afectada por el error aleatorio y las 3 fuentes a tener en cuenta al momento de efectuar las mediciones son: la variabilidad del observador, la variabilidad del observado y la variabilidad del instrumento de medición. La precisión de una variable se la puede evaluar estadísticamente a través del coeficiente de variación (desvío estándar/media).

Exactitud es el grado de representatividad real de una variable sobre lo que pretendemos estimar. Si efectuáramos mediciones de colesterol sérico usando un valor estándar incorrecto, el resultado será incierto no obstante que pudiera ser preciso.
La exactitud se halla influenciada por el error sistemático o sesgo y aquí también se distinguen tres fuentes, sesgo del observador, por ejemplo cuando el operador tiende a atribuir menos valores de presión arterial en los que reciben una medicación nueva; sesgo del sujeto, el que se da en la persona que participa de la investigación por el simple hecho de que está siendo estudiada; y sesgo instrumental, como el que se produciría por funcionamiento defectuoso de un aparato de medición mal calibrado.
La exactitud de una medición puede ser estimada al compararla con los resultados que se obtienen con técnicas de referencia, lo que en la literatura inglesa se define como *gold standard*.
Vale la peña señalar que precisión y exactitud son características que no conviene aplicar cuando se trata de variables que surgen de una escala donde los componentes subjetivos se hallan más presentes.

Los datos del archivo
En este caso el investigador tratará de responder a la pregunta en base a la utilización de datos que originalmente fueron recabados para otro objetivo. La relativa mayor disponibilidad de bases de datos en nuestros días ofrece una oportunidad para este tipo de investigaciones. La principal ventaja reside en la rapidez y bajo costo con que se obtiene la información. No obstante ello hay que tener en cuenta que el modo en que

fueron recogidos esos datos respondía a otros objetivos pre-determinados, lo cual en alguna medida constituye una limitación.

1.9 Cuestionarios o entrevistas

Respondiendo a la demanda cada vez más creciente de información y análisis para asistir en las políticas de decisiones en Salud, y el diseño de tratamientos efectivos, muchos profesionales de este campo se hallan ante el problema de tener que desarrollar o revisar cuestionarios. A menudo las preguntas sin demasiada elaboración resultan en datos de baja calidad, particularmente en exploraciones a nivel poblacional. Dado que las encuestas son una fuente necesaria de datos, la importancia de identificar los defectos en cuanto al armado de cuestionarios es fundamental.

Por medio de una entrevista, es posible obtener una amplia variedad de datos significativos respecto de los factores etiológicos posibles para una enfermedad. Dado que en general estas investigaciones recaban información en forma retrospectiva y en virtud de las dificultades para establecer la validez de lo referido por el encuestado, muchos investigadores coinciden en que las respuestas a estas clases de preguntas parten de alguna inferencia, lo cual puede entrañar cierto error. En paralelo y por el amplio rango de factores factibles de ser explorados a través de encuestas, las mismas son adecuadas para clasificar una gran variedad de hipótesis acerca de cuestiones de salud, tratamientos y resultados. Diseñar un cuestionario no es algo sencillo. En primer término cada pregunta debe proveer una medición válida y confiable. En segundo lugar, las preguntas deben comunicar claramente lo que se intenta indagar en el entrevistado. Por último, las preguntas tienen que estar ensambladas en una forma lógica y clara con fluidez natural como para mantener al encuestado interesado e inclinado a seguir cooperando. Para satisfacer estos requerimientos, el diseño del cuestionario debe comprender 3 fases: la planificación inicial del cuestionario, desarrollo y prueba de las preguntas específicas, y la construcción final del instrumento.

1.9.1 Planificación Inicial

Habitualmente el cuestionario se prepara, al menos en sus lineamientos generales, en forma conjunta con la redacción del protocolo a efectos de determinar con claridad cuáles son las preguntas más relevantes para los objetivos del estudio. Ello ayuda a establecer un límite en el número de preguntas a incluir ya que un interrogatorio prolongado puede ir en detrimento de la participación de las personas y restarle tiempo, a su vez, a las cuestiones que sí importan en la investigación. Una mala estimación del tiempo necesario para desarrollar el primer borrador de preguntas puede tener efectos muy desfavorables sobre la calidad del cuestionario. Normalmente, la persona responsable del desarrollo del instrumento es una de las tantas involucradas en el proceso. Sin embargo y dado que la encuesta no puede salir a campo sin que se haya completado el instrumento de recolección de información, muy a menudo se ejerce presión para desarrollar el cuestionario lo más rápido posible. Algunos sostienen que el armado del cuestionario concluye cuando ya no queda más tiempo.

Si el instrumento debe ser construido desde el inicio en lugar de tener que modificar uno existente, se debe generar una lista de tópicos o conceptos. Llevar estos tópicos a una serie de preguntas es laborioso. Consultas, reuniones, lluvias de ideas con expertos en la materia y metodólogos en encuestas son actividades muy útiles, previo a la redacción de cualquier pregunta. La revisión de encuestas relacionadas y los instrumentos de recolección de datos ayudan al momento de redactar las propias.

Una vez que se ha preparado el borrador de preguntas, es conveniente revisarlo y corregirlo tanto como uno lo considere necesario. Es un proceso obligatorio para el

desarrollo evolutivo del instrumento, puesto que el borrador nos ayuda a refinar y clarificar los objetivos de la investigación. Las revisiones deben asegurarnos que cada ítem del cuestionario está justificado como una contribución significativa al análisis que se intenta. Es aconsejable confeccionar una tabla donde se va especificando cómo serán analizados los datos de una serie de cuestiones. La extensión del cuestionario también es un tema de debate en la fase inicial. Los revisores del borrador están inclinados a incrementar la representación de los tópicos sobre los cuales se hallan interesados, puesto que ven el rico potencial que ofrecerán los datos. Sin embargo, es necesario limitar los objetivos del instrumento por una serie de razones. Más allá de las dificultadas que implica el trabajo con un cuestionario extenso hay que tener en cuenta el impacto que posee sobre la calidad de los datos. Una serie abultada de preguntas puede llevar a problemas en las respuestas y el consiguiente error. Contemplar el NO a una cuestión central eliminará una larga serie de preguntas subsiguientes. El entrevistador también puede sentirse agotado. A fin de completarlo puede que lea las preguntas rápidamente o puede guiar la respuesta del entrevistado.
Al momento de la revisión es preciso cotejar los objetivos del estudio con las preguntas para asegurarnos que las mismas son acordes. Los cuestionarios pueden contener una combinación de preguntas que intentan indagar acerca de conocimiento, conducta, práctica, actitudes y opinión. Consideremos las similitudes y diferencias entre las siguientes preguntas:
1. Conocimiento: ¿Cada cuánto recomienda la Sociedad de Ginecología que una mujer de 50 o más años se efectúe una mamografía?
2. Opinión: ¿Qué opina Ud. respecto del tiempo que debería dejar pasar una mujer mayor de 50 o más años para efectuarse una mamografía?
3. Conducta: ¿Después de los 50 años, cuántas mamografías se ha efectuado Ud.?
4. Proximidad: ¿Cuántas mamografías se ha efectuado su esposa, después que cumpliera 50 años de edad?

1.9.2 Desarrollo y prueba de las preguntas específicas

Tras haber completado el borrador de las preguntas, el próximo paso es refinarlas y probarlas con distintas personas para que respondan a ellas. Ello posibilita identificar problemas con las preguntas. En teoría al responder a la pregunta, el entrevistado procesa la misma en estados cognitivos tales como: comprensión, recupero de información, decisión o juicio acerca de la pregunta, y la respuesta que se brinda finalmente.

Los términos y conceptos deben ser familiares y fáciles de entender

Se suelen presentar problemas cuando las preguntas contienen términos o conceptos no familiares, vagos, o ambiguos. El entendimiento del entrevistado puede no coincidir con los propósitos de quien lo preparó. Como norma general, las preguntas deben redactarse en el idioma habitual. Las palabras específicas y concretas son más aptas para comunicar un significado uniforme. Es preferible utilizar palabras cortas y no incluir tecnicismos. Respecto de la estructura, la pregunta puede ser tan breve que el entrevistado se halla inseguro del propósito de la misma. Lo mismo puede ocurrir cuando es demasiado larga y compleja. Así las cosas, el entrevistado puede hacer una interpretación muy personal de lo indagado. Para el caso en que se lo busca deliberadamente, la pregunta debe solicitar además que dicha información sea volcada en la respuesta.

Las preguntas deben poseer un ordenamiento y suerte de guía que estimulen la capacidad de recordar

A menudo, los entrevistados son interrogados a rememorar eventos o datos. Este proceso puede brindar una respuesta rápida y adecuada a la pregunta formulada. Sin embargo, se pueden presentar problemas cuando el entrevistado debe recordar demasiada información, o su memoria no le permite acordarse de esos detalles. Rememorar algo puede ser particularmente dificultoso si el evento no revestía importancia o era algo trivial para la persona, si ocurrió hace mucho tiempo, o si la pregunta exige recordar demasiados pormenores. A raíz de esto, se pueden recurrir a estrategias facilitadoras de la capacidad de recordar. Una serie de preguntas relacionadas, por ejemplo eventos bien claros a lo largo del año (pasó cerca de las fiestas), usos de almanaques, herramientas visuales, pueden facilitar la memoria y poner los eventos en orden. De todos modos conviene incluir preguntas a las cuales el entrevistado puede responder rápidamente. A medida que el entrevistado va recordando la información, el mismo va a decidir cuán adecuada es. No perder de vista que el entrevistado recurrirá a estrategias estimativas si la pregunta demanda demasiados detalles o cálculos mentales.

El orden y el formato de las preguntas deben estar balanceados

El ordenamiento de las preguntas puede influir sobre la respuesta obtenida. En general los cuestionarios se arman por grupos de preguntas sobre un mismo aspecto lo cual puede promover una suerte de efecto por contexto. A veces las preguntas generales son tan amplias y como tal más abierta a distintas interpretaciones. El formato de la pregunta también puede ocasionar problemas. Un formato abierto permite al entrevistado ofrecer cualquier respuesta y el entrevistador registra la misma al pie de la letra. Sin embargo, los cuestionarios orientados a obtener información de un número importante de entrevistados habitualmente emplean un formato cerrado. Las respuestas están codificadas, se le asignan números, y así son proclives al análisis estadístico.
Veamos con un poco más de detenimiento las maneras para confeccionar el cuestionario.

Preguntas abiertas, tratan de obtener una respuesta con las propias palabras del individuo y pueden proveer futuras líneas de exploración, por ejemplo ¿qué factores cree Ud. que pueden favorecer el tener una úlcera en el estómago? En investigaciones preliminares o aquellas con un mayor componente psicosocial es preferible utilizar este tipo de preguntas a fin de obtener un panorama más abarcador de las respuestas posibles.

Preguntas cerradas, en este caso el tipo de respuestas ha sido establecido con anterioridad y el individuo seleccionará una o mas de las opciones ofrecidas. Basándonos en el ejemplo anterior, la pregunta sería: cuáles de los siguientes factores considera Ud. que pueden favorecer el desarrollo de una úlcera en el estómago
[ ] frituras
[ ] bebidas alcohólicas
[ ] exceso de trabajo
En estudios que involucran un gran número de personas es conveniente utilizar este tipo de cuestionarios, ya que se facilitará la realización del análisis estadístico.

Preguntas directas e indirectas. En el primer caso, como su nombre lo indica el entrevistador formula el interrogante sin rodeos. Por el contrario, si la información que

deseamos recabar tiene que ver con aspectos más "íntimos" o de mayor compromiso es conveniente utilizar un interrogatorio indirecto vale decir obtener el dato a través de una serie de preguntas relacionadas.

Asignación de valores

Los códigos numéricos y alfabéticos del cuestionario permiten transformar a las respuestas en variables que luego se analizan estadísticamente. En algunos casos se pueden asignar un puntaje a la respuesta del individuo, por ej., supongamos que la pregunta es ¿cuán importantes son los siguientes aspectos para preservar el estado de salud en varones adultos?

a. Examen cardiológico cada 6 meses 1 2 3
b. Examen urológico anual 1 2 3
c. Análisis de laboratorio anual 1 2 3

Donde las asignaciones de significado son las siguientes, 1= mucha relevancia, 2= mediana relevancia, 3= poca relevancia

Puede ocurrir que deseáramos tener una estimación sobre el conocimiento o imaginario que la persona tiene respecto de algún problema médico, para lo cual debamos efectuar una serie de preguntas como las señaladas más arriba. En ese caso habrá que obtener un único puntaje general en base a la utilización de una escala que atribuye un valor (que siempre será el mismo) para cada una de las preguntas que se formulan.

1.9.3 Articulando el cuestionario final

Luego que se revisaron y probaron las preguntas, sobreviene la integración en el instrumento final de recolección de los datos. Aunque cada ítem del cuestionario debe ser ponderado cuidadosamente sobre su mérito, en general es responsabilidad de la persona que diseña el cuestionario el armado de los ítems de un modo coherente como para que exista unidad. La progresión (flujo discursivo) entre las preguntas sobre los distintos tópicos debe ser natural y llevadera. Esto favorece la cooperación tanto del entrevistador como el entrevistado. Es útil tener en cuenta los siguientes aspectos al momento de revisar el cuestionario:

¿Existe una progresión lógica de las preguntas, como para que se despierte el interés del entrevistado en los tópicos?

¿Se preguntará en primer término sobre cuestiones simples de responder para luego proseguir con las más detalladas o complejas?

¿Se preguntarán inicialmente cuestiones objetivas, directas para posteriormente pasar a aspectos más personales?

¿Se guiará al entrevistado de una manera sencilla, se trabajará con elementos de transición para pasar a otra serie de preguntas afines entre sí?

La formulación del cuestionario también requiere que se tengan en cuenta los siguientes aspectos:

- Establecer la información que se pretende obtener con la pregunta y redactarla en forma concisa en un lenguaje claramente comprensible para la población que deberá completar el cuestionario. Para el caso de aspectos vinculados con la privacidad de las personas extremar todos los cuidados posibles al momento de redactar el cuestionario. No es aconsejable incluir preguntas que favorezcan un tipo de respuesta en particular.
- Determinar de antemano la validez de lo que se está preguntando. Dicho en otros términos establecer en qué medida la respuesta que da el entrevistado refleja la situación real.

- Procurar que las preguntas sean concretas y formuladas en un modo que mantenga la atención de la persona que lo completa. Las preguntas demasiado extensas no son convenientes.
- Formular las preguntas a los investigadores que participan del estudio o bien a personas que pudiesen ser representativas de la población que será entrevistada. Esta actividad ayudará a determinar cuál es la percepción que las personas tienen efectivamente de lo que nos interesa conocer.
- Realización de un estudio piloto de campo en una muestra acorde con el blanco poblacional a estudiar y posterior análisis para valorar si es necesario reformular el cuestionario antes de su utilización en la muestra definitiva. Puede ocurrir que algunas cuestiones muy correctas para nosotros resulten irritantes para los encuestados.
- En toda muestra de individuos, existen aquellos que exhiben una tendencia a acordar con los enunciados de la encuesta. A efectos de mitigar este error, conviene que los ítems se expresen alternativamente en forma positiva: “mi médico me informa correctamente sobre los riesgos de fumar” y negativa: “la publicidad de las marcas de cigarrillos es habitualmente engañosa”.

Cuando se trata de un estudio que involucra un número importante de personas lo aconsejable es que el llenado del cuestionario no lleve más de 30 minutos. Si es posible conviene valerse de un cuestionario validado y utilizado en investigaciones previas, que por tratarse de un mismo instrumento nos sirve además para contrastar los datos de ambos estudios.

Si se quisiera cotejar las ventajas del cuestionario respecto de una entrevista resulta claro que el primero de ellos es menos costoso de realizar y más fácil de estandarizar dado que las instrucciones escritas reducen el sesgo que puede generar el entrevistador. Por otro lado, la preservación del anonimato aumenta la probabilidad de que el participante responda con veracidad. Respecto de la entrevista, es evidente que a través de ella se puede clarificar la pregunta y obtener información más exacta o aumentar la amplitud de la misma extrayendo respuestas incluso más complejas. Una ventaja adicional es que de acuerdo al modo en que se va desarrollando la entrevista, el entrevistador puede controlar el orden de las preguntas para favorecer la recolección de la información.

<u>Consideraciones adicionales sobre el entrevistador y la entrevista</u>
En lo que se refiere al primero de los aspectos se debe prestar especial atención a las características de la persona encargada de efectuar la entrevista, tales como la confiabilidad, reserva, puntualidad. Estas condiciones facilitan la tarea a realizar y crean una mayor disposición por parte del entrevistado. Otras cuestiones a tener en cuenta son la edad, género, clase social, y nivel de instrucción del entrevistador. No es conveniente que interrogatorios dirigidos a adultos mayores sean efectuados por jóvenes. Para el caso en que el cuestionario incluye preguntas las cuales podrían plantear conflictos en el entrevistado, por ejemplo el alcoholismo, se suele aconsejar que el entrevistador sea una persona sin relación alguna con la comunidad donde se recaba la información.
A fin de asegurar la reproducibilidad del procedimiento, es necesario estandarizar la técnica para entrevistar, de modo que si un participante fuera entrevistado nuevamente (por la misma persona o bien otro individuo) la respuesta debería ser la misma. Al preparar los criterios para la estandarización tener presente que el sitio y la hora de la entrevista no debe presentar mayores variaciones entre los participantes.

Si bien es poco frecuente que en los estudios basados en cuestionarios se obtenga un 100% de respuestas, es necesario tener en cuenta que la exclusión de este grupo sobre todo cuando la proporción de no respondedores no es despreciable introducirá un sesgo importante al momento de extraer las conclusiones del estudio. Para apartarnos de una situación de esta naturaleza se puede efectuar un estudio piloto a fin de identificar las preguntas con mayor probabilidad de no respuesta. A partir de este dato se procede a explicar debidamente a los participantes sobre los objetivos del estudio o bien elegir entrevistadores que respetando las formas sean lo suficientemente hábiles para obtener la información a recabar.

También debe atenderse al modo de presentación en la revisión final del instrumento. Existen 3 maneras de recolección de datos: (1) auto-completado donde el entrevistado completa el formulario sin interactuar con el entrevistador; esto se presenta cuando las personas a interrogar presentan un nivel de educación aceptable y el cuestionario se entrega en un sitio prefijado; (2) diálogo telefónico entre el entrevistador y el entrevistado; (3) persona a persona. La elección del modo puede afectar la calidad de los datos. En el primero de los casos, las instrucciones deben ser claras, y fáciles de seguir como para prescindir de asistencia. Dado que el entrevistado no puede preguntar para mayor detalle, puede interpretar las preguntas a su manera. Si la pregunta es muy complicada o demasiado técnica, pueden generarse errores y algún grado de frustración del entrevistado. Cuando se incluye al entrevistador las explicaciones facilitan el entendimiento del real significado de la pregunta. El encuentro persona a persona provee de una excelente oportunidad para que el entrevistador utilice recursos recordatorios.

Para no perder de vista las dificultades con las cuales podemos tropezar conviene tener presente que el significado de un cuestionario no es lo mismo para el investigador y el participante. El docente-investigador persigue efectuar una presentación, publicación que redundará en un beneficio en su posición académica, informe favorable para el Programa de Incentivos, carrera docente, obtención de un subsidio, etc. Por la sobrecarga de actividades a que se halla sometido el colega, en algunos casos la tarea de recolección de la información o prueba piloto podría estar en manos de jóvenes entusiastas no muy bien entrenados en el tema. Por su parte, quien completa el cuestionario puede hacerlo por varias razones, un interés real de contribuir a un mejor conocimiento del problema, una cierta presión a colaborar, agradar al investigador (mucho más si es su docente) y a veces un modo de atemperar el aburrimiento.

Sin duda, un buen cuestionario parte de una hipótesis de trabajo que ha sido cuidadosamente estudiada y sopesada. La discusión del problema con los colegas y expertos en la materia es esencial para el desarrollo de buenas preguntas. Las mismas deben ser revisadas, corregidas, y probadas reiteradamente. Examinar al cuestionario en su totalidad es un paso crítico en el diseño del mismo.

1.10 Tipos de hipótesis

Si bien no es posible ajustarse a una única definición del término, en líneas generales se trata de una supuesta relación entre predictores y fenómenos de interés o bien de una explicación anticipada de éstos últimos a ser verificada. También es el nexo entre teoría e investigación y un interrogante cuya respuesta conducirá al planteamiento de nuevas hipótesis. Desde el plano relacional existen hipótesis de asociaciones causales (los valores que toma una variable dependen de la otra variable) o no. En este último caso, el comportamiento de una variable influye sobre la variación de la otra, o bien existe una relación de covariación; es decir una determinada correspondencia entre el comportamiento de ambas variables.

Como ya fuera mencionado a principios del capítulo, existen estudios observacionales meramente descriptivos que no parten o no requieren de una hipótesis de trabajo como la entendemos habitualmente, ya que el propósito de dicha exploración es identificar las características de un grupo poblacional, por ejemplo el hábito de ingerir alimentos conservados entre los pacientes con un determinado desorden metabólico. Por el contrario, los estudios que comparan variables con alguna vinculación fisiopatogénica, tal como determinar si existen mayores antecedentes familiares de tuberculosis entre los individuos que padecen esta enfermedad, parten indefectiblemente de una hipótesis.

Hipótesis nula y alternativa

La primera indica que no existe relación alguna entre la variable independiente y dependiente. Partiendo de la proposición de no asociación, las pruebas estadísticas determinarán la probabilidad de que la detección de una eventual asociación sea simplemente por chance. La hipótesis alternativa que sustenta los objetivos del estudio no puede aceptarse directamente sino por exclusión, vale decir cuando la prueba estadística nos autoriza a rechazar la hipótesis nula.

La hipótesis alternativa de una cola especifica la dirección de la asociación entre las variables independiente y dependiente, mientras que la hipótesis a dos colas sólo establece que existe una asociación, pero no especifica la dirección de la misma. En muchas ocasiones el investigador presume de antemano el sentido de la asociación en caso de que pueda ser detectada.

Error tipo I y error tipo II

En base a los conceptos vertidos precedentemente, el lector habrá advertido que la meta fundamental de la metodología de la investigación científica es reducir el error a su mínima expresión posible. Aún cuando hayamos prestado especial atención a los aspectos referidos al error aleatorio y al error sistemático, puede darse que los resultados de una investigación no reflejen lo que en realidad está ocurriendo en la población, y de ese modo extraer conclusiones desacertadas por no haber tenido en cuenta dos errores adicionales, el tipo I y el tipo II.

El error tipo I se presenta en el caso en que el investigador rechaza la hipótesis nula (HN) cuando en realidad es cierta. Una manera alternativa de definir a este error es la probabilidad de hallar que dos grupos son diferentes cuando en realidad no lo son (falso positivo). El otro error, tipo II o beta tiene a lugar cuando no podemos rechazar la HN, no obstante que es falsa. Dicho en otros términos, es la probabilidad de encontrar que dos grupos son iguales cuando en realidad son distintos (falso negativo).

Más adelante veremos que para el caso de las mediciones, el error sistemático puede llevar a falsos positivos o falsos negativos, pero esto no tiene nada que ver con el error tipo I o tipo II.

Para una mejor ilustración de los conceptos vertidos, la Tabla 1.4 brinda una confrontación teórica entre lo que ocurre en realidad (y desconocemos) y los hallazgos de la investigación. Partiendo de la premisa de que la HN sostiene que no existe asociación entre las variables predictiva y de salida, en la situación [a] los investigadores han rechazado equivocadamente la HN (error tipo I), mientras que en la situación [b] la misma ha sido aceptada como correcta cuando en realidad no lo era.

Tabla 1.4 Comparación entre lo que efectivamente se da en la población y los resultados de un estudio

		Variables predictiva y de impacto asociadas	**Variables predictiva y de impacto no asociadas**
Resultados del análisis Estadístico	Rechazo de la hipótesis nula	Decisión correcta	Error tipo I [a] no deberíamos rechazarla
	Aceptación de la hipótesis nula	Error tipo II [b], no deberíamos aceptarla	Decisión correcta

La probabilidad de cometer el error alfa (tipo I) es lo que mide el nivel de significación estadística (valor de *p*). Esto es, si establecemos un valor de $p = 0.05$, la máxima probabilidad de rechazar incorrectamente la hipótesis nula será del 5%.
El error beta (la probabilidad de cometer el error tipo II) tiene que ver con aceptar lo que en realidad no es cierto; $1 - \beta$ define el poder del estudio[2] y representa la posibilidad de evitar un resultado falso negativo. Podríamos definirlo como la probabilidad de detectar una diferencia ($p<\alpha$) asumiendo que la misma existe. Si el error β se establece al 0.10 estamos aceptando que trabajaremos con una chance del 10% de no encontrar la diferencia. Si fuésemos indagados de cuán seguros estamos de que esa diferencia no existe, la respuesta es un 90% de certeza. Idealmente uno debería trabajar con error = 0, pero eso requeriría emplear muestras demasiado grandes. Generalmente se parte fijando el error $\alpha=0.05$ y el error $\beta=0.20$; aunque puede no ser así. Suponiendo que deseamos introducir un nuevo tratamiento para una patología infecciosa donde existe una terapéutica antibiótica efectiva y con pocos efectos colaterales, la probabilidad de cometer el error alfa de este ensayo clínico deberá fijarse en el 0.01 para reducir la chance de un resultado falso positivo. Podríamos ubicar al error beta por debajo de 0.80 puesto que el nuevo tratamiento no nos interesa tanto, dado que ya contamos con uno muy satisfactorio.
En un sentido más amplio puede afirmarse que el poder de un estudio, en términos de la capacidad de detectar el efecto que plantea la hipótesis, no sólo tiene que ver con el error beta sino también con el peso o magnitud del efecto y el nivel de probabilidad por la cual esa diferencia será estadísticamente significativa. Ambos conceptos se tratan a continuación

Magnitud del efecto

La probabilidad de hallar una asociación entre las variables predictiva y de impacto dependerá de cuán grande es esa relación en la población de referencia. Suponiendo que el hábito de fumar condiciona un riesgo 10 veces aumentado de desarrollar enfermedad coronaria, será relativamente fácil detectar dicha asociación en la muestra seleccionada. La dificultad frecuente para el investigador es que en muchos casos no tenemos elementos como para establecer un valor aproximado de la magnitud del efecto. Por el contrario es eso lo que precisamente queremos saber. Planificar el estudio sin tener a mano una idea aproximada sobre este dato tampoco es posible porque es un elemento indispensable al momento de calcular el tamaño muestral. Si se hallan disponibles se

[2] El poder del estudio también se pone de manifiesto al indicar el intervalo de confianza (IC, ver capítulo 4) de la variable estimada, puesto que un IC muy amplio en torno al efecto calculado pone de relieve el bajo poder del mismo. De ese modo evitaremos aquello de que la "ausencia de evidencia no es evidencia de la ausencia"

pueden utilizar datos de estudios relacionados o pruebas pilotos. Existen ejemplos donde previamente al estudio en sí, se tuvo que efectuar un ensayo preliminar para tener una idea de la magnitud del efecto.

Puede darse una situación donde el investigador piense que existe un factor de confusión -FC- (en inglés "*confounder*") que incida sobre la asociación que se pretende establecer. Un "*confounder*" es un factor extrínseco que está vinculado a la variable independiente (predictiva) y que tiene efecto por sí mismo sobre la variable de impacto o resultante. El FC puede aumentar, disminuir o cambiar el sentido de la asociación. Para clarificar la cuestión, tomemos como ejemplo hipotético sobre un estudio en que los investigadores quieren determinar si la presencia de estrés modifica el curso de una enfermedad, dejando de lado que el estrés puede estar asociado con una mayor ingestión de alcohol, el cual también modifica la enfermedad en cuestión. Es necesario diferenciar, sin embargo, que cuando la variable predictiva ejerce algún efecto sobre el FC, éste no sería un "*confounder*" sino un eslabón en la cadena de vínculos causales. Valiéndonos de otro ejemplo hipotético, supongamos que deseáramos evaluar el efecto tóxico potencial del medicamento X sobre la producción de una anemia hemolítica, enfermedad que por otro lado está en parte vinculada a la presencia aumentada de anticuerpos anti-eritrocitos, que de por sí existen en muchas personas, el supuesto "*confounder*". De contar con evidencia experimental en el sentido de que el fármaco X aumenta significativamente los niveles de estos autoanticuerpos, el "*confounder*" será más bien el (o uno de los) mecanismo(s) que median el desarrollo de la anemia.

Planificar un estudio incluye prever los FC de antemano actuando desde el mismo diseño del estudio. La manera de controlar los FC es por medio de la randomización, la restricción y la comparatividad. Del primero de ellos nos ocuparemos al referirnos a los ensayos clínicos.

La restricción debe aplicarse en los criterios de inclusión para eliminar el FC o reducirlo sustancialmente. Si la variable de impacto también estuviese afectada por la edad, se puede restringir el rango de edades como para tener un control aceptable sobre el FC. Al adoptar esta decisión debemos ser conscientes que se reducirá la disponibilidad de candidatos potenciales para ser incluidos. Por otro lado, tampoco podremos determinar los eventuales cambios de la variable de impacto según la variación del FC.

A diferencia de la restricción donde se circunscribe al FC, la comparatividad permite que el mismo pueda variar siempre que la distribución de esta variación no difiera entre los grupos. Retomando el ejemplo de la edad podremos trabajar con un rango mucho más amplio respetando la condición de que no existan diferencias intergrupales. Una desventaja de esta modalidad reside en la dificultad que generalmente se presenta cuando son varios los factores a tener en cuenta en la comparatividad. Supongamos que existieran dos FC que necesitamos controlar, tales como género (2) y edad (en 4 categorías), lo que arrojará 8 combinaciones posibles para tener en cuenta. La comparatividad es un recurso habitualmente empleado en los estudios observacionales con un tamaño muestral más reducido (casos y controles) y en los ensayos clínicos. Se lo utiliza mucho menos en las investigaciones en gran escala (cohortes), dado las dificultades en términos de tiempo, dinero y sujetos que abandonan el estudio con el transcurrir de los años; factores todos ellos que terminan afectando a la comparatividad. El capítulo 4 referido al tratamiento de los datos, brinda información acerca de los métodos estadísticos para evaluar el efecto del FC.

Cálculo del tamaño muestral

Tras habernos referido a los tipos de hipótesis, los errores que nos llevan a aceptar como verdadero o falso situaciones que en realidad no lo son y la magnitud del efecto,

podemos abocarnos a determinar el número de individuos que se requiere para llevar a cabo nuestra investigación. Los elementos a tener en cuenta para estimar el tamaño de la muestra son los siguientes:

- Establecer la hipótesis nula y la alternativa (uni o bilateral, también llamadas una o dos colas)
- Determinar con el mayor certeza posible la magnitud del efecto
- Fijar el error α y el error β
- Seleccionar la prueba estadística que utilizaremos
- Proceder al cálculo utilizando las tablas apropiadas o algún programa estadístico.

A manera de guía simplificada, la Tabla 1.5 indica los procedimientos estadísticos que se utilizan con mayor frecuencia. Como puede advertirse la elección de la prueba estadística tiene que ver con las características de las variables predictiva y de impacto.

Tabla 1.5 Tipos de pruebas estadísticas que se requieren para el cálculo del tamaño muestral

Variable Predictiva	**Variable de impacto (o variable resultante)**	
	Dicotómica	Continua
Dicotómica	Tabla de 2x2	Prueba de la *t*
Continua	Prueba de la *t*	Correlación

Si el investigador tuviera fundadas razones para reducir el error alfa, indefectiblemente deberá incrementarse el tamaño muestral. En líneas generales una reducción del error alfa desde 0.05 a 0.01 ocasiona un incremento del 70% en el tamaño muestral (si el poder del estudio es 0.50) o del 50% si el error beta = 0.20. La Tabla 1.6 provee una idea aproximada del tamaño muestral con que se deber trabajar según el margen de error establecido. Esto es en términos relativos y a los fines ilustrativos, ya que la tabla no contempla el tercer elemento que se requiere para el cálculo exacto del tamaño muestral que es la magnitud del efecto.

Tabla 1.6 Tamaño relativo aproximado de un ensayo clínico según los diferentes niveles de error alfa y beta.

Error tipo 1 (α)	**Poder del estudio (1-β)**			
	0.50	*0.80*	*0.90*	*0.99*
0.05	100	200	270	480
0.01	170	300	390	630
0.001	280	440	540	820

Estimación del tamaño muestral para el caso de una variable continua

Supongamos que un investigador quiere determinar la ventaja de un hipotensor sobre otro para el tratamiento de la hipertensión arterial. La variable dependiente es la presión sistólica (PS). La información previa establece que la PS media en los hipertensos tratados es de 150 ± 30 mmHg (DS, desvío estándar). El investigador pretende detectar una diferencia del 10% más (es decir una mejoría del 10% sobre esa variable).
La hipótesis nula dirá que la PS en los tratados con cualquiera de los 2 hipotensores será la misma.
La magnitud del efecto = 15 mmHg (10% x 150 mmHg)
La magnitud del efecto estandarizado (magnitud del efecto / DS)= 0.5
Error $\alpha = 0.05$ (bilateral o a dos colas), error $\beta = 0.20$
Si observamos en la tabla 1.7 se advierte que necesitaremos 63 pacientes por grupo.

Tabla 1.7 Número de observaciones necesarias por cada grupo para comparar medias de variables continuas en base a la *t* de Student

Magnitud del efecto estandarizado	Error α 0.05 (unilateral) Error α 0.10 (bilateral)		Error α 0.025 (unilateral) Error α 0.05 (bilateral)		Error α 0.005 (unilateral) Error α 0.01 (bilateral)	
	β=0.10	β=0.20	β=0.10	β=0.20	β=0.10	β=0.20
0.10	1713	1237	2102	1570	2977	2337
0.20	428	309	526	393	744	584
0.30	190	137	234	174	331	260
0.40	107	77	131	98	186	146
0.50	69	49	84	63	119	93
0.60	48	34	58	44	83	65
0.70	35	25	43	32	61	48
0.80	27	19	33	25	47	36
0.90	21	15	26	19	37	29
1.00	17	12	21	16	30	23

Ejemplo de cálculo del tamaño muestral para una variable dicotómica

Imaginémonos ahora que la cuestión a estudiar es si la incidencia de patología bronquial inflamatoria en adultos con marcada exposición a contaminantes medioambientales es superior a la de los menos expuestos. La información disponible indica que para la región donde se prácticará el estudio la incidencia de enfermedad bronquial en los últimos 4 años es del 0.10. Los investigadores suponen que la alta exposición a los polutos ambientales podría duplicar esa incidencia.
La incidencia en los no expuestos (P2) es 0.10 y en los expuestos (P1) = 0.20 y la diferencia entre P1 y P2 = 0.10. Los investigadores fijan un error $\alpha = 0.05$ (unilateral o una cola) y un error $\beta = 0.20$. Si nos ubicamos en la Tabla 1.8 podemos ver que para el menor de los valores (en este caso P2=0.10) y una diferencia esperada de 0.10 necesitaremos 156 personas por grupo.
La decisión de utilizar un error α a una sola cola está fundamentada en la presunción que la contaminación ambiental incrementa el riesgo de bronquitis.

Tabla 1.8 Número de observaciones requeridas para la comparación de 2 proporciones (por razones de espacio la tabla está truncada y la diferencia entre P1 y P2 llega hasta 0.50)[3]

Valor menor de P1 y P2	Diferencia esperada entre P1 y P2				
	0.10	0.20	0.30	0.40	0.50
0.10	156 199	48 62	25 31	15 19	10 13
0.20	231 293	64 81	30 38	17 22	11 14
0.30	280 356	73 93	33 42	18 23	11 14
0.40	305 387	76 97	33 42	17 22	10 13
0.50	305 387	73 93	30 38	15 19	- -

P1 y P2 representan la proporción de personas en quienes se espera que presenten el fenómeno de interés, en los grupos 1 y 2 respectivamente.
Fila superior, Error α 0.05 (unilateral) o 0.10 (bilateral) y Error β 0.20
Fila inferior, Error α 0.025 (unilateral) o 0.05 (bilateral) y Error β 0.20

En la actualidad las tablas para el cálculo del tamaño muestral están siendo sustituidas por programas de computación que facilitan la tarea (por ejemplo el EPI INFO). También existen sitios en la web donde se puede acceder a diversos programas que calculan tamaños muestrales, por ej.: http://www.fisterra.com/mbe/investiga/index.asp
Otro factor a tener en cuenta en el cálculo del tamaño muestral es la intención de efectuar análisis interinos, para revisar los resultados que se están obteniendo sea cada 6 meses o anualmente. La fundamentación de efectuar comparaciones preliminares tiene que ver con el hecho de que si detectara la superioridad de un tratamiento, se puede finalizar inmediatamente el estudio y evitar que un grupo de pacientes sea sometido a un régimen terapéutico de menor eficacia. En estudios con intervenciones de 2-4 años de duración puede ser conveniente efectuar, en el ínterin, dos análisis más las comparaciones finales. En estos casos el tamaño muestral se incrementa alrededor de un 15% y el nivel de significación estadística se ubica generalmente en el 0.01.
A manera de recomendación final tener siempre presente si los recursos disponibles son adecuados para el número de personas que debemos reclutar en el estudio.

[3] Tablas adaptadas de Hulley & Cummings. Designing clinical research. Williams & Wilkins, Baltimore, 1988.

Bibliografía Recomendada

Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.

Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134: 663-694.

Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ 2004; 328: 1312-1315.

Boynton PM. Administering, analyzing, and reporting your questionnaire. BMJ 2004; 328: 1372-1375.

Bradburn N, Rips L, Shevell S. Answering autobiographical questions: the impact of memory and inference on surveys. Science 1987; 239: 157-161.

Detsky AS, Sackett DI. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. Arch Intern Med 1985; 145: 709-712.

Gordis L. Epidemiology. WB Saunders Co, Filadelfia 1996.

Grimaldi DA, Engel MS. Why descriptive science still matters. Bioscience 2008; 57:646-647.

Herrmann D. Reporting current, past, and changes in health status: what we know about distortion. Medical Care 1995; 33: AS89-AS94.

Hulley SB, Cummings SR. Designing clinical research. Williams & Wilkins, Baltimore, 1988.

Kelsey JL, Thompson WD, Evans AS. Methods in observational Epidemiology. Oxford University Press, Nueva York 1986.

Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Controlled Clin Trials 1981; 2:93-114.

Lipid research clinics program: recruitment for clinical trials. Circulation 1982; 66 (Suppl IV): 1-78.

Mc Donald CJ, Thierney WM. Research uses of computer-stored practice records in general medicine. J Gen Intern Med 1986; 1: S19-S24.

Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports or parallel-group trials. Lancet 2001; 357: 1191-1194.|

Piantadosi S. Clinical trials: a methodologic perspective. New Cork, John Wiley and Sons, 1997.

Pocock S. Clinical Trials. A practical approach. John Wiley & Sons, Londres, 1992.

Schoenbach VJ, Rosamond WD. Understanding the fundamentals of epidemiology, an evolving text. Fall 2000 Edition, NC, EEUU.

Schwarz N, Strack F, Hippler H, Bishop G. The impact of administration mode on response effects in survey measurement. Appl Cognitive Psychology 1991; 5: 193-212.

Smith HO. Guidelines for Clinical Trials. TDR-WHO, Ginebra 1992.

Streiner DL, Norman GR. ''Precision'' and ''Accuracy'': Two Terms That Are Neither. J Clin Epidemiol 2006; 59: 327-330.

Wahl RA, Sisk DJ, Ball TM. Clinic-based screening for domestic violence: use of a child safety questionnaire. BMC Medicine 2004; 2: 25.

Willis G, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. Applied Cognitive Psychology 1991; 5: 251-267.

CAPITULO 2. Estudios observacionales

Una gran parte del conocimiento médico que hoy poseemos se debe a los estudios observacionales efectuados a lo largo de las centurias. Los comienzos de la epidemiología como disciplina se remontan a mediados del siglo XX con las investigaciones sobre patologías que estaban cobrando una desusada relevancia como la enfermedad coronaria y el cáncer de pulmón. Los esfuerzos dieron lugar a estrategias muy fructíferas como lo son los estudios de cohortes y de casos y controles. Es claro, sin embargo, que los orígenes de la epidemiología hunden sus raíces mucho más atrás como un intento de la Medicina para detectar causas de enfermedades y así implementar de medidas de prevención y tratamiento. Fue crucial por ejemplo el estudio de John Snow sobre la epidemia de cólera en Londres y el indicio que la enfermedad se contraía al beber agua contaminada; lo cual proveyó del puntapié inicial para los posteriores descubrimientos en cuanto al rol de los microbios como causa de muchas enfermedades.
Mucho más recientemente, el estudio Framingham implementado hacia finales de los 40 sigue considerándose como el modelo de investigación comunitaria y no hay dudas que el diseño de cohorte es esencial en la investigación epidemiológica observacional. Por su parte el estudio de casos y controles también es útil para la prueba de hipótesis, aunque por las limitaciones inherentes a su formato algunos expertos lo consideran inferior al anterior (Epidemiol Rev 2004). De todos modos el fundamento metodológico y valor del estudio de casos y controles ha quedado firmemente establecido (Epidemiol Rev 1994). Le corresponden a estos estudios el haber brindado pruebas sólidas en cuanto a tabaco y cáncer de pulmón como así también dietilestilbestrol y adenocarcinoma de vagina. Hoy en día también se los aplica para analizar los beneficios y riesgos de tratamientos ya instituidos.
Los datos aportados por los estudios observacionales proveen una información muy valiosa al momento de delinear políticas de salud. Sin embargo y a raíz de que estas investigaciones enfocan cuestiones complejas respecto de las causas de la enfermedad, algunos autores cuestionan la validez de los resultados habida cuenta que también pueden participar otros predictores no contemplados durante su realización. Así las cosas resulta difícil lograr un consenso entre los epidemiólogos respecto del rol de la disciplina en el ejercicio de establecer pautas de prevención y control. Mientras algunos se abstienen de tomar parte en dicha controversia, otros por el contrario participan activamente de dichos debates. A pesar de la discusión, el uso de la epidemiología en políticas de salud es creciente, acorde con los principios de la medicina basada en la evidencia. El proceso que va desde la formulación de la hipótesis hasta el establecimiento de políticas no es simple ni mucho menos lineal. En el área de las enfermedades infecciosas los hallazgos se traducen rápidamente en acciones; en otros casos la evidencia se va obteniendo gradualmente (por ejemplo, cáncer y dieta) y el logro de un nivel crítico de certeza lleva más tiempo. Es obvio que la investigación y las acciones de políticas de salud son interactivas e iterativas, por lo que la segunda evoluciona en función de los cambios que brinda la primera.
Dada la relevancia de este tipo de investigaciones, en las secciones subsiguientes se expondrán los lineamientos centrales en los que se encuadran estos estudios.

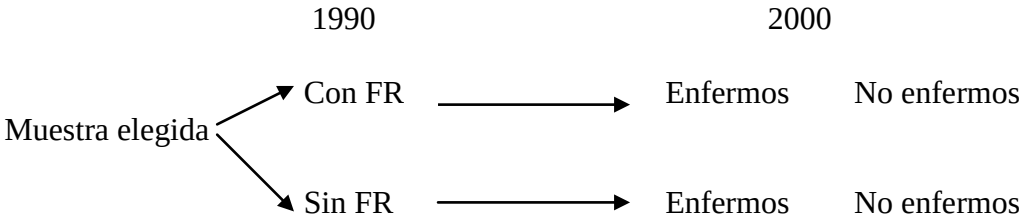
2.1 Estudios de cohortes

Las investigaciones de este tipo son aquellas en las que se efectúa un seguimiento de los participantes durante un período de tiempo estipulado. Ello puede ser a los fines descriptivos, tal como estimar la incidencia de alguna anormalidad o bien analíticos. Vale decir, determinar si existe asociación entre un supuesto factor de riesgo (FR) con una situación patológica, dado que los individuos se hallan libres de enfermedad al

momento de la inclusión. Una de las ventajas de los estudios de cohortes es que permiten examinar incluso múltiples efectos de una sola exposición. La limitación más importante consiste en el costo operativo que acarrean estos estudios, lo cual implica que la fundamentación para llevarlo a cabo debe apoyarse en alguna evidencia previa en favor de la hipótesis. En muchos casos la justificación se basa en los resultados de un estudio de casos y controles "*ad-hoc*" practicado con antelación.
Los estudios de cohortes se clasifican generalmente en prospectivos y retrospectivos, los cuales dependen de la relación entre el momento en que se inicia el estudio y la aparición de la enfermedad.

Tipos de cohortes

Estudios prospectivos: en este caso seleccionamos a un grupo de individuos que no presentan determinada anormalidad en cuestión, para analizar el (los) factor(es) de riesgo que podrían ejercer alguna influencia sobre el desarrollo de la misma. Luego de un período de tiempo estipulado constataremos quiénes desarrollaron la enfermedad y quiénes no, en función de la presencia o ausencia del FR.



Es un método muy poderoso para investigar las potenciales causas de una situación patológica, dado que el supuesto factor causal es estimado antes de que ocurra el fenómeno. Esa secuencia temporal fortalece la inferencia de que el FR tenga que ver con los determinantes de la enfermedad.
La mayor dificultad que ofrece este tipo de estudio es económica, más aún cuando la incidencia de la condición es baja y nos obliga a trabajar con una muestra demasiado grande. Otra limitación está referida a la existencia de algún "*confounder*"[4], el que nos llevaría a una conclusión errónea de no tenérselo en cuenta. Supongamos que nuestra propuesta es determinar la participación del hábito de sedentarismo sobre la patología coronaria, sin determinar si los que realizan actividad física o no difieren en la proporción de fumadores. Un problema que también puede plantearse es cuando la patología en estudio tiene una larga fase pre-clínica. En este caso se deben utilizar pruebas diagnósticas de alta sensibilidad o bien prolongar el tiempo de seguimiento.

Estudios retrospectivos: el diseño de este tipo de cohorte es parecido al anterior, ya que se trata de un grupo de sujetos sometidos a un seguimiento con evaluación de los FR en la inclusión y la subsiguiente aparición de la enfermedad, luego de un lapso de tiempo. La diferencia reside en que todo el proceso tuvo lugar en el pasado. Es útil en las enfermedades con largo período de incubación. La cuestión reside en que estos estudios sólo pueden hacerse si se tiene acceso a archivos donde consta toda esa información, lo cual no es siempre posible, dado que las historias clínicas no necesariamente suelen contener datos sobre los FR y "*confounders*" que sean objeto de nuestro estudio. Lo cierto es que si la información de interés está disponible, los diseños retrospectivos son

[4] También se los suele denominar factor de confusión (FC) o confusor

más fáciles de llevar a cabo y por supuesto mucho menos costosos. La limitación es que no tenemos control sobre la naturaleza y calidad de las mediciones que se efectuaron.
Recientemente se publicó un estudio efectuado en Finlandia que trataba de identificar el potencial efecto protector de la vacuna BCG sobre el desarrollo de alergias. Dado que en ese país existe un Sistema Nacional de Salud los investigadores fueron capaces de rastrear los antecedentes de adolescentes de 20 años (alérgicos o no) respecto de la vacunación con BCG en la primera infancia y las sucesivas consultas en servicios de pediatría.

Estudios en ambas direcciones: este tipo de diseño se aplica en los casos en que la variable predictiva puede tener efectos tanto a corto como largo plazo. Supongamos que existiera una sospecha bastante fundada de que un compuesto químico X no sólo produce efectos tóxicos a poco de tomar contacto con el mismo sino que también puede generar la aparición de neoplasias tras varios años de ocurrida la exposición. Para indagar sobre este problema se puede confeccionar una cohorte de individuos expuestos y no expuestos en quienes se analiza en forma retrospectiva los efectos tóxicos del corto plazo, estableciéndose un seguimiento a largo plazo para estimar la incidencia de diversos tipos de cáncer.

Estudios de cohorte doble: se utilizan para las enfermedades ocupacionales y medioambientales. En situaciones de este tipo se trabaja con dos muestras, una perteneciente a la población expuesta y la otra con un nivel de exposición bajo o nulo. Una variante de este tipo de estudio, es comparar los datos de los individuos expuestos con los resultados disponibles en un registro (controles externos).

Cohortes con un estudio de casos y controles anidado: este tipo de estrategia suele utilizarse para recabar datos cuyo procedimiento de recolección es oneroso y no puede practicarse en todos los participantes. Tomemos un caso hipotético en donde se desea saber si las personas infectadas con un determinado retrovirus tienen un riesgo aumentado para el desarrollo de un linfoma de células B. En el diseño prospectivo habitual uno debería obtener muestras de sangre al momento de la inclusión para el diagnóstico de la infección y esperar un período variable de tiempo, para luego analizar si existen diferencias en cuanto a la incidencia de este tumor entre los infectados y los no infectados. Si la realización de las determinaciones virológicas implica un costo imposible de afrontar, la alternativa es efectuar un estudio de casos y controles anidado en una cohorte que por su conformación ofrezca la posibilidad de contestar nuestra pregunta. En el momento en que los individuos son admitidos en el estudio se toman las muestras de sangre y el material procesado (plasma, células, ácidos nucleicos) se congela para su posterior utilización. Una vez que se ha detectado un número adecuado de casos, se confecciona una muestra de controles extraídos de la misma cohorte y en ambos grupos se efectúan los estudios virológicos planificados al inicio.

Elección de los participantes.

Los criterios que se establecen para la incorporación de los voluntarios están directamente vinculados a los propósitos del estudio, sin perder de vista la factibilidad de llevar a cabo efectivamente las evaluaciones. Generalmente se incluyen personas con mayor posibilidad para ser seguidas, entre quienes el FR es relativamente frecuente. Para el caso en que la variable predictiva está referida a factores medioambientales u ocupacionales, no cabe duda que la muestra debe extraerse de la población en la cual se produce dicha exposición. Como se observa en la literatura científica, una buena

proporción de los estudios sobre cohortes se efectúa en personas con ocupaciones estables, por la mayor probabilidad de obtener datos completos y confiables. Una investigación muy conocida a través de las publicaciones médicas se llevó a cabo en la comunidad de Framingham precisamente por una serie de condiciones socioeconómicas que favorecían el seguimiento de los participantes del estudio.

Con excepción del FR en estudio, es esencial que el grupo control sea lo más comparativo posible respecto del grupo de personas expuestas. Algunos estudios parten de una sola cohorte en la que los participantes son clasificados en diferentes categorías de exposición, donde las personas no expuestas constituyen el control interno. Sin embargo, es necesario tener presente que cuando se analizan varios FR el grupo control deberá estar constituido por personas que no denotan exposición a ninguno de los factores en consideración. Como ya fuera señalado, en los estudios donde se investigan agentes nocivos ocupacionales es necesario incluir un control externo (cohorte doble), ya que todos los miembros de la cohorte original tienen un mayor o menor grado de contacto con el FR. Este grupo control externo se conforma a partir de la población donde residen las personas expuestas. Sin embargo, no perder de vista que dicho procedimiento tiene sus limitaciones puesto que algunos miembros de población general pueden hallarse expuestos al mismo factor. Podría resultar, asimismo, que el grupo conformando terminara no siendo comparable a la cohorte original no sólo en lo que hace a las características demográficas sino también a la presencia de otro FR. Si el grupo control no resulta muy adecuado (¡y el subsidio es millonario!) se deberá conformar por lo menos un grupo adicional; de persistir la diferencia con cualquiera de las comparaciones efectuadas, la validez del estudio resultaría notoriamente reforzada.

Registro de la información

Al igual que con la elección de los participantes, los estudios de cohortes requieren una pormenorizada planificación respecto a la manera en que se recogerá la información vinculada, tanto a la exposición como con el fenómeno de interés. Para el caso de los FR, los datos pueden ser provistos de varias formas:

- Historias clínicas o archivos personales (que las empresas tienen de sus empleados). Esta manera de recabar información es menos costosa y brinda una objetividad aceptable en cuanto al modo que se consignaron los datos, dado que la enfermedad aún no se ha producido. Puede ocurrir, sin embargo, que los archivos no consignen información sobre todos los FR de interés o bien no existan datos sobre los "*confounders*" que pueden afectar la validez del estudio.
- Entrevistas y cuestionarios. Estos procedimientos pueden ser la única fuente de provisión de datos o bien constituir un complemento de la información que no es posible obtener a través de los registros personales. Si el caso lo requiere, los investigadores pueden armar un cuestionario que incluya además un examen físico o estudio complementario para recabar algún dato en particular.
- Mediciones en fluidos. Esto se emplea generalmente en los estudios donde se pretende evaluar el efecto de algún contaminante presente en el aire o el agua.
- Evaluaciones periódicas. Las mismas se utilizan en el caso de las cohortes con un largo período de seguimiento. El objetivo de este procedimiento es determinar si el nivel de exposición ha experimentado cambios, a la vez de indagar sobre la aparición de nuevos factores o la existencia de situaciones que inicialmente no se consideraron riesgosas.

En lo concerniente a la variable de impacto, la información puede provenir de diversas fuentes. Independientemente del criterio que se adopte, el procedimiento debe ser

exactamente el mismo, tanto para los expuestos como los no expuestos. Las principales fuentes de datos se detallan a continuación:

- Certificados de defunción. Este documento adquiere valor cuando el fenómeno de interés es la mortalidad ocasionada por enfermedades que generalmente se consignan como la causa primaria del deceso, por ejemplo cáncer.
- Datos consignados en las historias clínicas, registros de organismos gubernamentales, archivos de las obras sociales. Estos procedimientos son útiles en los estudios donde no está implicada la mortalidad como variable de impacto.
- Exámenes médicos periódicos. A pesar del costo y el tiempo demandado por este tipo de estrategia, es obvio que la utilización de medios objetivos y estandarizados de diagnóstico asegura una alta confiabilidad en la información obtenida.

Cuando se haya establecido el modo de recolección de los datos, el paso siguiente será fijar pautas para completar el seguimiento de la mayor parte de los participantes. Como veremos a continuación la falta de seguimiento de las personas expuestas y no expuestas es un sesgo muy importante en los estudios de cohorte.

Factores capaces de afectar la validez del estudio

No obstante que los estudios del cohorte son muy poderosos por su valor inferencial, es necesario tener presente de antemano los factores que pueden restarle valor a las conclusiones de la investigación. Los más destacados se señalan a continuación.

- Sesgo de selección. A diferencia de lo que ocurre en los estudios de casos y controles, este tipo de sesgo es mucho menos probable, ya que al momento de incorporar a los participantes no se conoce quiénes desarrollarán o no la enfermedad. Para el caso en que la cohorte sea retrospectiva tomar las debidas precauciones para no introducir un sesgo de esta naturaleza (ver Sección 2.3, error de muestreo).
- Grado de exactitud con que se categorizan los individuos respecto a la presencia de los FR y a la ocurrencia del fenómeno de interés. Aún cuando se tomen todos los recaudos, es casi seguro que en algunos casos se cometerán errores al clasificar a los participantes sea en cuanto al grado de exposición como a la presencia de la variable de impacto. Cuando la proporción de personas en quienes se efectuó una clasificación incorrecta de ambas variables es similar en ambos grupos (expuestos y no expuestos) decimos que se trata de un error de clasificación aleatorio o no diferencial. El error de clasificación no aleatorio o diferencial tiene a lugar cuando se producen disparidades en el modo en que se recoge la información entre los expuestos y no expuestos. En el caso hipotético de un estudio donde se evalúa el efecto potencial de un pesticida sobre el desarrollo de problemas neurológicos, es probable que la población expuesta efectúe consultas médicas con más frecuencia originando un exceso en la incidencia de tales anormalidades.
- Casos que no completan el seguimiento. La ausencia de información respecto de los controles a los que regularmente son sometidos los participantes, es una situación que en mayor o menor escala se produce en cualquier estudio de cohortes. El punto es que dicha pérdida no alcance proporciones significativas, por ejemplo un 20% o más. Como buenos planificadores, al momento de preparar el protocolo no perder de vista la posibilidad que las variables que pretendemos relacionar condicionen de alguna manera la aparición de pérdidas en el seguimiento. Si el objetivo fuera determinar la eventual influencia de un contaminante químico sobre enfermedades del tracto respiratorio, es muy probable que un número importante de personas abandonen el lugar por la preocupación que origina el estar expuesto a agentes

potencialmente nocivos o bien porque ya han experimentado problemas respiratorios. Cuando el investigador advierte de antemano que el mismo estudio puede determinar abandonos, procederá a preparar un menú de modos alternativos para obtener la información deseada. Entre los más utilizados se pueden incluir los registros de hospitales públicos, de obras sociales, visitas a los vecinos o parientes para ubicar a las personas por diversos medios.

- El efecto de la no participación. El hecho que no todas las personas a las cuales se les ofrece participar del estudio, acepten la proposición, plantea la posibilidad cierta que los participantes difieran en varios aspectos respecto de quienes no lo hacen. Esto no genera mayores problemas, siempre que el investigador no efectúe generalizaciones a partir de los resultados de su estudio. La dificultad surge cuando la posibilidad de no participar pueda guardar alguna relación con el FR o el fenómeno de interés. En ese caso se afectará la validez del estudio, puesto que se llega a una incorrecta valoración del riesgo real. Para reducir esta incertidumbre, algunos investigadores incluyen datos demográficos y características relacionadas con el estudio, pertenecientes a un grupo de no participantes.

Incidencia y riesgo relativo

Los estudios de cohorte describen los cambios de una variable continua que identifica a una enfermedad (por ej. glicemia) o la incidencia de una variable categórica (infarto de miocardio). La representación de los resultados se hace por medio de la estadística descriptiva (medias, proporciones) y la variabilidad de estas estimaciones, tales como el DS, los rangos, e intervalos de confianza. La asociación entre la variable independiente y dependiente se describe como riesgo relativo (RR). Es un coeficiente entre el riesgo de enfermedad en las personas con el FR respecto del riesgo en aquellos que no presentan tal factor. De ello se desprende que cuando el valor del RR es superior a 1, el factor en estudio incrementa el riesgo mientras que si es menor que 1 en realidad se trata de un factor protector.

La manera de calcular el RR se indica en la Tabla 2.1., partiendo del hecho que riesgo es igual al número de enfermos dividido por el número de personas en riesgo o no; el esquema que se representa en la tabla permite definir:

a) el riesgo para los individuos expuestos: a/(a + b)
b) el riesgo para los no expuestos: c/(c + d)

De este modo el RR surge de la siguiente ecuación

$$RR = \frac{a/(a + b)}{c/(c + d)} = \frac{a(c + d)}{c(a + b)}$$

Tabla 2.1 Elementos a tener en cuenta para el cálculo del RR

Factor de Riesgo	**Enfermos**	**No enfermos**	**Total**
Expuestos	a	c	a + b
No expuestos	c	d	c + d
Total	a + c	b + d	n

El ejemplo hipotético que se indica a continuación está referido a una situación simplificada en la que se buscó evaluar la influencia de la altura corporal sobre la

fractura de cadera en mujeres menopáusicas. El estudio se habría efectuado en 5020 mujeres que completaron un seguimiento de 10 años cuyas edades oscilaban entre 55 y 80 años. Además de la edad se tuvo en cuenta el índice de masa corporal, consumo de alcohol y tabaco como potenciales factores de confusión. Para la altura se asignaron dos categorías por debajo y por encima de 1.65 metros.

Tabla 2.2 Altura corporal y fractura de cadera en la muestra analizada

Altura corporal	**Fracturadas**	**No fracturadas**	**Total**
Por encima de 1.65	49	2161	2210
Por debajo de 1.65	32	2778	2810
Total	81	4939	5020

El riesgo de fracturas en las mujeres con mayor altura fue 49/2210=0.02217, mientras que en las de menor altura se obtuvo 32/2810=0.01139. En base a la fórmula provista, el RR es igual a 1.946. Como habrá de tratarse más adelante, existen procedimientos estadísticos que permiten calcular el intervalo de confianza de este RR. En nuestro caso hipotético el límite de confianza del 95% oscila entre 1.25 y 3.03; vale decir que existe un 95% de seguridad que el valor del RR se ubique entre esos valores. Cuando el límite inferior del intervalo es menor a 1, el RR estimado carece de valor dado que no es estadísticamente significativo.
En líneas generales, el RR puede definirse como una probabilidad que estima el número de veces en que se produce el fenómeno de interés, respecto del número total de las observaciones.

Riesgo atribuible

Si bien el RR nos habla de la importancia de un determinado factor sobre el desencadenamiento de una situación patológica, al momento de evaluar el impacto poblacional del mismo, el RR no es el estimador adecuado por cuanto no tiene en cuenta la prevalencia de dicho factor de riesgo. Vale decir que si el FR es poco frecuente no tendrá mayor repercusión en cuanto a la situación poblacional, aún cuando el mismo constituya un riesgo para el desarrollo de la enfermedad. Para tener una idea más cabal de cuán importante es este FR, se debe calcular el riesgo atribuible (RA), lo cual surge de la siguiente fórmula,

$$\mathrm{RA} = \frac{r - r_{\mathrm{E}}}{r}$$

Donde r significa el riesgo general y r_E es el riesgo del grupo no expuesto. Retomando nuestro ejemplo, el riesgo general es 81/5020=0.0161, mientras que el de los no expuestos era 0.01139. De acuerdo a la fórmula provista el riesgo atribuible nos da 0.29. Con ello se puede concluir que casi un 30% de las fracturas se deberían a la mayor altura de las mujeres, en el supuesto caso que el cálculo del RR ajustado para los "*confounders*" contemplados, siguiera siendo significativo y dentro de los mismos valores. Aún así no se puede descartar la existencia de algún otro "*confounder*" capaz de oscurecer la dimensión real de la variable altura, de tal modo que no es conveniente identificar riesgo atribuible a causalidad.

2.2 Estudios transversales

Se trata de un estudio observacional mucho más descriptivo que analítico, en el cual las variables predictiva y de impacto se evalúan al mismo tiempo. A diferencia de los estudios de cohorte donde la información sobre la presencia o no de la variable predictiva (factor de riesgo por ejemplo) se hallaba disponible antes que se produjera la enfermedad, los estudios transversales no siguen esa secuencia lógica puesto que se basan en una sola medición. Aún cuando se hubiera detectado una asociación entre ambas variables, la inferencia de causalidad tiene mucho menos fuerza dado que no es posible determinar si la variable predictiva precedió o fue posterior a la aparición del fenómeno de interés.

Generalmente se lo realiza por mediciones en una muestra aleatoria o muestra aleatoria estratificada de la población, aunque se puede utilizar una muestra por conveniencia. En líneas generales, estos estudios poseen una variable binaria como el quantum de interés primario, por ejemplo la prevalencia de cierta enfermedad o factor de riesgo. También pueden interesarnos las variables continuas, tales como el peso corporal o parámetros laboratoriales susceptibles de una posterior categorización. Los estudios transversales también se pueden llevar a cabo para monitorear cambios poblacionales a lo largo del tiempo por medio de una serie de exploraciones de este tipo. A partir de datos obtenidos de evaluaciones nutricionales en EEUU conducidas entre 1976–1980 y 1988–1991 donde se efectuaron mediciones sanguíneas de plomo, se pudo documentar una caída en los niveles sanguíneos a raíz de la eliminación gradual desde 1976, de la mayor parte del plomo en las naftas [JAMA 1994].

Dado que la distribución de la variable bajo análisis podría cambiar con el transcurso del tiempo, lo ideal es que el estudio transversal se llevara a cabo en forma instantánea. Sin embargo un estudio real requiere de un tiempo para su realización. En la práctica, el requisito básico es que los cambios de la variable bajo estudio sean prácticamente despreciables durante el curso de la investigación. Obviamente que si efectuamos un análisis acerca de la presencia de alguna anormalidad osteo-articular uno puede llevarlo a cabo durante un año con una relativa tranquilidad. Distinto sería si lo que nos interesa es una virosis respiratoria como la influenza.

En líneas generales, los estudios transversales pueden brindarnos información acerca de la prevalencia de la enfermedad o de la variable en estudio. La misma puede definirse como la proporción de la población estudiada que presenta el fenómeno de interés al momento de la medición. La diferencia con incidencia es que ésta mide la proporción de personas que desarrollan la enfermedad en un período de tiempo, por ejemplo en un año. Es obvio que la prevalencia de una enfermedad tiene que ver con la incidencia de la misma; sin embargo no debemos pasar por alto que la prevalencia como valor final también está influenciada por las características intrínsecas de la patología en cuestión, tales como mortalidad, remisiones espontáneas, recidivas o curas por tratamiento.

Supongamos que a partir de datos experimentales donde surge que la exposición a productos nitrogenados originaría una toxidermia en animales de laboratorio, un grupo de investigadores decide efectuar un estudio transversal para determinar si existe una asociación de este tipo en seres humanos. Para ello proponen la realización de un minucioso examen dermatológico a todos los operarios de una planta química, con o sin exposición a dichos compuestos.

La hipótesis de trabajo sugiere que debería existir una mayor proporción de pacientes entre los expuestos, respecto de los que no lo están. De verificarse la asociación, la evidencia sentará la base racional para la realización de un estudio analítico mucho más certero.

Factor de riesgo	Enfermo	No enfermo
Presente	más casos	menos casos
Ausente	menos casos	más casos

Si por el contrario las comparaciones estadísticas no revelaran diferencias en cuanto a la lesión cutánea según la exposición, cabría preguntarse si no han existido factores capaces de afectar a los hechos en estudio. Quizás algunos operarios expuestos al producto químico hayan presentado lesiones cutáneas de poca relevancia que se resolvieron por sí mismas y no referidas por la persona al momento de la exploración. Mucho peor aún: sería el caso de empleados que a raíz de experimentar el problema cutáneo hubieran sido trasladados a secciones de exposición mínima o nula y sus lesiones aún persisten. Estas personas han llegado allí a raíz de su afección, vale decir ¡por el efecto!

Por lo tanto al planear un estudio transversal es necesario cerciorarse que el fenómeno de interés no tenga efecto sobre la supuesta variable predictiva. Cuando exista un alto grado de confiabilidad que esto no es así, podemos seguir avanzando en la propuesta.

Los estudios transversales adquieren el valor de una investigación analítica cuando la variable predictiva es un carácter que no cambia a lo largo de la vida, por ejemplo las investigaciones de asociación entre determinada patología con los genes del complejo mayor de histocompatibilidad.

Aunque un estudio transversal sólo estima lo que ocurre en un determinado momento, a menudo el interés se centra en cuestiones más dinámicas dependientes del tiempo; la incidencia de la enfermedad puede ser un dato que nos interese. Una situación en la cual se puede medir la incidencia de una enfermedad a través de un diseño transversal es ante un evento efímero y de aparición reciente. Si la duración media, *w*, de este evento es conocida, la prevalencia de personas con ese estado puede convertirse en incidencia a través de la relación prevalencia/*w* = incidencia. Para una enfermedad de corta duración, el evento efímero puede constituir el curso entero del mismo. Para que el método sea válido, la escala de tiempo en el cual se producen cambios en la incidencia de la enfermedad debe ser más largo que *w*. Weinstock y col [2002] utilizaron esta estrategia para estimar la incidencia de HIV entre los pacientes de una clínica para ETS. Aunque el método es sensible al valor de *w*, es posible probar o estimar cambios en la incidencia en exploraciones transversales seriadas al comparar la prevalencia de individuos en el evento efímero en el tiempo. El método descrito es apropiado para situaciones en las cuales la incidencia cambia en el tiempo. Por ejemplo es posible estimar la incidencia de una enfermedad al comparar dos mediciones de prevalencia tomadas en diferentes momentos y que dan cuenta del envejecimiento. Así al comparar la proporción de jóvenes de 15 años con los de 19 que poseen una PPD+ con la proporción observada 5 años después (entre los que tengan entre 20 y 24 años), una podría estimar la tasa de infección con *M. tuberculosis*, asumiendo a ambos grupos como miembros de la misma "pseudo-cohorte". Sin embargo, la validez de estas estimaciones depende de la presunción que no existe una pérdida diferencial de las personas con el problema en cuestión (i.e. mortalidad, migraciones). De ser así se requieren datos externos para ponderar dichos efectos.

Por otro lado, para que los distintos grupos etarios sean considerados miembros de la misma pseudo-cohorte, la distribución de las covariables más relevantes (por ejemplo características demográficas) deben ser las mismas en los diferentes grupos, un

requerimiento que implica a veces un desafío en los estudios observacionales. Las conclusiones acerca de la incidencia se pueden extraer a veces al comparar prevalencias crudas (no edad-específica) a través de exploraciones transversales sucesivas. Sin embargo, las tendencias en cuanto a prevalencia pueden ser difíciles de interpretar dado que son el resultado de nuevos casos menos los casos viejos que se pierden o fallecen. Es así que antes de efectuar una inferencia sobre incidencia uno debe poseer un buen conocimiento acerca de la naturaleza de las pérdidas o muertes de los casos prevalentes. Una ausencia de cambio en la prevalencia cruda no implica necesariamente que la situación sea estable. La línea divisoria entre un estudio transversal y estudios retrospectivos es vaga cuando los datos longitudinales retrospectivos se obtienen de una exploración transversal.

La validez de los datos en retrospección depende de la capacidad de recolectar adecuadamente los eventos de interés por el período de tiempo en consideración.

A diferencia de los estudios de cohorte donde se mide el efecto de los factores de riesgo sobre la incidencia de la enfermedad en una población definida, los estudios transversales evalúan el efecto de los factores de riesgo sobre la prevalencia de la enfermedad. A raíz de ello, pueden surgir diferencias cuando se estudian las mismas asociaciones por ambos métodos. Los factores que afectan tanto la incidencia como la mortalidad tras la adquisición de la enfermedad mostrarán un distinto grado de asociación cuando se los analiza en forma transversal o longitudinal. Así, un factor de riesgo asociado con un incremento en incidencia y mortalidad entre los afectados denotará una asociación menor con la prevalencia de la enfermedad (exploración transversal) que con la incidencia de la misma en un estudio longitudinal, puesto que las personas con el FR tendrán menor chance de sobrevivir hasta el momento del estudio transversal. Incluso los factores no relacionados a la incidencia de la enfermedad podrían mostrar una asociación con la prevalencia si afectaran la sobrevida de los individuos con enfermedad en forma diferencial. Así las cosas, si la etiología de la enfermedad es el principal interés, los estudios de cohortes o casos y controles incidentes son preferibles, mientras que los estudios de prevalencia podrían brindar una información más útil para la caracterización de una población por cuestiones de Salud Pública.

Dado que los estudios transversales muestran más bien casos prevalentes que incidentes; el requisito de selección de sobrevida a la fecha en la cual se conduce la exploración puede generar diferencias entre la población de casos prevalentes y la población obtenida con el seguimiento de casos incidentes a lo largo del tiempo. La probabilidad de observarse el evento en un cierto período del estudio transversal es proporcional a la extensión de ese tiempo. Como consecuencia de la heterogeneidad en el curso de la enfermedad, los individuos con mayor tiempo de sobrevida tienen más posibilidades de ser incluidos en la muestra del estudio transversal respecto de aquellos que sobreviven menos. De forma similar, la población de casos prevalentes que tienen una característica dada puede diferir de la población de aquellos que nunca han presentado tal particularidad. Una estimación transversal de la prevalencia relativa de dos tipos de cáncer (uno agresivo y otro no tanto) podría ser distinta a las incidencias relativas, dado que la incidencia relativa de la forma agresiva sería mayor que la prevalencia relativa.

Aunque los estudios transversales pueden reflejar apropiadamente la distribución de las sobrevidas o los subtipos de enfermedad entre los individuos que poseen actualmente la enfermedad, estas estimaciones deben ser interpretadas como fotos instantáneas de la dinámica poblacional, y no necesariamente representativa de otra población, como aquellas con enfermedad incidente. Otra distinción entre estudios transversales y

longitudinales es que el estudio transversal o una serie de los mismos puede sólo contemplar cambios sumativos en la población, y no variaciones en el tiempo. Estos se obtienen con mayor certeza a través de estudios longitudinales.

Ejemplo de estudio transversal tradicional

Si bien es cierto que la infección con *Trypanosoma cruzi* es la causa esencial para el desarrollo de la miocardiopatía chagásica crónica, es probable que también existan factores adicionales que podrían contribuir al desarrollo del daño miocárdico. Sobre la base de esta presunción, hace unos años con un grupo de colegas decidimos llevar a cabo un estudio transversal en el que se analizaba si existían diferencias en cuanto a la presencia de alteraciones electrocardiográficas en individuos crónicamente infectados con este protozoario y que además presentaban factores de riesgo asociados (FRA) para el desarrollo de cardiopatías, tales como tabaquismo, alcoholismo, sobrepeso e hipertensión arterial. Todos los participantes concurrían a la misma institución de salud por cuestiones no vinculadas directamente a los objetivos del estudio. Dichas personas habían nacido y residido durante un lapso variable de tiempo en áreas endémicas para la Enfermedad de Chagas y posteriormente emigraron a la ciudad de Rosario. Para lograr un buen grado de certeza el diagnóstico de la infección se efectuó por 3 pruebas serológicas específicas, mientras que la evaluación cardiológica y el análisis de los eventuales FRA se realizó por medio de los estándares internacionalmente aceptados. De ese modo los 124 participantes se separaron en cuatro grupos, seropositivos y seronegativos a su vez subdivididos de acuerdo a la presencia o ausencia de los FRA. Como se indica en la Tabla 2.3 los grupos no difirieron en cuanto a las edades, aunque en las personas seronegativas sin FR hubo una mayor proporción de mujeres, lo cual se reflejó en una diferencia estadísticamente significativa al comparársela con la relación observada en los demás grupos. Las anormalidades electrocardiográficas (en su mayoría compatibles con los signos que se suelen observar en la miocardiopatía chagásica) prevalecieron en los grupos infectados y en los seronegativos con FRA sin diferencias entre estos 3 grupos. En las cuatro personas del grupo seronegativo sin FRA que presentaron electrocardiogramas anormales, las alteraciones eran de poca jerarquía. Estas diferencias en la proporción de electrocardiogramas anormales se mantuvieron aún cuando se practicó el ajuste para la disparidad en la distribución por géneros.

Tabla 2.3 Datos demográficos y presencia de anormalidades electrocardiográficas según la serología específica para *T. cruzi* y los FRA

Variables	Seronegativos Sin FRA	Seropositivos Sin FRA	Seronegativos con FRA	Seropositivos con FRA
Género (femenino/mascul)	18/11¶	15/12	11/16	11/30
Edad	39.3±11	39±12	43±7	42±7
ECG anormales/ ECG normales	4/25 §	14/13	14/13	18/23

¶ $p<0.02$ en comparación con los grupos restantes
§ diferencia con respecto a los demás grupos, $p<0.01$ (prueba de chi cuadrado 11.68)
Seronegativos vs. Seronegativos con FR, 7.62 $p<0.006$
Seronegativos vs. Seropositivos sin FR, 7.62 $p<0.006$
Seronegativos vs. Seropositivos con FR, 5.82 $p<0.016$

A la hora de extraer conclusiones, hemos debido ser cautos. Las mediciones efectuadas no permitían determinar si las anormalidades electrocardiográficas habían estado presentes en cierta medida antes de que tuviera lugar la infección o la exposición a los FRA. Sin perder de vista lo antedicho, los hallazgos indican que los sujetos sin evidencia de infección chagásica, pero con FRA, presentan anormalidades electrocardiográficas similares en proporción y tipo de trazado a las observadas en los chagásicos. Ello nos obliga a reflexionar que de no contar con un diagnóstico serológico confiable podríamos relacionar a la infección crónica por *T. cruzi* con fenómenos más bien vinculados a otros factores. El hecho que las personas seropositivas con y sin FRA presentaran un porcentaje similar de anormalidades electrocardiográficas plantea al menos dos posibilidades. Por un lado que los FRA son "eclipsados" por el efecto que la infección con *T. cruzi* es capaz de promover por sí misma. Como explicación alternativa podría plantearse que los seropositivos con FRA y afectación cardíaca tienen una mayor morbimortalidad que finalmente ocasiona un sesgo de selección. Para ahondar en torno a las hipótesis que generan los estudios transversales, y en particular al aquí comentado, debería llevarse a cabo un seguimiento a largo plazo (cohortes) en personas con distintos niveles de exposición (infección y/o FRA) sin miocardiopatía, para de ese modo determinar la incidencia y desenlace de la misma. Lamentablemente, el mayor obstáculo para su realización no es precisamente científico sino económico.

Tabla 2.4 Esquema de trabajo para el cálculo de la prevalencia relativa

Predictor	Fenómeno de interés presente	Fenómeno de interés ausente
Presente	14 (a)	13 (b)
Ausente	4 (c)	25 (d)

Los estudios transversales también nos permiten calcular la prevalencia relativa Tabla 2.4).
Para ello estimamos la prevalencia en las personas donde el predictor se halla presente = a/(a+b); y luego la prevalencia en los individuos no expuestos al predictor= c/(c+d). Dividiendo la primera por la segunda se obtiene la prevalencia relativa.
También se puede calcular el exceso de prevalencia
a/(a+b) – c/(c+d)
Si nos valemos de los datos provistos en la Tabla 2.3 y trabajamos con los valores correspondientes a los grupos sin factores de riesgo asociados –FRA- donde el predictor es infección chagásica o no, obtenemos los siguientes resultados

$$\text{Prevalencia relativa} = \frac{a/(a+b)}{c/(c+d)} = \frac{14/27}{4/29} = \frac{0.52}{0.14} = 3.71$$

Exceso de prevalencia = 0.52-0.14 = 0.38

Si bien no existe consenso en torno al real valor de ambas estimaciones, la prevalencia relativa y el exceso de prevalencia guardan alguna analogía con el riesgo relativo y el exceso de riesgo que se obtienen en los estudios de cohortes.

Una prevalencia relativa superior a uno significa que la posibilidad de enfermar es mayor en los expuestos que en los no expuestos y a la inversa. Cabe mencionar que a pesar de la similitud en el análisis estadístico de los estudios transversales y 1os de casos y controles, estos diseños difieren en cuanto a interpretación de medidas, sesgos y aproximación a la causalidad. Por ejemplo, un estudio transversal sobre el tabaquismo como factor etiológico de cáncer pulmonar comparando enfermos con no enfermos, apunta fundamentalmente a asociación entre variables. Por su parte, la estrategia de casos y controles es más robusta en términos causales pero aún así no debemos perder de vista que al recoger información sobre el predictor la historia previa de tabaquismo sería mas apropiada que el hábito tabáquico actual, y que también podría existir un sesgo en la incorporación de casos por cuanto podrían captarse enfermos con patología de mayor duración.
Debido a que frecuentemente la selección de sujetos se obtiene de la población general, el control de variables confusoras suele hacerse en el análisis de los datos (y no por diseño). Para ajustar el efecto de otras variables confusoras es posible realizar análisis estratificado o emplear técnicas de análisis multivariado -regresión lineal múltiple, regresión logística-.

Ventajas y desventajas de un estudio transversal
A manera de síntesis se puede afirmar que este tipo de estudio,

- Es relativamente rápido de llevar a cabo y no demasiado costoso.
- Es útil para examinar factores que podrían tener causalidad.
- Los hallazgos de esta investigación pueden constituir el paso preliminar para un estudio de cohorte.
- Definen las características demográficas y clínicas del grupo al momento de la medición.
- Plantea dificultades para establecer una relación causal.
- El hecho que sólo estima la prevalencia acota la información que brinda en cuanto a la historia natural y otros aspectos de la enfermedad.

2.3 Estudio de Casos y Controles

En "casos y controles", nuestro punto de partida es una muestra de personas que poseen la enfermedad (los casos) las cuales serán confrontadas con otras no afectadas (controles) para indagar si existen diferencias entre ambas, en cuanto a la presencia de la variable predictiva (independiente). Este tipo de estudios es particularmente útil para las enfermedades con un largo período de incubación ya que la estrategia de cohortes demandaría mucho más tiempo y dinero. También es adecuado para el caso de patologías poco frecuentes dado la gran cantidad de personas que se debería reclutar en una cohorte como para contar con un número suficiente de individuos que desarrollaran la enfermedad.
A diferencia de la cohorte donde las personas en riesgo son controladas por un período de tiempo a efectos de determinar quién desarrolla la enfermedad, en el estudio de casos y controles los hechos (exposición y enfermedad) ya se han presentado. Vale decir que

nuestra mirada es hacia atrás a efectos de determinar si existen diferencias en cuanto a la variable predictiva entre sanos y enfermos.

Predictor Patología

CASOS

La presencia del factor de riesgo debe ser preponderante	Mirada retrospectiva	**Enfermos**

CONTROLES (una población mayor a de los casos)

El factor de riesgo debe estar mucho menos presente	Mirada retrospectiva	**No enfermos**

De ser correcta la hipótesis de trabajo, debería verificarse que la presencia del FR es significativamente mayor en los enfermos (casos) respecto de los no afectados (controles). Los estudios de casos y controles no pueden estimar prevalencia o incidencia, en razón de que el número de participantes a incluir en la muestra se halla determinado arbitrariamente por el investigador.
Este modo de analizar asociaciones ha merecido algunas críticas en el sentido de no seguir la secuencia lógica de eventos que llevan a una enfermedad, es decir primero la exposición y luego el fenómeno. Muchos ejemplos indican, sin embargo, que el enfoque efecto-causa puede dar buenos resultados cuando el diseño y la implementación del estudio son adecuados. Por otro lado, la vida cotidiana está plagada de situaciones en las que se procede de esta manera. ¡Pensemos simplemente en las investigaciones policiales!
Por otro lado, si bien la estrategia predominante respecto del estudio de casos y controles es la mirada retrospectiva, ello no se da necesariamente en todas las situaciones. Supongamos que un investigador decide aplicar este tipo de diseño, estableciendo como criterio de inclusión la incorporación, solamente, de los casos nuevos que se detectan durante el período de reclutamiento. Obviamente que el enfoque es mucho más prospectivo.

Casos y controles anidado

Como ya se adelantara, esta estrategia es aplicable para situaciones donde el estudio de todos los participantes se torna complicado por problemas monetarios. Vale decir la elaboración de un estudio prospectivo ideal donde las mediciones del predictor, a lo largo de los distintos momentos de evaluación, en la cohorte íntegra son muy poco factibles. La opción para estas circunstancias es la elaboración de un estudio de casos y controles extraído de la muestra que conforma la cohorte.
Utilizando las bases de datos del sistema de salud en Quebec, Martel y colaboradores (BMJ 2005) efectuaron un estudio de este tipo con el objeto de determinar si el uso de corticosteroides inhalatorios durante el embarazo incrementa el riesgo de hipertensión y pre-eclampsia entre las mujeres asmáticas. Se trabajó con una cohorte de mujeres

asmáticas (3505) que totalizaban 4593 embarazos entre los años 1900 y 2000, donde pudieron detectar 302 casos de hipertensión por embarazo y 165 cases de pre-eclampsia. El uso de corticosteroides desde el momento de la concepción hasta el desarrollo del evento no se relacionó con un mayor riesgo a desarrollar hipertensión ni pre-eclampsia. Pero la administración de corticosteroides orales estuvo asociada con una mayor chance a presentar hipertensión y una tendencia a la pre-eclampsia.

Consideraciones específicas para el estudio de casos y controles

Al decidirnos por este tipo de estudios es necesario tener en cuenta que la validez de una eventual asociación entre las variables predictiva y de impacto requiere que casos y controles sean comparables. Para ello se debe prestar especial atención a los criterios de selección de los participantes y al modo en que se obtiene la información referente a las variables. Un punto fundamental es establecer una definición exacta de cuáles son los casos. Aunque nos propongamos estudiar pacientes con Psoriasis, puede que esta única designación no resulte suficiente y se deba especificar además la variedad de la afección. El paso siguiente es determinar si la investigación se conducirá en pacientes hospitalarios o en una muestra aleatoria. Cuando la financiación es suficiente conviene adoptar la segunda opción a fin de evitar el sesgo que se presenta con la inclusión de enfermos hospitalarios. La otra cuestión a definir es si trabajaremos con casos de reciente diagnóstico (incidencia) o de mayor antigüedad (prevalencia), lo cual está en relación directa con la frecuencia de la enfermedad en cuestión. Si se trata de una patología relativamente frecuente es preferible la primera alternativa puesto que la chance de establecer una relación temporal entre exposición y enfermedad es mayor cuando se incluyen pacientes nuevos o de reciente diagnóstico.

Al momento de planificar también habrá que decidir si trabajaremos con un grupo de pacientes limitado por algunas de sus características (por ejemplo edad o género) o deseamos conformar una muestra aleatoria representativa de la población de pacientes. Obviamente que esta alternativa permite extraer conclusiones más generalizables en caso que obtengamos una información confiable de todos ellos, de lo contrario nuestras pretensiones se harán añicos contra el muro de la validez. Si no estamos seguros de poder llevar a cabo grandes emprendimientos, es preferible un estudio menos ambicioso pero factible.

Así como la elección de los casos tiene sus bemoles, la conformación del grupo control puede ser aún más difícil. Por las características intrínsecas del diseño, la muestra de controles se confecciona a partir del perfil de los casos; vale decir que en sentido estricto los controles no representarán a la población de personas no enfermas, sino a los individuos que habrían sido considerados casos, de hallarse enfermos. De este modo existe la posibilidad que la muestra presente algunas diferencias respecto de la población general, entre otras la variable predictiva sobre la que estamos indagando. Es por eso que en determinadas situaciones se aconseja incluir más de una muestra de controles. En lo concerniente al modo de reclutamiento de los participantes, el grupo puede conformarse a partir de controles hospitalarios cuya causa de ingreso no guarda relación con la enfermedad en estudio. Este método ofrece una serie de ventajas tales como fácil acceso, bajo costo y posibilidad de obtener la información que se necesita. Al igual que los casos, los controles hospitalarios son más proclives a cooperar y es probable que exista bastante similitud en las condiciones socioeconómicas. Siempre que el hospital no sea un centro de referencia para la enfermedad bajo análisis, puesto que de ser así los casos suelen provenir de un área geográfica mucho más amplia. La desventaja que plantea la inclusión de controles hospitalarios, es que ellos también son

enfermos y la información recogida puede no corresponderse con lo que ocurre en los individuos sanos.
El otro punto a tener en cuenta es que la enfermedad por la cual fueron hospitalizados los "controles" no esté asociada, positiva o negativamente, con la variable predictiva.
Cuando la elección de controles hospitalarios pone de relieve demasiadas dificultades o los casos han sido extraídos de la población general de enfermos, los controles deberán pertenecer a una población de individuos sanos. Para ello se pueden utilizar varios procedimientos de selección aleatoria tales como la guía telefónica, padrón de votantes, seccionales de policía, etc. El tema es que aún cuando se haya contactado a un número parecido de personas por área no todas estarán dispuestas a participar y los que sí decidan hacerlo pueden constituir una sub-muestra que de entrada le resta validez a los resultados.
La muestra de controles también se puede integrar con parientes, amigos o vecinos de los casos. Estos individuos son sanos, proclives a cooperar y muy similares en algunas variables socioeconómicas; pero también puede serlo el hipotético factor de riesgo, con lo cual obtendremos una sub-estimación de su efecto real.
Otra cuestión a definir está referida a la cantidad de controles a incluir por cada caso. Cuando se trata de una muestra importante de individuos es aceptable trabajar con una relación 1:1. En las situaciones donde el número de casos es reducido es recomendable incluir más controles (2:1 o 3:1) dado que esto incrementa el poder del estudio.
Como corolario puede afirmarse que la planificación de un estudio de casos y controles demanda muchas horas de discusión y reflexión, sin perder de vista que lo esencial de un estudio es la validez. Si está bien hecho, este tipo de estudios puede darnos información en cuanto a la fuerza de la asociación entre la variable predictiva y la presencia o ausencia de enfermedad. Este valor recibe el nombre de coeficiente de desigualdad relativa o índice de disparidad (Odds Ratio -OR-). Estima la relación entre el número de veces en que el evento se produce o no y es una aproximación del RR. La manera de calcular los OR se halla ejemplificada en la Tabla 2.5. En ella se consigna el número de enfermos y controles que se hallan o no expuestos, al FR en cuestión. Si nos atenemos a la ecuación, veremos que finalmente se extrae un cociente cuyo numerador surge de *multiplicar los valores concordantes*, mientras que el divisor es el producto de los datos que se ubican en las *celdas discordantes*. Al resultado que arroja dicho cociente se lo denomina OR crudo. Algunos procedimientos estadísticos mas sofisticados como la regresión logística permiten ponderar los efectos de otras variables sobre la relación y de ese modo se obtienen los OR ajustados (de acuerdo al género y la edad, entre otras variables posibles). Con ello se consigue mejorar la inferencia, puesto que podremos saber si el FR en estudio interacciona con otros factores (ver capítulo 4).

Tabla 2.5. Cálculo de los OR

	Enfermos	No Enfermos
Expuestos al FR	a	c
No expuestos al FR	c	d

$$OR = \frac{a/b}{c/d} = \frac{a.d}{b.c}$$

Si los casos presentan una mayor frecuencia del predictor respecto de los controles tendremos una asociación de riesgo, mientras que si la frecuencia del predictor es menor en los casos se trata de una asociación protectiva.
Para estimar el intervalo de confianza del OR, se calcula el logaritmo del OR y también se obtiene el error estándar del logaritmo del OR, el cual surge de la siguiente fórmula:

$$\text{Error estándar –ES- (log OR)} = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

Dado que el IC = ±1.96 x ES, se calcula sus valores y finalmente se procede a hallar el antilogaritmo. Para ilustrar el procedimiento, tomemos los datos representados en la Tabla 2.4.

Predictor	Con ECG alterado	Con ECG normal
Expuestos	14 (a)	13 (b)
No expuestos	4 (c)	25 (d)

Con ello de obtiene un OR= 6.73, mientras que el $\log_e(6.73)$= 1.9066

$$\text{ES (log OR)} = \sqrt{\frac{1}{14}+\frac{1}{13}+\frac{1}{4}+\frac{1}{25}} = 0.662$$

De este modo el IC = 1.9066 ± 1.96 x 0.662 (1.2975), lo que arroja los siguientes valores 0.6091 - 3.2041 que al obtenerse el antilogaritmo nos termina dando el IC del OR, 1.839 – 24.633

A diferencia del RR que era una probabilidad, el coeficiente de disparidad en realidad especifica una chance. La relación aquí se establece entre la cantidad de veces que se presenta el fenómeno de interés, respecto de la cantidad de veces que ello no ocurre. Al igual que en el RR, un índice mayor a 1 implica que la exposición a determinado FR aumenta la chance de que ocurra el fenómeno de interés y viceversa. Los OR pueden ser utilizados como una aproximación al RR mayormente cuando el fenómeno de interés no es tan frecuente (menos del 20% en ambos grupos de análisis). Si la proporción del evento es alta en alguno de los grupos, es desaconsejable efectuar dicha derivación.
Partiendo de la tabla de contingencia 2.5, también se puede calcular la fracción etiológica (FE, representado con λ), vale decir la proporción de casos en la población que es atribuida al factor, y por ende en similitud con el riesgo atribuible que surgía del RR.

	Enfermos	No Enfermos	
Expuestos	a	b	m_1
No expuestos	c	d	m_2
	n_1	n_2	n

El cálculo surge de la siguiente expresión $\lambda = 1- (cn_2/dn_1)$

Alcances del estudio

Una de las ventajas del estudio de casos y controles es el caudal de información que puede obtenerse con un número relativamente bajo de personas. La mirada retrospectiva y la posibilidad de examinar una o más variables predictivas permiten la generación de nuevas hipótesis fisiopatogénicas en torno de la enfermedad. Sin embargo debemos tener en cuenta que la información extraída es limitada y no permite calcular prevalencia ni incidencia. Por otro lado, contamos con una sola variable de salida, es decir la enfermedad.

Una limitación difícil de superar totalmente, es la probabilidad de sesgo que puede existir en este tipo de estudio. Las fuentes de sesgo están relacionadas con el hecho que el muestreo de casos y controles es por separado y que la medición de la variable predictiva es retrospectiva. Retomando lo antedicho para las cohortes, no podemos dejar de considerar que la relación *exposición-enfermedad* detectada en el estudio podría ser diferente de aquella que se da en los individuos que, siendo elegibles para participar, no se prestaron a ello o no fueron seleccionados por el investigador.

Dado que el estudio de casos y controles es un diseño muy amenazado por el error sistemático, vale la pena considerar los puntos más críticos donde puede producirse un sesgo.

Error de muestreo

Básicamente este tipo de sesgo se genera cuando los casos y los controles se seleccionan de un modo no representativo de lo que es la distribución del predictor en la población base.

Especificación de la base del estudio

En este sentido es fundamental asegurarse que los casos y los controles son extraídos de la misma población base; puesto que si los grupos fueran extraídos de dos poblaciones en las cuales la prevalencia de la exposición al predictor es desigual, el sesgo es inevitable. Para ello tener siempre presente que el estudio debe cimentarse en una población. Todos los casos serán identificados y seleccionados; o bien se puede conformar una muestra aleatoria representativa. Sea cual fuere la elección, la serie de controles se constituirá por “no casos” muestreados desde la misma base del estudio. Cuando la población base está bien definida, y todos los casos son identificables, en principio es posible muestrear a los mismos sin caer en sesgos. Aunque ello podría originarse de producirse una ausencia de respuesta (nos referiremos más adelante).

A veces no es posible especificar la base primaria. Por ejemplo pensemos en una serie de casos seleccionados sin un conocimiento pormenorizado de la población (personas con Síndrome Metabólico de un servicio de Endocrinología detectados durante un determinado período de tiempo sin mayor información respecto de la población asistida en ese servicio). En una situación de esta naturaleza debemos “concebir” una suerte de segunda base del estudio de la cual se habría seleccionado el caso con la patología que nos ocupa. A fin de ser coherentes, los controles también deberán ser muestreados de la misma base hipotética. En términos operativos, tales controles frecuentemente se obtienen de la misma fuente que se extrajeron los casos (personas sin alteraciones metabólicas, que concurrían a dicho servicio). Dado la imposibilidad de identificar explícitamente los miembros de la base secundaria, es necesario recurrir al buen juicio y experiencia en el tema a fin de seleccionar controles con alta probabilidad de ser representativos de lo que sería la exposición al predictor en la población base hipotética. Pensemos en un estudio sobre exposición al alcohol y hepatopatía crónica llevado a cabo en un hospital. El hospital tiene un servicio de hepatología muy importante y admite hepatópatas de toda la ciudad; en algunas áreas de la ciudad el alcoholismo es

alto, mientras que en otras no tanto. En el hospital también se internan casos con otras patologías pero generalmente de zonas adyacentes, y en esa zona los niveles de exposición al alcohol son al menos diferentes. En una situación así, los casos y los controles provienen de distintos basamentos, y la exposición es diferente. Una solución para este problema es que la serie de casos se restrinja a aquellos residentes en la misma área de los controles.

Sesgo de selección debido a no respuesta

Aún cuando hayamos especificado correctamente la base del estudio, podría presentarse un sesgo de selección de producirse un muestreo o identificación diferencial de los casos y controles. Por ejemplo una pérdida sustancial en el reclutamiento de casos y controles, originalmente elegibles, debido a una ausencia de respuesta. En un estudio poblacional, todos los casos originados en la población, o una muestra representativa de ellos, son incluidos. Alternativamente, si el estudio asienta sobre una base secundaria, los casos deberían ser representativos de los surgidos de dicha base (digamos de un único hospital). En cualquiera de las dos situaciones, los controles potencialmente elegibles deberían constituir una muestra representativa de todos los "no casos" de la misma base. En la práctica, sin embargo, es casi inevitable que algunos de los casos o los controles especificados inicialmente como elegibles no sean incorporados debido a problemas en su seguimiento, poca disposición a participar, enfermedad severa o muerte. Si entre los casos o los controles, la exposición al predictor para aquellos que resultan no reclutados es sistemáticamente diferente se producirá un sesgo de selección por no respuesta. Este sesgo es prácticamente despreciable cuando todos o casi todos los casos y controles seleccionados son finalmente reclutados. Cuanto mayor sea la proporción de no reclutados, más deberemos preocuparnos por este sesgo. Los estudios donde se toma como base el hospital tienden a presentar un mayor porcentaje de respuestas respecto de los poblacionales, más aún cuando la investigación implica la toma de muestras biológicas.

Un cierto reaseguro para el caso de no respuesta es demostrar que la distribución de las características poblacionales (edad, sexo, o residencia) son similares entre los reclutados o no. El mejor remedio sigue siendo incrementar la tasa de reclutamiento.

Sesgo de selección en la identificación de los sujetos del estudio

El sesgo de selección también se puede producir si los casos y los controles son identificados de un modo que no es independiente de la exposición al predictor. Consideremos un situación hipotética en dos personas que presentan un cuadro de gastritis; una ingiere anti-inflamatorios y la otra no recibe medicación alguna. Quien se haya medicada tiene información acerca del riesgo de padecer este problema, y obviamente consulta al médico quien hace los estudios acordes para arribar a un diagnóstico. Quien no ingiere medicamentos permanece en la casa y sin diagnóstico. Ambas finalmente se recuperan del problema. Evidentemente la identificación de los casos es incompleta y dependiente del factor de riesgo lo que puede conducir a una sobre estimación de la asociación entre enfermedad ácido-péptica y anti-inflamatorios, puesto que esta medicación aumenta la chance de inclusión en el estudio. En un estudio respecto a cardiopatía isquémica y tabaco, podría observarse un sesgo de detección si los fumadores fueran sometidos más frecuentemente a estudios cardiológicos respecto de los no fumadores.

En el mismo sentido, si bien más de un 60% de los hombres mayores a 60 años tienen un cáncer de próstata sintomático, la determinación del antígeno prostático hace posible la detección de estos casos, pero ello depende en buena medida de la cobertura social

del individuo. Como regla general los estudios donde el reclutamiento se efectúa a partir de programas de detección suelen presentar un sesgo de selección no despreciable.
Es muy importante asegurarse que todos los casos dentro de una base (sea primaria o secundaria), tengan la misma chance de ser identificados e incluidos en el estudio, independientemente del factor de riesgo al que se hayan expuestos. En conexión con esto es muy atingente proveer datos sobre el grado de ocurrencia de los casos no expuestos y la frecuencia de exposición en los controles elegidos adecuadamente.

Sesgo debido a la selección de controles no representativos de una base de estudio secundaria

Los pacientes hospitalizados siguen constituyendo los controles más utilizados en los estudios de casos y controles. Los problemas potenciales que ello plantea sirven para ilustrar las cuestiones suscitadas cuando se selecciona a partir de una base de estudio secundaria e hipotética. Se debe prestar mucho cuidado en el sentido que los pacientes hospitalizados y seleccionados para su inclusión como controles hayan sido internados por enfermedades no relacionadas con el factor de riesgo bajo análisis. Más allá de este sesgo potencial, los estudios hospitalarios siguen siendo un bastión en la investigación de casos y controles, fundamentalmente cuando están bien conducidos. En líneas generales los datos sobre entrevistas tienden a presentar menos sesgos entre los controles hospitalarios, y por otro lado son más fáciles de llevar a cabo; la tasa de respuesta es usualmente mayor, y también es más fácil obtener muestras biológicas de ser necesario. En términos generales, la selección de los controles hospitalarios no presenta sesgos si los seleccionados para su inclusión son representativos de la distribución del predictor en la base secundaria hipotética. En la práctica, usualmente no hay razones para que tales controles no puedan ser identificados cuando el motivo de internación (diagnóstico primario) es independiente del factor de riesgo bajo estudio. De existir dudas al respecto es mejor excluirlos. Frecuentemente los pacientes internados (casos o controles), tienen más de 1 diagnóstico, y es importante valorar que los diagnósticos secundarios son irrelevantes a la selección de los controles, a menos que la enfermedad secundaria también haya influido sobre la selección de los casos (inusual).
Como regla general, las personas cuyo diagnóstico primario son condiciones agudas que requieren una internación (traumatismo) reúnen el requisito de independencia, al igual que con otras situaciones (internación selectiva por cirugía de cataratas). Sin embargo conviene reflexionar in extenso en función de la hipótesis bajo análisis. Si pensamos en un estudio sobre riesgo de cáncer de ovario y anticonceptivos en mujeres internadas por traumatismos, la razón de internación en general es independiente del factor de riesgo. Sin embargo, si el estudio es sobre cáncer de mama en relación a la ingesta de alcohol, el traumatismo no es un control adecuado. El reaseguro que la identificación de los controles hospitalarios no está sesgado puede conseguirse si la tasa de exposición entre las principales categorías diagnósticas (traumatismo, infecciones agudas) es uniforme; en ese caso el sesgo sólo es posible si la selección de todo el grupo de controles está sesgada, y en la misma medida para cada categoría diagnóstica.
Algunos autores sostienen que los controles hospitalarios casi siempre son sub-representativos de la exposición en la población. Este argumento ignora la premisa que cuando los casos representan una base de estudio secundaria (frecuente en estudios hospitalarios), la única serie válida de controles son los pacientes internados en el mismo hospital donde se asisten los casos.
Una desventaja teórica respecto a la inclusión de controles hospitalarios es el supuesto que la condición incluida es independiente de la exposición.

Sesgo debido a una selección no representativa de controles de la base primaria del estudio

Un ejemplo típico de este sesgo puede darse cuando el plan de muestreo se basa en una cobertura incompleta de la población base (a partir del registro de dueños de vehículos), por el cual se puede sub-representar personas de un estado socioeconómico bajo. De manera similar la selección de otros controles, tales como amigos de los casos, o colegas, pueden ocasionar problemas. Por ejemplo, los amigos no expuestos de un caso sí expuesto podrían ser más proclives a participar porque desean ayudar. Insistimos en reflexionar acerca de si los controles elegidos son representativos de la población control.

Tener en cuenta además que la respuesta en la población de controles tiende a ser menor de la obtenida en controles hospitalarios, más aún si se deben obtener muestras biológicas.

Sesgo de selección en casos y controles anidado

En un estudio de casos y controles anidado, los casos integran una cohorte que desarrolla una condición dada, y los controles son una muestra de no casos seleccionada de la misma cohorte, seguidos el mismo lapso de tiempo. Ambos provienen de la misma base de estudio; si el seguimiento ha sido satisfactorio la tasa de no respuesta es baja; y el sesgo de información resulta casi nulo dado que el estado de exposición se determina habitualmente antes de que el sujeto se convierta en caso. Una ventaja adicional es que podría ser más fácil obtener muestras biológicas puesto que los individuos ya están participando del estudio. La desventaja es que a veces no es posible reunir un número suficiente de casos, a menos que se disponga del seguimiento de todos los participantes.

No obstante ser conscientes de los problemas que acabamos de comentar, no siempre es posible superarlos porque la realidad difícilmente constituye la situación ideal; sin embargo, el haberlos advertido hará que tratemos de incluir controles que resulten lo más representativo posible de los casos.

Error de medición

Este sesgo se presenta cuando el modo, en que los casos o los controles refieren el FR es disímil, o bien la información es requerida diferencialmente. La probabilidad de presentación de este sesgo es mayor cuando el sujeto de estudio, o aquellos a cargo de la recolección de los datos están en conocimiento de la hipótesis. También se puede incurrir en el sesgo de memoria; vale decir los casos conocen la hipótesis bajo análisis y tienden a referir sus exposiciones más que los controles, y con ello se sobreestima el OR. Aún cuando los controles estén en conocimiento de la hipótesis al igual que los casos, si ellos son más sanos tienen menos motivos para chequear su bagaje de recuerdos. Los casos pueden ser más proclives a recordar más detalladamente sus exposiciones si su memoria ha sido "cebada" a raíz de una serie de preguntas previas formuladas por el médico en relación con las causas de su padecimiento. El desconocimiento de los objetivos del estudio reduce este sesgo pero no lo elimina. Los casos hospitalarios suelen recordar las exposiciones respecto de una muestra de controles entrevistados en sus casas. Por ello la entrevista de controles hospitalarios aminora el sesgo. A veces pueden darse situaciones diferentes. En un estudio sobre traumatismo y alcohol, los casos podrían sub-estimar su consumo respecto de los controles ante el remordimiento de haber contribuido a su propia enfermedad. Las personas a cargo de la recolección de los datos tampoco están exentas de este sesgo cuando conocen los objetivos del estudio.

Ocasionalmente es posible evitar o minimizar el sesgo de medición. En un estudio de cáncer de útero en relación con el uso de estrógenos conjugados, se examinaron las historias clínicas acerca de las prescripciones sin que la información, sobre si era caso o control, fuese revelada. Lamentablemente el diseño de casos y controles no es muy adecuado como para descartar con toda seguridad el sesgo de medición. Por eso uno debe barajar todas las alternativas posibles, por ejemplo; (1) no revelar la hipótesis de trabajo a los participantes y entrevistadores; (2) utilizar un cuestionario no ambiguo y muy bien estructurado; (3) recurrir a ayuda memorias para maximizar la recordación; y (4) un entrenamiento riguroso del entrevistador. Aún cuando hayamos empleado el mejor de los diseños posibles, el interrogante acerca de si existe o no un sesgo de medición lo suficientemente poderoso como para invalidar la asociación detectada es algo que hace al buen tino del investigador.

Errores de clasificación

Los mismos están referidos a procedimientos que llevan a una evaluación errónea de cualquiera de las variables (predictiva o de impacto). En conexión con lo especificado en la sección 2.1 cuando el porcentaje de casos y controles en quienes se efectuó una clasificación incorrecta de ambas variables es similar estamos frente a un error de clasificación no diferencial o aleatoria; por ejemplo historias clínicas donde la información no es clara pero se da tanto de un lado como del otro. Si por el contrario la proporción es diferente entre grupos, estamos frente a un error diferencial o no aleatorio. Éste sería el caso de las personas que tienen mucho más presente al FR porque experimentaron reacciones adversas.

El error de clasificación no diferencial determina una subestimación de la asociación real, lo que favorece a la hipótesis nula; mientras que el diferencial puede exagerar o reducir el efecto que estamos midiendo, según el caso. Tengamos en claro que de todos modos los resultados se verán afectados.

El estudio de casos y controles es un ejemplo cabal de los recursos metodológicos que ofrece la investigación para “administrar” el error y obtener conclusiones válidas. El Dr. Rabasa -quien mucho nos enseñó- solía decir: “no sé si los investigadores somos los más inteligentes, pero sí los más obsesivos”.

2.4 Pautas para la publicación de estudios observacionales

Atento a la necesidad de unificar criterios y contar con una información los más fehaciente posible, hace unos años se desarrollaron las recomendaciones STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*); las cuales apuntan a mejorar la calidad de los trabajos referidos a estudios observacionales. El documento consiste en un listado de 22 ítems, 18 son comunes a estudios de cohorte, casos y controles y estudios transversales y cuatro son específicos para cada uno de los 3 diseños. Tales lineamientos realzan la publicación y a la vez facilitan el análisis e interpretación crítica de los estudios por los revisores, editores y lectores. Para más información las pautas se pueden consultar al sitio http://www.strobe-statement.org/).

2.5 Causalidad

Desde tiempos inmemoriales la Medicina viene intentando brindar respuestas razonables a un par de interrogantes que ocupan un espacio significativo de la consulta médica, ¿a qué se debe la enfermedad que me diagnosticaron y de qué modo la contraje? Un “*racconto*” del recorrido de las ciencias médicas en torno a causalidad excede a los alcances del texto. No obstante ello, a *prima facie* uno puede adelantar una

definición por la cual causa es un agente que trae aparejado un evento o resultado. Adentrándonos en la disciplina, se la define como un factor que en determinadas circunstancias hace que la enfermedad sea más o menos probable según que el mismo se halle presente o ausente, respectivamente. El estudio en torno a las causas que determinan una enfermedad no sólo permite un mejor conocimiento en cuanto a la etiología, patogénesis y factores de riesgo sino que facilita la práctica médica, tanto en la prevención, diagnóstico y tratamiento como así también en investigación en servicios de salud. Cuando nos referimos a causalidad es necesario dejar en claro desde el principio que raramente la variable de impacto (enfermedad) a la cual se orientan nuestros estudios está ocasionada por una sola causa. A raíz de ello, el conocimiento íntimo y acabado sobre los condicionantes que llevan al desencadenamiento de un trastorno es una tarea harto difícil, sino irrealizable. No por ello vamos a desistir de nuestro intento de entender algo de esa compleja realidad, y seguramente hallaremos agentes que cuando se presentan aumentan la probabilidad de que el fenómeno se produzca. A partir de este concepto, podemos arriesgar una definición de causa como un hecho o circunstancia que antecede a un desenlace y sin la cual este fenómeno puede no presentarse o bien surgir por otra razón. Nos referimos a causa necesaria cuando el hecho o circunstancia es indispensable para que se produzca el fenómeno, y denominamos causa suficiente al evento o circunstancia que da lugar al fenómeno, más allá de su indispensabilidad o bien al conjunto de hechos o situaciones que ineludiblemente llevan a la anomalía. Aunque la exposición a una causa necesaria es una condición *sine qua non* para que se produzca el evento puede que no sea suficiente. Por ejemplo, no todas las personas infectadas con *Mycobacterium tuberculosis* desarrollan la enfermedad.

No obstante que ciertos factores tienen un efecto causal sobre determinadas enfermedades no siempre es posible ubicarlo a qué tipo de causa pertenecen. Un ejemplo típico es el tabaquismo, el cual no encaja dentro de las necesarias ni suficientes, puesto que algunos no fumadores pueden desarrollar cáncer de pulmón mientras que determinadas personas que han fumado durante toda su vida no presentan la patología.

A medida que nos adentramos en el estudio de una enfermedad rápidamente surge que existe una cadena causal, vale decir que una causa encierra a su vez sus propios determinantes. Retomando el ejemplo de la Tuberculosis uno podría establecer que el desarrollo de la enfermedad es el resultado de una compleja interacción entre la micobacteria y el huésped. Si consideramos el componente “individuo” es fácil entrever que la respuesta inmune es un elemento esencial para la contención del proceso, la cual por su parte depende del grado de determinación genética del individuo y de las influencias medioambientales. Cada uno de estos 2 componentes está ligado a su vez a factores propios y así sucesivamente. A poco de andar uno advierte que como la identificación de todas las causas potenciales se torna muy difícil conviene individualizar aquellas con un alto grado de influencia.

Tan importante como conocer las causas es caracterizar las circunstancias en que ellas actúan. A este respecto puede que se trate de un proceso estocástico, vale decir sabemos cuán probable es que el fenómeno se produzca pero no conseguimos comprender porqué en algunos se da y en otros no. Otros autores sostienen, en cambio, que muchos sistemas biológicos están regulados por procesos determinísticos y aperiódicos los cuales requieren un enfoque no lineal y caótico. Al respecto de ello hace unos años se ha desarrollado un método para el control del caos (PNAS 2001; 98: 5827) que podría tener aplicabilidad en la prevención de arritmias.

Siempre que se detecta una correlación significativa entre dos variables surge una especulación en torno a la causalidad de una sobre la otra. No obstante el importante

grado de asociación, lejos estamos a esa altura de poder afirmar que el hecho "x" determina el fenómeno "y". Bien podría ser que se tratara de una asociación falaz debida simplemente al azar, o algún factor externo en paralelo con las variables bajo análisis que introduce un sesgo en nuestros resultados. Por supuesto que existirán casos en que la relación entre ambas variables encierra una real causalidad y sienta las bases para la realización de una investigación explicativa o analítica, donde confrontaremos nuestra hipótesis con la realidad. Lo ideal sería llevar adelante un estudio experimental en el cual uno puede en mayor o menor grado controlar la situación a punto tal que algunos factores se mantienen constantes y otros pueden variar. En la investigación clínica, el análisis de la hipótesis no puede sustentarse en la provocación de la enfermedad o fenómenos relacionados (recuperación, muerte, etc.) sino simplemente en aguardar a que la misma se produzca espontáneamente. El estudio también es analítico pero no experimental en sentido estricto por cuanto no ha sido provocado.

Al trabajar este tipo de cuestiones, desde el punto de vista médico conviene ocuparse por separado de las enfermedades de etiología conocida respecto de aquellas que no la tienen. Si nos referimos por ejemplo a la Enfermedad de Chagas, cae de maduro que la enfermedad no se produce si previamente no estamos infectados con el *Trypanosoma cruzi* (causa necesaria). Si por el contrario quisiéramos precisar las causas que determinan la obesidad, es probable que aparezcan varios factores, el componente genético (seguramente poligénico), la ingesta, el sedentarismo y trastornos psicológicos, entre otros. Tomados en forma independiente todos ellos podrían ser una condición necesaria pero no suficiente para que ocurra el fenómeno. Si por el contrario todos los factores operan en conjunto existirá una altísima probabilidad de que la persona sea obesa. Dicho de otro modo, el agregado de todas las condiciones necesarias (predictores) actúa como condición suficiente y cuando ella se halla presente el fenómeno se produce.

Si volvemos a nuestro ejemplo de la Enfermedad de Chagas de existir un predictor que tuviese un efecto directo sobre los mecanismos que median la génesis de la miocardiopatía chagásica crónica (MCC) -independiente del efecto de otras causas- podemos asignarle el valor de causa suficiente. Esto que desde el punto de vista teórico aparece relativamente simple suele ser bastante complicado en la práctica, puesto que en la realidad surgen otros tipos de causas (más todavía!) que se las designa como condiciones contribuyentes y contingentes. Una causa contribuyente es aquella que incrementa la probabilidad de que se produzca el fenómeno, aunque no significa que indefectiblemente sea así. Si la Enfermedad de Chagas afecta al miocardio uno podría pensar que factores que han demostrado tener una acción nociva sobre este tejido (tabaquismo, alcoholismo) también podrían operar como causas contribuyentes para la generación de la MCC. Más aún si en la franja de la población infectada ambos hábitos son harto frecuentes (condición contingente) estas causas contribuyentes adquieren mayor peso.

A lo largo de las páginas precedentes hemos señalado que la relación entre dos variables puede verse influida por factores no contemplados en ese enfoque binario. Dentro de estas terceras variables podemos mencionar a los FC, modificadores de efecto, variables intermedias, variables no controladas y variables posteriores. Los FC ya han sido presentados en sociedad y baste con mencionar que son variables capaces de falsear la relación entre el predictor y la variable de impacto. Buena parte del trabajo del investigador se centra en el análisis de esta posibilidad para eliminarla o tenerla bajo control. Un FC peculiar es el que aparece en la denominada paradoja de Simpson. Se trata de un FC tan importante que es capaz de invertir la verdadera relación existente

entre dos variables. Es un fenómeno que se cita clásicamente en todos los textos y revisiones sobre este tema, pero que aparece muy pocas veces en la realidad.

Por su parte, los modificadores de efecto alteran la magnitud de la relación entre el predictor (causal) y la variable de impacto, que en la jerga estadística se lo conoce como interacción. En otras palabras la fuerza de la relación entre el predictor y desenlace se modifica ante la presencia de una tercera variable. Bien podría ser que el riesgo para el desarrollo de cáncer a raíz de la exposición a un determinado contaminante medioambiental fuese distinto para varones respecto de las mujeres, lo cual pone sobre el tapete el concepto de heterogeneidad del efecto.

En cuanto a las variables intermedias, las mismas forman parte de la cadena causal y se ubican entre el predictor y la variable de impacto. En el nexo entre la infección con el VIH y el cuadro de inmunodeficiencia que resulta luego de varios años de transcurrida la misma, la desaparición de linfocitos cooperadores (por procesos vinculados al virus o no) encaja en este tipo de variable.

Las variables no controladas son aquellas que ejercen una acción sobre la variable de impacto, pero que no guardan relación con el predictor (a diferencia del FC). Es un buen ejercicio metodológico-fisiopatogénico tratar de identificar dichos factores a fin de lograr una correcta estimación de la relación entre el predictor y el fenómeno de interés de nuestro estudio.

Como su nombre lo indica, las variables posteriores son ulteriores a la aparición del fenómeno de interés y pueden tergiversar nuestra inferencia en cuanto a causalidad sobre todo en los estudios donde no existe seguimiento y por lo tanto es muy difícil conocer si un factor es anterior o posterior a la aparición del fenómeno de interés.

Para el caso de estudios observacionales habrá que reflexionar en profundidad acerca de dichas variables al momento de diseñar la investigación. Puede que debamos establecer criterios de exclusión, asegurar la comparatividad de los grupos o incluir estratos si es necesario y prever el tipo de análisis estadístico que nos permita tener un cálculo ajustado de las diferencias. Si el FC constituye una situación harto frecuente de la realidad, no es aconsejable excluirlo del estudio porque nos estaríamos apartando de lo que ocurre en el entorno. Es mucho mejor planificar la utilización de pruebas de ajuste para la etapa de análisis de los resultados. En los estudios intervencionistas, el factor de confusión desaparece mediante la randomización y un tamaño muestral “generoso”.

Parafraseando a un desaparecido político argentino que solía decir “calma radicales” la metodología provee una serie de estrategias que nos permiten ejercer un cierto control y modo de desglosar la participación de las terceras variables. Las estrategias a utilizar y el significado de las mismas se resumen en la Tabla 2.6.

Tabla 2.6 Pautas para el análisis de datos con miras a ponderar causalidad

Pasos a seguir	Pruebas a utilizar	Diferencia Significativa	Diferencia No Significativa
Estadística descriptiva			
Tablas de contingencia	χ^2, asociación, riesgo (R) e IC		
Análisis/estratos	χ^2 Mantel-Haenszel R e IC		
Homogeneidad de estratos		Existe interacción	No hay interacción
Cálculo de riesgo ajustados (de no existir interacción)	Análisis multivariado	Si R crudo≠R ajustado Presencia de FC	Si R crudo=R ajustado Ausencia de FC

IC intervalo de confianza
FC factor de confusión

Los criterios de causalidad de Bradford Hill

Como se señalara anteriormente una correlación entre dos variables no es un dato suficiente para hablar de causalidad. Aún cuando los estudios sobre los cuales se asientan los resultados son metodológicamente correctos existen una serie de exigencias adicionales que deben satisfacerse para otorgar la fuerza de causa; ellos son los 9 criterios de causalidad de Bradford Hill: coherencia, consistencia, especificidad, evidencia experimental, fuerza de la asociación, gradiente biológico, plausibilidad, razonamiento por analogía y secuencia temporal. No obstante que ninguno de ellos constituye una prueba irrefutable de causalidad hay acuerdo en que la temporalidad es una condición imperativa puesto que el predictor debe preceder en el tiempo al fenómeno de interés. Los rasgos más salientes de cada uno de los criterios se exponen brevemente.

COHERENCIA. Nuestra lógica e inferencia en torno al predictor y sus efectos no puede presentar inconsistencias o contradicciones con las características que distinguen a la enfermedad o fenómeno en cuestión.

CONSISTENCIA O REPRODUCIBILIDAD. Ello requiere que la asociación entre las dos variables haya sido verificada en otros estudios, donde las poblaciones, las circunstancias y los investigadores que llevaron a cabo las pruebas eran diferentes. Es un criterio de cuantía a la hora de sopesar.

ESPECIFICIDAD DEL EFECTO. Este criterio no ofrece dificultades cuando trabajamos con factores (causas) que generan un único efecto, por ejemplo la relación entre hiperlipidemia y arterioesclerosis. La cuestión se torna más complicada para el caso en que un efecto puede obedecer a causas variadas o bien un predictor que impacta sobre distintos blancos.

EVIDENCIA EXPERIMENTAL. La comprobación experimental que el predictor en estudio provoca el efecto imputado es un elemento que confiere un enorme peso a nuestra presunción. En sentido inverso la eliminación del factor debería reducir o suprimir la ocurrencia del fenómeno.

FUERZA DE ASOCIACIÓN. Esto depende del tipo de investigación que se ha efectuado. Cuando nos referimos a los estudios de cohorte hacíamos mención al RR el cual

nos daba una estimación del grado de relación entre el predictor y la variable de impacto. Cuando los hallazgos provienen de un estudio de casos y controles el estimador de la intensidad de la asociación es el coeficiente de disparidad (Odds Ratio). También se pueden utilizar los datos de investigaciones de tipo más mecanísticas tales como análisis de correlación, regresión y covariancia, entre los más destacados.

GRADIENTE BIOLÓGICO. En analogía con un análisis de regresión donde uno coteja el efecto de los valores asignados a la variable independiente sobre la dependiente (los ejes x e y) aquí se trata de establecer si existe una relación directamente proporcional entre la magnitud que alcanza el predictor (o el tiempo de exposición al mismo) y la presentación del fenómeno de interés. Como para todo existen excepciones, los sujetos con un terreno muy susceptible no estarían comprendidos en las generales de la ley, puesto que pueden presentar el evento ante una mínima exposición al factor de riesgo.

PLAUSIBILIDAD BIOLÓGICA. Llegados al punto de querer atribuir causalidad debemos contar con un sustento científico que brinde una base racional sólida al efecto que deseamos explicar.

RAZONAMIENTO POR ANALOGÍA. Al igual que en el momento de enunciar una hipótesis de trabajo para un proyecto de investigación, aquí tratamos de analizar si el conocimiento de asociaciones causales de naturaleza similar puede ser aplicable a nuestro caso.

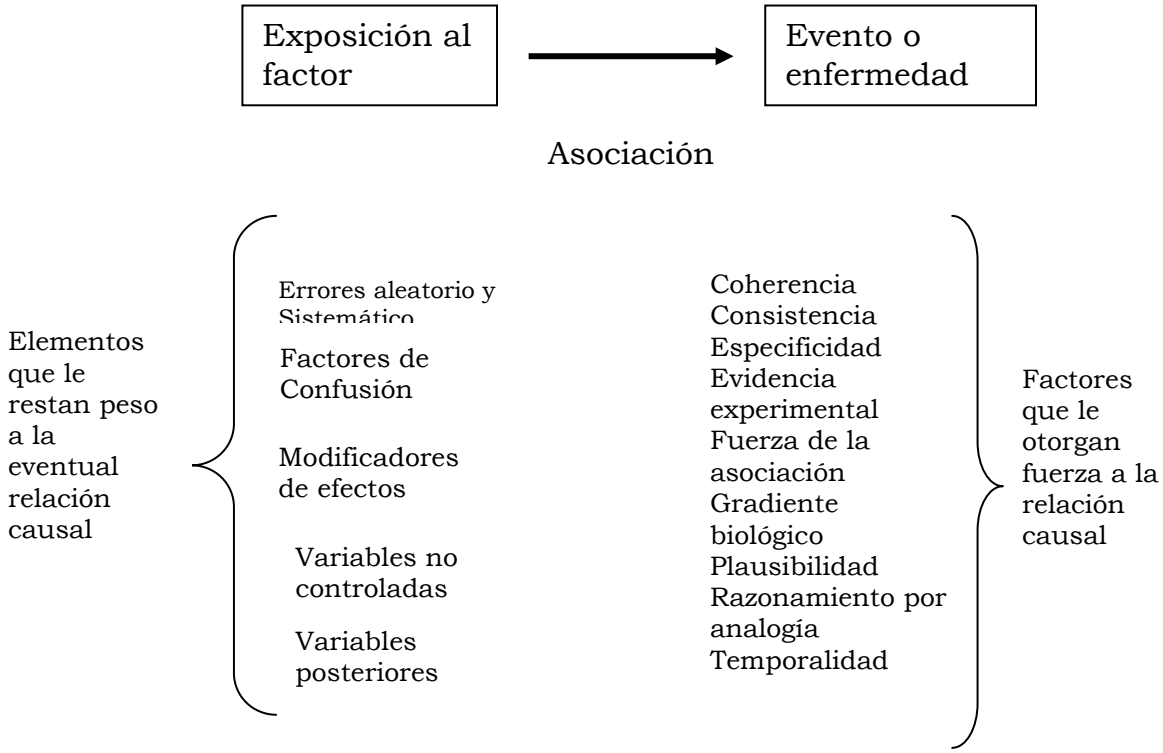
SECUENCIA TEMPORAL. Como fuera expresado, es de fundamental importancia demostrar que el factor de riesgo ha precedido al fenómeno de interés. De ahí la relevancia de los estudios de cohortes, donde al momento de inclusión de los participantes (todos libres de enfermedad) uno puede asegurar que el pretendido factor de riesgo existió antes. Lamentablemente no siempre estamos en condiciones de hacer una afirmación tan categórica.

Uno de los desafíos más importantes de la epidemiología moderna es el análisis de la relación causa-efecto para el cual es primordial un juicio de valor acerca de la calidad de la evidencia científica. A modo de herramienta de trabajo, en la figura 2.1 se representa un marco conceptual que nos ayuda a reflexionar en torno a esta cuestión. Partiendo de la relación entre predictor y fenómeno de interés, tendremos que cotejar entre los factores que le confieren o restan validez a la asociación verificada para lograr una valoración lo más acertada posible de cuán causal es el predictor bajo análisis. En la parte izquierda de la gráfica se indican los elementos capaces de sustraer peso a la relación y en primer término se ubican los errores aleatorio y sistemático. El primero de ellos es inevitable puesto que deberíamos incluir a toda la población blanco para eliminarlo. Contentémonos con minimizarlo todo lo que se pueda y cuantificar la posibilidad de que la variabilidad aleatoria de cuenta de los resultados observados. En otras palabras uno recurre al valor de *P* para saber qué probabilidad existe de que un efecto observado se deba al azar, si no existiera relación alguna entre la exposición y la enfermedad. El error sistemático, por su parte, lleva a una estimación incorrecta de la asociación entre predictor y resultado. Las diferencias radican en la manera en que los grupos del estudio son seleccionados, seguidos y evaluados. Aunque parezca obstinado, hay que preverlos en la fase de diseño y realización del estudio ya que son muy difíciles de corregir cuando ya se recogieron los resultados. Brevemente repasaremos los distintos tipos de sesgos que pueden presentarse. Sesgo de selección: se refiere a las distorsiones que ocurren en los procedimientos utilizados para la elección de los sujetos. De existir un sesgo de este tipo, la relación entre el factor de exposición y el resultado a

evaluar es distinta en los participantes del estudio respecto de los teóricamente elegibles. En razón de ello, la asociación observada en realidad es una combinación entre las fuerzas que determinan la participación de un tipo particular de sujetos y aquellas que operan sobre el fenómeno de interés. Los estudios prospectivos en cohortes son los menos proclives a presentar este tipo de dificultades. Sesgo de información: se trata de un error en la recolección de la información deseada. Puede que el problema se deba al investigador (por Ej., sesgo de entrevistador) o al participante del estudio (individuos con marcadas diferencias en su capacidad de relatar eventos patológicos o exposiciones a factores de riesgo). Sesgos de medición y de clasificación: el hecho de trabajar con un importante número de sujetos, torna muy difícil asegurar que no se producen errores de esta naturaleza. Más que nada para el último de ellos, ya que en el primer caso uno puede tomar una serie de precauciones a fin de que la persona encargada de la obtención de los datos no conozca a qué grupo pertenece el participante. Al referirnos al diseño de casos y controles hacíamos mención al error de clasificación no diferencial o aleatoria, el cual minimizaba las diferencias entre grupos ocasionando una subestimación de la asociación real; mientras que en el error de clasificación diferencial o no aleatoria, el error es distinto según los grupos con lo que el sesgo puede darse en una u otra dirección. En los estudios de corte transversal o de casos y controles hay que estar muy atentos al fenómeno de efecto-causa, vale decir el resultado es lo que lleva al predictor, sobre todo cuando el supuesto predictor es un dato de laboratorio que bien puede ser la consecuencia del proceso patológico y no la causa. No abundaremos acerca de los FC, modificadores de efectos, variables no controladas y variables posteriores, por cuanto ya han sido tratados en la sección anterior.

Confrontando con lo antedicho, en el otro extremo de la figura hemos ubicado a los 9 criterios de causalidad de Bradford Hill más arriba expuestos. La presencia de estos factores otorga un sustento racional a los supuestos causales formulados.

Figura 2.1 Marco conceptual en causalidad



El concepto de causalidad sigue siendo muy debatido no sólo desde la visión científica sino también filosófica. No todas las asociaciones entre exposición y enfermedad se deben a relaciones causales.

En un estudio reciente sobre 159 compuestos clasificados por la Agencia Internacional para la investigación en Cáncer (tomando los carcinógenos 1 o 2A) se aplicaron los 9 criterios de Bradford Hill. El empleo del análisis discriminante para ponderar el peso de cada uno de estos nueves criterios demostró que los criterios de fuerza, consistencia y evidencia experimental eran los de mayor impacto. De todos modos no existe acuerdo sobre qué peso darle a cada criterio individual. Debe otorgarse peso a cada uno de los 9 criterios expresando la importancia relativa de cada criterio para la evaluación de la causalidad global. Es obvio que no todos los criterios contribuyen igualmente a la evidencia general para causalidad y algunos criterios serán más útiles que otros. Al sopesar el valor de la evidencia para la inferencia causal es necesario considerar que la probabilidad de satisfacción para un cierto criterio debe ser estimada en función de los datos disponibles, epidemiológicos o de otra índole.

Bibliografía recomendada

Altman DG, Matthews JNS. Interaction 1: heterogeneity of effects. BMJ 1996; 313: 486.

Altman DG, Deeks JJ, Sackett DL. Odds ratios should be avoided when events are common. BMJ 1998: 317:1318.

Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? .BMJ 1999; 318: 1209.

Armenian H, Lilienfeld DE. Overview and historical perspective, Epidemiol Rev 1994; 16: 1-5.

Austin H, Hill HA, Flanders WD, Greenberg RS. Limitations in the application of case-control methodology. Epidemiol Rev 2004; 16: 65-76.

Bland JM, Altman DG. The odds ratio.BMJ 2000; 320: 1468.

Bradford Hill A. The environment and disease: Association or Causation? Proc Royal Soc Medicine 1965; 58: 295-300.

Brennan P, Croft P. Interpreting the results of observational research: chance is not such a fine thing. BMJ 1994; 309: 727-730.

Christini DJ, Stein KM, Markowitz SM et al. Nonlinear-dynamical arrhythmia control in humans. PNAS 2001; 98: 5827-5832.

Coggon DIW, Martín CN. Time and chance: the stochastic nature of disease causation. Lancet 2005; 365: 1434-1437.

Cole P. The evolving case-control study. J Chronic Dis 1979; 32: 15-27.

Doll R. Sir Austin Bradford Hill and the progress of medical science. BMJ 1992; 305: 1521-1526.

Fletcher R, Flechter S, Wagner E. Clinical Epidemiology. Baltimore, Williams and Wilkins, 1988.

Gordis L. Epidemiology. WB Saunders Co, Filadelfia 1996.

Hulley SB, Cummings SR. Designing clinical research. Williams & Wilkins, Baltimore, 1988.

Julious SA, Mullee MA. Confounding and Simpson's paradox. BMJ 1994; 309: 1480-1481.

Kelsey JL, Thompson WD, Evans AS. Methods in observational Epidemiology. Oxford University Press, Nueva York 1986.

Martel MJ, Rey E, Beauchesne MF, Perreault S, Lefebvre G, Forget A, Blais L Use of inhaled corticosteroids during pregnancy and risk of pregnancy induced hypertension: nested case-control study. BMJ 2005; 330: 230.

Pirkle JL, Brody D, Gunter EW, Paschal DC, Flegal KM, Matte TD. The decline in blood lead levels in the United States: The National Health and Nutrition Examination Surveys. JAMA 1994; 272: 284-291.

Rothman KJ. Causes. Am J Epidemiol 1976; 104: 587-592

Rothman KJ, Greenland S. Modern epidemiology. Philadelphia, Lippincott-Raven 1998.

Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32: 51-63.

Schlesselman JJ. Case-Control studies. Oxford University Press, Nueva York, 1982

Schoenbach VJ, Rosamond WD. Understanding the fundamentals of epidemiology, an evolving text. Fall 2000 Edition, NC, EEUU.

Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet 2002; 359: 431-434.

Simpson EH. The interpretation of interaction in contingency tables. J Royal Stat Assoc 1951; 13: 238-241.

Smith HO. Guidelines for Clinical Trials. TDR-WHO, Ginebra 1992

Swaen G, van Amelsvoor L. A weight of evidence approach to causal inference. J Clin Epidemiol 2009; 62: 270-277.

Swan SH, Shaw GM, Schulman J. Reporting and selection bias in case–control studies of congenital malformations. Epidemiology 1992; 3: 356–363.

Vandenbroucke1 JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M for the STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLoS Medicine 2007 Vol 4 Issue 10| e297

Vernick LJ, Vernick SL, Kuller LH. Selection of neighborhood controls: logistic and fieldwork. J Chronic Dis 1984; 37: 177-82.

Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. Am J Epidemiol 1992; 135: 1019-1028.

Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. II. Types of controls. Am J Epidemiol 1992; 135: 1029-1041.

Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options Am J Epidemiol 1992; 135: 1042–1050.

Weinstock H, Dale M, Gwinn M, Satten GA, Kothe D, Mei J, Royalty J, Linley L, Fridlund C, Parekh B, Rawal BD, Busch MP, Janssen RS. HIV seroincidence among patients at clinics for sexually transmitted diseases in nine cities in the United States. J Acquired Immune Def Syndr 2002; 29: 478–483.

Weiss JN, Garfinkel S, Spano ML, Ditto WL. Chaos and chaos control in biology. J Clin Invest 1994; 93:1355-1360.

West DW, Sehaman KL, Lyon JL, Robison LM, Allred R. Differences in risk estimation from a hospital and a population-based case-control study. Int J Epidemiol 1984; 13: 235–239

White C. Research on smoking and lung cancer: a landmark in the history of chronic disease epidemiology. Yale J Biol Med 1990; 63: 29-46.

Woodeard M. Epidemiology. Study design and data analysis. Chapman & Hall, Londres 1999.

CAPITULO 3. Estudios con intervenciones o ensayos clínicos

Sin temor a equivocarnos, podemos afirmar que los grandes objetivos de la investigación médica propenden a tres logros fundamentales: prevenir enfermedades, desarrollar mejores herramientas diagnósticas y lograr tratamientos más eficaces y menos tóxicos. Para la concreción de dichos propósitos la investigación clínica recurre a un tipo especial de diseño por el cual el médico no sólo observa sino que también interviene sobre la enfermedad en cuestión. Dentro de este contexto, los estudios intervencionistas se pueden agrupar de la siguiente manera:

- Ensayos preventivos; en este grupo se incluyen fundamentalmente los protocolos basados en la utilización de vacunas, sea para proteger a las personas en riesgo de adquirir la infección, desarrollar la enfermedad en los que ya están infectados y al mismo tiempo reducir la transmisibilidad del agente infeccioso.
- Ensayos terapéuticos; ellos están orientados a lograr un mejor control de la enfermedad.
- Pruebas diagnósticas; si bien este tipo de estudios no tiene efecto directo sobre la enfermedad en términos de prevención y tratamiento, la disponibilidad de métodos diagnósticos adecuados es indispensable para la definición de la entidad nosológica y el curso de la misma.
- Intervenciones educativas; se basan en la utilización de campañas cuyo objetivo es conseguir un cambio más saludable en la conducta o hábitos de vida.

3.1 Pruebas diagnósticas

El principal objetivo por el cual se lleva a cabo un ensayo diagnóstico es para establecer su aplicabilidad clínica. En los casos en que ya existe una prueba diagnóstica la nueva herramienta debe ser validada en relación a los estándares disponibles del momento y el diseño que se utiliza es un ensayo aleatorizado a ciego. En este tipo de estudios la variable predictiva es la prueba diagnóstica, mientras que la variable de salida o de impacto es la enfermedad. Respecto a las características de la primera de ellas, la variable predictiva puede ser dicotómica, ordinal o continua. Una prueba diagnóstica ideal sería aquella que arroja un resultado positivo en todos los enfermos, siendo 100% negativa en las personas que no lo son.

Los ensayos diagnósticos requieren que la definición del grupo de enfermos se efectúe en base a los criterios que reúnen consenso (*gold standard*). Si fuese el caso de una prueba pronóstica, como para decidir quiénes se recuperan o no de la enfermedad, debemos tener claras definiciones de lo que ello significa. Dado que en buena parte de la patología médica no se dispone de un "*gold standard*" de absoluta certeza diagnóstica, el investigador debe utilizar un estándar que sea lo más representativo del estándar ideal. En razón de ello en la actualidad es preferible hablar del estándar referencial en lugar del "*gold standard*".

Supongamos que un grupo de investigadores ha observado que los pacientes con una enfermedad infecciosa viral presentan un anticuerpo específico en sangre y deciden efectuar un estudio orientado a determinar qué diferencias existen en cuanto a la positividad de esta prueba, respecto de las personas sin la enfermedad.

Tabla 3.1 Resultados de la prueba serológica

Resultado de la prueba	**Participantes**	
	Enfermos	No enfermos
Positiva	98 (**a**)	30 (**b**)
Negativa	42 (**c**)	90 (**d**)

Los resultados de la Tabla 3.1 indican que la prueba es mayormente positiva en la muestra de enfermos y que ello se traduce en una diferencia estadísticamente significativa ($p<0.0001$) respecto de los individuos sanos. Aún así, el método no será muy útil en términos clínicos ya que en uno de tres pacientes el resultado es negativo.

Sensibilidad y especificidad

De acuerdo a lo representado en la Tabla 3.2, al evaluar una prueba diagnóstica puede ocurrir que no todos los enfermos arrojen un resultado positivo (es decir que se produzcan falsos negativos) y que por el contrario en algunas personas sanas la determinación sea positiva, vale decir falsos positivos.

Tabla 3.2 Agrupamiento para el cálculo de sensibilidad y especificidad

Enfermedad Resultado de la prueba	Presente	Ausente
Positiva	Positivo (P)	Falso P (FP)
Negativa	Falso negativo (FN)	Negativo (N)
	P+FN	FP+N

Dado estas posibilidades, el análisis de una prueba diagnóstica requiere que nos manejemos con dos tipos de estimaciones:

- Sensibilidad (S) = P/P+FN (**a/a+c**); estima cuántos de los enfermos son positivos, lo que equivale a determinar la proporción de positivos verdaderos
- Especificidad (E) = N/FP+N (**d/b+d**); establece cuántos de los sanos son negativos, o sea la proporción de negativos verdaderos

Los ensayos diagnósticos permiten obtener, asimismo, otro tipo de estimadores cuyas principales características se hallan explicitadas en la Tabla 3.3. Una buena estimación en cuanto a la utilidad de una prueba es la razón de probabilidad –RP- o cociente de verosimilitud (*likelihood ratios*); vale decir cuántas veces es más probable hallar un resultado positivo en un sujeto afectado respecto de una persona que no padece la afección. Muchas veces una prueba diagnóstica no arroja hallazgos categóricos, por el contrario son mucho más frecuentes las zonas grises. No obstante ello, la RP de una prueba diagnóstica puede indicarnos en qué medida se ha incrementado o reducido la chance de que el individuo padezca el problema y como tal puede sernos de utilidad en la práctica clínica. Cuando la RP se ubica por encima de 1 nos indica que el resultado de la prueba se asocia con la presencia de la enfermedad y a la inversa si el resultado se halla por debajo de 1. En líneas generales las RP+ mayores de 10 significan que una prueba positiva es buena para inclinarnos hacia el diagnóstico propuesto, mientras que una RP- inferior a 0.1 sugiere que es más apropiada para descartar el diagnóstico presuntivo.

Hace 30 años Fagan (N Engl J Med 293:257-61, 1975) proporcionó un nomograma que permite calcular la probabilidad "*post-test*" cuando se conoce la probabilidad "*pre-test*" (ligada a la prevalencia de la enfermedad) y la RP para la prueba en cuestión. Para ilustrar el tema, la probabilidad previa se construye primordialmente en base a la

frecuencia con que se presenta la enfermedad. Si se tratara del caso de Asma Extrínseca donde alrededor de 1 entre 10 niños la padece es valor que podemos asignarle a la probabilidad previa es 0.10. En el supuesto caso que efectuáramos un estudio diagnóstico "*ad hoc*" cuya RP = 6, la probabilidad posterior será aproximadamente de 0.40, lo cual le confiere mayor fuerza a la presunción diagnóstica inicial. Para el lector interesado en cálculos más ágiles se puede consultar al siguiente sitio http://www.fisterra.com/mbe/investiga/pruebas_diagnosticas/pruebas_diagnosticas.asp.

Para quienes prefieren el cálculo manual y partiendo del ejemplo de Asma extrínseca, el procedimiento es el siguiente:

Probabilidad previa (pretest) = $p_1 = 0.1$
Disparidad previa (pretest odds) = $p_1/(1 - p_1) = 0.1/0.9 = 0.11$
Disparidad posterior (post-test odds, o_2) = disparidad previa × RP = $0.11 \times 6 = 0.66$
Probabilidad posterior = $o_2/(1 + o_2) = 0.66/1.66 = 0.397$

Si retomamos los datos consignados en la Tabla 3.1, los cálculos de sensibilidad y especificidad arrojan los siguientes valores: 0.70 y 0.75. En cuanto a la RP se la calcula para el caso de la prueba positiva (RP+): (98/140)/(30/120)= 2.8 y también para el resultado negativo (RP-): (35/100)/(70/100)= 0.4. Como ya se adelantó, estas pruebas adquieren significación clínica cuando la RP positiva supera ampliamente el valor 1 y la RP negativa por su parte se sitúa bien por debajo de esa cifra. Es obvio que los resultados derivados del ejemplo presentado en la Tabla 3.1, no apuntan a una prueba diagnóstica relevante.

Tabla 3.3 Otros estimadores que pueden obtenerse de los ensayos diagnósticos

Designación	Qué estima	Cálculo
Razón de probabilidad para una prueba positiva*	Cuánto más probable es encontrar un resultado positivo en una persona enferma respecto de la sana	S/(1-E)
Razón de probabilidad para una prueba negativa**	Cuánto más probable es encontrar un resultado negativo en una persona sana respecto de la enferma	(1-S)/E
Adecuación	Qué proporción de todas las pruebas han arrojado un resultado correcto	(a+d)/(a+b+c+d)

*Es el % de un positivo verdadero dividido por el % de un falso positivo
** Es el % de un falso negativo dividido por el % de un positivo verdadero

Valor predictivo.
Los estudios diagnósticos tratan de conseguir dos tipos de mediciones: una de discriminación para separar quiénes están enfermos de quiénes no, y la otra de predicción. Visto a *prima facie*, el valor predictivo positivo –vpp- constituye la probabilidad *post-test* de una prueba positiva; vale decir si una persona tiene resultado positivo, cuál es la posibilidad de que tenga la enfermedad. Tomando los datos de la Tabla 3.1, su estimación se obtiene de la siguiente manera, a/(a+b). Por su parte el valor predictivo negativo –vpn- es la probabilidad post-test de una prueba negativa. En otras palabras, si una persona posee un resultado negativo, cuál es la probabilidad de que no

exista la enfermedad. De acuerdo con la misma tabla su cálculo es como sigue, d/(c+d). En función de los datos presentados en la Tabla 3.1, debemos advertir sin embargo que el valor predictivo varía en función de cuán prevalente es la enfermedad en la muestra con la que se efectúa el estudio. Si nos atenemos a los valores allí representados obtendremos los siguientes resultados, vpp = 0.76 y vpn = 0.68. Supongamos ahora que la prueba sigue mostrando la misma sensibilidad y especificidad, pero el estudio se realizó en un número diferente de personas (150 enfermos y 90 no enfermos, ver tabla), con lo cual el vpp = 0.82 y el vpn = 0.60. Queda claro, entonces, que ambos estimadores, sólo tienen "valor interno".

Resultado	Participantes	
	Enfermos	No enfermos
Positiva	105	23
Negativa	45	67

A la par de estos hechos, es necesario tener presente que cuando se desea trabajar el valor predictivo en la situación clínica su real estimación se torna más complicada por cuanto el mismo está ligado a la probabilidad previa, la cual a su vez puede variar entre pacientes. La probabilidad previa se extrae en base a las características clínicas y demográficas de la persona y la prevalencia de la enfermedad en cuestión. Por ejemplo la probabilidad de enfermedad coronaria es muy baja en un adolescente y muy alta en un fumador con angina de pecho. Aparece así un método adicional para el cálculo del valor predictivo que surge de relacionar sensibilidad, especificidad y probabilidad previa; donde

$$VP+ = \frac{\text{Sensibilidad x probabilidad previa}}{[\text{Sensibilidad x probabilidad previa}] + [(1 - \text{especificidad}) \text{ x } (1 - \text{probabilidad previa})]}$$

$$VP- = \frac{\text{Especificidad x } (1 - \text{ probabilidad previa})}{[\text{Especificidad x } (1 - \text{ probabilidad previa})] \ + \ [(1 - \text{ sensibilidad}) \text{ x probabilidad previa}]}$$

Obviamente que este tipo de cálculo guarda una estrecha relación con el ejemplo representado más arriba de probabilidad previa y posterior para el caso de asma.

Insistimos que cuando se desea determinar la probabilidad previa es imperioso tener presente que el valor de la misma varía según las características del servicio donde se lleva a cabo la exploración. Valiéndonos de lo señalado más arriba, la probabilidad previa de enfermedad coronaria en un centro de atención primaria será muy inferior a la registrada en el servicio donde se derivan pacientes con sospecha de esta patología. Un modo de obviar este inconveniente sería disponer de una hoja tabulada de probabilidades previas para las patologías más significativas según el perfil de la persona.
Por las razones operativas que pueden plantear estos procedimientos no perder de vista que la RP puede utilizarse independientemente de esta circunstancia. Aunque no sea explícito, al momento de solicitar la prueba diagnóstica el médico se mueve con una

probabilidad previa, que al confrontarse con la RP permitirá estimar la probabilidad posterior.
También es posible estimar el poder discriminante de una prueba diagnóstica a través del cálculo del OR diagnóstico (ORD). El mismo es el cociente entre la razón de una prueba de resultado positivo en los enfermos y la razón de un resultado positivo entre los no enfermos. Su cálculo se obtiene de la siguiente manera:

$$\text{ORD}=\frac{Sensibilidad\ /(1-sensibilidad)}{(1-especificidad)\,/\,especificidad}$$

En cierta medida indica la exactitud del procedimiento diagnóstico, vale decir cuanto más frecuente es que ocurra un resultado positivo entre las personas con el fenómeno en cuestión respecto de aquellos que no lo poseen. Como la fórmula lo indica el ORD tiene la ventaja de que su valor no depende de la prevalencia de la patología en la muestra confeccionada. El mismo toma valores que partiendo desde 0 pueden llegar en teoría hasta el infinito. Cuanto más alto es el valor, mejor será el poder discrimatorio de la prueba. Un resultado = 1 indica que el método es totalmente incapaz de discriminar entre grupos.
Asimismo es posible estimar su IC. Para ello se debe obtener el error estándar del ORD (trabajando con el logaritmo)

$$\text{ES (log ORD)} = \sqrt{\frac{1}{P}+\frac{1}{N}+\frac{1}{FP}+\frac{1}{FN}}$$

Tras lo cual se calcula el intervalo (1.96 ± ES) y luego se procede a extraer al antilogaritmo de dicha expresión.

Planificación de un estudio diagnóstico

A la hora de planear una investigación de esta naturaleza hay que poner especial cuidado en los criterios de admisión dado que, de incluirse pacientes con enfermedad severa y sujetos en perfecto estado de salud, es muy probable que se produzca una sobrestimación tanto en sensibilidad como especificidad respecto de lo que se obtendría en la clínica diaria.
La decisión de establecer la positividad de una prueba diagnóstica plantea algunas dificultades cuando la variable es continua. Los datos de la Figura 3.1 están referidos a un estudio hipotético en pacientes tuberculosos y controles de la misma región en quienes se midieron los niveles circulantes de anticuerpos hacia glicolípidos micobacterianos por enzimo-inmunoensayo, a efectos de determinar si esta prueba de laboratorio podría ser utilizada como elemento adicional para el diagnóstico de la Tuberculosis.
En estos casos para decidir cuándo el valor es positivo o negativo debemos elegir el punto o título de corte. Tomando el ejemplo representado en la Figura 3.1, la decisión de establecer un valor límite de 320 Unidades de Densidad Optica -UDO- (valores en que se expresan los resultados) haría que todos los controles fuesen clasificados correctamente como negativos, pero también lo serían algunos pacientes que no deben ser asignados a esta categoría. Si por el contrario el título de corte se ubicara en 260 UDO, la prueba sería positiva en el 100% de los pacientes y en un número mucho mayor de controles poblacionales que obviamente no están enfermos. En este punto debemos ser conscientes que una decisión en favor de aumentar la sensibilidad irá en detrimento de la especificidad y viceversa. Para establecer el título de corte hay que

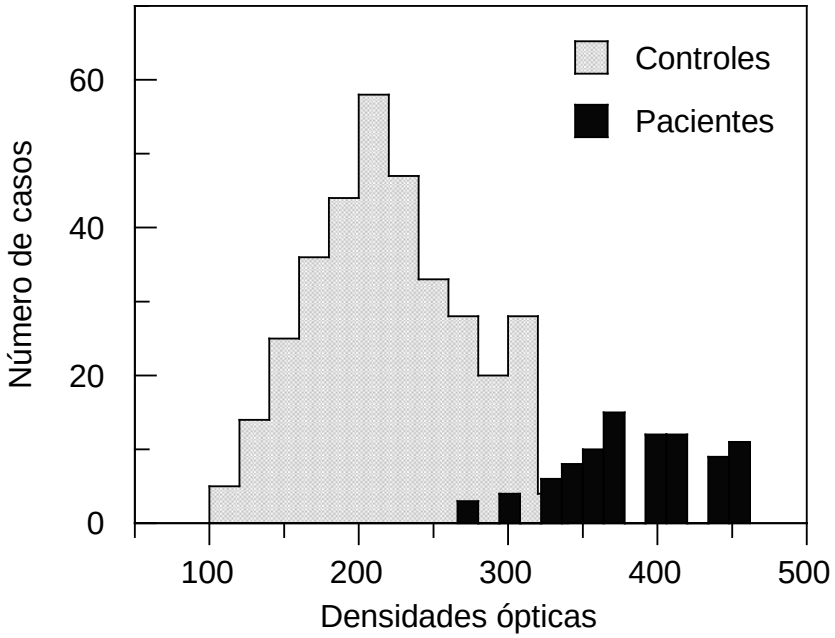


Figura 3.1 Distribución de los niveles de anticuerpos contra glicolípidos micobacterianos en controles poblacionales y enfermos tuberculosos

considerar las posibles implicaciones de los dos errores, el falso positivo y el falso negativo. Si se trata de evitar los falsos positivos (puesto que todos los que arrojen resultados positivos deberán ser sometidos a tratamiento) habrá que maximizar la especificidad.

Un procedimiento muy útil para establecer el punto de corte es la curva ROC (receiver operator characteristic). El mismo se asienta en una metodología desarrollada en el seno de la teoría de la decisión en los años 50 y cuya primera aplicación fue motivada por problemas prácticos en la detección de señales por radar (aunque el detalle pueda parecer anecdótico, la equivalencia entre el operador que interpreta los picos en la pantalla del radar para decidir sobre la presencia de un misil y el médico que emplea el resultado de una prueba diagnóstica para decidir sobre la condición clínica del paciente, es muy alta).

En el ejemplo de la Figura 3.2, el investigador elige varios puntos de corte (P1, P2, P3, P4, P5), determina la sensibilidad y especificidad para cada uno de ellos y procede a graficar sensibilidad en función de 1-especificidad (proporción de falsos positivos). Una prueba ideal es aquella que alcanza un 100 % de sensibilidad y 100 % de especificidad (extremo superior izquierdo de la gráfica); mientras que una determinación carente de valor nos daría una diagonal que va desde el extremo inferior izquierdo hacia el extremo superior derecho, vale decir cada incremento en sensibilidad se acompaña de una pérdida equivalente en especificidad. En el ejemplo de la Figura el título de corte más apropiado estaría entre P3 y P4.

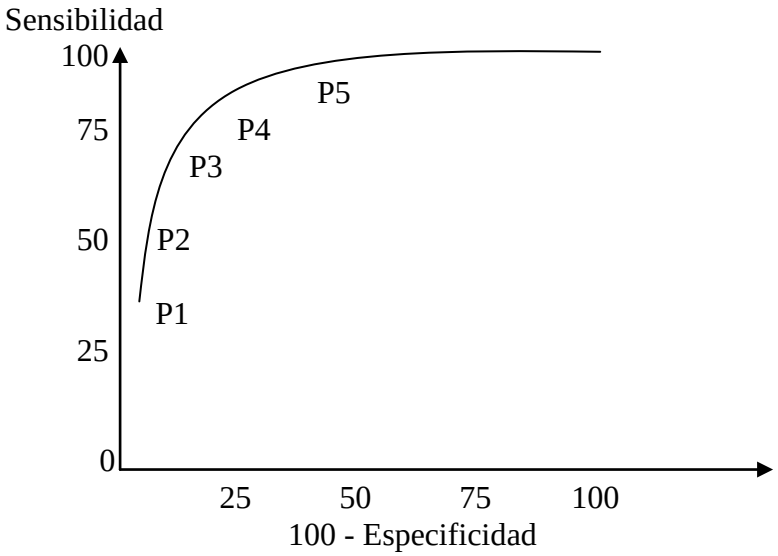


Figura 3.2 Representación de la curva ROC

Teniendo en cuenta las dificultades que puede plantear la elección del título más adecuado, hace unos años Magder y Fix (J Clin Epidemiol 2003; 56: 962) propusieron una estrategia que facilita la toma de decisiones en este sentido, particularmente cuando se trata de un estudio donde la dicotomización que brinda la prueba de laboratorio es utilizada para estimar riesgos (RR u OR).
La curva ROC también permite el cálculo del área bajo la curva (ABC), que para el caso en que se utilizan varias pruebas diagnósticas en una misma patología ayuda a decidir cuál de ellas es la mejor. El ABC de la curva ROC se puede emplear como un índice conveniente de la exactitud global de la prueba: el grado máximo correspondería a un valor de ABC de 1 mientras que el mínimo a uno de 0.5. La aplicación de este procedimiento en pacientes con pleuresía tuberculosa (Okamoto et al., 2005), demostró que los niveles de interferón gamma en los exudados pleurales tuberculosos son de una significativa utilidad diagnóstica.
Desde el aspecto estadístico la cuestión es hasta qué punto la diferencia observada entre las ABC obtenidas es función de variación relacionada con la muestra o diferencias propias de la pruebas.

En los meta-análisis de pruebas diagnósticas se pueden trabajar los distintos ORD obtenidos para proceder al cálculo general y asimismo extraer un resumen de las curvas ROC que brindan los estudios incluidos en el análisis (sROC). Este procedimiento grafica los pares de sensibilidad y especificidad que arrojaron los estudios contemplados a efectos de obtener una curva que pase por tales puntos. Cuando se han empleado diferentes procedimientos diagnósticos para una misma patología, las sROC obtenidas permiten ponderar el valor de las técnicas aplicadas. En una publicación recientemente aparecida (Reitsma et al., 2005), un grupo de investigadores holandeses proponen un modelo bivariado el cual parece superar algunas limitaciones de la sROC.

Dado que el desarrollo de una prueba diagnóstica está orientada finalmente hacia su utilidad clínica no debemos perder de vista que la realización de un estudio donde se comparan pacientes con una enfermedad perfectamente definida respecto de un grupo de personas normales es un hecho “forzado” que no necesariamente representa lo que se da en la situación real. Para ejemplificar el caso, imaginemos un estudio en individuos

con fibrosis pulmonar y controles sanos en los cuales la medición de autoanticuerpos hacia un componente de la matriz extracelular arrojara títulos significativamente superiores para el primer grupo. A partir de esta demostración el punto siguiente será determinar si dicha prueba puede diferenciar casos con y sin fibrosis pulmonar en un grupo de pacientes en quienes hay datos clínicos sugestivos de este tipo de afección.

No está demás insistir que el valor de una prueba diagnóstica no sólo depende de la sensibilidad y especificidad sino también de cuán prevalente es la enfermedad. Si se trata de una patología rara es menos probable que una persona con resultado positivo tenga efectivamente la enfermedad. Esto implica que para patologías poco frecuentes el método debe tener alta especificidad.

Limitaciones y estrategias

También aquí debemos tener cuidado con el error aleatorio y el sistemático. El primero de ellos es casi inevitable, puesto que un paciente bien puede presentar una prueba negativa. La manera de saber cuán imprecisas pueden ser la sensibilidad y la especificidad, se obtiene a través del intervalo de confianza. Si queremos que la dispersión sea baja (±5%), obviamente deberemos incrementar el tamaño de la muestra. Supongamos que nuestra intención para una prueba x, es que la misma tenga un 80% de sensibilidad y 90% de especificidad, en este caso debemos calcular dos tamaños muestrales, enfermos y no enfermos, respectivamente.

El error sistemático tiene que ver por su parte con dos tipos de sesgos. El primero es el sesgo de muestreo, por el cual los pacientes elegidos en un centro de referencia pueden no ser representativos de la población real (por ej. casos más severos o "superpacientes"), o bien los controles son excesivamente sanos. Para evitar esta falencia debemos tratar que la muestra resulte lo más parecida posible a la población sobre la que se aplicará la prueba. El otro tipo de sesgo está relacionado con las mediciones, mucho más si la persona que interpreta los resultados de la prueba diagnóstica tiene acceso a los datos.

En situaciones donde la prueba diagnóstica es primordial para la definición del caso, la sensibilidad y la especificidad de la misma revisten suma importancia al momento de diseñar un ensayo clínico. En este caso la sensibilidad nos indicará cuántos de los casos considerados como tales, lo son en realidad; mientras que la especificidad tendrá que ver con la proporción efectiva de controles que se incluyan bajo esta categoría en el estudio.

Una prueba diagnóstica de baja sensibilidad también reducirá los cálculos de la incidencia de la enfermedad, con lo cual se verá afectado el poder del estudio siendo necesario incrementar el tamaño muestral. Por su parte, la utilización de una prueba diagnóstica de baja especificidad puede ser mucho más perjudicial aún, ya que producirá una sobreestimación de la incidencia de la enfermedad. Al generarse una situación de esta naturaleza tendremos más casos tanto en el grupo con intervención como aquel no sometido a dicho procedimiento, y de ese modo un descenso en la eficacia de lo que pretendemos evaluar, por ejemplo una vacuna.

Paralelamente a lo expuesto, una justa valoración de la utilidad de una prueba diagnóstica requiere un análisis que abarca no sólo los aspectos meramente numéricos sino también el contexto del estudio. Al respecto de ello debemos tener en cuenta si la prueba fue confrontada con un indicador fidedigno de la enfermedad (*gold standard*, o la prueba de referencia que incluso puede ser una combinación de varias pruebas) y si el espectro de individuos que se eligieron era el apropiado. Por ejemplo es importante conocer si se trataba de enfermos de distinta severidad que sumado a ello presentaban

otros trastornos, o cómo eran las características generales de los pacientes y la población control. Otras cuestiones que conviene tener en cuenta son la reproducibilidad de la prueba, los intervalos de confianza y el rango de normalidad para los casos en que se trate de una variable cuantitativa. Para el lector interesado en torno a los estudios con potencial aplicabilidad clínica, existen recomendaciones preparadas recientemente por expertos calificados en la materia (J Natl Cancer Inst 2005; 97: 1180**)**.

Hace poco tiempo dos investigadores Daneses propusieron dividir a los estudios diagnósticos en 4 fases. La primera está destinada a determinar el rango de valores normales de la prueba en cuestión por medio de estudios observacionales en la población sana, a fin de examinar la potencial influencia de variables como edad, sexo, ingesta de fármacos. La fase II debe orientarse a establecer la adecuación diagnóstica de la prueba a través de un diseño de casos y controles. Esta etapa se desgranaría en tres subfases, IIa es la comparación entre enfermos bien definidos y controles; IIb a los efectos de examinar si existe una relación entre la prueba y la severidad de la enfermedad; IIc en este nivel deben incluirse personas en quienes se sospecha la enfermedad con el propósito de analizar el valor predictivo de la prueba. La ejecución de esta fase exige una evaluación a ciego de la prueba y una adecuada elección del punto de corte y los límites de la normalidad para el caso de pruebas cuantitativas. Tras ello, sobrevendría la fase III, en la cual se llevarían a cabo ensayos randomizados donde a un grupo se le aplica la prueba y a otro no. La información obtenida por la determinación sería analizada en términos de su utilidad en cuanto al curso clínico de la enfermedad y las pautas terapéuticas para lograr una valoración más acertada de las consecuencias clínicas que acarrea la introducción de la prueba. Finalmente se arriba a la fase IV de “diagnovigilancia” (se nos ocurre este neologismo) llevada a cabo en cohortes constituidas por un importante número de personas, a fin de evaluar si en la práctica clínica la prueba diagnóstica es tan exacta como apuntan las predicciones de los estudios de fase III.

Consistencia de las pruebas diagnósticas

Dada la importancia de las pruebas diagnósticas no sólo a los fines investigativos sino también para su aplicación clínica, es necesario asegurarse que los resultados arrojados por la técnica sean reproducibles. Mas allá de la variación lógica observada en un mismo individuo al efectuar mediciones en distintas ocasiones, lo importante es determinar el grado de concordancia que existe entre observadores respecto de los resultados de la prueba. Supongamos que se trata de un estudio donde se pretende determinar la proporción de trabajadores de una empresa que presentan electrocardiogramas anormales. A tal efecto se prepara una tabla (3.3) donde se confrontan las categorizaciones (anormal y anormal) efectuados por dos cardiólogos. El porcentaje de acuerdo general surge de aplicar la siguiente fórmula

$$\% \text{ de acuerdo} = \frac{A + D}{Total} \text{x}100$$

en nuestro caso 105/117 x 100 = 89.7%

Tabla 3.3 Ordenamiento de los datos para el cálculo de acuerdo entre dos observadores

Primer Cardiólogo	**Segundo Cardiólogo**	
	Anormal	**Normal**
Anormal	**A (10)**	B (7)
Normal	C (5)	**D (95)**

En estudios donde la mayor parte de las evaluaciones arrojan resultados que se ubican dentro de lo normal, el porcentaje de acuerdo está un tanto sobredimensionado por esta situación. En esos casos y a efectos de lograr una estimación más acertada del grado de similitud entre los observadores, el cálculo se efectúa sin tener en cuenta las evaluaciones cuyos resultados fueron normales para los dos evaluadores. Vale decir que no tendríamos en cuenta las cifras de la celda D. En el denominador se ubican las evaluaciones que fueron consideradas anormales por alguno de los observadores como mínimo. De este modo se obtiene la siguiente fórmula: (a/a+b+c) x 100; vale decir: (10/22) x 100 igual a 45%.
Cuando los observadores que deben efectuar las interpretaciones tienen bastante disparidad de criterio es probable que el porcentaje de acuerdo esté influenciado en cierta medida por una simple coincidencia entre las dos evaluaciones. Para no descartar esta posibilidad nuestros amigos los estadísticos han diseñado la prueba *kappa*. La misma estima en qué medida el acuerdo observado supera al que cabría esperarse simplemente por chance. La fórmula de la prueba *kappa* se describe a continuación,

$$\frac{\text{Porcentaje de acuerdo observado - Porcentaje de acuerdo esperado por chance}}{\text{100\% - Porcentaje de acuerdo esperado por chance}}$$

Primer Cardiólogo	Segundo Cardiólogo		Totales del 2do
	Anormal	Normal	
Anormal	10	7	17 (14.5%)
Normal	5	95	100 (85.5%)
Totales del 1ro	15 (12.8 %)	102 (87.2 %)	117

Recordemos que el porcentaje de acuerdo que habíamos observado inicialmente era de 89.7%. Para evaluar el porcentaje de acuerdo por chance se debe tener en cuenta que el 2do Cardiólogo interpretó como anormales al 12.8% de los ECG. En base al criterio que el mismo utilizó para definir la anormalidad se puede asumir que el 2do Cardiólogo asignaría como anormal al 12.8% de los ECG clasificados como anormales y normales por el primer Cardiólogo. De este modo podemos esperar que un 12.8% de los ECG clasificados como anormales por el 1er Cardiólogo (2.17) también fuesen anormales y que otro tanto de los ECG señalados como normales (12.8) por el Cardiólogo primero sean clasificados como anormales. Vale decir que el porcentaje de acuerdo por chance es 2.17+12.8/117 = 12.8%.

$$\text{Kappa} = \frac{89.7 \;-\; 12.8}{100 \;-\; 12.8} = \frac{76.9}{87.2} = 0.88$$

En líneas generales se admite que los valores que superan el 75% denotan un grado de acuerdo importante.

Cuando se trata de variables continuas se han utilizado diferentes estrategias para analizar los niveles de acuerdo o concordancia. Los más familiares por así decirlo son los estudios de correlación y el coeficiente de correlación intraclase. En razón de las limitaciones de ambos procedimientos, Altman y Bland propusieron en 1983 un método de graficación, donde en el eje de Y se indica la diferencia entre las dos mediciones (A-B) y en el de X el promedio de ambas mediciones (A+B/2). Partiendo de este enfoque, hace unos años se propuso un nuevo método (J Clin Epidemiol 2003; 56: 963) donde se confronta la magnitud del desacuerdo entre observadores respecto de la frecuencia conque se presenta cada uno de los valores que asumen las diferencias.

3.2 Ensayos terapéuticos

Udite, udite, o rustici; atenti, non fiatate.
Io già suppongo e immagino che al par di me sappiate
ch'io sono quel gran medico, dottore enciclopedico,
chiamato Dulcamara, la cui virtù preclara,.....
Benefattor degl'uomini, riparator de'mali,
in pochi giorni sgombero, io spazzo gli spedali,
e la salute a vendere per tutto il mondo io vo.....
Per questo mio specifico, simpatico, prolifico,
un uom settuagenario e valetudinario,
nonno di dieci bamboli ancora diventò,
di dieci o venti bamboli fin nonno diventò.
Per questo "tocca e sana" in breve settimana
più d'un'afflitta vedova di piangere cessò.
O voi matrone rigide, ¿ringiovanir bramate?
Le vostrte rughe incomode con esso cancellate.
Volete voi, dozelle, ben liscia aver la pelle?
Voi giovani galanti, per sempre avere amanti?
Comprate il mio specifico, per poco io ve lo do.
Da brave giovanotti, da brave vedovette,
comprate il mio specifico per poco io ve lo do.
Ei move i paralitici, spedisce gli apopletici,
gli asmatici, gli asfitici, gl'isterici, i diabetici,
guarisce timpanitide, e scrofole e rachitidi,
persin'il mal di fegato, che in moda diventò.....

Extracto de la cavatina de Dulcamara, Opera *L´elisir d´amore*

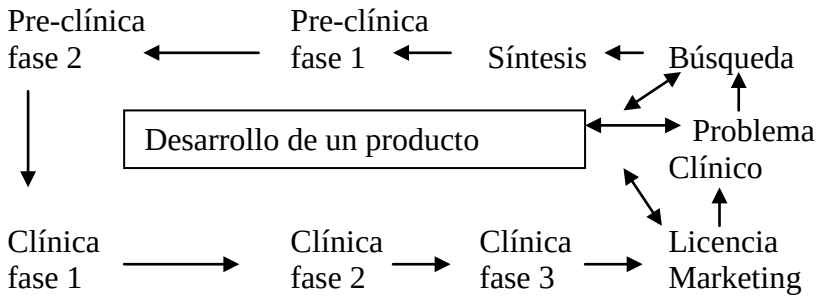
El propósito fundamental de un ensayo clínico es determinar si un fármaco es bueno o no para el tratamiento de una determinada enfermedad y proveer una evidencia sólida sobre la validez de los resultados. Si bien la opinión de los médicos es un dato a tener en cuenta, la utilización de una droga no puede depender de este criterio puesto que se trata de una impresión clínica. Como tal es probable que muchos factores puedan incidir sobre el resultado restándole valor a las conclusiones. Para no cometer esta equivocación se han diseñado los ensayos comparativos, controlados, aleatorizados y doble ciego. Todo estudio que involucra una intervención, en este caso un tratamiento, requiere la comparación con un grupo control ya que desde el punto de vista estadístico se lo trata como un factor de riesgo.
Al igual que en los estudios observacionales, la validez de las conclusiones extraíbles de un ensayo clínico no sólo depende de la precisión estadística sino también del grado de representatividad de la muestra respecto de la población blanco.

<u>Indicaciones para llevar a cabo un ensayo clínico</u>
Así como los estudios observacionales requieren el financiamiento por parte de las entidades públicas o privadas que promocionan la investigación, la realización de ensayos clínicos no sería posible sin la participación de la industria farmacéutica. En la mayor parte de los casos, es precisamente el interés del laboratorio comprometido con el desarrollo de un fármaco quien promueve los estudios que determinarán su utilidad

clínica. En líneas generales, el sustento racional para la ejecución de un ensayo clínico se resume en 4 aspectos fundamentales:

- Que exista una buena chance de obtener un resultado positivo, a juzgar por los fundamentos previos (resultados *in vitro*, experimentación animal, información preliminar).
- Investigación de una práctica que si bien está aceptada no ha sido rigurosamente evaluada.
- Pérdida de eficacia de los tratamientos utilizados hasta ese momento.
- Evaluación en mayor profundidad de regímenes en uso (toxicidad, farmacocinética).

Figura 3. 3 Fases de la investigación farmacológica



Las distintas etapas que atraviesa el desarrollo de un medicamento

Como se ilustra en la Figura 3.3, el desarrollo de una nueva modalidad terapéutica es un proceso gradual en el cual no se pueden saltar etapas. Muy por el contrario, el cumplimentado de una serie de pautas es el requisito fundamental para proceder al paso siguiente. La secuencia de dichos procesos se resume a continuación.

Síntesis (1-2 años). A partir de un determinado problema médico y teniendo en cuenta los tratamientos disponibles en ese momento, un laboratorio farmacéutico encarga al departamento de estudios básicos la síntesis de compuestos (8 a 10.000 de ellos), que en teoría podrían tener algún efecto terapéutico específico. Los distintos aspectos que abarca esta etapa incluyen:

- Procedimientos de biología molecular y tecnología recombinante para los casos en que se pretende realizar terapia sustitutiva con un mediador naturalmente sintetizado por el ser humano (por ejemplo, el interferón alfa).
- Química de proteínas que permiten aislar el producto.
- Química analítica para la caracterización del producto y síntesis en gran escala.
- Estudios microbiológicos para determinar la esterilidad y ausencia de pirogenicidad.
- Estudios inmunológicos en el caso que debamos determinar su inmunogenicidad (vacunas, antígenos) y pruebas de potencia.
- Estudios sobre la eventual toxicidad de los compuestos.
- Preparación de los documentos que exigen las agencias reguladoras para un medicamento nuevo (“investigational new drugs”).
- Cuestiones legales vinculadas al patentamiento del producto, contratos y licencias con alguna otra entidad, a los efectos de profundizar y continuar los estudios.

Estudios pre-clínicos (2-3 años)

Los compuestos que hayan reunido los requisitos indispensables de aceptabilidad serán utilizados en investigaciones pre-clínicas que comprenden 2 fases. En líneas generales,

un número aproximado de 18 compuestos arriban a la fase 1 (el examen de ingreso es extremadamente difícil) y en la misma se llevan a cabo los siguientes estudios:

- Toxicidad aguda por medio de la administración a dos especies distintas de animales (mayormente roedores).
- Investigaciones farmacológicas detalladas sobre su acción, duración y efectos colaterales.
- Análisis y estabilidad de la sustancia activa.

Transcurridos estos estudios, aproximadamente dos tercios de los compuestos, pueden ser utilizados en los estudios de fase 2, dirigidos a evaluar los siguientes ítems:

- Farmacocinética, absorción, distribución, metabolismo y excreción.
- Toxicidad subcrónica por administración repetida a dos especies animales (generalmente roedores).
- Estudios toxicológicos reproductivos tales como fertilidad y teratogenicidad.
- Pruebas de mutagenicidad.
- Desarrollo de un esquema de dosificación.

Una vez que se han cumplimentado las investigaciones básicas se inician los *estudios clínicos* en seres humanos, los cuales tienen una duración aproximada de 3 a 4 años. Esta etapa se subdivide a su vez en tres fases. Menos de la mitad de los compuestos que se emplearon en los estudios pre-clínicos 2 son aceptables para su empleo en la fase I (4-5 fármacos). Los objetivos fundamentales que persiguen estas investigaciones tienen que ver con el conocimiento acerca de,

- La farmacocinética del compuesto.
- La tolerabilidad en voluntarios sanos, tendiente a caracterizar la dosis máxima tolerada y mínima efectiva, la relación dosis-efecto, la duración del efecto y las eventuales acciones indeseables.

También existen ensayos oncológicos de fase I. En ellos se incluyen pacientes con cánceres avanzados y refractarios a los tratamientos disponibles con el objeto de evaluar la inocuidad y toxicidad de los nuevos compuestos terapéuticos, para establecer la farmacocinética y determinar el rango de dosis más seguros a emplear en los futuros estudios. La legislación en EEUU establece que se permite la exploración de estos regímenes. Algunos autores sostienen que los participantes de estudios fase I en cáncer pueden tener la chance de experimentar un beneficio clínico significativo, aunque más no sea un pequeño grupo de pacientes y que los comités de revisión deberían permitir estas investigaciones. En el mismo sentido se suman los ensayos con agentes biológicos (por ejemplo anticuerpos monoclonales). Es necesario no perder de vista sin embargo que la seguridad del paciente siempre será el primer objetivo de la investigación clínica. El respeto por las personas, la beneficencia y la justicia son 3 principios éticos fundamentales y que deben ser sopesados con los beneficios potenciales tanto para el participante como la sociedad.

Lo central en estos casos es que ni el médico ni el paciente interpreten erróneamente el objetivo de un estudio fase I, al creer consciente o inconscientemente que la participación tendrá un beneficio directo sobre el participante. Algunos análisis sistemáticos muestran alguna evidencia de regestión tumoral en un 5 a 6% de los pacientes que participan de ensayos fase I con compuestos anti-neoplásicos. Volvemos a insistir que para estos casos lo mejor es que el paciente entienda que el propósito de esta fase I es ganar un conocimiento con posibilidades de ayudar a otros pero probablemente

no a su persona; aunque en algunas situaciones puede ocurrir un beneficio clínico objetivo y subjetivo por la misma participación en el estudio.

Concluida la primera fase clínica, sobreviene la Fase II (2-3 sustancias). La característica fundamental de esta fase es que en ella se llevan a cabo los primeros ensayos controlados para evaluar la eficacia del fármaco en los pacientes con la enfermedad en cuestión. La planificación de estos estudios requiere tener en cuenta una serie de aspectos esenciales. Aún cuando la información preexistente indique que el fármaco puede ser efectivo, esta posibilidad aún no ha sido evaluada en pacientes. De ello se desprende lo anti-ético de administrar un nuevo medicamento a personas que pueden ser tratadas con los fármacos habitualmente en uso. Dicho en otros términos: el nuevo esquema terapéutico sólo se podrá probar en aquellos pacientes no respondedores al tratamiento disponible o bien que presenten efectos tóxicos con la consecuenta suspensión del mismo. El requisito de trabajar a partir de enfermos con fracasos terapéuticos o intolerancias medicamentosas está orientado a que sólo se introduzcan fármacos de mayor eficacia y/o inocuidad. En lo referente a este último aspecto, los ensayos clínicos de fase II también incluyen las evaluaciones de toxicidad crónica por administración repetida y los estudios de carcinogenicidad.
Por las características señaladas es evidente que este tipo de ensayos son abiertos y en muchas oportunidades a un solo brazo (es decir un único grupo) con un número reducido de pacientes. El nivel de eficacia que se obtiene a partir de estas investigaciones nos permite tener una idea bastante aproximada de la magnitud del efecto. Este elemento es utilizado para el cálculo del tamaño muestral (ver Capítulo 1), que demanda la realización de los estudios de fase III. A modo de ejemplo, en los inicios de los años 90 nos tocó coordinar un ensayo abierto Fase II, en el cual se evaluaba la eficacia terapéutica del interferón gamma recombinante en la leishmaniasis mucocutánea. Como el tratamiento estándar para esta enfermedad infecciosa son los antimoniales pentavalentes, por razones éticas el protocolo establecía que sólo se podían incluir pacientes clínicamente resistentes a la medicación.
Cuando nos trasladamos a los estudios preventivos basados en el empleo de vacunas, la fase II puede adoptar modalidades distintas. Si quisiéramos determinar la eficacia de una vacuna sobre determinada enfermedad viral, un estudio fase II no sería lo suficientemente potente como para proveer una idea preliminar acerca de si los vacunados tienen menos chance de enfermar. Este inconveniente se supera fácilmente si se cuenta con una prueba de laboratorio fehaciente sobre el papel protector de la vacuna, tal como la medición de los niveles de anticuerpos neutralizantes específicos que induce la inmunización.

Arribamos así a la Fase III (1 sustancia) en la cual finalmente se evaluará si el nuevo compuesto puede ser aplicado o no. Son los típicos ensayos multicéntricos en gran escala que deberán probar la eficacia e inocuidad de la intervención y determinar el perfil terapéutico de la misma. Dicho en otros términos se establecerán las siguientes pautas:

- Indicaciones
- Dosis y tipo de administración
- Contraindicaciones
- Efectos colaterales
- Precauciones
- Interacción con medicaciones concomitantes

Si los resultados fueran exitosos (es difícil imaginar que el laboratorio patrocinante haya destinado tanto tiempo y dinero para un fracaso), se procederá al Registro en las dependencias sanitarias correspondientes, para su posterior lanzamiento y ventas; es un trámite que demanda de 2 a 3 años. Las instituciones más prestigiosas del mundo para otorgar las autorizaciones correspondientes son las comisiones de la Comunidad Económica Europea y la administración de drogas y alimentos (FDA) de los EEUU. El laboratorio deberá someter toda la documentación relevante de la sustancia para que sea analizada por comités de expertos en ensayos clínicos, en estudios farmacológicos y en toxicología. Tras haberse obtenido la aprobación, se prepara el lanzamiento al mercado, se establecerán los planes de comercialización y se procede a entrenar a los agentes de propaganda. Es responsabilidad del laboratorio efectuar los controles de calidad para la producción del fármaco.
Existe una fase adicional (4 o de fármaco-vigilancia) que está orientada a detectar efectos tóxicos no constatados en la etapas previas. De vez en cuando nos enteramos que determinado laboratorio retira de circulación un medicamento a raíz de un efecto adverso severo, que se presenta a las perdidas (por ejemplo cada 50.000 pacientes). Si el tamaño muestral del ensayo fase III con el que se obtuvo la aprobación hubiese sido de 7.500 enfermos, habría habido una muy baja probabilidad de observar dicho efecto puesto que el error beta era bastante elevado para detectarlo. Recordar que el error beta es la probabilidad de no encontrar la diferencia (para este caso en términos de inocuidad) cuando en realidad la misma se halla presente.

Consideraciones para el diseño de un ensayo clínico
La planificación de un estudio de estas características abarca desde la fundamentación y presentación de los objetivos de la investigación hasta la preparación de los formularios para la recolección de datos. La redacción de un protocolo pormenorizado es esencial por varias razones. Una de ellas es facilitar la evaluación por parte del Comité de Bioética. Un protocolo exhaustivo da la pauta que se ha reflexionado en profundidad acerca de las distintas etapas del estudio a efectos de su concreción. Si bien es difícil prever todas las dificultades que pueden surgir al llevar a cabo un estudio (fundamentalmente en investigaciones de campo) una planificación detallada puede proveer de antemano la solución de eventuales problemas de ejecución. Por otro lado la formulación del pedido de subsidio no podría especificarse si no establecemos de entrada lo que queremos hacer y cómo pretendemos alcanzar esos objetivos.
Las recomendaciones sobre los principales ítems a tener en cuenta en el momento de planificar el estudio se resumen a continuación:
1) Qué criterios adoptaremos respecto al conocimiento sobre el tipo de tratamiento que recibe cada paciente, abierto, simple ciego, doble ciego, o cerrado para algunos observadores.
2) Cómo será la configuración de los grupos a tratar, en paralelo o cruzado (cross-over).
3) De qué manera se constituirá el grupo control,

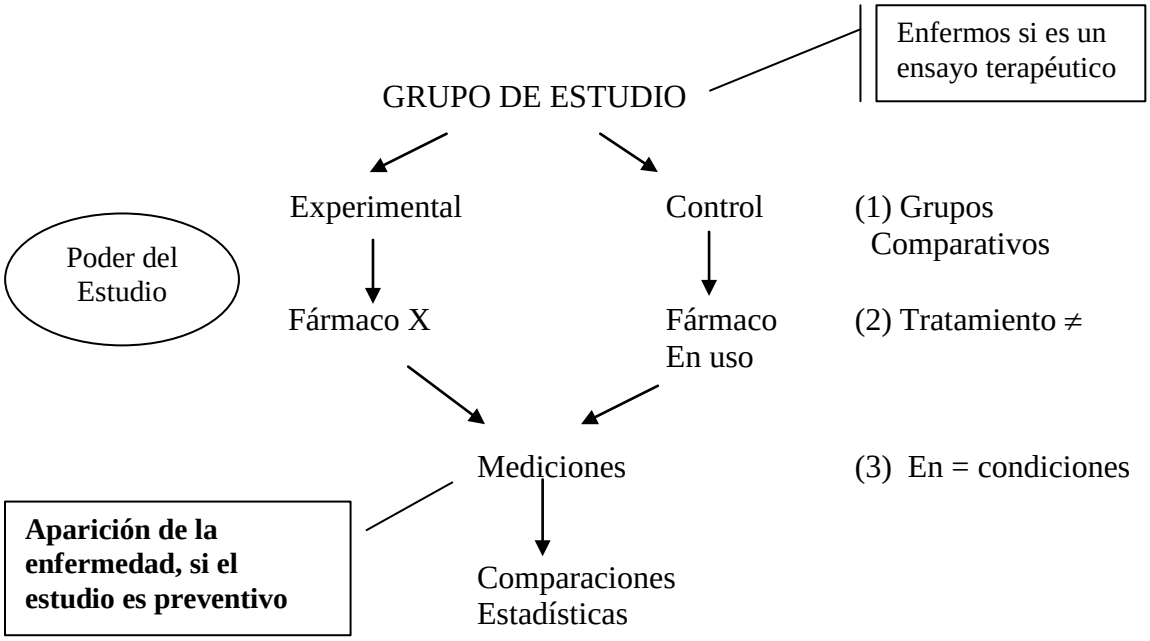
- simultáneos con placebo
- simultáneos y sin tratamiento (control negativo)
- simultáneos pero con tratamiento (control positivo)
- simultáneos con dosis diferentes
- históricos (se lo trata en el párrafo siguiente)

4) Cuál es el valor delta, vale decir en cuál nivel la diferencia de eficacia entre fármacos es considerada clínicamente significativa.

5) Cuáles serán las variables de eficacia con las que trabajaremos
   categóricas (curado / no curado)
   escalas numéricas
6) De qué manera se recogerán los efectos adversos para el análisis de la inocuidad.

Un ensayo con controles históricos es un estudio en el que se comparan los resultados en un grupo tratado con datos de una serie de casos ya publicada o con pacientes previamente tratados en la misma institución, a menudo un ensayo clínico aleatorizado (ECA) y estudios observacionales. Dado que el ECA es el estándar para la incorporación de nuevos tratamientos, qué ventajas tendría la comparación con controles históricos. Por un lado las cuestiones éticas y las relacionadas al tamaño muestral, puesto que se requiere un número menor de pacientes sin reducción en el poder del estudio. Si la respuesta del grupo control es conocida con un buen grado de certeza o si el número de controles supera en gran medida al grupo experimental, el número de pacientes a incluir en el nuevo grupo se reduce hasta la mitad de lo calculado (vale decir una cuarta parte de lo que requeriría un ECA). Para el caso de enfermedades infrecuentes o ensayos muy costosos la ventaja es clara. Los ECA se llevan a cabo con el propósito de demostrar que un nuevo tratamiento es superior al estándar, lo cual implica que muchos pacientes recibirían un tratamiento supuestamente inferior.
Un estudio con controles históricos en realidad es un estudio observacional. La estimación del efecto del tratamiento puede estar distorsionada si los controles difieren sistemáticamente del grupo experimental. Las diferencias pueden darse en la selección de los pacientes o bien en la evaluación subsiguiente del tratamiento. El análisis demandará que las covariables basales con potencialidad para actuar como confusoras estén distribuidas igualmente en los 2 grupos, o bien que exista un procedimiento que permita ajustar para tales diferencias.
Otra cuestión a tener en cuenta es cómo han sido los criterios de inclusión para ambos grupos, incluso es probable que los pacientes con menor riesgo no sean asignados al nuevo tratamiento, o bien que aquellos con un mal pronóstico sí ingresen porque se busca mejorar la chance de éxito. El estadío de la enfermedad también es otro factor a tener en consideración, ni que hablar si al momento en que se estudiaron los controles no existían las técnicas que se disponen cuando se analiza el grupo experimental.
Los estudios con controles históricos difícilmente resultan convincentes y la comunidad científica seguramente aguardará la confirmación a través de un ECA.
La recomendación de efectuar un ECA y a doble ciego, vale decir que ni el paciente ni el médico conozcan el tratamiento asignado, también apunta a descartar el efecto que podría ejercer el conocimiento de la medicación administrada sobre la forma en que el paciente es controlado por su enfermedad y la respuesta del individuo ante la terapéutica instituida.
Existen situaciones en las que a raíz del tratamiento indicado es fácil reconocer quién esta recibiendo la intervención y a quién se le administra placebo únicamente, por ejemplo los protocolos donde se utilizan interferones que pueden ocasionar fiebre. De ocurrir una situación de estas características hay que propender a que la evaluación sea lo más objetiva posible utilizando observadores externos al estudio, como una alternativa.
De acuerdo con lo ya expresado, antes de abocarnos al diseño propiamente dicho del estudio, nuestro propósito deberá haber satisfecho afirmativamente 3 cuestiones fundamentales, si existen indicaciones claras para la realización del ensayo, si es posible llevarlo a cabo y quiénes serán las personas a estudiar. Habiendo avanzado en torno a

dichos interrogantes procederemos a establecer los criterios para confeccionar una muestra representativa de la población blanco, para aplicar el siguiente esquema:



Si los grupos son comparativos (1) y los procedimientos del estudio se hallan controlados (3), podemos inferir que de observarse alguna diferencia significativa en las evaluaciones la misma tiene que ver con el tipo de tratamiento instituido (2).

En la mayor parte de los ensayos clínicos existe una variable de salida (dependiente) que reviste el mayor interés para los investigadores. Para el caso de una intervención terapéutica será el número de curados y no curados tras el tratamiento, mientras que si se trata de una práctica preventiva (vacuna) lo importante será determinar la incidencia de la enfermedad infecciosa en un período de años estipulado. Dependiendo de los aspectos operativos, éticos y fundamentalmente económicos puede ocurrir que los alcances del estudio vayan más allá de los objetivos mencionados y también quisiéramos determinar si el procedimiento de inmunización consigue reducir la transmisibilidad de la infección (no la enfermedad) en la población. Continuando con el ejemplo de las enfermedades infecciosas, si nuestro estudio estuviera centralizado en una patología que tiene un desenlace fatal en un 20% de las personas afectadas, por razones éticas es obvio que la variable de impacto sobre la que planearemos el estudio no puede tener en cuenta esta situación sino la aparición de la enfermedad, aún cuando el tamaño muestral y el costo de la investigación se vean notablemente incrementados.

<u>El ensayo de no inferioridad o equivalencia</u>

El ECA típico esta dirigido a demostrar que el nuevo tratamiento es mejor que el placebo o control (ensayo de superioridad). Por el contrario un ensayo de no inferioridad con control activo busca probar que el nuevo tratamiento no es inferior al estándar en función de una magnitud del efecto predefinida y clínicamente aceptable conocida con el margen de no inferioridad (NI) o equivalencia. Si se establece la NI, la utilidad del nuevo tratamiento puede adoptarse en función de ventajas adicionales como inocuidad, conveniencia y costo. El ensayo de NI con control activo es cada vez más utilizado, especialmente en patología cardiovascular u oncológica donde el ECA con placebo es anti-ético.

De acuerdo a la Conferencia Internacional para la armonización de los lineamientos, el término *no inferiordad* se utiliza al referirnos a un ensayo una sola dirección; *equivalencia*, cuando se trata ensayo en dos direcciones (no es superior ni inferior al otro tratamiento). Se espera que el nuevo tratamiento tenga una eficacia no inferior o equivalente respecto del tratamiento estándar y que a la vez tenga las ventajas ya señaladas. El nuevo tratamiento también puede representar una terapia alternativa o de segunda línea. Estos estudios implican una planificación y análisis especial. El margen de NI debe ser tenido en cuenta al efectuar el cálculo del tamaño muestral. Dicho margen debe ser menor o igual al valor más pequeño que representaría una diferencia clínicamente significativa, o el valor más grande que significaría una diferencia clínicamente intrascendente. Para ello debemos basarnos en criterios tanto clínicos como estadísticos. Se requieren tres atributos para establecer la NI: 1) el tratamiento en consideración debe mostrar una no inferioridad terapéutica respecto del control activo; 2) el tratamiento exhibiría eficacia terapéutica en un ensayo controlado con placebo si se efectuara tal estudio; y 3) el tratamiento ofrece ventajas adicionales en inocuidad, tolerabilidad, costo, o conveniencia. Los ensayos designados para mostrar NI requieren de una población de referencia apropiada, un control activo probado, un alto grado de adherencia al tratamiento, y un adecuado poder del estudio.
Cuando los investigadores suponen que el compuesto en cuestión provee de otras ventajas además de una mayor eficacia, se puede recurrir al estudio de NI. Para llevar a cabo un estudio de NI válido, debe existir evidencia de estudios históricos que proveen estimaciones confiables, reproducibles y precisas del efecto del control activo, comparado con placebo sobre los puntos de interés específicos que se piensa analizar en el NI. Este efecto debe tener una magnitud sustancial, y la estimación del efecto del control activo proveniente de los estudios históricos debe ser tenida en cuenta en el ensayo de NI en ciernes de incluirse un grupo placebo. Estos requisitos permiten formular un margen de NI tal que, si el estudio establece una efectividad de la intervención experimental no menor a la efectividad del control activo, el régimen experimental (1) preserva una fracción sustancial del efecto del control activo y (2) que no resultará en una caída de la efectividad clínicamente relevante.
Para resumir digamos que un estudio de NI requiere establecer la mayor diferencia desfavorable que sigue siendo compatible con NI. Un margen de NI demasiado amplio puede llevar a la conclusión que un nuevo tratamiento es equivalente al tratamiento estándar cuando en realidad es inferior.

Tipos de randomización

La razón fundamental para la asignación aleatoria de los tratamientos es evitar sesgos potenciales. El término aleatorio no debe confundirse con azar. El primero está referido a una chance igualitaria que no puede ser predicha. Azar tiene que ver más bien con la ausencia de un plan u orden. Una traducción aproximada sobre el significado de randomización podría ser "procedimiento orientado a evitar la animosidad". Para el caso de un estudio clínico, "descartar la subjetividad del investigador" al momento de asignar qué tipo de tratamiento recibirá el paciente. No siempre es fácil para el médico a cargo del control de los participantes sustraerse de la tentación de administrar el tratamiento supuestamente más efectivo, fundamentalmente en los casos donde se siente más inclinado a tomar una decisión de esta naturaleza. Para neutralizar este factor subjetivo, conviene que exista una asignación aleatoria preparada por una persona ajena al estudio, como podría ser el personal del Departamento de Estadística del laboratorio patrocinante. Los tipos de randomización que pueden utilizarse se resumen a continuación

- Secuencia de sobres cerrados conteniendo una tarjeta que indica el tipo de tratamiento. Esta modalidad sólo es aconsejable cuando no existe una oficina centralizadora de asignación y el clínico a cargo del paciente no tiene otro referente.
- En el caso de ensayos doble-ciego la persona a cargo del Departamento de Provisión de Fármacos, asigna los tratamientos (frasco A o frasco B) utilizando un criterio similar al anterior.
- Para ensayos multicéntricos la Oficina de Asignación del laboratorio que lleva adelante el estudio transmite el tipo de tratamiento por teléfono, al clínico o bien al Director del Departamento de Farmacia.

Sea cual fuere el procedimiento utilizado para indicar qué tipo de tratamiento recibirá el paciente, este acto se lleva a cabo tras la incorporación del enfermo al estudio según los criterios de inclusión y exclusión. De ese modo se excluye la posibilidad de una admisión del paciente influida por el hecho que el médico sabe de antemano la medicación a administrar.

Tipos de randomización

Una vez que se haya elegido el modo de informar al médico sobre el tratamiento a aplicar, resta por definir el tipo de randomización a utilizar. Las alternativas existentes se resumen a continuación.

Randomización simple, se prepara una tabla de números al azar (desde el 0 al 9) y luego se establece un criterio de asignación. Por ejemplo los dígitos 0-4 corresponden al tratamiento A, mientras que los 5-9 indican tratamiento B. Un ejemplo para indicar el tipo de tratamiento en un estudio con dos modalidades terapéuticas es el siguiente:

1 6 3 9 8 0 4 5 2 7 1 0 9 6 5 2 6 3
A B A B B A A B A B A A B B B A B A

Otra manera de asignar el tipo de tratamiento podría ser: droga A, para los números pares, y droga B, para los impares. La longitud de la lista depende de la cantidad de pacientes que se necesite incorporar. En los casos en los cuales el tamaño de la muestra es reducido (n = 20 o n = 40) puede ocurrir que el número de pacientes medicados con uno u otro tratamiento sea desigual. Para evitar esto conviene chequear la lista, a efectos de determinar si hay desigualdad en las asignaciones. En caso de ser así se procederá a preparar una nueva lista.

Bloques permutados. Es un método que permite obtener una muestra igual de pacientes. El tamaño de los bloques debe ser múltiplo del número de grupos a tratar. Suponiendo que tenemos 2 tipos de tratamientos (A y B) podemos tomar bloques de 4 pacientes, asignar un número a cada tipo de combinación y posteriormente elegir uno de estos números en forma aleatoria.

ABBA	1
BABA	2
AABB	3
BBAA	4
BAAB	5
ABAB	6
Sin asignación	7,8,9,0

Tomando la secuencia de números ejemplificados en el párrafo anterior, los dígitos 4 y 1 nos permitirán establecer la siguiente secuencia de tratamientos: BBAA ABBA
Se pueden establecer bloques mayores, según el tamaño muestral.

Randomización estratificada

Al llevar a cabo un ensayo aleatorizado es muy conveniente que los grupos bajo diferente tratamiento sean similares en cuanto a características de significación para el tipo de pacientes que se estudiarán. Por ejemplo, no sería deseable que en un ensayo donde se prueban dos hipolipemiantes uno de los grupos tuviera una proporción significativamente mayor de obesos. Si se trata de un estudio lo suficientemente grande lo más probable es que no ocurran desigualdades importantes. Para los casos en que existe una posibilidad cierta de la existencia de diferencias en el perfil de los pacientes, se emplea la randomización por estratos. Para ello debemos tener en claro qué factores tomaremos en cuenta al efectuar la estratificación. Si la idea es vaga o no fundada conviene no utilizar esta alternativa.
Supongamos que dos factores importantes en la dislipidemia son el peso y el tabaquismo. Para el peso establecemos tres categorías (según las tablas confeccionadas en base a estatura y edad), normal, sobrepeso y obeso; mientras que para el tabaquismo atribuimos dos categorías, fumador y no fumador. De este modo tenemos 6 estratos, con lo que deberemos preparar 6 listas de randomización en bloques permutados.

Normal y fumador:	BABA	ABAB	ABBA	BBAA	BAAB	AABB
Normal y no fumador:	ABAB	ABBA	BBAA	BAAB	AABB	BABA
Sobrepeso y fumador:	ABBA	BBAA	BAAB	AABB	BABA	ABAB
Sobrepeso y no fumador:	BBAA	BAAB	AABB	BABA	ABAB	ABBA
Obeso y fumador:	BAAB	AABB	BABA	ABAB	ABBA	BBAA
Obeso y no fumador:	AABB	BABA	ABAB	ABBA	BBAA	BAAB

Minimización para la asignación de los tratamientos

La randomización por bloques o por estratos tiende a asegurar la comparatividad entre grupos de tratamiento. Sin embargo en los estudios donde el tamaño muestral es más bien reducido la estratificación no puede aplicarse. Para estos casos la alternativa que puede utilizarse es la minimización.
En este tipo de procedimiento el primer paciente es asignado en forma aleatoria. Para cada participante subsiguiente se debe determinar qué tratamiento llevaría a un mejor balance entre los grupos respecto de las variables que nos interesan. Siguiendo el ejemplo provisto pensemos en peso y tabaquismo (ver Tabla 3.4).

Tabla 3.4 Perfil de los participantes del estudio

Características de los pacientes	Droga A (n=25)	Droga B (n=25)
Peso		
Normopeso	12	12
Sobrepeso	11	10
Obesidad	2	3
Fumadores	8	9

Los investigadores han incluido un total de 50 pacientes (25 para cada tipo de tratamiento). Supongamos que el nuevo candidato para incorporar al estudio es una

persona no fumadora y con sobrepeso, si a la misma se le asignara la droga A tendríamos una persona más con sobrepeso (11+1) y otro no fumador en el grupo (17+1). Tomando los guarismos correspondientes a sobrepeso y hábito de fumar en el grupo A tenemos 11+17=28; mientras que en el B es 10+16=26. La asignación del tratamiento B restauraría el balance. Si las sumas en los 2 grupos arrojan cifras similares, el nuevo tratamiento se asignará aleatoriamente.

Randomización desigual

En los ensayos donde se compara un nuevo tratamiento con el utilizado en ese momento, puede ser que pretendamos incluir más pacientes en el brazo de la nueva modalidad terapéutica. Cuando estemos frente a este tipo de situaciones, lo aceptable es trabajar en una relación 2:1 o 3:2, puesto que si la proporción es mayor se afecta ostensiblemente el poder del estudio

Aunque este procedimiento ofrece una serie de ventajas el método es poco utilizado como para reducir los costos de un estudio. La randomización desigual debería ser más considerada en los diseños especialmente cuando existen diferencias importantes entre los costos de tratamiento. Algunos han bregado por una asignación desigual a fin de exponer menos pacientes a los riesgos del tratamiento en estudio. Este enfoque presupone que existe un alto grado de certeza respecto de los riesgos y beneficios que resultará del tratamiento en evaluación. No obstante, como muchos ECA se efectúan cuando hay un equilibrio entre ambos componentes, este enfoque no puede ser generalmente utilizado como una justificación para la asignación desigual. Sin embargo, pueden darse casos, donde si bien se conoce la naturaleza benéfica de un tratamiento, los efectos adversos desfavorables están aún siendo investigados. Aquí la asignación desigual podía estar potencialmente justificada desde el punto de vista ético.

<u>El índice de reclutamiento (IR)</u>

Se trata de un índice propuesto recientemente, el cual indica el número de días que demanda incorporar un paciente analizable en un centro típico para ensayos multicéntricos. El mismo se obtiene de la siguiente manera,

$$IR = (PVUP - PVPP) \times C/P$$

Dónde:

- PVUP es la fecha de la primera visita del último paciente (el final de la fase de reclutamiento del estudio).
- PVPP es la fecha de la primera visita del primer paciente (el comienzo de la fase de reclutamiento del estudio).
- C corresponde al número de centros que participan de la investigación.
- P es el número de pacientes que efectivamente completaron el estudio.

A partir de la fórmula presentada también es posible calcular la duración del período de reclutamiento (PVUP – PVPP). El cual surge de la siguiente ecuación,

$$PVUP – PVPP = IR \times P/C$$

O bien del número de centros que se requerirán para llevar a cabo el estudio,

$$C = IR \times P /(PVUP – PVPP)$$

Teniendo en cuenta los valores obtenidos en estudios previos (referidos siempre a la misma patología), el IR puede ser utilizado para evaluar,

- la eficiencia de los distintos centros para reclutar pacientes,
- el impacto de algunas características del protocolo (criterios de inclusión y exclusión), sobre la duración del período de reclutamiento,
- la duración del período de reclutamiento para un nuevo estudio o el número de centro que participarán,
- el IR promedio para un determinado tipo de tratamiento y enfermedad,
- el IR específico para determinada región geográfica.

También permite efectuar comparaciones dentro y entre proyectos.

Efectos adversos

La identificación de las acciones indeseables que puede tener un fármaco es un aspecto tan importante como la eficacia. La verificación de los potenciales efectos colaterales que puedan experimentar los pacientes será mucho más fácil si al momento de conducir el estudio los investigadores cuentan con información en este sentido, sea de las investigaciones llevadas a cabo en modelos animales o estudios efectuados con compuestos que guardan similitud estructural. En líneas generales no es aconsejable preparar un interrogatorio dirigido a recabar efectos adversos por la posibilidad de introducir un sesgo en el sentido de que el paciente comenzara a referir una serie de malestares para los cuales no es fácil establecer una relación causal con el tratamiento. Como elementos alternativos se puede recurrir a la información obtenida del grupo que recibe el tratamiento estándar o bien placebo, aunque lo más valioso es el dato espontáneo referido por el paciente sobre tal signo o manifestación. De producirse esta eventualidad los responsables del estudio deberán implementar todas las medidas que sean necesarias tanto para la certificación del efecto adverso como para la resolución de dicha complicación. Muchas compañías farmacéuticas suelen adjuntar una serie de recomendaciones a seguir para cada uno de los efectos adversos más relevantes.
En el año 2008 apareció una publicación de la Alianza Mundial para la Seguridad de los pacientes (WHO) que provee de una información muy elaborada respecto a los efectos adversos que pueden presentarse en el cuidado de los enfermos, sus causas, el modo de comunicarlos, y cómo prevenirlos, entre otros aspectos. El documento es un consenso al que arribaron expertos internacionales altamente calificados en el tema y también provee de los lineamientos fundamentales para el manejo de estas contingencias, sobre todo cuando el cuadro requiere que se actué rápidamente.

Tengamos en cuenta que un efecto adverso grave es aquel que genera una lesión o impedimento definitivo para el paciente (o su progenie) pudiendo resultar en cualquiera de las siguientes eventualidades:

- Muerte
- Disminución en la expectativa de vida
- Una situación que amenaza la vida del paciente (aunque sea superada incluso sin secuelas)
- Incapacidad prolongada o permanente
- Deterioro en la capacidad para superar problemas de salud que se susciten en el futuro

Los ECA hacen hincapié en la eficacia del compuesto pero mucho menos en cuestiones de inocuidad. Un estudio reciente basado en la bibliografía disponible en MEDLINE

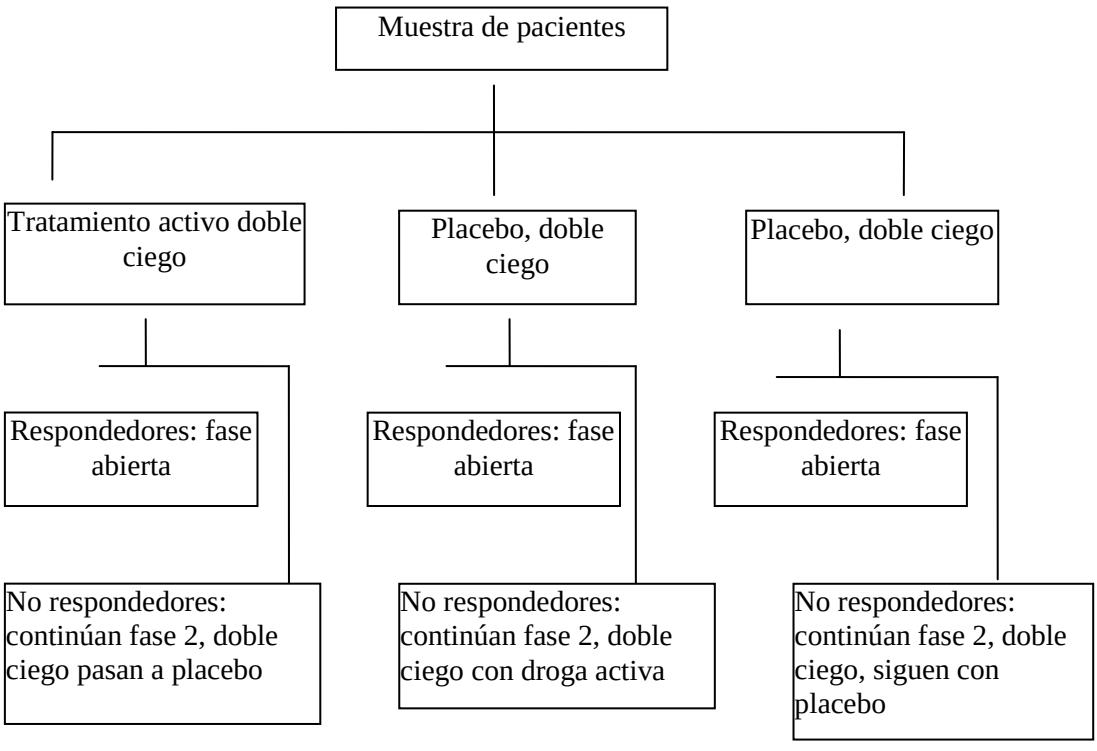
respecto de estudios publicados entre el 01/01/2006 y el 01/01/2007, en 6 revistas médicas con alto impacto, indica que un 88.7% de 133 trabajos se mencionan los efectos adversos [Arch Intern Med 2009]. En 43 artículos también se observaron restricciones respecto a los datos referidos a daño para el paciente (32.3%); con una descripción de los efectos adversos más comunes solamente (n=17), únicamente los severos (n=16) o bien una combinación (n=5).

Otros tipos de ensayos clínicos

Un diseño ad hoc para la utilización de placebo

La utilización de placebo en la investigación clínica es una cuestión muy debatida que ha merecido un sinnúmero de comentarios, investigativos y éticos; cuya consideración escapa a los propósitos de este manual (ver citas al final del capítulo). En lo concerniente a la investigación farmacológica para trastornos psiquiátricos el tema se torna aún más álgido dado que las respuestas al placebo son elevadas reduciendo el margen de la magnitud del efecto para lo cual se necesitan más pacientes y con ello un costo mayor para el desarrollo de nuevos tratamientos.

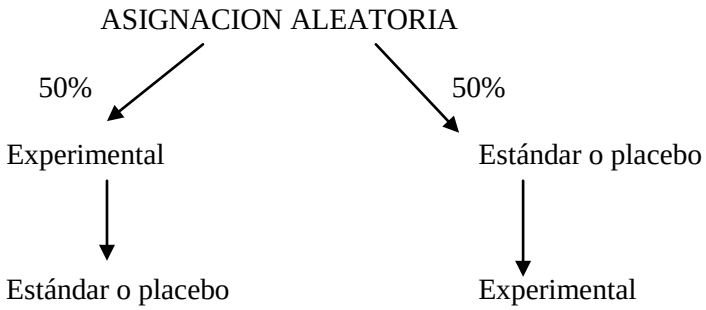
A raíz de ello se ha propuesto un nuevo tipo de diseño orientado a reducir la respuesta al placebo. Los autores (*Psychother Psychosom* 2003; **72:** 115–27) lo han bautizado comparación secuencial en paralelo, suerte de diseño cruzado parcial que se expone brevemente. En la 1er fase del estudio se establecen 3 grupos randomizados de asignación, dos reciben placebo y el restante tratamiento activo. Al final de esta primera fase, los que hayan respondido al nuevo tratamiento ingresan a un ensayo abierto (segunda fase), mientras que los no respondedores pasan a un tratamiento con placebo a doble ciego. Por su parte los respondedores al placebo (de la fase 1) también ingresan al ensayo abierto (segunda fase), y los no respondedores por su parte participarán de una segunda fase (doble ciego) donde podrán cambiar a droga activa o permanecer con placebo.



Sin desmedro del esfuerzo que han hecho los investigadores para pergeñar este tipo de diseño es necesario tener en cuenta que: ello implica un número desigual de pacientes en cada grupo a efectos del poder del estudio; entre quienes no respondan en el grupo placebo existe una chance de 50/50 de continuar con placebo aún cuando hayamos visto la ausencia de respuesta. Qué comité de Bioética aprobaría tal actitud y si lo hiciera exigiría un consentimiento escrito donde va de suyo que la chance de recibir placebo es muy alta. Otra objeción importante es que no todos los datos generados se incluirán en el análisis. Por otro lado al no existir un grupo donde los que reciben droga activa continúan con la misma medicación, puede que perdamos la oportunidad de visualizar un efecto retardado del fármaco.
Del mismo modo en que se han obtenido respuestas en los pacientes tratados con placebo, una revisión muy reciente sobre ensayos con estatinas señala que entre un 4-26% de los casos medicados con sustancia inactiva abandonaron los estudios a raíz de experimentar efectos adversos. Para tener una mejor dimensión del tema inocuidad seguramente habrá que cotejar la frecuencia de efectos adversos en cada ensayo (por grupo de tratamiento) respecto de la que ocurre en la población general.
Resta por comentar que un estudio en curso subvencionado por el Instituto Nacional de la Salud de EEUU está explorando un nuevo diseño para el análisis del efecto placebo en pacientes con colon irritable (Contemp Clin Trials 27: 123, 2006).

El ensayo clínico cruzado (*cross-over*)

A diferencia de los diseños paralelos en los cuales ambos esquemas de tratamiento se prueban en dos grupos independientes de enfermos, para el estudio cruzado el mismo grupo de enfermos recibe los dos tratamientos. Este tipo de estrategias es factible de ser realizada en pacientes con enfermedades crónicas (degenerativas, psoriasis, terapia sustitutiva en la menopausia, entre otras), en los cuales se administran dos tipos de tratamiento, uno tras otro durante un lapso de tiempo similar. De ese modo se puede evaluar ambos tratamientos en los mismos pacientes, lo que en teoría permite una comparación más precisa al eliminar diferencias que en mayor o menor grado siempre se presentan entre dos grupos de enfermos. A pesar de esta importante ventaja, no siempre es posible llevar a cabo un ensayo cruzado por factores inherentes a la enfermedad, los pacientes y el tratamiento. Si el curso de la enfermedad no es estable (por ejemplo se producen remisiones espontáneas) será dudoso atribuir la mejoría al tratamiento instituido. Es necesario tener en claro que los fármacos a utilizar en este tipo de ensayos son aquellos que logran mejorar el cuadro del paciente pero no curar la enfermedad. Si se tratara de un caso de este tipo debe diseñarse un ensayo paralelo. El otro factor a considerar es el nivel de cooperación de los enfermos; no nos olvidemos que los pacientes serán sometidos a dos tipos de tratamientos, con todo lo que ello implica.
La realización de ensayos cruzados sigue dos tipos de estrategias. En la primera de ellas los pacientes son asignados al azar y doble ciego a dos grupos de tratamiento. Los mismos consistirán en la administración de la droga experimental seguida del tratamiento estándar y viceversa.



El lapso de administración está en relación directa con el tipo de enfermedad y el tiempo que requiere el fármaco para ejercer un efecto clínicamente evidente. En algunos casos conviene fijar un período sin medicación entre ambos tratamientos ("*wash-out*") a fin de eliminar la posibilidad de efectos residuales.

Aún cuando la literatura especializada indique que la enfermedad en cuestión presenta una estabilidad razonable, la incertidumbre de no saber cuán estables son los pacientes del estudio ha derivado en una estrategia orientada a incluir enfermos que reúnen dicho requisito. En este caso los pacientes reciben a simple ciego, únicamente placebo por un período de tiempo similar al empleado en los tratamientos posteriores. Los pacientes que no experimentan mejoría son admitidos en el estudio real. Esta alternativa asegura en gran medida que se trata de enfermos estables y/o no susceptibles al efecto "escoba nueva". La cuestión de difícil resolución está relacionada con los planteos éticos del período de prueba con placebo.

Por todo lo antedicho, los ensayos cruzados están reservados para situaciones bastante bien definidas y se emplean con más frecuencia en los estudios de fase II.

El análisis estadístico de la información que brindan los ensayos cruzados requiere de algunas consideraciones especiales, que se describen a continuación.

Supongamos que se trata de un estudio en enfermos diabéticos insulino-independientes donde se prueban dos hipoglucemiantes orales (A y B). El primer paso será efectuar un análisis global de los niveles de glucemia al finalizar el tratamiento A, menos los valores obtenidos al concluir la terapia con el compuesto B y viceversa. Supongamos que las mediciones respectivas en uno de los pacientes indican 124 y 119 gr/dl, con lo cual la diferencia será de + 5gr/dl. Tras efectuar esta misma operación para todos los participantes se obtiene la media y el desvío estándar de las diferencias entre A y B. Con ello es posible evaluar el significado estadístico de los cambios mediante el empleo de la *t* de Student para datos apareados (ver Capítulo 4). Si la media de diferencias fuese positiva y estadísticamente significativa estaríamos en condiciones de concluir que el hipoglucemiante B es superior al compuesto A. Este modo de analizar los datos no tiene en cuenta, sin embargo, la influencia potencial del efecto período. Dicho en otros términos la posibilidad que la condición de los enfermos no haya sido la misma en los períodos en que recibió cada uno de los dos tratamientos. Para indagar acerca de esta posibilidad procederemos a ordenar los datos como se muestra en la Tabla 3.4.

Tabla 3.4 Disposición de los datos para la estimación del efecto período

	A seguido de B	B seguido de A
Media de la diferencia entre glucemias	AB	BA
Desvío estándar de esa media	ds1	ds2
Número de pacientes	n1	n2

De este modo se puede efectuar la prueba de la *t* para el efecto período

$$t = \frac{AB - BA}{\sqrt{\frac{ds_1^2}{n_1} + \frac{ds_2^2}{n_2}}}$$

Si el resultado de la prueba arroja un valor estadísticamente significativo, no estaremos tan seguros de que un tratamiento es superior al otro, dado que las diferencias obtenidas dependen en buena medida del momento en que se administra el hipoglucemiante.
Para ensayos donde la variable de impacto es categórica o en escala ordinal se deben utilizar otros tipos de comparaciones estadísticas, por ejemplo la prueba de rangos de Wilcoxon (ver Capítulo 4).

Recapitulando sobre todo lo expuesto, en alguna medida un ensayo clínico es la confrontación entre dos puntos de vista. Por un lado el hipotético, según el cual lo existente en la información previa deberá observarse en el estudio.

Existente → Observación

El punto de vista observacional señala por el contrario que lo observado en la investigación es lo que en realidad existe.

Observación → Existente

Para no caer en falsas antinomias y obtener conclusiones valederas, es aconsejable reducir el nivel de error a su menor expresión posible ya que de ese modo aumentaremos el nivel de certeza de ambos puntos de vista.

Ensayos clínicos en pequeña escala tanto mecanísticos como de eficacia terapéutica

En líneas generales dichos estudios están referidos a ensayos clínicos de 100 pacientes o menos aún en un único centro hospitalario. Habitualmente, serían las primeras investigaciones sobre un nuevo tratamiento para una enfermedad (fase II). Su propósito es mayoritariamente explicativo a fin de lograr una mejor comprensión en cuanto a la fisiopatología y terapéutica. Los mecanismos subyacentes pueden ser variados y no necesariamente uno solo, independientemente de la patología en estudio. Es uno de los grandes desafíos de la medicina. En parte porque el sustento proviene de estudios funcionales o pruebas laboratoriales, y por otro lado debido a que no todos los pacientes afectados por una enfermedad compartirían un mecanismo fisiopatogénico similar. Muchas veces los estudios previos son trabajos en progreso no totalmente desarrollados, y si el paciente responde a la medicación estaremos logrando un avance en el tema.

En estudios de esta naturaleza es aconsejable aplicar la última tecnología al alcance, en pos de discernir del mejor modo posible los mecanismos fisiopatogénicos que operan en cada uno de los pacientes. Por las dificultades de trasladar a la práctica investigaciones con un alto nivel de sofisticación, en paralelo se deberían incluir variables de medición más sencillas y subrogantes de los procesos subyacentes.
Veamos algunas de las estrategias existentes.

Maximizar el efecto del Tratamiento y Minimizar la variación
Los elementos para el cálculo del tamaño muestral ya han sido comentados. Si la estimación establece que debemos incluir 60 pacientes por brazo quizás no exista suficiente dinero como para llevar adelante el estudio, en virtud de las pruebas laboratoriales requeridas. Las alternativas son incrementar el efecto del tratamiento (maximizar la dosis u optimizar la elección de los pacientes incluyendo aquellos en quienes hay sospecha de una mayor chance de lograr el efecto) o bien reducir la variancia experimental; lo cual no es tan sencillo.
En este sentido se podría reducir la variabilidad fármacocinética. Al respecto es aconsejable calcular los parámetros farmacocinéticos del paciente tras una única dosis de la droga y utilizar dichos valores para diseñar un régimen de dosis repetidas o medir las concentraciones plasmáticas durante administraciones reiteradas y utilizar esos datos como estrategia de ajuste. Este diseño es atractivo para compuestos que pueden tener una toxicidad elevada (quimioterápicos e inmunosupresores).
Otra estrategia es disminuir la variabilidad en la medición del punto final. Si existe un "*gold standard*" para este propósito obviamente que deberemos adherir al mismo. Independientemente de ello valen aquí las consideraciones efectuadas acerca de precisión y exactitud como así los controles de calidad en cuanto a la forma de evaluar, si existe más de un operador. De ser factible conviene mejorar la estimación del efecto y consecuentemente la variabilidad, lo que redundará en una reducción del tamaño muestral. Finalmente y cuando está a nuestro alcance se puede recurrir al ensayo cruzado; del cual ya nos hemos referido. A pesar de sus ventajas, este diseño puede plantear algunos problemas por ejemplo si los cambios inducidos por el tratamiento no revierten rápidamente tras la discontinuación del mismo. Las modificaciones en las características de la enfermedad de base a lo largo del tiempo pueden introducir variabilidad en la respuesta de los pacientes, restando valor a las ventajas de esta estrategia. Esto se supera en parte si el ensayo tiene una duración corta. Por otro lado cuando el estudio es muy prolongado, tarde o temprano algunos pacientes se retirarán del mismo. En razón de ello algunos expertos son bastante críticos en cuanto a la aplicabilidad de los resultados en un estudio que llega a completarse en una minoría de los pacientes enrolados inicialmente. Otro punto crítico es la posibilidad de sesgo producida por los efectos "persistencia" desigualmente distribuidos. Por ejemplo cambios en la eficacia de los tratamientos resultantes de tratamientos administrados en el primer período; probablemente por la persistencia de los fármacos o sus metabolitos, como así también cambios tisulares ocasionados por el tratamiento, factores conductuales o psicológicos. El cálculo por efecto período es útil en estos casos. Los más críticos señalan no obstante que los resultados pueden ser difíciles de interpretar siempre que el tratamiento difiera en ambos períodos. En estos casos no se puede distinguir con certeza si ello se debe a efecto persistencia o interacción "período-tratamiento" (el pasaje del tiempo afecta la eficacia relativa de los tratamientos).
En razón de ello las instituciones que regulan la utilización de medicamentos ponen bastantes reparos respecto de los datos derivados de este tipo de diseño.

Diseño de reclutamiento enriquecido

Es una variante del diseño cruzado, que puede ser útil para estudiar tratamientos a los cuales responde una minoría de los pacientes. Si las diferencias no son estadísticamente significativas en un ensayo convencional, no se puede retrospectivamente imputar a los respondedores y argumentar que el tratamiento ocasionó la mejoría. Es posible, sin embargo, incluir a los respondedores en una posterior comparación prospectiva, o una serie de comparaciones entre tratamiento y placebo. Si los resultados del segundo ensayo son estadísticamente significativos, podemos suponer que la respuesta inicial del paciente no fue por simple chance. Aunque es defendible desde el punto de vista estadístico, el diseño es criticado ya que la exposición previa al tratamiento podría anular el procedimiento doble ciego (en particular con tratamientos que tienen efectos colaterales distintivos) y dar a lugar a resultados positivos espurios. Otra precaución es que los resultados positivos de una población enriquecida de respondedores no pueden generalizarse a toda la población de pacientes; sólo sugiere que existe una subpoblación de respondedores. El reclutamiento enriquecido puede tener un rol valioso en el desarrollo de un fármaco puesto que uno de los primeros pasos en este proceso es detectar un grupo en el cual la droga tiene un efecto favorable.

N=1 o diseño de un único caso

Es otra variante del estudio cruzado. Los investigadores interesados en enfermedades raras o aquellos capaces de elaborar distinciones mecanísticas pueden desear estudiar la respuesta de pacientes individuales. Sin embargo en un diseño cruzado simple de una droga vs. placebo, el poder para distinguir efecto real de chance por variación en un único paciente es bajo. Uno puede incrementar ese poder al randomizar al paciente a múltiples rondas de cada tratamiento. Una estrategia utilizada es darle 3-5 pares de droga vs. placebo, y randomizar el orden dentro de cada par para luego ver que pasa en cada uno de los ensayos sucesivos.

Algunos autores enfatizan sobre su utilidad clínica como para guiar la práctica de un paciente en particular. Otros sostienen que está muy mal basar el tratamiento sólo en los datos de un paciente, dado que la variación aleatoria puede ocasionar tanto falsos positivos como falsos negativos.

Diseño en Paralelo

Si la realización del estudio cruzado plantea serias dudas o si la historia natural de la enfermedad puede ir cambiando a lo largo del estudio, es preferible realizar un estudio paralelo. La variabilidad entre pacientes es una limitante para el diseño en paralelo y existen algunas estrategias como para limitar en parte dicha desventaja. Por ejemplo se pueden tener en cuenta los puntajes registrados al momento de la inclusión para luego sustraerlo del puntaje logrado por el tratamiento y trabajar con la diferencia o bien tomar el puntaje de inicio como co-variable. Ello elimina una buena parte de la variación e incrementa el poder del estudio. También se puede tratar de balancear los grupos para las variables con algún efecto potencial sobre la respuesta, si se sospecha de ello. Si se desea evaluar la respuesta por subgrupos, las asignaciones deben estar balanceadas en función de ello. Se puede utilizar la randomización por estratos. En un estudio con un tamaño muestral de 20- 40 pacientes por grupo, este método incrementa el poder del estudio si las variables pronósticas están bien seleccionadas. Con una muestra mayor a 50 pacientes por grupo, el proceso de randomización por sí mismo tiende a balancear la mayoría de las variables. Si la estratificación no fuera posible deberemos pensar en un análisis *post hoc* para las co-variables si las variables en cuestión muestran una distribución no tan dispar entre los grupos.

Diseño quasi-*experimental*

Los *quasi*-experimentos son estudios en los cuales existen diversos tratamientos, variables de impacto, pero no se cuenta con un procedimiento de asignación al azar como para inferir que las diferencias que arrojan las comparaciones se deben a los tratamientos instituidos. En un experimento de verdad las asignaciones se efectúan por randomización. Este tipo de diseño se utiliza cuando no es posible aplicar un procedimiento de randomización para el tipo de tratamiento a instituir. En principio cada ensayo randomizado podría tener una contraparte *quasi* experimental. La designación de "*quasi*-experimento" puede incluir diseños de un único grupo que obviamente no poseen una contraparte de ensayo randomizado, como sería una serie de casos. Los datos que provee este tipo de estudio son esenciales para ponderar la magnitud del efecto, que es uno de los elementos centrales para el cálculo del tamaño muestral.

Dado que este tipo de diseño no utiliza la randomización para la conformación de los grupos, los mismos (experimental y control) pueden llegar a diferir respecto de otros factores (estimados o no) capaces de influir sobre el desenlace. De ahí que las diferencias en cuanto a los resultados en ambos grupos bien pueden deberse a una mixtura entre los efectos de la intervención bajo análisis y efectos surgidos de las diferencias pre-existentes entre grupos. Así las cosas, la remoción de los confusores aparece como un elemento central en este tipo de diseño a fin de obtener una estimación lo menos sesgada posible del efecto de la intervención. En la etapa del diseño del estudio, estos factores pueden ser controlados por medio de la restricción y/o emparejamiento. Si existiera la firme sospecha que el efecto depende de la edad, uno puede acotar el estudio a un determinado grupo etáreo. Ambos procedimientos pueden no obstante resultar insuficientes, especialmente cuando el número de potenciales confusores es importante. En estos casos, lo recomendable es considerar a estos factores al momento de analizar los resultados por medio de las técnicas de análisis multivariado.

El ensayo de comienzo demorado

En enfermedades de curso crónico, no curables, es necesario discriminar las modificaciones que el tratamiento puede ejercer sobre la enfermedad respecto de los efectos a corto plazo relacionados con la mejoría sintomática. Un procedimiento en tal dirección es el diseño de comienzo demorado. Al comienzo se recaban los datos iniciales del paciente y se establece un puntaje de su enfermedad, en función de algún criterio ya consensuado. A medida que la enfermedad evoluciona el compromiso irá en aumento. Al momento de la inclusión los pacientes son asignados aleatoriamente para recibir tratamiento activo o placebo (control). Durante la fase 1, los pacientes reciben los tratamientos asignados. A partir de la fase 2, a todos los pacientes se les asigna tratamiento activo, y los datos obtenidos durante esta fase son utilizados para evaluar los efectos modificadores sobre el curso de la enfermedad de la intervención. La duración de la fase 1 debe comprender un período de tiempo suficiente para que los efectos del tratamiento sobre los síntomas sean perceptibles.

Concluido este período todos los pacientes que recibieron placebo cambian a tratamiento activo, y los que habían recibido dicha terapia continúan con el mismo. Completada la segunda fase se analiza el grado de separación en los puntajes de los pacientes que comenzaron a recibir tratamiento activo al inicio del estudio respecto de aquellos que a quienes se les administró en la fase 2. Esa separación indicaría los efectos modificadores del compuesto sobre la evolución de la enfermedad. Este tipo de

estudios requiere un conocimiento muy pormenorizado del curso de la enfermedad y métodos estadístico acordes a lo que se pretende detectar. La diferencia en el escalado de la enfermedad en función del tiempo se evalúa por medio de pruebas para mediciones repetidas. Resulta bastante trabajoso decidir cuántos datos introducir para el análisis como así también cuántas mediciones deben ser hechas durante la fase 1. Tampoco hay que perder de vista los ajustes del error alfa ante las múltiples comparaciones.
La fase 2 comprende el análisis de 2 hipótesis y dos pruebas estadísticas como mínimo. Una capaz de determinar si existe un menor desmejoramiento en los que siempre recibieron tratamiento respecto de los que fueron tratados con placebo y principio activo; esto permite ponderar si el efecto beneficioso persiste aún. También deberá establecerse que los efectos en el grupo bajo tratamiento activo todo el tiempo y el mixto no convergen; a fin de determinar si la modificación sobre la enfermedad persiste y no cae con el tiempo. De existir un efecto fijo del tratamiento sobre el mecanismo patogénico subyacente la diferencia en las pendientes de la fase 2 debería permanecer constante tras la remoción de los datos del período transitorio al comienzo de la fase 2. Si las pruebas para la fase 1 muestran diferencias significativas en favor del tratamiento activo y los resultados de la fase 2 no son significativos, se puede inferir que el tratamiento es beneficioso desde el punto de vista sintomático. Un tema no menor son los datos faltantes.

Ensayos clínicos vs. Cohortes para la evaluación del efecto terapéutico

Si bien los ensayos clínicos randomizados, comparativos, controlados y a doble ciego son una estrategia aceptada para la incorporación de nuevas modalidades terapéuticas, este tipo de diseño no esta exento de limitaciones. Debe tenerse en cuenta que las conclusiones a las que arriba se basan en un período de tiempo relativamente corto, con individuos muy seleccionados y en condiciones estrictamente controladas.
Con el propósito de obtener una información que vaya más allá de este círculo constreñido se suele recurrir a un estudio de cohortes, donde es posible evaluar una muestra mucho mayor de personas y seguirlas por un período de tiempo más prolongado, en el cual las condiciones experimentales están más próximas a la situación real. Las bondades y limitaciones de cada diseño se indican en la Tabla 3.5. Para el caso de la cohortes también se conformarán dos grupos de sujetos quienes recibirán o no la intervención. La diferencia estriba en que en este caso la asignación de los tipos de tratamientos no es aleatoria (como en los ensayos clínicos), sino que depende de la decisión clínica adoptada por el profesional a cargo del paciente, la elección del mismo enfermo o pautas consensuadas por los Servicios de Salud. Como tal en la cohorte puede que los grupos no sean comparativos en sus condiciones de base y por lo tanto exista un sesgo de selección. En la gran mayoría de los casos de la práctica médica la elección de un tratamiento en particular depende de las características del paciente las que a su vez pueden influir sobre la evolución o desenlace de la enfermedad. En razón de ello la identificación de los potenciales FC que pueden estar presentes en los grupos a comparar cobra una importancia fundamental. Una manera de reducir el sesgo de selección es incorporar sujetos que se ciñan a un diagnóstico bien definido y con determinadas características. Ello tiene un costo, la reducción en el tamaño muestral de la cohorte con la consecuente disminución en el poder del estudio y generalización de los resultados.

Tabla 3.5 Características distintivas entre ambos diseños

	Cohortes	Ensayos randomizados
Población estudiada	Diversas poblaciones de pacientes, observado en un rango de situaciones	Personas bien seleccionadas de acuerdo a los criterios establecidos
Asignación de tratamientos	Basada en la decisión adoptada por los médicos que controlan a las personas	Basada en la chance y controlada por los investigadores
Desenlace	Pueden definirse luego de aplicada la intervención	Se define de antemano los beneficios y los riesgos
Seguimiento	Proveen la posibilidad de un seguimiento mayor	Generalmente es corto en razón de los costos
Análisis	Técnicas multivariadas complejas	Métodos más directos

Tras la identificación de los factores que pueden llevarnos a error, el paso siguiente será evaluarlos, para lo cual necesitaremos contar con una base de datos apropiada (cruzad los dedos). Una vez que hayamos conseguido registrar la información habrá que analizar si estos factores están distribuidos desigualmente entre los grupos. Los métodos habituales para verificar situaciones de este tipo consisten en recurrir a las pruebas estadísticas clásicas para variables cualitativas o cuantitativas. Debemos tener presente, sin embargo, que estamos trabajando con tamaños muestrales grandes y podemos encontrar diferencias que desde el punto de vista estadístico son significativas pero poco relevantes en el campo clínico. Como procedimientos alternativos, es más adecuado calcular el tamaño del efecto (cuán desigual es la distribución en términos cuantitativos) o las diferencias estandarizadas. La estimación de ésta última se obtiene dividiendo la diferencia entre grupos por el desvío estándar general (de los 2 grupos). Los valores superiores a 0.1 se asumen como significativos. Asimismo, se puede recurrir al análisis multivariado.

En la definición de los FC mencionábamos que los mismos ejercen un efecto de por sí sobre el fenómeno de interés. A raíz de esto es necesario reducir el sesgo que ellos pueden introducir. Las estrategias que se utilizan a tal efecto son la regresión y estratificación (Tabla 3.6). La primera estima la asociación de cada variable independiente con la dependiente, tras ajustar el efecto de las variables (independientes) restantes. Los métodos de regresión están ligados a las características de la variable dependiente, si la misma es binaria se utiliza la regresión logística, si es cuantitativa la multivariada o el modelo de los riesgos proporcionales (ver capítulo siguiente).

El modelo de regresión asienta sobre el supuesto que el efecto de la intervención (o el predictor) es invariable en los distintos subgrupos que se conforman según las características de base de los participantes. Si existiera interacción (ver página 118), por ejemplo, los efectos de la intervención son distintos entre mayores y menores de 50 años se debe tener en cuenta esta circunstancia y efectuar cálculos por separado.

Por su parte, la estratificación es un procedimiento donde la muestra se descompone en subgrupos (estratos) según los factores que se creen capaces de introducir confusión. Es importante determinar si la asociación entre la intervención aplicada y el fenómeno de interés es distinta en los estratos conformados. Si los efectos son iguales es válido efectuar una estimación sumaria. La desventaja de esta estrategia es que al crear subgrupos se reduce el poder del estudio y por otro lado es que los estratos dispuestos podrían diferir en algunas características de base no tenidas en cuenta para la

conformación de los mismos. Por esos es conveniente trabajar con ambos métodos (regresión y estratificación).
Cuando hayamos obtenido los cálculos crudos y ajustados habrá que detenerse a observar en qué medida difieren. Si el ajustado se aparta demasiado de la estimación cruda, las características de base están ejerciendo una acción importante sobre el fenómeno de interés.

Tabla 3.6 Diferencias entre ambos procedimientos

Estrategias	Ventajas	Desventajas
Regresión	El investigador está familiarizado con la práctica Su cálculo utiliza todos lo datos	Es difícil determinar la comparatividad de los grupos Requiere extrapolar
Estratificación	Permite que nos concentremos en factores confundentes claves Evalúa fácilmente si el factor modifica el efecto de la intervención	Los estratos confeccionados pueden estar desbalanceados Se reduce el poder del estudio

A pesar de lo expuesto y para ilustrar una vez más el carácter provisional de la ciencia, no podemos dejar de lado, que en el escudriñar sobre los eventuales FC trabajamos sobre los hechos conocidos que distan mucho de ser todos los posibles.

A riesgo de ser denominados reiterativos una vez más volvemos a la carga con dos factores de error que como venimos viendo, no se apartan nunca de nuestro lado.

Sesgo de selección: es un error sistemático que se produce al conformarse los 2 grupos sobre los cuales se intervendrá, en modo tal que se presentan diferencias en cuanto a las características pronósticas de base de los mismos.
Factor de confusión: es el caso en que la estimación del efecto de la intervención aplicada se halla sesgada en razón de alguna diferencia entre los grupos a comparar (características de base, factores pronósticos, tratamientos u otras medidas concomitantes que se hayan adoptado). Para que un factor alcance la jerarquía de confusor el mismo debe estar distribuido desigualmente entre los grupos (cuali o cuantitativamente) y tener efecto sobre el fenómeno de interés.

3.4 Lineamientos éticos

La adhesión de un protocolo a los principios éticos que rigen la investigación biomédica es una cuestión tan crucial como los aspectos metodológicos a los cuales nos hemos referido. En nuestros días es casi imposible que un investigador logre obtener financiación para su proyecto o pueda publicar los resultados del mismo si la documentación no ha sido aprobada por un Comité de Bioética. Aún cuando se contemplaran todas las reglas éticas que deben tenerse en cuenta en la investigación médica (ver Declaración de Helsinki), existen situaciones que pueden plantear "incomodidades" al momento de decidir qué tipo de diseño aplicar. Tomemos como ejemplo un ensayo de vacunación para prevenir una enfermedad que en algunos casos es letal y 1/3 de los participantes debe recibir placebo, ¡cuando la información previa sugiere que la vacuna tiene una chance bastante buena de ser efectiva!

Para no ser invadidos por el desaliento o asegurarnos que nuestra propuesta es consistente en sus fundamentos éticos, es necesario poner en primer plano algunas recomendaciones.

- Un proyecto que carece de mérito científico (vaguedad de la pregunta, diseño incorrecto, datos que no contestarán a los interrogantes), va de suyo que se aparta de la ética.
- Otra cuestión que debemos sopesar siempre es la relación riesgo/beneficio. Ello implica que en algunos casos aparezca justificado llevar a cabo un ensayo terapéutico con un fármaco de conocida toxicidad, si existieren razones para suponer que el mismo servirá para tratar una enfermedad grave para la cual las terapéuticas disponibles son de una eficacia escasa o nula. Para el caso en que se produzca un desbalance desfavorable de los riesgos y beneficios de una investigación estaremos frente a la presencia de abuso. Puede darse aún cuando la persona haya consentido y se beneficie de participar. El punto es que el beneficio es mucho mayor para el investigador que para el paciente. Esta falta ética ya está presente en el diseño del estudio.
- Un aspecto que también hace a la ética de la investigación médica es la confidencialidad. Se trata del compromiso del investigador de no revelar la identidad de la persona a quien pertenecen los datos. Para las situaciones en que la enfermedad por sí misma ya es problemática (por ej. SIDA o Lepra) conviene que sólo un integrante del grupo de investigadores sea el depositario de esa información.
- Referido a la coerción es necesario una profunda autocrítica del investigador en torno al ejercicio de alguna forma de presión sobre el paciente para hacerlo participar del estudio en aras de un bien mayor. Es la mayor intimidación a la voluntariedad, y constituye una amenaza directa y real del investigador o alguien vinculado al estudio, sobre el participante.
- Dos situaciones no menos preocupantes se presentan cuando los participantes del ensayo son personas carenciadas: la selección desigual de los candidatos y el abuso. Del primero ya nos hemos ocupado. La segunda se presenta cuando el investigador selecciona la población en estudio por rezones ajenas a lo que es científicamente correcto y apropiado o el punto de vista de los beneficios del paciente. Se da independientemente de que el paciente haya otorgado el consentimiento.

La Comisión de revisión presta especial atención a la forma que ha sido preparado el consentimiento escrito informado (CEI) que debe otorgar el paciente antes de ser incluido en el estudio. Dependiendo de las características de la investigación y sobre todo en los que se aplica una intervención, el Comité de Bioética puede exigir que el CEI lleve la firma del paciente o su tutor legal, el investigador principal y en algunos casos un testigo. La información que debe proveer el CEI se puede resumir de la siguiente manera:

- Explicar al participante que se trata de una investigación, poniéndolo en conocimiento de los objetivos como así también de los potenciales riesgos y beneficios del estudio.
- Describir los procedimientos que contempla el estudio y la duración del mismo.
- Referir las posibles incomodidades que acarreará su participación.
- Indicar qué tratamientos adicionales o alternativos existen al presente.
- Especificar qué medidas se adoptarán de observarse toxicidad y a quién contactar de producirse dicha eventualidad.
- Asegurar la confidencialidad de los datos.

- Indicar que su participación es voluntaria y puede retirarse en cualquier momento del estudio sin que ello afecte su derecho a seguir siendo tratado y controlado de su enfermedad.
- Dejar en claro que de surgir alguna información nueva sobre el tratamiento administrado capaz de afectar su decisión de seguir participando del estudio será notificado inmediatamente.
- Mencionar si existen circunstancias por las cuales podrá ser excluido del estudio e indicar qué actitud se adoptará en ese caso.

Muchas veces se ha debatido hasta qué punto las poblaciones marginales son capaces de otorgar el consentimiento. En la vida de las personas, quien más quien menos se deben adoptar decisiones forzadas por situaciones o necesidades del momento, pero ello no implica que las mismas se tomen involuntariamente. El problema se agudiza cuando se trata de personas que al no poder desarrollar sus capacidades, y en consecuencia autonomía, no tienen las libertades básicas que hacen a su dignidad. En una sociedad con desigualdades las personas privadas de educación carecen de autonomía. Sin ello el individuo no puede desenvolverse correctamente ni puede tener una capacidad plena para tomar las decisiones más adecuadas para él y los suyos. Su dignidad se vuelve frágil y peligra el modo en que deben asegurarse sus derechos en la sociedad. Un consentimiento es realmente válido cuando la persona ha podido sopesar los riesgos y beneficios de aceptar el tratamiento que se le propone y en consecuencia está en condiciones de efectuar una elección. Para ello es fundamental que la información sea "plenamente entendible". Un individuo debe ser capaz de comprender y retener la información concerniente al tratamiento que se le propone y qué ocurrirá si decide no ser tratado.
En 2003 la revista Lancet publicó un estudio (361: 918-22) donde sólo el 22% de aproximadamente 400 pacientes que participaron de un ensayo clínico estaba en condiciones de entender cabalmente la información que se proveía y sólo un 18% había leído el formulario previo a otorgar su consentimiento para participar o no del estudio.
En un Hospital Escuela del Reino Unido se envió un cuestionario a los pacientes dentro de un mes de habérseles practicado una cirugía (ginecológica/obstétrica), para indagar en torno al grado de comprensión de las implicaciones legales del consentimiento escrito. De 732 participantes casi la mitad de las pacientes (46%, IC 95% 43-50%) creía que la función primaria del consentimiento era proteger al hospital y 68% (65-71%) pensaba que este documento permitía a los médicos asumir el control. Buena parte de los enfermos no asimilan que el consentimiento escrito sirve primariamente a sus intereses.
Paralelamente a estas consideraciones aparece otra justificación no poco frecuentada "el mejor control de la enfermedad que recibirá el paciente por participar del estudio". Este argumento ha sido denominado "mal entendido terapéutico" y apunta a una confusión entre participación en los ensayos clínicos con atención médica personalizada. Se trata de una falsa concepción que torna más vulnerable al paciente a ser explotado pero tampoco deja afuera al investigador. La pregunta que uno se formula en estos casos es: no más buenos cuidados cuando el ensayo haya concluido→ *back to reality*?
Nunca será ético actuar en pos de un propósito que no es el de aquellos que no tienen poder.

Utilización de Placebo

La expresión proviene del latín y significa “daré placer o complaceré”. Fue una práctica utilizada durante siglos cuando no se sabía de qué manera tratar a los enfermos o bien cuando no existía un tratamiento útil. En las últimas 4 décadas hemos asistido a cambios fundamentales en los avances de la medicina incluido los nuevos tratamientos surgidos de la puesta en marcha del ECA. Fue en ese terreno donde surgió la noción de un control con placebo como para tener una correcta estimación de los efectos benéficos o perjudiciales que no podían ser atribuidos en forma directa al tratamiento bajo estudio. En la práctica clínica es sabido que muchos procedimientos pueden tener un valor terapéutico y así influir sobre la respuesta del paciente.

En términos concretos la palabra *placebo* significa una sustancia inerte o procedimiento simulado diseñado como para que aparezca idéntico a la sustancia activa o el procedimiento a analizar pero carente de acción terapéutica en sentido estricto. La información disponible apunta a que el efecto placebo es un evento psicobiológico atribuible al contexto terapéutico global. Este contexto comprende factores del paciente, del clínico y el ambiente terapéutico; donde a su vez confluyen otros cuantos como la naturaleza del compuesto, el modo de administración y la relación médico-paciente.

Los procesos por los cuales el placebo consigue ejercer un efecto sobre el paciente se agrupan en dos categorías: psicológicos y neurobiológicos. La primera comprende una serie de situaciones que operan sobre el sujeto tales como la expectativa, condicionamiento, aprendizaje, memoria, motivación, cuestiones somáticas, reparación, reducción de la ansiedad y significaciones. Es claro que la expectativa generada en el paciente, en función de las estrategias utilizadas por el médico, da a lugar a un efecto sobre la enfermedad. El otro aspecto bastante bien identificado es el condicionamiento el cual es conocido desde hace mucho tiempo. Al parecer este último sigue a la expectativa, donde la relación médico-paciente parece ser esencial para el desarrollo de un efecto placebo significativo. Cuanto más grande es la expectativa mayor el efecto y obviamente la posterior acción de condicionamiento.

Respecto de los mecanismos neurobiológicos, y en lo concerniente a la analgesia las acciones pueden ser de naturaleza opioide y no-opioide.

Es necesario tener en cuenta que la administración de placebo también puede tener consecuencias negativas, denominadas efecto *nocebo*. Hoy en día se efectúa una diferenciación entre efecto placebo real y efecto placebo percibido. Este último incluye el real y otras acciones no específicas, incluidos el curso natural de la enfermedad, efectos dependientes de la circunstancia o efectos no identificados. Los pacientes incorporados en un estudio pueden experimentar distintos desenlaces al implementarse intervenciones paralelas no identificadas (co-intervenciones), por ejemplo comenzar con un programa de adelgazamiento mientras se participa de un ECA con hipolipemiantes o anti-diabéticos.

Durante mucho tiempo se pensó que la respuesta al placebo era similar en los distintos tipos de enfermedades, digamos una mejoría aproximada del 35%. En los últimos años se ha visto que esto no es necesariamente así, sino y que tanto el efecto real como el percibido son variables. Una revisión sobre 75 estudios sobre antidepresivos señala que la respuesta al placebo varía entre un 12.5 a 51.8% y que se ha ido incrementando a lo largo del tiempo. Otra revisión sobre ECAs con placebo y grupos no tratados, concluye que existe un efecto placebo real leve sobre algunas variables subjetivas (dolor y ansiedad) pero no sobre variables objetivas (presión sanguínea, pérdida de peso) [Hrobjartsson & Gotzsche, 2001]. Si bien existen muchas críticas sobre este efecto en los ECA, puesto que los pacientes son informados de poder recibir o no el tratamiento activo, el efecto placebo percibido podría ser más disimulado respecto de lo observado

en la práctica clínica. La demostración que el grupo placebo no difiere de quienes no reciben tratamiento alguno no descarta la posibilidad de un efecto placebo real producido igualitariamente en ambos grupos. Es importante tener en cuenta que los placebos acarrean el riesgo de demorar el acceso a terapias efectivas para la enfermedad bajo investigación, sobre todo cuando sus efectos sobre síntomas no específicos enmascaran la situación de fondo, para la cual existe tratamiento específico.
La actitud y conducta del clínico respecto del tratamiento y el paciente, la percepción del enfermo hacia su propia patología, el tratamiento, como así también las pautas culturales, familiares y sociales son capaces de ejercer una influencia sobre el efecto placebo. Las variables relacionadas con la terapia, por ejemplo invasividad, impresión, plausibilidad, experiencias anteriores, y costo también podrían desempeñar un papel al respecto. En el mismo sentido, se pueden mencionar cuestiones propias al ejercicio de la profesión, aspecto del médico, interés, empatía, autoridad, compasión, manera en que se lleva a cabo la consulta, motivación y habilidades del galeno. En un estudio efectuado en el Reino Unido sobre 200 pacientes que ocurrían por síntomas sin un diagnóstico definido, los pacientes fueron asignados aleatoriamente para recibir una respuesta en positivo o en negativo como así también tratamiento con placebo y no tratamiento [Brody & Waters, 1980]. La respuesta en positivo consistía en dar un diagnóstico firme y el reaseguro de sentirse mejor en unos pocos días mientras que en la respuesta en negativo, el médico admitía que no tenía certeza sobre lo que estaba pasando. Dos semanas después, 64% de los que recibieron una respuesta en positivo experimentaron mejoría en comparación a un 39% de los que habían recibido respuesta por la negativa. El porcentaje de mejoría en los que recibieron placebo no fue significativamente diferente del grupo no tratado. Las características del producto, más allá de lo estrictamente farmacológico, también ejercen efecto placebo. Respecto de las vías de administración se ha observado que las inyecciones son más potentes que las píldoras.

Ética del empleo de Placebo e investigación
Una de las críticas más prominentes respecto del uso de placebo señala lo anti-ético de asignar a un paciente para ser medicado con un tratamiento mucho menos eficaz que la terapia actual. Los que disienten con ello señalan, que los ECA con placebo son un primer paso muy necesario (con el menor grupo de pacientes posible) como para establecer si se puede continuar con la investigación de la droga en cuestión. En enfermedades amenazantes de la vida del paciente y donde existe un tratamiento efectivo probado, incluir placebo es inadmisible. Un área donde existe un gran debate respecto al placebo es en la depresión. Quienes criticaban su uso señalaban el riesgo de un mayor número de suicidios dentro de los tratados con placebo. Las revisiones al respecto demostraron no obstante que las proporciones en cuanto a los intentos de suicidios o suicidios reales eran similares dentro de los grupos con placebo, antidepresivos estándares (imipramine, amitriptilina, trazodona), y los compuestos bajo estudio (fluoxetina, sertralina, paroxetina); y que la respuesta clínica tampoco era tan diferente entre los grupos [Kahn y col. 2000]. Si bien en el CEI el participante indica su libre complacencia a participar de un ECA, siempre persiste la duda de cuán informados están los pacientes de los riesgos y beneficios de las distintas opciones terapéuticas. El paciente puede confundir tratamiento con cuidado médico, sobreestimar los beneficios de participar en el ensayo, y subestimar los riesgos. En cierta medida una especie de efecto placebo está ya presente desde el inicio, antes de instituirse el tratamiento.
Hay coincidencia que no debe existir coerción y el consentimiento debe brindar plena información respecto al estudio. La Declaración de Helsinki en su artículo 32 establece

Los posibles beneficios, riesgos, costos y eficacia de toda intervención nueva deben ser evaluados mediante su comparación con la mejor intervención probada existente, excepto en las siguientes circunstancias:
- El uso de un placebo, o ningún tratamiento, es aceptable en estudios para los que no hay una intervención probada existente.
- Cuando por razones metodológicas, científicas y apremiantes, el uso de un placebo es necesario para determinar la eficacia y la seguridad de una intervención que no implique un riesgo, efectos adversos graves o daño irreversible para los pacientes que reciben el placebo o ningún tratamiento.
Se debe tener muchísimo cuidado para evitar abusar de esta opción.
En la conferencia internacional sobre armonización, un comité con representantes de la industria y entidades reguladoras ha publicado las "Guideline for Choice of Control Group and Related Issues in Clinical Trials [Tri-Council Policy statement. Ethical conduct for research involving humans]. En ellas se permite el uso de placebo aún cuando existiera tratamiento efectivo a menos que los pacientes estuvieran expuestos a un riesgo inaceptable de muerte o daño permanente o si la toxicidad del tratamiento estándar fuese tan severa que muchos pacientes rehusaran recibirlo. El comité apoya el uso de diseños especiales siempre que sea posible, por ejemplo diseños factoriales o "escape precoz" de tratamientos ineficaces. En algunos países como Canadá la persona que recibe el subsidio para el estudio debe adherir a la Declaración de Helsinki, que prohíbe el uso de placebo si hay disponibilidad de aplicar un tratamiento efectivo. Se admiten sin embargo circunstancias excepcionales donde el tratamiento efectivo no está disponible debido al costo, provisión por un tiempo limitado, uso en pacientes refractarios, pacientes que rehúsan del tratamiento estándar o cuando el impedimento de tal tratamiento no implique un sufrimiento innecesario o la posibilidad de daño irreversible. LA FDA también acepta los ensayos con placebo pero abogan muy fuertemente por la necesidad del consentimiento informado [U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry: E 10 Choice of Control group and Related Issues in Clinical Trials.
http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125912.pdf
No obstante los lineamientos actuales para el uso de placebo en investigación, existen varias cuestiones a tener en cuenta. El argumento de eficacia de muchas drogas en uso al comparar con placebo sigue siendo debatible. Las consideraciones acerca del placebo en la Declaración de Helsinki apuntan a que no se termine utilizando a poblaciones empobrecidas sin acceso a tratamientos probados. Apoyar el uso de placebo sólo cuando no existe tratamiento estándar deja la eficacia de la droga en ciertos grupos de pacientes en buena medida desconocida.
En coincidencia con lo señalado por la FDA, los participantes de un estudio donde se incluye placebo, deben tener plena conciencia de esta situación y que la decisión de participar sea un acto de total autonomía.
Más allá de lo que concierne a la investigación, la otra cuestión a considerar es si se debe promover el efecto placebo en la práctica clínica. De momento no se dispone de evidencia como para avalar esta supuesta iniciativa, pero aunque si la hubiera; el punto crítico es que no se puede promover esa acción incurriendo en engaños. El médico debe actuar con honestidad y señalar que el procedimiento a implementar no ha demostrado ser efectivo para la enfermedad del paciente pero bien podría producirse una mejoría de la patología, vía de algún proceso psicobiológico.

El detenimiento de un ensayo clínico

La detención de un ECA en razón de un beneficio aparente es cada vez más frecuente. Para proteger o promover el interés de los participantes de un estudio, los investigadores pueden sentirse obligados a detener el ensayo tempranamente en razón de un beneficio aparente del nuevo tratamiento. Existen, no obstante, serios problemas éticos al proceder de ese modo. Los estudios truncados sobreestiman sistemáticamente los efectos del tratamiento; y en los casos donde el número de los eventos acumulados es pequeño esto puede ser muy grande. Las connotaciones éticas de ello son evidentes (ver tabla). Los investigadores deben asegurarse que existe un número importante de eventos antes de detener el estudio. Es necesario proteger a los participantes de un estudio con los requisitos éticos de la validez científica, el valor social y una relación riesgo-beneficio favorable.

Violaciones éticas que resultan del detenimiento temprano de un ensayo por beneficio aparente

Requisito ético	**Violación potencial**
Validez científica	La teoría estadística y la evidencia empírica sugieren que los ensayos detenidos precozmente sobre-estiman los efectos del tratamiento
Valor social o científico y relación riesgo/beneficio favorable	Las sobreestimaciones de los beneficios llevan a recomendaciones erróneas y una práctica clínica sub-óptima
Consentimiento del participante y respeto por los participantes potenciales y reclutados	Al informar a los pacientes sobre los primeros resultados se compromete la validez científica del estudio y no se asegura que los pacientes potenciales reciban la terapia beneficiosa imputada, de cara a un beneficio dudoso e insuficientemente información en cuanto al riesgo
Revisión independiente	Los investigadores, agencias de financiación, revistas, y pacientes están incentivados al detenimiento precoz; la revisión ética y los descuidos del comité de monitoreo de datos son cruciales

<u>La relación entre el investigador clínico y el sujeto de investigación</u>

La realización de un ECA plantea una serie de circunstancias en la cual el médico se confronta con cuestiones éticas que no siempre son de fáciles de allanar; tales como la utilización de placebo en el grupo control, randomización, detención del estudio, acceso a la medicación tras la realización del estudio, la información a los sujetos respecto de hallazgos imprevistos, provisión de cuidados accesorios, entre los más salientes.

Las posturas respecto a cómo debe enmarcarse esta relación son variadas.

Entroncado con la declaración de Helsinki, el punto de vista **fiduciario** sostiene que si los deberes de un investigador entran en conflicto con las obligaciones de brindar al paciente un cuidado óptimo, debe prevalecer esto último. Casi en las antípodas, quienes abogan por la **tesis de la diferencia**, señalan que los investigadores no están obligados a brindar cuidados óptimos. Los adherentes a este pensamiento puntualizan que el investigador clínico no debe adherir a las mismas premisas que el clínico asistencial, por cuanto la investigación tiene una orientación científica mientras que la clínica es terapéutica y en función de ello la ética es distinta. Lo que no parece advertirse es que el paciente sigue siendo el mismo independientemente de donde se lo encasille.

Otra corriente de interpretación propone el **balance clínico**, por el cual un investigador puede incluir a un paciente en un ECA sólo si los diferentes grupos de tratamiento están

contrabalanceados, vale decir los méritos relativos de ambos tratamientos son aún inciertos. Una vez que se ha establecido cuál es la terapia eficaz no más discusiones; de lo contrario se estaría violando el deber de ofrecer el mejor cuidado médico. Los críticos de esta propuesta aducen que el investigador no necesita proveer de cuidado óptimo siempre que el paciente haya consentido en recibir el disponible según el protocolo de estudio y entienda que esta participando de una investigación. Ante formulaciones de este tipo uno no puede menos que preguntarse si la cesión de un derecho nos da vía libre para violarlo. Por otro lado, nadie desconoce que el consentimiento escrito informado dista de ser un documento perfecto y el proceso de consentir tiene sus claroscuros. Dentro de la problemática, el enfoque de **no explotación** establece que el objetivo primordial de la investigación es desarrollar conocimiento, por lo tanto los clínicos no deben proveer al sujeto del cuidado óptimo. Según esta doctrina, la principal obligación ética es no explotar a los seres humanos, pero establecer los estándares de no explotación también tiene sus bemoles. Por su parte la tesis **de la confianza depositada** indica la existencia de dos justificaciones respecto de las obligaciones terapéuticas del investigador. Por tratarse de un agente moral el médico debe rescatar/hacer un esfuerzo razonable para la persona en riesgo. A la par de ello coexisten las obligaciones morales surgidas de la misma relación con el paciente. Finalmente también se presenta el **enfoque contextual**, que pone el acento sobre las circunstancia clínico-investigativa como hecho esencial para determinar los alcances del investigador en cuanto a la obligación de brindar cuidados. La investigación clínica tiene muchas facetas, desde lo meramente descriptivo, lo analítico (cohorte, casos y controles) hasta las intervenciones o ensayos clínicos. En función de ello las obligaciones éticas para con el paciente pueden ser distintas, sin perder de vista que existe “un piso” innegociable. Si nos atenemos a los distintos momentos de una investigación es claro que existen diferencias en función del diseño y las líneas de acción a ejecutar. De acuerdo al protocolo uno puede ir balanceando los riesgos y beneficios de la intervención. Al respecto hay que ser muy cautos para no incurrir en una concepción terapéutica errónea, por la expectativa que se genera en el paciente. Otros aspectos sobre el tapete tienen que ver con la vulnerabilidad y motivaciones de los enfermos, como así también la relación previa que el investigador tiene con el paciente; en el caso que el médico asume los roles de tratante e investigador.

Una estrategia global para la bioética

Los dilemas que suscitan los ECA en el mundo subdesarrollado, en particular los de fase III donde se requieren muchos participantes, son variados e incluyen la problemática del consentimiento informado, la elección del grupo placebo, sesgo por la publicación de los resultados y las diferencias en los estándares de cuidados clínicos. Sumado a ello, el propósito económico o académico puede que no resulte en un beneficio directo para el país, y ni hablar si el tratamiento resulta ser exitoso pero no está al alcance de los ciudadanos. Por las limitaciones surgidas en el mundo desarrollado, la cantidad de procedimientos regulatorios, y el surgimiento de una amplia gama de terapias innovadoras, se ha producido un desplazamiento en la realización de ensayos clínicos hacia el mundo subdesarrollado. En el intento de encarrilar la situación, han surgido una serie de emprendimientos tales como La iniciativa estratégica para el desarrollo de capacidades en el análisis ético (Strategic Initiative for Developing Capacity in Ethical Review –SIDCER-), la iniciativa bioética panafricana (Pan-African Bioethics Initiative –PABIN-) y la asistencia a los Comités de Bioética (Assisting Bioethics Committees –ABC-) de la UNESCO. Singer y Benatar (BMJ 2001) proponen un modelo que parte del desarrollado por la red de Epidemiología Clínica Internacional

(International Clinical Epidemiology Network –INCLEN-). Esta red global deberá proveer a la OMS y los respectivos gobiernos de lineamientos más claros en cuanto a las líneas de acción a seguir a fin de promover una redistribución de los recursos para la investigación y provisión de cuidados médicos.
En su informe acerca de la ética de la investigación en los países en desarrollo el Consejo de Bioética Nuffield, señala lo siguiente: el deber de aliviar el sufrimiento, el deber de mostrar respeto a los seres humanos, el deber de ser sensible a las diferencias culturales, el deber de no explotar al vulnerable. Todo ello teniendo en cuenta el contexto social, cultural, político y económico de cada país. Para el caso de una investigación planificada en otro país cuyos objetivos no encajan con las prioridades de la nación, deberá existir una clara relevancia que justifique su realización, amen de la adecuada capacitación ética para la investigación asociada a cuidados médicos.
Es necesario que los médicos que llevan adelante investigaciones clínicas tengan en claro que el diseño del ensayo debe estar en consonancia con la realidad del sitio, ser relevante para la población de la cual se conformará la muestra, y poseer un núcleo irreductible de estándares éticos aplicables a cualquier territorio. En este proceso es de importancia central el papel de las instituciones académicas que participan de dichas investigaciones a fin de que los hallazgos de los estudios se puedan trasladar a la práctica médica del país. Para los casos donde las muestras son enviadas al país patrocinante para su análisis, ello debería contemplar el entrenamiento de profesionales de los centros que participaron a fin de lograr un significativo desarrollo de capacidades a nivel local, a modo de mutuo enriquecimiento.

La Comisión de Bioética

Los objetivos de la Comisión de Bioética pueden circunscribirse a dos aspectos fundamentales: por un lado evaluar la relación riesgo-beneficio que resulta de la utilización de un nuevo tratamiento o intervención, y por el otro proteger a los pacientes para que no participen en una investigación mal diseñada. De acuerdo con dichos propósitos, la Comisión podrá desaprobar un proyecto, requerir y/o aconsejar modificaciones, como así también exigir informes mientras se lleva a cabo el estudio. Es aconsejable que todas las decisiones adoptadas queden registradas en un libro de actas, el cual lleva la firma de los asistentes a las reuniones. La Comisión suele constituirse con un mínimo de 5 integrantes (con voto) siendo conveniente que se agreguen personas no pertenecientes al área de la salud, por ejemplo abogados, miembros de la comunidad o de organizaciones no gubernamentales, todos ellos de reconocida idoneidad. Se debe procurar que ambos géneros estén representados en cada reunión y asegurarse que ninguno de los integrantes tenga un conflicto de intereses con algún proyecto en particular. Las decisiones de la Comisión también tienen en cuenta las cuestiones locales, respecto de aquello que sea aceptable o no para la comunidad.
A inicios de los años 90 se crea en la Facultad de Ciencias Médicas de la UNR una Comisión de Bioética orientada a evaluar toda la investigación llevada a cabo en las distintas unidades académicas existentes. Dependiendo de las características del proyecto, a continuación se detallan los **principales documentos que tienen en cuenta estas Comisiones al evaluar los aspectos éticos de un ensayo clínico**

- Protocolo y modificaciones posteriores que se hayan introducido si las hubiere.
- Nota de invitación del laboratorio al responsable del centro.
- *Curriculum Vitae* del investigador principal.
- *Curriculum Vitae* de los co-investigadores.
- Declaración de los investigadores respecto de su aceptación a ser integrantes del grupo.

- Formulario de CEI que deberá firmar el paciente.
- Certificación de los valores normales del laboratorio con firma del director.
- Antecedentes del fármaco ("*Drug Investigator Brochure*"). Es un documento confidencial provisto por la firma farmacéutica donde constan todos los datos bioquímicos y farmacológicos de la droga como así también las fases pre-clínicas y clínicas. Deberá proveer las citas bibliográficas de las publicaciones pertinentes.
- Existencia de un seguro para cubrir los eventuales riesgos de los participantes.
- Cuadernos de observación
- Formulario para la comunicación de efectos adversos.

Aún cuando todos estos requerimientos hubieran sido perfectamente cumplimentados, podría ocurrir que la realización del estudio planteara un problema ético de difícil solución, si la muestra se hubiera conformado con personas no representativas de la población que eventualmente habrá de beneficiarse con el nuevo tratamiento. Supongamos que para probar la eficacia e inocuidad de un antibiótico de última generación se efectúa un ensayo fase III en Instituciones Públicas, resultando que una vez obtenida la licencia para la comercialización del producto, dicho antibiótico no lograría ser utilizado en esos hospitales por su alto costo.

Si può? Si può?
Signore! Signori! Scusatemi
Se da sol mi presento. Io sono il Prologo.
Poichè in iscena ancor
Le antiche maschere mette l´autore.
In parte ei vuoi riprendere
Le vecchie usanze, e a voi
Di nuovo inviarmi.
Ma non per dirvi come pria
"Le lacrime che noi versiam son false!
Degli spasimi e dei nostri martir
Non allarmatevi" No.No.
L´autore ha cercato invece pingervi
Uno squarcio di vita.
Egli ha per massima sol che l´artista
È un uom, e che per gli uomini
Scrivere ei deve. Ed al vero isparavasi.
Un nido di memorie in fondo all´anima
Cantava un giorno, ed ei con vere lacrime
Scrisse, e i singhiozzi il tempo gli battevano!
Dunque, vedrete amar si como s´amano
Gli esseri umani, vedrete dell´odio
I tristi frutti. Del dolor gli spasimi,
Urli di rabbia, udrete, e risa ciniche!
E voi, piuttosto che le nostre povere
Gabbane d´istrioni, le nostr´anime
Considerate, poichè siam uomini
Di carne e d´ossa, e che di quest´orfano
Mondo al pari di voi spiriamo l´aere!
Il concetto vi dissi. Or ascoltate
Com´egli è svolto.
Andiam. Incominciate!

Prólogo de *"I pagliacci"*
Ruggero Leoncavallo

Las reglas de la buena práctica clínica.

Como su nombre lo indica se trata de una serie de normas internacionales de calidad ética y científica a tener presente en el diseño, implementación, conducción, registro y divulgación de un ensayo clínico, a fin de otorgar credibilidad a los resultados de la investigación y al mismo tiempo asegurar que los derechos, integridad y confidencialidad de las personas se han respetado. Están dirigidas fundamentalmente a la industria farmacéutica y todos aquellos que se hallen involucrados en la generación de resultados clínicos, que luego se incorporan en la solicitud de licencia para la comercialización del nuevo compuesto terapéutico o preventivo.

Su designación es inapropiada ya que las normas no están relacionadas a la práctica clínica sino a la conducción de investigaciones clínicas. Como tal sería mucho más adecuado denominarla "Buena Práctica de la Investigación Clínica". Este cuerpo normativo pone un especial énfasis a la labor de auditoría para asegurar que los datos vertidos a los cuadernos de observación son fidedignos y detectar la existencia de algún fraude en cuanto a la información que va surgiendo del estudio.

Los lineamientos generales que establecen las normas se pueden desglosar en los siguientes aspectos:

- Protección de las personas que participan del estudio
- Evaluación y dictamen por parte del Comité de Bioética, o Comité de Revisión Institucional.
- Consentimiento informado que otorga el paciente.
- Responsabilidades que le competen a la entidad patrocinante del estudio, generalmente un laboratorio Farmacéutico.
- Responsabilidades del monitor del estudio, quién es el nexo entre el Laboratorio y los investigadores.
- Responsabilidades del investigador principal.
- Estrategias de diseño.
- Procedimientos para la recolección y almacenamiento de los datos.
- Procesamiento estadístico de la información.
- Controles de calidad para los procedimientos a utilizar y los resultados que se obtendrán.
- Manejo y registro del producto farmacéutico.
- Reglamentaciones que establecen las entidades gubernamentales para los ensayos clínicos que se llevan a cabo en el país.

Bibliografía Recomendada

Pruebas diagnósticas

Bland JM, Altman DG. The odds ratio.BMJ 2000; 320: 1468.

Browner WS, Newman TB. The analogy between diagnostic tests and clinical research. JAMA 1987; 257: 2459-2463.

Cochrane Methods Group on Screening and Diagnostic tests. Recommended methods [http:/www.cochrane.org/Cochrane/sadt.htm]

Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004; 329: 168-169.

Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet HCW, van der Windt DAWM Dick Bezemer P. Conducting systematic reviews of diagnostic studies: didactic Guidelines. BMC Medical Research Methodology 2002, 2: 9.

Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003; 56: 1129-1135.

Gluud C, Gluud LL. Evidence based diagnostics. BMJ 2005; 330: 724-726.

Grimes DA, Schulz KF. Uses and abuses of screening tests. Lancet 2002; 359: 881-884.

Griner PF. Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. Ann Intern Med 1981; 94: 553-600.

Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. QJM 1998; 91: 247-258.

Hanley JA. McNeil BJ. The meaning and use of the area under a receiver operating chacracteristic (ROC) curve. Radiology 1982; 143: 29-36.

Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839-843.

Hulley SB, Cummings SR. Designing clinical research. Williams & Wilkins, Baltimore, 1988.

Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994; 271: 389-391.

Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? JAMA 1994; 271:703-707.

Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 2003; 56: 1118-1129.

Luiz RR, Leal Costa AJ, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: a new graphical approach. J Clin Epidemiol 2003; 56: 963-967.

Littenberg B, Moses LE: Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Dec Making 1993, 13: 313-321.

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). J Natl Cancer Inst 2005; 97: 1180-1184.

Magder LS, Fix AD. Optimal choice of a cut point for a quantitative diagnostic test performed for research purposes. J Clin Epidemiol 2003; 56: 956-962.

Midgette AS, Stukel TA, Littenberg B: A meta-analytic method for summarising diagnostic test performances: Receiver-operating-characteristic-summary point estimates**.** Med Dec Making 1993, 13: 253-257.

Moses LE, Shapiro D, Littenberg B: Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations**.** Stat Med 1993, 12: 1293-1316.

Okamoto M, Kawabe T, Iwasaki Y, Hara T, Hashimoto N, Imaizumi K, Hasegawa Y, Shimokata K. Evaluation of interferon-γ, interferon-γ-inducing cytokines, and interferon-γ-inducing chemokines in tuberculosis pleural effusions. J Lab Clin Med 2005; 145: 88-93.

Read MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. JAMA 1995; 274:645-651.

Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005; 58: 982-900.

Robertson EA, Zweig MH. Use of receiver operating characteristic curves to evaluate the clinical performance of analytical systems. Clin Chem 1981; 27: 1569-1574

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science fro clinical medicine. 2da edición. Boston: Little Brown and Co. 1991.

Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 2002; 21: 1237-1256.

Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993; 39: 561-57

Ensayos Terapéuticos

Agrawal M, Emanuel EJ. Ethics of phase 1 oncology studies: reexamining the arguments and data. JAMA 2003; 290:1075-1082.

Altman DG, Bland JM. Treatment allocation in controlled trials: why randomize. BMJ 1999; 318: 1209.

Altman DG, Bland JM. How to randomise BMJ 1999; 319: 703-704.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134: 663-694.

Amdur RJ, Biddle CJ. An algorithm for evaluating the ethics of a placebo-controlled trial. Int J Cancer 2001; 96: 261–269. JAMA 2005; 294: 1411-1414.

Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: consent to research and the therapeutic misconception. Hasting Cent Rep 1987; 17: 20-24

Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. N Engl J Med 2000; 342: 1878-1886.

Buyse M, George SL, Evans S, Geller NL, Ranstam J, Scherrer B, Lesaffre E, Murray G, Edler L, Hutton J, Colton T, Lachenbruch P, Verma BL. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. Stat Med 1999; 18: 3435-3451.

Charney DS, Nemeroff CB, Lewis L, Laden SK, Gorman JM, Laska EM, Borenstein M, Bowden CL, Caplan A, Emslie GJ, Evans DL, Geller B, Grabowski LE, Herson J, Kalin NH, Keck PE Jr, Kirsch I, Krishnan KR, Kupfer DJ, Makuch RW, Miller FG, Pardes H, Post R, Reynolds MM, Roberts L, Rosenbaum JF, Rosenstein DL, Rubinow DR, Rush AJ, Ryan ND, Sachs GS, Schatzberg AF, Solomon S; Consensus Development Panel. National Depressive and Manic-Depressive Association consensus statement on the use of placebo in clinical trials of mood disorders. Arch Gen Psychiatry 2002; 59: 262–270.

Conboy LA, Wasserman RH, Jacobson EE, Davis RB, Legedza ATR, Park M, Rivers AL, Morey EB, Nam BH, Lasagna L, Kirsch I, Lembo AJ, Kaptchuk TJ, Kerr CE. Investigating placebo effects in irritable bowel syndrome: A novel research design. Contemp Clin Trials 2006; 27: 123-134.

D'Agostino RB. The Delayed-Start Study Design. N Engl J Med 2009; 361: 1304-1306.

Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981; 114: 906-914.

Dumville JC, Hahn S, Miles JNV, Torgerson DJ. The use of unequal randomisation ratios in clinical trials: A review. 2006; Contemporary Clinical Trials 27: 1- 12.

Ellenberg SS. Randomization designs in comparative trials. N Engl J Med 1984; 310: 1404-1408.

European Medicines Agency. Guideline on the choice of the non inferiority margin. European Medicines Agency Web site. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf

Gayet-Ageron A, Agoritsas T, Combescure C, Bagamery K, Courvoisier DS, Perneger TV. What differences are detected by superiority trials or ruled out by noninferiority trials? A cross-sectional study on a random sample of two-hundred two-arms parallel group randomized clinical trials. BMC Medical Research Methodology 2010, 10:93.

Grimes DA, Schulz KF. Bias and causal associations in observational studies. Lancet 2002; 359: 248-252.

Grunberg SM, Cefalu WT. The integral role of clinical research in clinical care. N Engl J Med 2003; 348: 1386-1388.

Horstmann E, McCabe MS, Grochow L, Yamamoto S, Rubinstein L, Budd T, Shoemaker D, Emanuel EJ, Grady C. Risks and Benefits of Phase 1 Oncology Trials, 1991 through 2002. N Engl J Med 2005; 352: 895-904.

Joffe MM, Rosenbaum PR. Propensity scores. Am J Epidemiol 1999; 150: 327-333.

Jones B, Kenward MG. Design and analysis of cross-over trials. Chapman & Hall, Londres, 1989.

Markman M. Therapeutic Intent in Phase 1 Oncology Trials. A Justifiable Objective. Arch Intern Med 2006; 166: 1446-1448.

Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. BMJ 1948; 2: 769-782.

Millar FG, Rosenstein DL. The therapeutic orientation to clinical trials. N Engl J Med 2003; 348: 1383-1386.

Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of Safety Results in Published Reports of Randomized Controlled Trials. Arch Intern Med 2009; 169: 1756-1761

Pocock S. Clinical Trials. A practical approach. John Wiley & Sons, Londres, 1992.

Rief W, Avorn J, Barsky AJ. Medication-Attributed Adverse Effects in Placebo Groups Implications for Assessment of Adverse Effects. Arch Intern Med 2006; 166:155-160.

Rojavin MA. RECAuitment index as a measure of patient rECAuitment activity in clinical trials. Contemporary Clinical Trials 2005; 26: 552-556.

Rothman KJ, Michels KB. The continuing unethical use of placebo controls. N Engl J Med 1994; 331: 394-398.

Sanjay Kaul, MD, and George A. Diamond, MD. Good Enough: A Primer on the Analysis and Interpretation of Non inferiority Trials. Ann Intern Med. 2006;145:62-69.

Schwarz RP Jr. Maintaining integrity and credibility in industry-sponsored clinical research. Control Clin Trials 1991; 12: 753-760.

Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. Control Clin Trials 2002; 23: 662-674.

Smith HO. Guidelines for Clinical Trials. TDR-WHO, Ginebra 1992.

Treasure T, MacRae KD. Minimisation: the platinum standard for trials. BMJ 1998; 317: 362-363.

Ensayos clínicos vs. Cohortes

Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, Normand SLT, Anderson GM, Rochon PA. Reader´s guide to critical appraisal of cohort studies: 1. Role and design. BMJ 2005; 330: 895-897.

Mamdani M, Sykora K, Li P, Normand SLT, Streiner DL, Austin PC, Rochon PA, Anderson GM. Reader´s guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. BMJ 2005; 330: 960-962.

Normand SLT, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Reader´s guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. BMJ 2005; 330: 1021-1023.

Lineamientos éticos

Akkad A, Jackson C, Kenyon S, Dixon-Woods M, Taub N, Habiba M. Patients' perceptions of written consent: questionnaire study. BMJ 2006; 333: 528 (doi:10.1136/bmj.38922.516204.55)

Benson H, Friedman R. Harnessing the power of the placebo effect and renaming it "remembered wellness". Annu Rev Med 1996; 47: 193–199.

Brody H, Waters DB. Diagnosis is treatment. J Family Practice 1980; 10: 445-449.

Brody BA, McCullough LB, Sharp RR. Consensus and Controversy in Clinical Research Ethics. JAMA 2005; 294: 1141-1144.

Costello A, Zumla A. Moving to research partnerships in developing countries. BMJ 2000; 321:827-829.

Emanuel EJ. A phase I trial on the ethics of phase I trials. J Clin Oncol 1995;13:1049-1051.

Emanuel EJ, Miller FG. The ethics of placebo-controlled trial—a middle ground. N Engl J Med 2001; 345: 915–919.

Ernst E, Resch KL. Concept of true and perceived placebo effects. BMJ 1995; 311: 551–553.

European Economic Community note for guidance: Good Clinical Practice for trials on medicinal products in the European community, 1991.

Fava M, Evins AE, Dorer DJ, Schoenfeld DA. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychotherapy Psychosomatics 2003; 72: 115–127.

Finnis DG, Kaptchuk TJ, Millar F, Benedetti F. Biological, clinical and ethical advances of placebo effects. Lancet 2010; 375: 686-695.

Food and Drug Administration, Institutional Review Board information sheets. Department of health and human services. Rockville, EEUU, 1989.

Gracely RH, Dubner R, Deeter WD, Wolskee PJ. Clinicians expectations influence placebo analgesia. Lancet 1985; 325: 43.

Grimes DA, Hubacher D, Nanda K, Schulz KF, Moher D, Altman DG. The Good Clinical Practice guideline: a bronze standard for clinical research. Lancet 2005; 366: 172-174.

Hahn RA. The nocebo phenomenon: concept, evidence, and implications for public health. Prev Medicine 1997; 26: 607–611.

Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001; 344: 1594–1602.

Hrobjartsson A. What are the main methodological problems in the estimation of placebo effects? J Clin Epidemiol 2002; 55: 430–435

Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. Arch Gen Psychiatry 2000; 57: 311-317.

Kaptchuk TJ. Powerful placebo: the dark side of the randomised controlled trial. Lancet 1998; 351: 1722-1725.

Kaptchuk T. Intentional ignorance: a history of blind assessment and placebo controls in medicine. Bull Hist Med 1998; 72: 389-433.

Kaptchuk TJ, Kelley JM, Conboy LA et al. Componentes of placebo effect: randomized controlled trial in patients with irritable bowel syndrome. BMJ 2008; 336: 999-1003.

Kaptchuk TJ, Kerr CE, Zanger A. Placebo controls, exorcisms, and the devil. Lancet 2009; 374: 1234-1235.

Kent DM, Mwamburi M, Cash R, Rabin T, Bennish ML. Testing therapies less effective than the best current standard: ethical beliefs in an international sample of researchers. Am J Bioeth 2003; 3:W11.

Lurie P, Wolfe S. Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. N Engl J Med 1997; 337: 853-856.

Millar FG, Rosenstein DL. The therapeutic orientation to clinical trials. N Engl J Med 2003; 348: 1383-1386.

Miller FG, Wendler D, Swartzman LC. Deception in research on the placebo effect. PLoS Medicine 2005; 2: 0853-0858.

Moerman DE, Jonas WB. Deconstructing the placebo effect and finding the meaning response. Ann Int Med 2002; 136: 471-476.

Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical Issues in Stopping Randomized Trials Early Because of Apparent Benefit. Ann Intern Med. 2007; 146: 878-881.

National Research Act. 1974 [http://history.nih.gov/research/downloads/PL93-348.pdf]. Pub. L. 93-348

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research: *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* 1979

Normas éticas internacionales para las investigaciones biomédicas con sujetos humanos. Publicación científica No. 563, OPS-OMS, 1996.

Nuffield Council on Bioethics. The ethics of research related to healthcare in developing countries. London: Nuffield Council on Bioethics; 2005. http://www.nuffieldbioethics.org/sites/default/files/HRRDC_Follow-up_Discussion_Paper.pdf

Pollo A, Amanzio M, Arslanian A, Casado C, Maggi G, Benedetti F. Response expectancies in placebo analgesia and their clinical relevance. Pain 2001; 93: 77-84.

Resnik D. Liability for institutional review boards: from regulation to litigation. J Legal Med 2004; 25: 131-184.

Resnik DB. The clinical investigador-subject relationship: a contextual approach. Philosophy, Ethics, and Humanities in Medicine 2009; 4: 16 doi: 10.1186/1747-5341-4-16

Singer PA, Benatar SR. Beyond Helsinki: a vision for global health ethics. BMJ 2001; 322:747-8.

Streiner D. Alternatives to placebo-controlled trials. Canad J Neurol Sci 2007; 34 (Suppl 1): S37-41.

Studdert DM, Mello MM, Brennan TA. Financial Conflicts of Interest in Physicians' Relationships with the Pharmaceutical Industry — Self-Regulation in the Shadow of Federal Prosecution. N Engl J Med 2004; 351:1891-1900.

Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002; 287: 1840–1847.

Williams BF, French JK, White HD, for the HERO-2 consent substudy investigators. Informed consent during the clinical emergency of acute myocardial infarction (HERO-2 consent substudy): a prospective observational study. Lancet 2003; 361: 918-922.

WHO guidelines for good clinical practice (GCP) for trials on pharmaceutical products, Ginebra 1992.

World Alliance for Patient Safety Summary of the Evidence on Patient Safety: Implications for Research. The Research Priority Setting Working Group of the World Alliance for Patient Safety, WHO 2008.

í

APENDICE. LOS CONTENIDOS DE UN PROTOCOLO DE ENSAYO CLINICO

1. TITULO
El mismo deberá ser claro, sin ambigüedades, y brindará una breve descripción del estudio que se planea llevar a cabo incluyendo el tipo de tratamiento, la patología en estudio y el diseño que se propone.

2. SITIO DE REALIZACION
Consignar el nombre y la dirección de la Institución donde se efectuará el ensayo. Si correspondiera, mencionar las unidades auxiliares que tomarán parte en el mismo.

3. INVESTIGADORES
Indicar el nombre del investigador principal como así también de los co-investigadores.

4. COMITE DE REVISION INSTITUCIONAL
Identificar el Comité de Expertos responsable de la aprobación del proyecto de investigación. Ver ítem 15

5. RESUMEN
Proveer un resumen técnico que incluya los objetivos, definición de pacientes, tratamiento, diseño, y variables de impacto (clínicas y de laboratorio).

6. JUSTIFICACION
Efectuar una detallada presentación de las razones por las cuales el estudio debe realizarse, la cual estará basada en la información disponible al momento. En los ensayos de fase I y II, se deberá proveer una breve descripción de los estudios de seguridad e inocuidad llevados a cabo en animales. En los estudios de fase III consignar los resultados obtenidos de la fase II. Debe especificarse que el laboratorio dispone del "*Confidential Investigator Brochure*", para quien desee consultarlo.

7. OBJETIVOS
Definir con toda claridad qué es lo que intenta indagar el estudio en cuestión.

8. DEFINICION DE PACIENTE
En esta sección se deberán indicar los criterios de selección de los pacientes a ser reclutados en el estudio. Es aconsejable proveer una definición precisa de los pacientes o individuos a ser estudiados, incluyendo los criterios diagnósticos de la entidad a ser tratada, o bien de los criterios que se utilizarán como evidencia de susceptibilidad y exposición a la condición sobre la que se desea prevenir (ensayos preventivos). También se deberán consignar el grado de severidad de la enfermedad, edad y género de los pacientes, como así también exámenes bioquímicos confirmatorios. Incluir una lista, lo más exhaustiva posible de los criterios de inclusión y exclusión.

9. DISEÑO DEL ESTUDIO Y DEFINICION DEL TRATAMIENTO
Explicitar el diseño experimental a utilizarse, empleando la terminología habitual para estos casos, por ej. aleatorizado en bloques, cruzado, multicéntrico, etc. Indicar si el estudio es controlado, doble ciego, simple ciego o abierto. Establecer el número de pacientes o sujetos necesarios para lograr los objetivos del estudio. Describir la manera por la cual los enfermos van a ser asignados a los distintos grupos y cómo se ha previsto

reducir el sesgo. De ser necesario, indicar el modo de estratificación para asegurar la comparatividad de las variables pertinentes (edad, género, severidad, duración de la enfermedad, tratamiento adicionales, etc.).
Especificar el modo de tratamiento, aclarar si existe un período de "*wash-out*", estudios basales y posteriores, como así también seguimiento, si se lo incluye. Establecer la fase de la investigación, la dosis o bien el rango de dosis, la formulación, vía de administración, frecuencia y duración. Definir claramente todas las drogas que se proveerán a los pacientes, tanto el fármaco en evaluación, como el placebo.

10. TRATAMIENTOS CONCOMITANTES
Indicar la(s) medicación(es) que no podrá(n) utilizarse durante el estudio, en razón de las interacciones o incompatibilidades medicamentosas, o bien del algún efecto sobre las variables en estudio. Establecer cuál podrá ser la terapia concomitante a emplear, tratamientos suplementarios, o bien fármacos que pueden utilizarse para trastornos secundarios.

11. MEDICIONES CLINICAS Y BIOQUIMICAS DE EFICACIA Y SEGURIDAD
Describir breve pero claramente cómo se evaluará la efectividad e inocuidad del fármaco. Se deberá brindar información acerca de cada uno de los parámetros de evaluación (también designados como variables de impacto "*outcome variables*, o *end-points*"), cómo, cuándo y a quiénes, como así también la frecuencia de tales mediciones. Es muy conveniente la preparación de un esquema en Tabla que no sólo ayudará al investigador sino también a los demás participantes. Está Tabla deberá incluirse como Apéndice. De existir instrucciones especiales en cuanto a la forma de evaluación, definirlas con toda claridad.

Criterios de eficacia
Indicar si los mismos se basarán en mediciones objetivas (cuali y/o cuantitativas). Si la evaluación tomara en cuenta respuestas subjetivas, explicar cómo se graduarán las mismas.

Criterios de seguridad o inocuidad
En esta sección se deberán detallar todas las mediciones tendientes a detectar efectos tóxicos. El investigador deberá estar informado de los potenciales efectos adversos y cómo manejarse con ellos. Es conveniente advertir al paciente de los mismos, a efectos de una pronta intervención. Toda esta información tiene que estar contenida en el Manual del Investigador ("*Investigator Brochure*").

12. ANALISIS DE LOS DATOS
Efectuar una breve descripción de los métodos estadísticos a emplear. Consignar el modo de recolección y procesamiento de los datos. Indicar la diferencia que podrá detectarse con el tamaño muestral propuesto, y la probabilidad de hallar la diferencia que se planea demostrar (error beta).

13. CUESTIONES ADMINISTRATIVAS
Etiquetado de todo el material de estudio. Es muy importante en la provisión de los fármacos y el placebo, como así también para la recolección de muestras biológicas cuya medición se efectuará en un laboratorio de referencia.

Preparación del Cronograma de Actividades y Cuadernos de Observación (CO)
Brindar una descripción acerca de la utilización del Cronograma y el correcto llenado de los CO. Indicar la frecuencia conque el auditor clínico visitará el(los) centro(s). El investigador deberá firmar la primera página del CO.

Duración del estudio
Establecer el tiempo calculado para la realización del ensayo. Si el protocolo es multicéntrico establecer la duración para cada uno de los servicios involucrados.

Instrucciones para la Secretaría del Centro o Responsable
Las responsabilidades y labores concernientes a los mismos incluirán

- Auditoría del material provisto y devolución de lo no utilizado, llevando la contabilidad de la droga
- Conservación de las historias clínicas
- Chequeo de la información que se va recolectando para evitar inconvenientes a la hora de confeccionar el CO
- Comunicación con el paciente a efectos de recordarle su concurrencia para el control y/o seguimiento

Preservación de los documentos pertenecientes al estudio
Tener acceso al documento fuente (o bien fotocopia de la historia clínica), y conservar una copia del CO de cada uno de los pacientes con su cronograma individual.

14. CONSENTIMIENTO INFORMADO
El protocolo deberá proveer el formulario de consentimiento escrito y la hoja de información para el paciente.

15. COMITE DE REVISION INSTITUCIONAL O BIEN COMITE DE ETICA
Los protocolos requieren indefectiblemente la aprobación por parte de estos Comités.

16. CONFIDENCIALIDAD
Muchos laboratorios exigen que el investigador no divulgue la información precedente y posterior a la realización del estudio sin su autorización. En caso de que el Comité de Revisión Institucional exija un conocimiento más pormenorizado del protocolo y los antecedentes que lo fundamentan, el investigador podrá proveer tal información al Comité, quien deberá respetar la confidencialidad de la misma.

17. EFECTOS ADVERSOS
Todos los pacientes que experimenten efectos adversos severos deben ser seguidos.

18. FIRMA DE LOS INVESTIGADORES
El protocolo deberá llevar la firma del investigador principal y los co-investigadores como así también la fecha en que las mismas fueron rubricadas.

CAPITULO 4. Procesamiento y análisis de los resultados

"Ninguna prueba que se base en la teoría de la probabilidad puede por sí misma proveer una evidencia apreciable de la verdad o falsedad de una hipótesis. Pero podemos mirar a las pruebas desde otro punto de vista. Sin que esperemos saber si cada hipótesis por separado es verdadera o falsa, podemos investigar las reglas que dirijan nuestras acciones respecto de ellas, y haciéndolo así nos aseguramos que tras largas experiencias, a menudo no estaremos equivocados"
Neyman J, Pearson E. *Philos Trans Roy Soc A* 1933; 231: 289.

4.1 Generalidades

La estadística se desarrolla como una disciplina encargada de proveer información relevante para el Estado. Con la revolución industrial y los cambios democráticos del siglo XIX, los gobiernos europeos se vieron desbordados con cuestiones numéricas, y de allí el término de estadística. A lo largo del siglo XIX se fue consolidando la concepción numérica de la estadística y comenzaron a aparecer entidades que reunían a los distintos actores de ese campo como la Sociedad Estadística de Londres fundada en 1834. Una figura saliente de aquellos tiempos fue un astrónomo belga devenido en estadístico Adolphe Quetelet quien sostenía que la sociedad poseía una suerte de características, casi leyes, susceptibles de ser reveladas a través de datos estadísticos. Con el tiempo los médicos empezaron a acercarse a estos ambientes, probablemente a raíz del surgimiento de la Salud Pública en su intento por mejorar las condiciones de vida y los nuevos rumbos de la investigación médica la cual abundaba en datos cuantitativos. Un medico Francés, Louis René Villermé (1782-1863), quien compartía con Quetelet el interés en describir a la sociedad en términos numéricos, publica en 1828 un artículo donde señala la relación entre mortalidad y situación económica. Por su parte, el medico Inglés William Farr, quien en 1839 fuera asignado a la flamante Oficina del Registro General de Gran Bretaña, efectuó grandes contribuciones al campo de la estadística. Basado en datos disponibles en ese Registro, John Snow pudo demostrar que el cólera era una enfermedad transmitida a través del agua.
Durante el siglo XIX, los estadísticos efectuaban un trabajo mayoritariamente descriptivo. Esta orientación cambio fundamentalmente tras la creación de la Escuela de Biometría en el University College London bajo la dirección del matemático Karl Pearson (1857-1936). El mismo contaba con un fuerte apoyo, académico y financiero, de Francis Galton, quien abrazaba la idea del rol esencial de la herencia en el desarrollo de un individuo; y Pearson debía proveer un sustento científico a este planteo por medio del análisis estadístico. A la par de ello, el matemático también fue desarrollando un pensamiento por el cual si la inferencia se basa en la asociación de antecedentes y consecuentes, todo el razonamiento científico tiene en definitiva un núcleo estadístico.
Pearson se ocupó asimismo del entrenamiento en biometría del medico Major Greenwood (1880-1949). La valía de sus contribuciones hizo que Charles James Martin, director del Instituto Lister de Medicina Preventiva, le ofreciera el puesto de médico estadístico en el Instituto y con ello se pudo legitimar el uso de las técnicas biométricas en el análisis de datos. En 1927 Greenwood fue nombrado director de la unidad de estadística del Consejo de Investigaciones Médicas.
Durante el período comprendido entre las dos guerras se lograron posteriores desarrollos teóricos en los procedimientos de inferencia estadística, promovido inicialmente por Ronald Fisher, quien estaba muy abocado a la utilización de datos para la prueba de hipótesis. Fisher había sido nombrado en la Estación Experimental de Rothamsted, donde estudio la productividad de distintos tipos de granos en agricultura.

A partir de esto, publicó una serie de libros sobre metodología estadística. En uno de ellos (*The Design of Experiments*, 1935), estableció la importancia central de la randomización para asignar los distintos tipos de granos a las tierras de cultivo a fin de eliminar el sesgo del experimentador. Al final de la segunda guerra, Austin Bradford Hill (el sucesor de Greenwood en la jefatura de la división estadística del Consejo de Investigaciones Médicas) diseñó un ensayo clínico el cual analizaba la eficacia de la estreptomicina en TB pulmonar, y utilizó los números aleatorios para determinar qué pacientes recibirían el fármaco y quiénes serían los controles. Puede afirmarse sin lugar a dudas que este fue el nacimiento de los ECA.

El período de post-guerra proveerá de avances fundamentales no sólo para la estadística sino también para la Epidemiología. Así como una noxa infecciosa era capaz de ocasionar una enfermedad, la investigación clínica intentaba determinar los agentes específicos y el análisis de factores ambientales vinculados a otras enfermedades. Dado que podían existir múltiples actores que en definitiva nos llevarían a cometer errores en la apreciación de causalidad, la estadística debía brindar una ayuda valiosísima al respecto. Así aparecen investigaciones en las que se señalaba un mayor riesgo de desarrollar cáncer de pulmón entre fumadores, aunque la mayoría de estos estudios eran retrospectivos. En 1947, EEUU lanza el estudio de Framingham, el ejemplo paradigmático de la cohorte prospectiva, en el cual una muestra de residentes de Framingham-Massachusetts (30-59 años) iba a ser estudiada cada 2 años para analizar el papel de factores como el colesterol, actividad física, dieta, y estrés sobre el desarrollo de enfermedad cardíaca. Mucho más recientemente ha resurgido el interés por los procedimientos Bayesianos, tanto en la planificación de ECA como el análisis de los resultados (ver recuadro).

En nuestros días, la bioestadística es una herramienta indispensable a la hora de valorar cuestiones básicas de la fisiología de los seres vivos, las interacciones del hombre con el medioambiente, prevención de enfermedades y tratamientos, como así también sistemas de salud su organización y financiamiento. El entrenamiento en bioestadística en facultades de Medicina es una necesidad cada vez más creciente e insoslayable, también extensibles a otras carreras del campo salud, como odontología, enfermería, administración y políticas en salud.

La inferencia estadística es el procedimiento por el cual los supuestos acerca de una población se efectúan en función de los resultados obtenidos de una muestra extraída de esa población. Al inferir deseamos efectuar formulaciones no acerca de las personas que se analizaron en el estudio, sino de la población de sujetos de donde se extrajeron los participantes.

La prueba sobre la validez de una hipótesis requiere esencialmente la constatación de algún tipo de relación entre las variables predictiva y de impacto. Dicha asociación no significa que forzosamente exista una relación de causa-efecto, pero indudablemente constituye un buen punto de partida. La **significación estadística** en definitiva estima cuán probable es que las diferencias aparentes entre los grupos sean reales y no simplemente por chance. Las medidas más utilizadas para presentar los resultados de las comparaciones son el valor de *p* y el IC. El primero tiene que ver con lo enunciado en la frase anterior. El segundo tipo de estimación se ocupa de cuantificar la incertidumbre o variabilidad asociada con lo estimado. Vale decir que tendremos un 95% de confianza que el efecto está comprendido entre los dos valores extraídos. El intervalo nos da una idea de la versatilidad del tratamiento. Sin embargo, uno no puede determinar si un intervalo en particular contiene o no la media real del efecto, sino la probabilidad de que realmente la contenga. La prueba de hipótesis posee una perspectiva complementaria.

Cuando un estudio clínico arroja un resultado negativo frecuentemente uno se pregunta si una investigación en un mayor número de sujetos no llevaría a un resultado positivo. Ante este interrogante conviene mirar el extremo superior del IC al 95%. Existe una chance de 1 en 40 (2.5% en el extremo superior y el otro 2.5% en el inferior) de que el resultado real pudiera ser superior a lo observado. Admitiendo este supuesto surge la siguiente pregunta "es clínicamente relevante este nivel de diferencia". Si no lo es no existe necesidad de llevar adelante otro estudio y el **no** suena más definitivo.

Un valor de $p < 0.05$ significa que una comparación entre 20 podría ser significativa sólo por coincidencia. Cuando en un estudio se presentan muchas comparaciones estadísticas la chance de que aparezcan diferencias estadísticas se incrementa. Por lo tanto el valor de *p* debe ser corregido. El método más difundido es el de Bonferroni. La formula para el cálculo del error que se aplica para la corrección es $1-(1-\alpha)^n$, bastante aproximado α/n, donde n es el número de comparaciones que se efectúan. La corrección de Bonferroni se asienta sobre el supuesto de que todas las hipótesis nulas son ciertas. Si bien es lógico utilizar este tipo de corrección no conviene aplicarla en investigaciones donde se analiza una hipótesis muy puntual donde la integración entre la presunción previa con los datos recabados se logra mejor con los métodos Bayesianos (ver recuadro sobre las características del Bayesianismo y lo expresado en ensayos diagnósticos – razones de probabilidad).

No obstante que el ajuste de Bonferroni reduce la chance de error tipo I, no hay que perder de vista el incremento en la posibilidad de cometer el error tipo II, a menos que se trabaje con un mayor tamaño muestral lo cual significa más costo. Como alternativas menos onerosas algunos autores proponen la utilización de puntos finales compuestos o bien elegir un punto final primario y varios secundarios, o directamente pocas variables de impacto.

Para quien esté interesado en el tema Sterne y Smith (BMJ 2001; 322: 226) proveen cálculos sobre el valor predictivo de los distintos valores que asume *p* (=0.05, =0.01 e =0.001) bajo diferentes presunciones acerca del poder del estudio y la proporción de hipótesis con fundamento. Como es de esperar la proporción de resultados que caen en la zona de los falsos positivos se reduce ostensiblemente cuanto menor es el valor de *p* elegido.

A los problemas interpretativos que plantean las comparaciones estadísticas también debemos prestar atención si los cálculos se hallan correctamente expresados en su significado.

El Bayesianismo

La estadística utiliza dos enfoques al momento de inferir, uno es el que habitualmente conocemos "frecuentista" y el otro es el bayesiano. Este último interpreta en qué medida un resultado positivo o negativo se ajusta a la probabilidad de una determinada enfermedad. En otros términos el resultado actúa como un modificador de nuestra opinión.

Una definición de los métodos Bayesianos podría ser la utilización cuantitativa explícita de la evidencia externa en el diseño, supervisión, análisis, interpretación y comunicación de la evaluación de la tecnología aplicada a la salud.

El teorema de Bayes es una fórmula donde se demuestra en qué modo nuestras "presunciones" (la probabilidad previa) se modifican tras el aporte de un dato nuevo para generar una nueva probabilidad (posterior). El ejemplo más típico en la medicina es la opinión que tiene un médico acerca de si determinado paciente posee tal o cual enfermedad (teniendo en cuenta el cuadro clínico y la frecuencia con que la enfermedad se presenta en la población), la cual se verá modificada por los resultados de los estudios complementarios que se solicitan. El enfoque bayesiano no brinda una respuesta categórica Si/No, sino una probabilidad relacionada con el contexto en que se presenta la situación.

En una exploración reciente donde se revisaron todos los trabajos que Nature publicó en los volúmenes 409-412 (2001) y 12 trabajos de una muestra aleatoria extraída del BMJ (vols. 322-323, 2001), se observó que el 11.6% y 11.7% de los casos, presentaban incongruencias, al parecer originadas en errores de redondeo, transcripción o tipográficos. En el mismo sentido, una revisión aparecida en PLoS One (2009) da cuenta que muchas publicaciones sobre investigaciones en animales en revistas con referato no proveen información sustantiva sobre los métodos experimentales y procedimientos para el análisis estadístico de los datos.
Aún cuando hayamos detectado una asociación significativa entre variables, es nuestra responsabilidad como investigadores cerciorarnos de que no existe otra explicación posible para lo que hemos detectado. En párrafos anteriores ya habíamos señalado la necesidad de considerar que los resultados obtenidos pertenecen a una muestra y no al total de la población, y en definitiva lo observado podría haber sido simplemente por chance.
Esta incertidumbre adquiere mayor relevancia cuando el tamaño de la muestra es reducido. El valor de *p* depende de dos cuestiones fundamentales, la magnitud de la diferencia entre grupos y el tamaño muestral. De este modo, una diferencia poco importante en una muestra grande de pacientes puede ser estadísticamente significativa mientras que resultados mucho más dispares podrían no serlo si el tamaño muestral fuese insuficiente. Afortunadamente la estadística nos permite calcular el IC el cual provee un rango de valores plausibles para el resultado de nuestro estudio.

Dado que uno de los propósitos fundamentales de la investigación médica es encontrar causas para los efectos, el lector se preguntará si es posible arribar a conclusiones valederas con tantas falacias potenciales en derredor. Aún cuando la ciencia no pueda brindarnos una tabla de salvación, de existir la supuesta melodía la misma se tornará perceptible a medida que el ruido de fondo descienda a un nivel donde ya es posible distinguirla. Reducir esa "contaminación sonora" requiere la adhesión a una serie de recomendaciones metodológicas y un aporte personal que suele designarse como creatividad.

Como fuera ampliamente analizado en el Capítulo 2, la verificación de causalidad es un trámite dificultoso y requiere una demostración, generalmente corroborada por otros grupos, de que un cambio cuali o cuantitativo de la variable predictiva se corresponde con una modificación de la variable de impacto. No está demás insistir en los principales elementos que le otorgan peso a la variable predictiva como agente causal del efecto verificado se indican a continuación: (1) Poder de la asociación, (2) Conocimiento básico que apoya a la presunción, (3) Obtención de resultados similares por otros grupos de investigadores bajo condiciones distintas, (4) Secuencia temporal entre la presencia del factor y la aparición de la alteración y (5) Relación dosis-respuesta.

4.2 Planificación y análisis

Una vez que el estudio ha concluido nos enfrentamos a un cuerpo de resultados, los cuales debemos analizar estadísticamente. Para tal fin es deseable que la investigación se haya conducido según lo planeado y de ese modo efectuar las comparaciones de acuerdo a lo establecido previamente. Con esto queremos significar que es un error muy lamentable comenzar por la planificación del análisis de los datos cuando éstos ya se encuentran a nuestra disposición. Muy por el contrario, el tratamiento estadístico de los resultados es algo que se decide al preparar el protocolo. Trabajar desde el inicio en

estas cuestiones nos llevará a definir el tipo y la forma en que deberemos recolectar la información a fin de que el análisis conteste las preguntas planteadas.
La disponibilidad de computadoras y el cargado de la información en registros electrónicos, posibilitan graficar los resultados al momento de efectuar el análisis e identificar, de entrada a través de este procedimiento, los datos que se apartan del patrón general; los cuales pueden estar señalando errores de medición, sesgos introducidos, o la existencia de subgrupos (estratos) en la muestra.
La manera de representar y analizar los datos está ligada inexorablemente al tipo de variable con la que nos manejamos (ver Tabla 1.3). Cuando se trata de variables discontinuas (nominales u ordinales) la información se presenta indicando el número y la proporción de sujetos que se agrupan en cada categoría. Para el caso de las variables continuas se deben proveer por lo menos dos parámetros. Uno de ellos expresa la tendencia central de los datos y se puede escoger la media o bien la mediana, dependiendo en líneas generales que la variable adopte una distribución normal o no, respectivamente. En algunas situaciones, si la variable es continua pero las mediciones arrojan valores extremos, es muy probable que la media aparezca distorsionada por los valores fuera de rango, siendo preferible utilizar la mediana para la representación de la posición central de los datos. Supongamos que se trata de una serie de 13 pacientes con reciente diagnóstico de artritis reumatoidea (AR) en quienes se evalúan los niveles de un mediador proinflamatorio involucrado en el desarrollo de la lesión articular, el factor de necrosis tumoral alta (FNT). Las mediciones en suero arrojan los siguientes valores (picogramos/ml)

144 203 197 151 189 170 158 176 183 163 192 149 161

La media de estas evaluaciones es 172, mientras que el ordenamiento en orden creciente de los datos para el cálculo de la mediana (el valor que se sitúa exactamente en el medio) indica que el séptimo resultado de esa escala es 170, lo que evidencia una correspondencia muy aproximada entre ambas estimaciones.

144 149 151 158 161 163 **170** 176 183 189 192 197 203

Si en lugar de 203 el nivel más alto fuese 493 la mediana seguiría siendo la misma, mientras que la media alcanzaría un valor un tanto apartado de la posición central (191).
El otro parámetro que se requiere para tener una idea cierta sobre el comportamiento de la variable está referido a la dispersión que presenta la misma, vale decir la posición relativa de los demás datos de la muestra. El modo más simple de indicar la variabilidad es por medio del rango que surge de la resta entre el valor máximo y mínimo, en nuestro ejemplo 59. La desventaja de esta representación es que sólo se brindan los puntos extremos sin dar una idea de cómo es la distribución de los datos restantes. Para ello es necesario indicar la variancia de los resultados o un valor que deriva de la misma, el desvío estándar (DS)[5], el cual representa la dispersión de cada valor en torno de la media. Para distribuciones normales, vale decir una curva unimodal y simétrica como se representa en la Figura 4.1, los valores que están comprendidos dentro de 1 DS

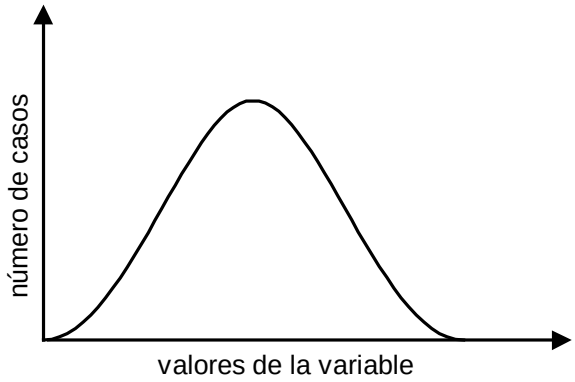
[5] Los procedimientos estadísticos que se utilizan tanto para estimar estos valores como para efectuar las comparaciones entre grupos escapan a los objetivos de este manual. El lector interesado en los mismos puede consultar los textos sobre Estadística que se mencionan en la bibliografía recomendada.

alrededor de la media (±) abarcan el 68% de las observaciones que hemos efectuado, alcanzando una proporción del 95% cuando se extiende a ± 2 DS. Al referirnos a la precisión con que se efectúan las mediciones (Capítulo 1), hicimos mención al coeficiente de variación (CV) como un indicador de la misma. Dicho coeficiente surge de dividir al DS por la media y el resultado que se obtiene indica el grado de variabilidad del dato que se recoge.
Tomando como ejemplo los datos de la página anterior, las mediciones séricas de FNT arrojaban una media de 172 con un DS de 19.5 pg/ml. Si procedemos a calcular el CV el resultado es el siguiente [19.5/172] x 100 = 11.3, un valor muy tranquilizante por cierto.
Frecuentemente los trabajos científicos consignan el error estándar de la media (ES), en lugar del DS, como medida de dispersión. Si bien el ES deriva del DS, puesto que surge de dividir este último por la raíz cuadrada del tamaño muestral (n), no estima el mismo parámetro. Siguiendo con el ejemplo de los pacientes con AR, bien podría ser que nos interesara medir los niveles de FNT en dos grupos adicionales de enfermos, extraídos de la misma población. De ese modo obtendríamos 3 medias que si bien son estimaciones de una sola media poblacional difícilmente sean iguales. Del mismo modo en que calculábamos el DS para saber cuál era la variabilidad de las observaciones individuales en torno a una media muestral, ahora nos interesa conocer la variabilidad de las 3 medias respecto de la media poblacional real. Tomando el ejemplo enunciado el ES surge de la siguiente manera:

$$19.5/\sqrt{13} = 5.41$$

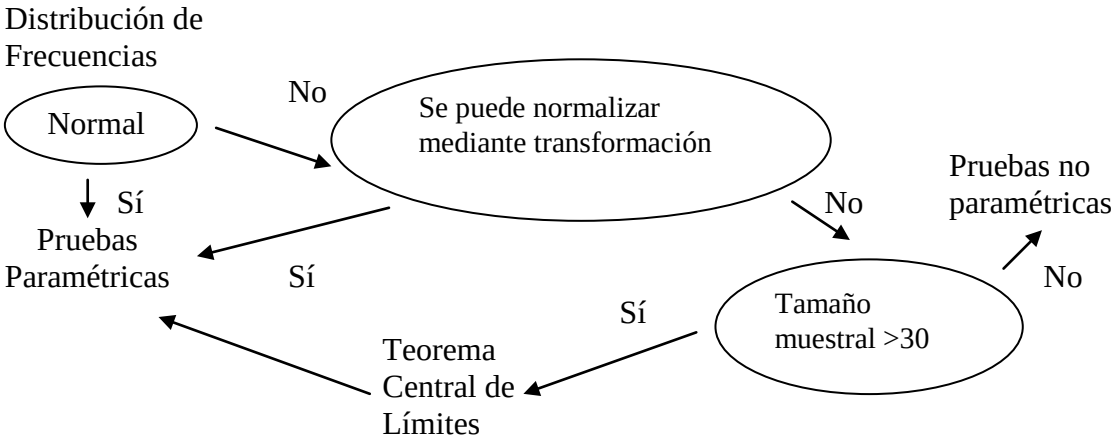
A partir de este dato se puede calcular el intervalo de confianza (IC) al 95% de la media poblacional real. Esto es la media muestral ± 2 x ES, en nuestro caso 172 ± 2 x 5.41, vale decir desde 161.18 hasta 182.82 pg/ml. Ello significa, que los valores que podría haber tomado la media (con un 95% de confiabilidad) se ubican entre 161.18 y 182.82.

Figura 4.1 Gráfica de distribución normal



Vale decir que el ES es útil para indicarnos la “zona” dentro de la cual existe un cierto nivel de certeza acerca de la ubicación de la verdadera media poblacional. Por el contrario, el DS es la medida más apropiada para indicar la distribución de observaciones individuales en una muestra.
Para variables de tipo continua o semicontinua existen dos tipos de pruebas estadísticas las paramétricas (variables de distribución normal) y no paramétricas (variables que se apartan de la normalidad en su distribución), dependiendo de lo observado en el gráfico

de distribución de frecuencias. Como se advierte en la Figura 4.1, la distribución normal es unimodal y simétrica con una alta proporción de valores en proximidad a la media, y muy pocos datos en los extremos (colas) de la curva. El hecho que muchas variables cuantitativas de la investigación biomédica sigan este tipo de distribución, no implica que ello sea indefectiblemente así. Si al efectuar una gráfica de distribución de frecuencias, los datos no parecen tener una distribución normal, existen transformaciones que tratan de normalizar a los mismos, por ejemplo la transformación logarítmica de los títulos de anticuerpos utilizados como estudios complementarios en patologías con algún sustrato inmunológico.
Cuando se trata de una variable que no sigue una distribución normal, pero el tamaño muestral es superior a 30, algunos investigadores emplean igualmente las pruebas paramétricas, al tomar como base lo establecido por el teorema central de límites. El mismo establece que las medias tienden a distribuirse normalmente (independientemente de la distribución poblacional) si el tamaño de la muestra es lo suficientemente adecuado.
Si no existen demasiados datos y no contamos con un buen método de transformación para normalizar los datos conviene utilizar las pruebas no paramétricas. A manera de resumen, el siguiente algoritmo puede servirnos de ayuda.



La medida de dispersión que puede experimentar una variable no es sólo aplicable a las de tipo continua, sino también a las nominales. Supongamos que un investigador desea determinar cuán frecuente es la presencia de enfermedad inflamatoria intestinal (EII) en pacientes con AR. A partir de ese interrogante se realiza una serie de estudios por imágenes en enfermos artríticos (n = 104) llegando a determinar que 23 de ellos tienen alteraciones compatibles con la EII, esto es el 22.1%. Asumiendo que la serie de pacientes con AR es representativa de la población real de individuos con esa enfermedad, el paso siguiente será determinar el IC al 95% de dicha proporción. Para ello se debe estimar en primer término el ES del porcentaje (p),

$$= \sqrt{p(100-p)/n} = \sqrt{22.1(100-22.1)/104} = 4.06$$

Tras ello se procede al calcular el IC al 95% de esa proporción aplicando la fórmula % ± 2 x ES, que arroja el siguiente resultado, 22.1 ± 2 x 4.06 = 12.98 - 30.22
Ello significa que existe un 95% de seguridad que en la población de pacientes con AR de la cual se obtuvo la muestra, la proporción de enfermos que presenta EII tome un

valor situado entre el 12.98 y 30.22 %. Suponiendo que la muestra de enfermos artríticos donde se efectuó el estudio no hubiese sido 104 sino 48 (10 de ellos con EII), la proporción habría sido muy parecida 20.8% pero el IC más amplio, de 9.08 a 32.25%. Cuanto mayor sea la muestra, más próximo se estará del valor que existe en la población real.
Vale la pena señalar que en los estudios donde el tamaño muestral es reducido no es aconsejable indicar el porcentaje de casos que presentan el fenómeno de interés, puesto que el IC presenta una amplia variación.

Criterios para la elección de la prueba estadística

La elección de la prueba paramétrica depende del número de grupos entre los cuales se deben efectuar las comparaciones. Si se trata de dos grupos independientes, pacientes tratados con el esquema estándar y pacientes que reciben la nueva medicación, y la variable es de tipo continua es correcto utilizar la prueba de la *t* de Student. Si fueran más de dos grupos (enfermos que fueron aleatorizados para recibir el esquema estándar, la nueva medicación, o terapia combinada con ambos compuestos) deberá emplearse el análisis de la variancia. Esta prueba nos informa sobre si existe una diferencia global significativa. Para determinar cuáles son los grupos responsables de la diferencia que hemos detectado (por ejemplo el esquema estándar en relación con la nueva medicación o bien con la terapia combinada), se debe recurrir a las pruebas de comparación post-hoc entre las cuales puede hallarse la prueba de la *t*. El empleo de este recurso es lícito cuando el análisis de la variancia arroja un resultado estadísticamente significativo.
Existen situaciones en las que se da un único grupo de pacientes factibles de ser evaluados previamente, durante y tras la finalización del tratamiento. Retomando el ejemplo de las mediciones de FNT en los pacientes con AR, puede que nos interese determinar si la administración de fármacos que reducen la actividad de la enfermedad consigue modificar, asimismo, los niveles plasmáticos de esta citocina. Para ello analizaremos si existen diferencias entre las mediciones efectuadas, previo a la instauración del tratamiento (tiempo 0) y los valores obtenidos a lo largo del tratamiento. Vale decir que cada paciente es su propio control y si bien se termina conformando 2, 3 o más series de mediciones todas pertenecen a un mismo grupo. A diferencia de la situación anterior, las muestras de valores que disponemos son dependientes y la comparación se efectúa en base al empleo de pruebas para observaciones ligadas: *t* de Student para datos apareados o análisis de la variancia para mediciones repetidas, según que se trate respectivamente de dos o más grupos de resultados.
Para el caso en que la variable no siguiera una distribución normal o bien el tamaño de la muestra fuese reducido, la estadística no paramétrica brinda una serie de pruebas que son la contraparte de lo expuesto. En lo que hace a muestras no relacionadas o independientes, las pruebas de Mann-Whitney y de Kruskal-Wallis sustituyen a la *t* de Student y el análisis de la variancia paramétrico, respectivamente. Si se trata de datos ligados los procedimientos equivalentes son la prueba de Wilcoxon y el análisis de la variancia de Friedman.
Retomando la situación que venimos utilizando como ejemplo, bien podría interesarnos analizar si las mediciones plasmáticas de FNT guardan relación con las características clínicas de la AR, para lo cual deberemos determinar la eventual asociación entre los niveles de este mediador con variables indicadoras de la magnitud y/o actividad de la enfermedad; por ejemplo la velocidad de la eritrosedimentación y un estimador de la severidad de la artritis (peso factor). Al igual que en las comparaciones entre grupos, existen pruebas de correlación (r) paramétricas y no paramétricas, según la distribución

de la variable. En el caso que el coeficiente r sea estadísticamente significativo, el cuadrado del mismo (r^2) provee una medida del grado de asociación que existe entre ambas variables. Supongamos que al comparar los niveles de FNT con el peso factor se determinara que existe una correlación positiva y significativa entre ambos, r = 0.70, $p < 0.01$. El r^2 que también se denomina variancia explicada, nos indica que el nivel de asociación es del 49%.

Cuando tenemos que manejarnos con variables nominales, el análisis está encaminado a determinar si la proporción de individuos que se ubican en una categoría en particular difiere entre grupos. Los procedimientos estadísticos más empleados para responder a este interrogante son las pruebas de la probabilidad exacta de Fisher y chi cuadrado (χ^2). Para determinar cuál es el procedimiento a adoptar, situémonos en una tabla de contingencia (4.1) donde se vuelcan los datos a comparar. Para el caso de una muestra pequeña, la opción correcta es la prueba de Fisher; esto es: cuando el número de individuos esperados para alguna de las celdas (por lo menos) es menor de 5. Si la cantidad de personas que se ubican en las 4 celdas es igual o superior a dicha cifra, se puede utilizar la prueba de χ^2.

Tabla 4.1 Tabla de contingencia 2 x 2

Variable predictiva	Grupos o Variable de Impacto		Total
	Presente	Ausente	
Presente	*a*	*b*	*a+b*
Ausente	*c*	*d*	*c+d*
Total de sujetos	*a+c*	*b+d*	*N*

Para ejemplificar la aplicabilidad de la prueba de χ^2, en la Tabla 4.2 se consignan datos de un estudio llevado a cabo en contactos de enfermos con Lepra. En la misma se representa la distribución por sexos de acuerdo a la característica de contactos (familiar u hospitalario). Al extraer las proporciones, el porcentaje de varones y mujeres en ambos grupos de contactos muestra disimilitudes que requieren un análisis para determinar si es estadísticamente significativo.

Tabla 4.2 Distribución por sexos en dos grupos de personas en contacto con pacientes de Lepra

Distribución de sexos	Tipos de Contactos		Total
	Familiar	Hospitalario	
Femenino	*211* (60.3)	*177* (75.3)	*388*
Masculino	*139* (39.7)	*58* (24.7)	*197*
Total	*350*	*235*	*585*

(porcentajes)

La fórmula para el cálculo de χ^2, se expresa en la siguiente ecuación,

$$\chi^2 = \frac{(a \times d - b \times c)^2 \times N}{(a+b)(c+d)(a+c)(b+d)}$$

Atendiendo a los valores de la Tabla procedemos a su estimación,

$$\chi^2 = \frac{(211\times 58 - 177\times 139)^2 \times 585}{(211+177)(139+58)(211+139)(177+58)} = 14.2$$

Tradicionalmente el significado de ese valor se busca en la tabla de valores críticos de χ^2, que para el caso en cuestión arroja un valor de $p < 0.001$. En la actualidad los programas estadísticos brindan el valor exacto de donde $p = 0.00023$. Sea cual fuere el valor representado queda claro que la distribución por sexos es estadísticamente diferente entre grupos.

Existen situaciones donde la comparación no es entre dos grupos, sino entre personas del mismo grupo pero en distintas ocasiones, es decir muestras relacionadas. Hace unos años nuestro laboratorio llevó a cabo una investigación en contactos de enfermos de Lepra en los cuales se quería evaluar la capacidad que tenían distintas estrategias de vacunación con BCG (sólo o más el agregado de una micobacteria no patógena muerta por calor), para inducir una respuesta positiva a la intradermorreacción con Leprosina. En este caso se compararon el resultado cualitativo a esta prueba intradérmica en los mismos individuos, previo y 2-3 años después de haber sido vacunados. La prueba estadística a utilizar es la de McNemar. Los resultados que se muestran en la Tabla 4.3 señalan que cada una de las vacunas utilizadas confiere una mayor proporción de respuestas positivas ante el desafío con Leprosina, con diferencias estadísticamente significativas respecto de la evaluación inicial. Si además del control a los 2 o 3 años siguientes a la vacunación, se hubiera efectuado una prueba adicional a los 5 años de la inmunización, el procedimiento estadístico a emplear habría sido la prueba de la *Q* de Cochran.
Valiéndonos del ejemplo citado, también podemos determinar si alguna de las vacunas utilizadas es más eficaz en su capacidad de promover respuestas positivas a la Leprosina. Para ello es necesario efectuar una comparación entre los 3 grupos (después de vacunar) por medio de la prueba de chi cuadrado, la cual indica que no existen diferencias significativas $\chi^2 = 2.07$, p=0.35. Vale decir que los tres procedimientos de vacunación son igualmente efectivos respecto de la positivización a la Leprosina.

Tabla 4.3 Respuestas positivas a la Leprosina en contactos de enfermos de Lepra, previo a la vacunación y 2-3 años después de inmunizados

Tipos de Vacunas	Respuestas positivas/total de vacunados		Valor de *p* (prueba de McNemar)
	Antes de vacunar	Después de vacunar	
A (BCG únicamente)	13/65	26/41	<0.001
B (BCG + *M. vaccae*)	14/66	26/47	<0.001
D (*M. vaccae* sólo)	70/381	132/256	<0.001

A modo de guía básica, en la Tabla 4.4 se indican las pruebas estadísticas más frecuentadas en la investigación biomédica. Lo reflejado en la Tabla no intenta reemplazar en modo alguno la amplia oferta de procedimientos que contienen los textos de estadística. Más aún, existen situaciones en que la decisión sobre la prueba a utilizar

surge de la entrevista con el experto en estadística. Unos colegas muy estimados que amigablemente solemos denominar Ulricas o Ulricos[6].

Tabla 4.4 Principales pruebas estadísticas de acuerdo al tipo de variable y modo de recolección de los datos[7]

Características de las mediciones	Variables continuas o discretas		Variables cualitativas
	Distribución normal	Distribución no normal	
Muestras no relacionadas			
2 grupos	*t* de Student	U de Mann-Whitney	Prueba de Fisher*
>2 grupos	Análisis de la variancia	Prueba de Kruskal-Wallis	Prueba de χ^2
Muestras relacionadas			
2 mediciones	*t* de Student para datos apareados	Prueba de Wilcoxon	Prueba de McNemar
>2 mediciones	Análisis de la variancia	Análisis de la variancia de Friedman	*Q* de Cochran

*La prueba de χ^2 también se puede aplicar para el caso de 2 grupos siempre que el número de individuos esperados para una de las celdas, como mínimo, no sea inferior a 5

En los ensayos controlados y para los casos en que existen distintas mediciones cuantitativas a lo largo del tiempo (inclusión y seguimiento) el mejor método de comparación entre grupos es el análisis de la covariancia.

Interacciones

En un estudio observacional o intervencionista, donde los grupos difieren entre sí muchas veces uno está tentado a indagar si las diferencias persisten cuando se confeccionan subgrupos (por encima o por debajo de tal edad, sexo, peso, antigüedad del padecimiento, entre otros). Cuando existe un sustento sólido y contemplado ya en la preparación del protocolo es muy lícito efectuar este tipo de comparaciones. De otro modo es desaconsejable puesto que incrementa la posibilidad de encontrar falsos positivos. El resultado global del estudio es mucho más conveniente a la hora de comprender el mensaje que el mismo nos transmite.

Sin embargo se presentan situaciones donde existen razones muy bien fundadas para examinar si alguna variable externa al predictor puede modificar el fenómeno de interés. Dicho en otros términos deseamos averiguar si existe heterogeneidad en los efectos, que trasladado al lenguaje estadístico se denomina interacción (no independencia de efectos). Desde este punto de vista se pretende analizar la diferencia en los efectos de la intervención según los distintos subgrupos conformados sobre la base de una hipótesis nula por la cual los efectos del tratamiento son siempre iguales.

Una prueba de interacción que arroje significado estadístico permitirá asignarle más peso a los hallazgos de un subgrupo en particular. Si por el contrario no se presenta heterogeneidad, lo que pesa es el resultado general.

[6] Por Ulrica, el personaje de la ópera "Un ballo in maschera", quien predecía el futuro con increíble exactitud.

[7] Una publicación reciente (Fleming et al. J Neuroimmunol 2005; 170: 71) provee un detalle muy adecuado de los procedimientos estadísticos a seguir en investigación experimental.

También debemos tener en cuenta que las interacciones puede ser cualitativas (la intervención es efectiva en un subgrupo y no así en otro) o bien cuantitativa. En este último caso el tratamiento es efectivo en todos los subgrupos pero el grado de efectividad varía entre ellos.
Para ilustrar el modo en que debemos proceder recurriremos a un ejemplo hipotético en el cual un grupo de investigadores ha desarrollado una nueva vacuna hacia una bacteria encapsulada con el propósito de generar una mejor respuesta de anticuerpos. El estudio se lleva a cabo en niños puesto que es un blanco poblacional muy afectado por el patógeno. Se conforman dos grupos de estudio (de 238 y 233 niños) que reciben la vacuna estándar o experimental. La variable de impacto es la presencia de niveles de anticuerpos de tipo IgG específicos por encima de un título de corte considerado como protector.

Tabla 4.5 Ejemplo para el análisis de interacción

Grupos	Niños respondedores		
	< 3 años	> 3 años	Acumulado
Experimental	53/104 [50.9]	74/134 [55.2]	127/238
Estándar	40/103 [38.8]	53/130 [40.8]	92/233
Valor de *p*	<0.11	=0.025	< 0.005

En los menores de 3 años la proporción de respondedores con la vacuna experimental es el 50.9% mientras que en aquellos que recibieron la vacuna tradicional es 38.8%, la diferencia de efecto es 50.9 – 38.8= 12.1. En el grupo por encima de 3 años los guarismos son 55.2 y 40.8 y la diferencia de efecto 14.4. Vale decir que la inmunogenicidad de la nueva vacuna para mejorar la proporción de niños respondedores es prácticamente la misma en ambos grupos. La p = 0.11 en realidad no significa que no existe una diferencia sino que no tenemos evidencia de que esa diferencia sea real. Este valor de *p* no logra significado porque la manera en que hemos estimado el efecto no era apropiada, la muestra de niños por debajo de 3 años debería haber sido superior.
El método para evaluar la interacción es el siguiente. Primero se debe extraer el error estándar para cada uno de los efectos observados.

$$ES = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

En los niños por debajo de 3 años teníamos 50.9 – 38.8= 12.1 (ES = 6.86)

$$ES = \sqrt{\frac{50.9(49.1)}{104} + \frac{38.8(61.2)}{103}}$$

En los mayores era 55.2 - 40.8= 14.4 (ES =6.08).

$$ES = \sqrt{\frac{55.2(44.8)}{134} + \frac{40.8(59.2)}{130}}$$

De acuerdo con esto la diferencia de tratamientos es 14.1-12.1 = 2. Para obtener el ES de esta diferencia recurrimos a la siguiente fórmula, que nos arroja un valor de 9.17.

$$ES_d = \sqrt{ES_1^2 + ES_2^2}$$

A partir de estos datos se obtiene el cociente (2/9.17 = 0.22), que brinda un valor de $p > 0.70$ el cual no apoya la presunción de que la mayor respuesta de anticuerpos que se obtiene con la vacuna experimental esté relacionada con la edad. Si el fenómeno de interés fuese una variable continua, procederíamos a calcular la diferencia entre las medias de los 2 tratamientos y el error estándar de esa diferencia para luego obtener el valor de *p*.

Más allá de esta estrategia, Altman y Brand (BMJ 2003; 326: 219) han provisto recientemente de un método para calcular si los RR obtenidos entre subgrupos independientes de un estudio difieren entre sí. El procedimiento se puede utilizar para comparar OR. Para ello es necesario obtener los logaritmos de los RR y sus intervalos de confianza. El ejemplo que se utiliza para el cálculo compara dos RR (en realidad factores de protección) sobre fracturas no vertebrales de ensayos aleatorizados sobre terapia hormonal de reemplazo uno estadísticamente significativo: 0.67 (0.46-0.98) y otro no: 0.88 (0.71-1.08), donde la comparación entre ambos indica que los mismos no son diferentes ($p = 0.2$). Esta herramienta certifica la presunción de que 2 valores no difieren entre sí cuando los IC se superponen en buena medida.

4.3 Análisis ajustados para evaluar el efecto potencial de un factor de confusión (FC)

En la sección 3.3 del capítulo precedente habíamos descrito brevemente las estrategias que se utilizan cuando se sospecha, o incluso se tiene certeza, sobre de la existencia de FC con distribución desigual entre grupos. De tratarse de un único factor es necesario determinar si el resultado de las comparaciones se halla afectado cuando se tiene en cuenta a dicho FC. La técnica que se utiliza para evaluar la eventual influencia del "*confounder*" es el análisis estratificado.

Supongamos que al analizar el efecto de dos fármacos (A respecto de B) sobre una enfermedad de curso letal verificamos que existen diferencias en cuanto a la supervivencia de las personas, pero al comparar la distribución por género advertimos que en aquellos tratados con la droga A había un 80% de varones mientras que en los medicados con el compuesto B la cifra era de un 20%; siendo la mortalidad (por cualquier fármaco) muy superior en los hombres (50%) respecto de las mujeres (10%).

Tabla 4.6 Resultados de un ensayo clínico, desagregado por género, donde se comparó la supervivencia de los enfermos según el tipo de tratamiento

	Compuesto A		Compuesto B	
Desenlace	*Varones*	*Mujeres*	*Varones*	*Mujeres*
Muertos	400	20	100	80
Vivos	400	180	100	720
Total	800	200	200	800

Si efectuamos un análisis con los datos crudos que se muestran en la Tabla 4.7, la prueba estadística de χ^2 arroja una diferencia altamente significativa, $p < 10^{-7}$.

Tabla 4.7 Comparación general

Desenlace	Droga A	Droga B
Muertos	420	180
Vivos	580	820
Total	1000	1000

No obstante la diferencia que hemos detectado, el hecho que la distribución por género era diferente entre los grupos de tratamiento, requiere efectuar un análisis estratificado por género. Para ello recurrimos a la prueba de Mantel Haenszel, utilizando el mismo criterio (supervivencia según el tipo de medicación) pero desagregado para varones y mujeres. La Tabla 4.8 indica con toda claridad que los resultados del análisis por separado y acumulativo (Mantel-Haenszel) no son estadísticamente significativos.

Tabla 4.8 Comparaciones sobre el resultado de los tratamientos según género

VARONES

Desenlace	Droga A	Droga B
Muertos	400	100
Vivos	400	100
Total	800	200

P = 0.99

Mantel-Haenszel P=0.9459

MUJERES

Desenlace	Droga A	Droga B
Muertos	20	80
Vivos	180	720
Total	200	800

P = 0.99

Evidentemente en nuestro caso hipotético, la variable género ejercía una influencia sobre el efecto de los tratamientos y, al hallarse distribuida desigualmente entre aquellos que recibían la droga A o la droga B, se generaba una diferencia espuria. Dado que no es tarea fácil determinar el umbral en que una desigualdad en la distribución del FC comienza a modificar ostensiblemente los resultados obtenidos, se recomienda efectuar el análisis estratificado de todas maneras. ¿Para qué arriesgarnos cuando podemos ajustar los cálculos?

Pasemos a considerar una situación diferente. En este estudio se analiza el efecto protector de un compuesto A (comparado con B, que es el placebo) sobre la mortalidad que puede ocasionar un enfermedad viral para la cual aún no existe tratamiento. Como se observa en la Tabla 4.9 la distribución por género es la misma y se registran diferencias en los varones respecto a los 2 esquemas utilizados

Tabla 4.9 Resultados generales del ensayo clínico

	Compuesto A		Compuesto B	
Desenlace	*Varones*	*Mujeres*	*Varones*	*Mujeres*
Muertos	100	40	200	40
Vivos	300	360	200	360
Total	400	400	400	400

El análisis de los datos crudos que se resumen en la Tabla 4.10, muestra una diferencia altamente significativa, $p = 10^{-7}$.

Tabla 4.10 Comparación global de la supervivencia

Desenlace	**Droga A**	**Droga B**
Muertos	140	240
Vivos	660	560
Total	800	800

Al practicar el análisis estratificado por género (prueba de Mantel-Haenszel) se obtienen los siguientes resultados

Tabla 4.11 Comparaciones en la supervivencia por tratamientos según género

VARONES

Desenlace	**Droga A**	**Droga B**
Muertos	100	200
Vivos	300	200
Total	400	400

$p = 10^{-7}$

Prueba de Mantel-Haenszel $p = 10^{-7}$

MUJERES

Desenlace	**Droga A**	**Droga B**
Muertos	40	40
Vivos	360	360
Total	400	400

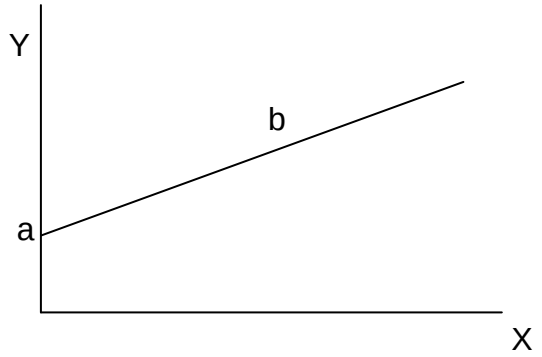
$p = 1$

Vale decir cuando hay similitud en la distribución entre grupos del supuesto FC (género), será ínfima la discordancia, si la hubiere, entre la estimación cruda y la obtenida a través de la prueba de Mantel-Haenszel. Para el caso en que un FC adopte más de dos categorías (por ejemplo: raza) se puede estratificar el análisis de acuerdo con el número de clasificaciones que se haya asignado para dicha variable.

De lo expuesto queda claro que el análisis estratificado es un procedimiento útil para el control de los eventuales efectos del FC sobre la variable de impacto. Es fácil de realizar, y hace que el investigador y los lectores del trabajo puedan tener una mejor idea de la interrelación entre las variables predictiva y de impacto en presencia de factores adicionales. La desventaja que posee este procedimiento reside en que si en el estudio existieran varios FC el análisis estratificado no permitirá una evaluación "en simultáneo" del efecto de todos ellos, puesto que los toma por separado. Supongamos que se tratara de un ensayo clínico donde se compara el efecto de dos hipolipemiantes y decidimos incluir cuatro FC, género, edad (4 grupos etarios), tabaquismo (no fumador,

ex-fumador, fumador), índice de masa corporal (3 clasificaciones). Al efectuar el cálculo (2 x 4 x 3 x 3) obtendremos 72 estratos de combinaciones posibles. Imaginemos la abultada cantidad de participantes que tendríamos que incluir en el estudio, como para lograr una muestra suficiente de individuos en cada una de los estratos.
Para superar este inconveniente nuestros amigos los estadísticos han desarrollado el análisis multivariado, una técnica que permite evaluar el grado de asociación entre las variables predictivas y de impacto, controlando en forma simultánea a dichos FC.
El método comúnmente utilizado para el ajuste en simultáneo de varios FC es el modelo de regresión múltiple. Este procedimiento se basa en el modelo de regresión lineal, el cual se emplea para determinar la relación entre las variables predictiva (independiente, X) y de impacto (dependiente, Y) donde $Y = a + bX$. En dicha ecuación "a" es el punto en que la recta intercepta a la ordenada y representa el valor que adopta Y cuando $X = 0$, mientras que el cociente "b" indica la pendiente de la recta y define el grado de asociación entre X e Y; vale decir: el cambio que se produce en Y cada vez que X se modifica en una unidad (Ver Figura 4.2). Si deseáramos analizar la relación entre la glucemia y la edad, el coeficiente b nos indica el cambio medio de la glicemia por cada unidad de variación de la edad, por ejemplo años.

Figura 4.2



A diferencia de lo anterior, en el modelo de regresión múltiple se incluyen un número adicional de variables independientes determinando la siguiente ecuación

$$Y = a + b_1X_1 + b_2X_2 + \ldots . b_nX_n$$

Dado que el modelo parte de una relación lineal entre ambas variables, se debe controlar estadísticamente que este supuesto se cumpla. Para el caso que la asociación fuera "no lineal" se puede transformar la variable para conseguir la linealidad. El coeficiente para cada variable independiente nos determinará el cambio que experimenta Y, controlando el efecto de las demás variables X que incluimos en el modelo. Utilizando el ejemplo de la glicemia en relación con la edad, se pueden incluir variables adicionales como género, peso, dieta, actividad física y tabaquismo entre otras.
Llegado a este punto es probable que el lector se haya preguntado cómo proceder cuando la variable independiente no es continua sino categórica, tal como "enfermo" (diabético) y "no enfermo". En situaciones de este tipo se utiliza el análisis de la regresión logística; el mismo es una variante del modelo de regresión múltiple, en el cual la chance de ocurrencia de la variable de salida, es una función de las distintas variables predictivas. Tomando a Y como la probabilidad de presentación de la enfermedad, Y/(1-Y) como la desigualdad relativa para que se produzca o no dicha eventualidad, la variable dependiente queda definida como el logaritmo natural (ln) de esta desigualdad relativa que es una función lineal de una serie de variables predictivas

$$\ln\ [Y / 1- Y] = a + b_1X_1 + b_2X_2\ + \ldots b_nX_n$$

Los valores que se obtienen a través del modelo de regresión logística oscilan entre 0 y 1. Los mismos indican la magnitud del aumento o descenso en el ln de la desigualdad relativa por cada cambio unitario en la variable predictiva, con las demás variables independientes bajo control. A diferencia de la Tabla de contingencia 2 x 2 de donde se extraían los OR crudos (ver Tabla 2.4), los coeficientes que ofrece la regresión logística nos permiten calcular los OR ajustados para los potenciales "*confounders*".
Podemos afirmar que el análisis multivariado es una herramienta muy útil para las investigaciones tendientes a profundizar en torno de los potenciales factores que contribuyen a generar una situación patológica. En los estudios de cohorte también adquieren un valor predictivo ya que es posible estimar el riesgo de una persona, con determinados atributos de la(s) variable(s) predictiva(s), de desarrollar el fenómeno de interés. Esto se ve posibilitado porque estas investigaciones se planifican y llevan a cabo teniendo en cuenta la incidencia del fenómeno en la población. Por el contrario, cuando se adopta la estrategia de casos y controles la muestra es fijada arbitrariamente por el investigador con lo cual no es posible atribuir a los OR un valor de riesgo.

4.4 Estrategias particulares para ensayos clínicos

Sumado a todo lo expuesto, resta por considerar algunas situaciones especiales en torno al manejo de los datos de un estudio clínico, que se exponen brevemente.

Cálculo de eficacia

En los ensayos terapéuticos, una vez que se ha comprobado la superioridad del nuevo tratamiento aplicado, es necesario otorgarle un valor real a dicha demostración mediante el cálculo de eficacia. La medición de la eficacia de la nueva modalidad terapéutica (experimental) se efectúa en relación con los resultados que arrojó el tratamiento estándar y se obtiene a través de la siguiente fórmula

$$\text{Eficacia} = \frac{\text{Fracasos con el tratamiento estándar} - \text{fracasos con la droga experimental}}{\text{Fracasos con el tratamiento estándar}}$$

Una manera alternativa de estimar eficacia es a partir de la siguiente ecuación
Eficacia = 1- RR

Donde RR es la relación entre $\frac{\text{Tasa de Fracasos con la droga experimental}}{\text{Tasa de Fracasos con el tratamiento estándar}}$

Supongamos que se trata de un estudio pequeño en enfermos con Hepatitis C sometidos a dos tipos de tratamientos, uno el que se viene utilizando hasta el presente (estándar) y por el otro lado el alternativo que se supone mejor (experimental). Los resultados al cabo de un tiempo estipulado y en base a los criterios que permitieron definir fracasos o éxitos terapéuticos se resumen en la Tabla 4.12

Tabla 4.12 Resultados del ensayo clínico en pacientes con Hepatitis C

Fármacos	*Fracasos*	*Éxitos*
Droga B (experimental)	6	14
Droga A (estándar)	13	7

Puede observarse que la tasa de fracasos en B es 6/20, mientras que en los tratados con A la proporción es 13/20.

Si partimos de la primer fórmula es cálculo es como sigue,

$$\frac{0.65 - 0.30}{0.65} = 0.54$$

Si procedemos a calcular RR, 0.30/ 0.65= 0.46 con un intervalo de confianza al 95% (0.22 - 0.97). La estimación de eficacia arroja un valor similar al que se obtuvo de acuerdo con el primer procedimiento = 0.54 intervalo de confianza al 95% (0.03 - 0.78)

El valor obtenido es la eficacia de B respecto de A, para reducir los fracasos de tratamiento.

Eficacia en vacunas

El método tradicional para determinar la eficacia de una vacuna surge de relacionar la ocurrencia de enfermedad en el grupo vacunado respecto del que no lo fue.

$$\text{Eficacia de la vacuna} = 1 - \frac{\textit{probabilidad de enfermedad en el grupo vacunado}}{\textit{probabilidad de la enfermedad en el grupo control}}$$

Para traducir esta expresión a términos reales, tomemos como modelo las investigaciones sobre la vacunación con BCG en la prevención contra la Lepra y la Tuberculosis. En líneas generales, los estudios llevados a cabo con este propósito adoptan dos tipos de diseños. El clásico, por así decirlo, es el longitudinal donde se conforman dos poblaciones de individuos, vacunados o no, y tras un período de seguimiento de varios años se compara la incidencia de los casos que han ido apareciendo en dicho lapso. Con este tipo de estudios se pueden confeccionar las curvas de Kaplan-Meier que se tratará en la sección siguiente. Cuando el estudio longitudinal no es factible se puede recurrir a un diseño alternativo, no tan poderoso como el anterior, que es el de casos y controles. De acuerdo a lo especificado en el capítulo 2, aquí se seleccionan individuos que ya presentan la enfermedad y una población de controles, lo más parecida posible a la de los casos. En ambos grupos se estima la frecuencia de vacunados que permite obtener el OR. A partir del OR procedemos a calcular la efectividad de la vacuna en base a la siguiente fórmula:

Efectividad = (1-OR) x 100%

También es posible extraer la fracción preventiva (FP) y de ese modo estimar la proporción de nuevos casos potenciales que pudieron prevenirse con el procedimiento.

La FP en definitiva no es otra cosa que la contraparte del la fracción etiológica (ver capítulo 2, sección), para el caso en que la exposición no implica riesgo sino protección.

A partir del modo de ordenar los datos para el cálculo de los OR, se toman los resultados de un estudio de casos y controles efectuado en nuestra ciudad acerca de la protección conferida por la vacuna BCG en el desarrollo de Tuberculosis extrapulmonar en niños. En dicho estudio se obtuvo un OR de 0.21 (0.06-0.79, IC 95%) que arrojó una efectividad de 79 (21-94), mientras que la FP fue igual a 0.74 (Tabla 4.13). También se puede extraer el IC de esta fracción, aunque el método difiere del utilizado para el OR.

Tabla 4.13 Cálculo de la fracción preventiva

	Enfermos	*No Enfermos*	
Expuestos	a	b	m_1
No expuestos	c	d	m_2
	n_1	n_2	n

	TB	*No Enfermos*	
Con BCG	11	739	750
Sin BCG	3	43	46
	14	782	796

$FP = (bc - ad)/cn_2$

Cuando existe una prueba de laboratorio que estima en forma adecuada el grado de protección inmune que posee la persona, se ha consensuado que el modelo de no inferioridad es una herramienta importante para comparar la inmunogenicidad de una vacuna nueva o reformulada. Para el caso de infecciones virales o bacterianas donde el estado protectivo se asocia a una buena respuesta de anticuerpos este procedimiento es aplicable. La cuestión más crítica para este tipo de estimación es definir el nivel que deben tener los anticuerpos para ser protectores.

$$\text{Eficacia de la Vacuna} = 1 - \frac{\%\ de\ vacunados\ con\ Ac < niveles\ protectores}{\%\ de\ controles\ con\ Ac < niveles\ protectores}$$

Análisis de puntos finales cualitativos

En muchos ensayos clínicos tanto preventivos como terapéuticos, el punto central a determinar, es si la vacuna ofrece protección contra la enfermedad o el tratamiento reduce las cifras de mortalidad. Como se adelantara, los elementos a considerar en el análisis tienen que ver con la presencia de la variable de impacto (enfermo o no enfermo, vivo o muerto, etc.), el tiempo transcurrido entre la fecha de admisión al estudio y el momento en que se produce el fenómeno de interés. Para ejemplificar la cuestion pensemos en una investigación en 48 voluntarios sanos donde se estudia el potencial efecto protector de una vacuna contra una enfermedad infecciosa como la Leishmaniasis en personas con riesgo de contraerla. La mitad de los participantes recibe la vacuna en estudio y el resto el vehículo solamente (controles). En este caso no podemos utilizar el cálculo de eficacia presentado anteriormente por cuanto el estado de protección no está vinculado a los niveles de anticuerpos. Aquí, la cuestión a evaluar es cuántos desarrollarán la enfermedad y en qué lapso de tiempo. Al cabo de un período de seguimiento (variable según los casos) se obtienen los siguientes resultados:

Tabla 4.14

CONTROLES Tiempo de aparición (semanas)		VACUNADOS Tiempo de aparición (semanas)	
13	**CE**	13	**CE**
15	**CE**	17	**CE**
16	**CE**	25	**CE**
18	**CE**	41	SE
22	**CE**	55	SE
31	**CE**	73	**CE**
42	**CE**	90	**CE**
44	**CE**	97	**CE**
47	**CE**	103	**CE**
54	**CE**	126	SE
60	**CE**	129	**CE**
61	**CE**	136	**CE**
63	**CE**	140	SE
68	**CE**	141	SE
72	**CE**	146	SE
83	**CE**	150	**CE**
91	**CE**	153	SE
110	SE	156	**CE**
117	SE	160	SE
121	SE	168	SE
138	SE	173	**CE**
151	SE	181	SE
178	SE	184	SE
185	SE	194	SE

CE: con enfermedad; SE: sin enfermedad

Si efectuamos un análisis global (CE, SE) mediante la prueba de χ^2, la diferencia no es significativa. El punto es que esta prueba no tiene en cuenta el tiempo de aparición de la enfermedad y el período de seguimiento de los participantes. Observando la Tabla 4.14 resulta claro que los individuos vacunados estuvieron mucho más tiempo libres de desarrollar la enfermedad, de modo que debemos elegir otra técnica estadística de comparación. Para ello conviene construir una gráfica con las curvas de supervivencia (no enfermo), las que permiten visualizar el efecto de ambas intervenciones. Con los mismos datos de la Tabla 4.14 se confecciona otra tabla (4.15) donde se consigna la proporción de personas que permanecen libres de enfermedad para un momento dado de la evaluación. Para ejemplificar la cuestión, en el grupo de no vacunados 23/24 personas (95.8%) estaban sin enfermedad transcurridos 13 semanas de seguimiento, mientras que 22/24 (91.67%) permanecían sanos luego de 15 semanas de control. El cálculo para estimar el porcentaje de supervivencia por encima de determinado período (T), surge de la siguiente fórmula

$$p(\mathrm{T}) = \frac{\#\ de\ personas\ que\ sobreviven\ más\ allá\ del\ período\ evaluado\ (T)}{\#\ de\ sobrevivientes\ más\ allá\ del\ período + nro\ de\ fallecidos\ en\ T} \text{ x } 100$$

A partir de estos porcentajes se confeccionan las curvas de supervivencia de Kaplan-Meier (Figura 4.3). El método utilizado para este tipo de comparación es la prueba del intervalo logarítmico (logrank test) el cual nos permitirá determinar si las diferencias entre los dos tipos de tratamientos son estadísticamente significativas o bien se de

muestras de una misma población (hipótesis nula). Para el caso que nos ocupa la prueba arroja que la diferencia es estadísticamente significativa ($p < 0.031$).

Tabla 4.15. Disposición de los datos para graficar las curvas de supervivencia

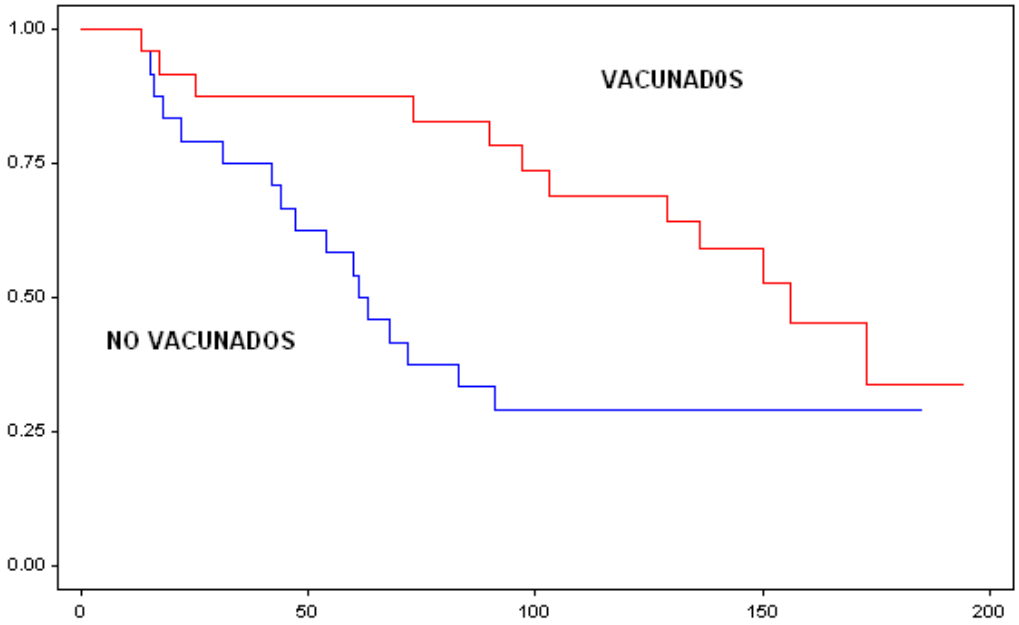
NO VACUNADOS				VACUNADOS			
Sujeto	Tiempo de observación	Estado	Porcentaje sin enfermedad	Sujeto	Tiempo de observación	Estado	Porcentaje sin enfermedad
1	13	CE	,9583	1	13	CE	,9583
2	15	CE	,9167	2	17	CE	,9167
3	16	CE	,8750	3	25	CE	,8750
4	18	CE	,8333	4	41	SE	
5	22	CE	,7917	5	55	SE	
6	31	CE	,7500	6	73	CE	,8289
7	42	CE	,7083	7	90	CE	,7829
8	44	CE	,6667	8	97	CE	,7368
9	47	CE	,6250	9	103	CE	,6908
10	54	CE	,5833	10	126	SE	
11	60	CE	,5417	11	129	CE	,6414
12	61	CE	,5000	12	136	CE	,5921
13	63	CE	,4583	13	140	SE	
14	68	CE	,4167	14	141	SE	
15	72	CE	,3750	15	146	SE	
16	83	CE	,3333	16	150	CE	,5263
17	91	CE	,2917	17	153	SE	
18	110	SE		18	156	CE	,4511
19	117	SE		19	160	SE	
20	121	SE		20	168	SE	
21	138	SE		21	173	CE	,3383
22	151	SE		22	181	SE	
23	178	SE		23	184	SE	
24	185	SE		24	194	SE	

Para poder estimar la magnitud de cuán diferentes son los grupos esta diferencia debemos recurrir al "*hazard ratio*". Esta prueba confronta el número de eventos observados –O- respecto de lo esperado –E- (en el supuesto que la hipótesis es cierta) para cada grupo y luego se extrae el cociente.

$$\text{HR} = \frac{O_1 / E_1}{O_2 / E_2}$$

En el ejemplo presentado el HR arroja un valor de 0.43 (p <0.03) y eso nos estima el grado de protección conferido por la vacuna.

Figura 4.3 Curvas de Kaplan Meier de acuerdo a los datos consignados en la Tabla 4.15



Si se quiere efectuar un cálculo ajustado para la eventual influencia de factores pronósticos la prueba de elección es el análisis de los riesgos proporcionales o la regresión de Cox. El lector interesado en los fundamentos y procedimientos referidos a esta metodología puede consultar a la bibliografía recomendada, (Pagano y Gauvreau 2001; Schoenbach y Rosamond, 2000) ya que la descripción de los mismos escapa a los objetivos de este manual.
Para el caso de supervivencia en transplantados, Fu y col. proponen un método alternativo de análisis.

Una prueba bien simple

En muchos ECA el punto de interés más saliente es un evento ligado al curso de la enfermedad (muerte, infarto, o accidente CDV). Los métodos estándares explicitados en los párrafos precedentes toman en cuenta la variación durante el seguimiento de los pacientes. Con las salvedades del caso, si uno fijara la evaluación en un determinado momento alternativamente se puede aplicar χ^2. Recientemente la literatura nos brinda un nuevo método para evaluar los efectos de la intervención sobre un determinado evento. Ello permite obtener una información bastante rápida respecto de si existen diferencias de tratamiento, pero de ningún modo implica que pueda reemplazar a las pruebas convencionales.
Consideremos un ECA con dos grupos de tratamiento (randomización igual), similar número de pacientes, un evento clínico como punto final, y la cantidad de pacientes que lo experimentan para su análisis. Se calcula la diferencia en las dos cantidades de eventos y se lo divide por la raíz cuadrada de su suma; el número resultante es z. A partir de tablas de distribución normal, z puede ser convertido en el valor de *p*. Así si z > 1.96 = $p < 0.05$ y z > 2.58 = $p < 0.01$. La prueba requiere un número casi igual de pacientes por grupo y el mismo tiempo de seguimiento.

$$\mathbf{Z}= \frac{a-b}{\sqrt{a+b}}$$

Valor de z	**Valor de *p***
1.28	0.2
1.64	0.1
1.96	0.05
2.05	0.04
2.17	0.03
2.32	0.02
2.58	0.01
3.29	0.001
3.89	0.0001

El procedimiento tiene la limitación de que si la tasa de eventos es alta el valor de *p* es mayor de lo que debería ser.

Análisis interinos

La realización de un ensayo clínico puede enmarcarse en una situación que por sí misma constituye un dilema médico, como lo sería la aplicación de un tratamiento de sospechada eficacia en una enfermedad letal o invalidante para la cual aún no existe una terapia efectiva. En circunstancias de esta naturaleza y otras que si bien no son tan extremas, se nos presenta una chance real de ofrecer una herramienta superadora, es razonable plantear que exista un número aceptable y justificado de análisis interinos en lugar de esperar a la conclusión del estudio. Si vamos a proceder en esta dirección habrá que asegurar que las normas del cegado a las asignaciones de tratamientos se respetan con todo rigor, puesto que el enterarnos de antemano sobre la tendencia de los resultados puede llevarnos a inmiscuirnos indebidamente en la conducción de un ensayo al detectar que uno de los grupos que reciben determinada intervención está experimentado un curso distinto del fenómeno bajo análisis. Pensemos por un momento en un estudio en donde se contempló un análisis interino, y tras la conclusión del ensayo constatamos que los resultados pre y post-análisis interino difieren entre sí. ¿Podrán creernos que esta estrategia no contenía intencionalidad alguna, o que no se ha introducido un sesgo en la selección y evaluación de los nuevos casos?

La supervisión de un ensayo clínico puede incluso llevar a suspender el estudio si se advirtiera que uno de los tratamientos es claramente superior o inferior. Un recurso fundamental para este tipo de "vigilancia" es precisamente la realización de estos análisis interinos. Dado que la aplicación de este tipo de análisis incrementa la chance de detectar hallazgos falso-positivos, la literatura provee de una serie de cálculos que redefinen el valor de *p* de acuerdo al número de análisis que se planifican (Geller y Pocock, 1987; Piantadosi, 1991). La intención de aplicar este tipo de estrategias debe estar incorporada al protocolo, para llevarse a cabo sobre mediados de su ejecución, o bien tras un número pre-especificado de inclusiones o puntos finales. En consonancia con lo expresado los objetivos del análisis interino incluyen (1) evitar la exposición a un riesgo para los pacientes subsiguientes de detectarse en perjuicio; (2) no continuar exponiendo a más pacientes cuando los datos disponibles dan cuenta de lo que pretendíamos saber; (3) evitar costos innecesarios; (4) proveer una respuesta valedera en el menor tiempo posible. Obviamente que esta estrategia cobra más relevancia cuando se trata de una intervención invasiva, potencialmente riesgosa, onerosa, o el evento bajo análisis es grave de por sí. Si se tratara de un ECA donde se espera que el efecto de la intervención se produzca en un corto período de tiempo, deberá planearse un período de inclusión más prolongado como para que sea posible el análisis.

En cuanto a los lineamientos para el análisis interino se debe establecer un criterio estadístico de detención del ensayo de existir una diferencia notoria en el punto final en favor de una de las intervenciones, ya en la primera parte del estudio (inclusión o seguimiento). Si por el contrario no se observara una diferencia como para interrumpir el ensayo ni tampoco se vislumbrara una tendencia a diferir, es poco probable que se llegue a detectar la pretendida diferencia aun tras la compleción de todo el estudio. En estos casos no es atinado continuar con el ECA. Una situación distinta se da cuando a pesar de no constatarse una diferencia estadística, los datos apuntan a que ello puede verificarse de completar el estudio planificado. En esta circunstancia nos enfrentamos a dos posibilidades, que la tendencia esté a favor del grupo experimental o bien del control. En el primer caso, está justificado completar los seguimientos como para determinar si el nuevo tratamiento es realmente superior. Esto se torna mucho menos probable si los datos favorecieran al grupo control; y no existe necesidad de continuar con la investigación para probar que la intervención resulta ser peor.
Atento a la no revelación de los códigos, los datos del análisis interino deben ser evaluados por un comité de expertos "externo y no cegado". Así, el protocolo deberá pre-especificar que los análisis interinos para un ECA donde se compara un nuevo tratamiento con otro inactivo o estándar, estarán disponibles a un grupo de monitores independientes con conocimiento de los códigos de randomización, para el caso en que se detectara una "diferencia prometedora". Sumado a estas consideraciones, otros autores señalan que a la hora de decidir sobre la continuidad o no de un ECA también debe tenerse en cuenta el balance riesgo/beneficio, la consistencia interna de los resultados, la concordancia con los hallazgos de otras investigaciones relacionadas, y la probabilidad que los resultados pudiesen llegar a modificar la práctica clínica.

Intención de tratar (ITT)
Es una estrategia que confronta los resultados del estudio de acuerdo a los grupos conformados aunque haya individuos que no recibieron o completaron el tratamiento. Por el contrario en el análisis por protocolo (PP) sólo se incluyen los que completaron el tratamiento y el seguimiento. El primero estima el efecto de recomendar un tratamiento a los participantes del estudio, mientras que el segundo la eficacia del tratamiento realmente recibido.
Al considerar la adherencia a la intervención asignada hay que distinguir efectividad del uso (estima el evento en las condiciones habituales de administración), y efectividad del método/eficacia, que analiza el método en condiciones ideales y sin desviaciones. Si los abandonos en el ensayo son similares a lo que puede darse en el futuro, un análisis ITT podría estimar bastante bien la efectividad del uso. Si los abandonos tienen que ver con el desenlace y originados por diferentes causas según el tipo de tratamiento, la estimación PP que excluyen desviaciones del protocolo estará sesgada, especialmente si el porcentaje de abandonos es significativo.
El argumento para el análisis ITT es que si bien sólo se debería comparar a las personas que realizaron el tratamiento, los pacientes que adhieren al mismo son de alguna manera diferentes de aquellos que no lo efectúan o bien lo abandonan. Dicho en otros términos adherentes y no adherentes emerge como una variable capaz de distorsionar el efecto de la intervención que pretendemos investigar. A pesar que permite superar este inconveniente, el método tiene sus desventajas ya que la inclusión de las personas que no efectuaron el tratamiento en el análisis general modificará el poder del estudio y el efecto potencial de la intervención. Contrariamente a este reparo, si uno tiene en cuenta que en la situación real muchas veces los pacientes son renuentes a cumplir con el

tratamiento, la estrategia ITT en definitiva no se aparta tanto de lo que ocurre en la práctica diaria.
Hace unos años Porta y col (2007) efectuaron una revisión sistemática de ECA a dos brazos publicados desde 2001 a 2003 para cuantificar el grado de desacuerdo entre los ambos métodos (ITT y PP). Tras una selección de 74 trabajos, la estrategia PP en promedió proveyó mejores estimaciones pero existía una gran variación en los límites de concordancia entre ambos métodos. También se observó que los datos faltantes podían causar un sesgo sistemático e impredecible. En definitiva ante la presencia de desviaciones al protocolo, las conclusiones de ECA no pueden asentarse ni en el ITT o PP por sí solos.
Una cuestión no menor es que cuando los pacientes no adhieren al protocolo establecido se puede producir una contaminación de tratamientos y la consecuente generación de error. Dado que el análisis de ITT ignora el tema de contaminación de tratamientos, recientemente se ha descrito una técnica bastante simple "el análisis de ITT ajustado por contaminación" el cual complementa la estrategia original y produce una mejor estimación de los beneficios y riesgos de recibir el tratamiento. Los ECA que reportan la estrategia de ITT modificado son cada vez más frecuentes en la literatura.

Análisis por subgrupos

El análisis por subgrupos puede ser una estrategia valiosa pero conviene tomar algunos recaudos. Consideremos un ECA típico con un punto final primario en el que se efectúan análisis por subgrupos para determinar si las subclases de pacientes responden diferencialmente al tratamiento. A *prima facie* el planteo no parece problemático, pero pueden presentarse varias fuentes de confusión e incertidumbres respecto a esto. Supongamos que se los divide en función del sexo, los resultados no dan diferencias significativas globales, pero podría ser mejor en las mujeres. O bien existen diferencias, pero ello varía de acuerdo con el sexo. Estas cuestiones podrían investigarse por medio del test de interacción (cuantitativa cuando es superior en ambos grupos o relativa cuando difiere entre los subgrupos) que ya fue tratado. Subdividir los datos en subgrupos reduce el poder del estudio para detector diferencias de tratamiento, dado que no sólo se reduce el tamaño muestral, sino que el número de pruebas estadísticas necesarias es más del doble de lo que se requiere para probar una interacción. Cuando se efectúan múltiples análisis de subgrupos la probabilidad de un resultado falso positivo está abultada. Aquí hay que establecer un valor de *p* distinto, por ejemplo se van a efectuar 10 comparaciones *p* deberá ser menor a 0.005. Cuando un trabajo presenta los resultados de análisis por subgrupos *post hoc*, no siempre está claro el por qué de haber elegido tales subgrupos y cuántos otros se conformaron. Los autores y los editores deben asegurarnos la transparencia del análisis por subgrupos. Al momento de presentar análisis por subgrupos, no es bueno presentar el valor de *p* para la comparación dentro de subgrupos, sino más bien brindar una estimación de la magnitud de la diferencia de tratamiento y el IC correspondiente. Estos IC no deben ser utilizados para inferir que una diferencia de tratamiento en un subgrupo es significativa, basándonos en si el intervalo excluye la hipótesis de igualdad entre los grupos de tratamiento; dado que tales análisis adolecen de los mismos problemas suscitados por la utilización de múltiples pruebas estadísticas. Es mejor interpretarlo como un rango plausible de diferencias de tratamiento, en consonancia con los resultados globales del ensayo.

Puntos finales compuestos

Este procedimiento trata de englobar a los pacientes que desarrollan uno o más de los eventos investigados. Por ejemplo en la enfermedad coronaria uno podría considerar las

muertes, infartos no fatales u hospitalización por un síndrome coronario agudo. Ello puede incrementar la precisión estadística y la eficiencia del ensayo, como así también ayudar a la interpretación cuando esta "composición" tiene que ver con el proceso global de la enfermedad. Su uso e interpretación plantea sin embargo algunos desafíos. No es fácil decidir lo que se va a combinar. Si los componentes que integran los puntos finales compuestos presentan una gran variación entre ellos no es conveniente recurrir a esta estrategia. De existir un grado razonable de similitud de los pacientes en cada brazo de tratamiento y en los efectos de los tratamientos el uso de puntos finales compuestos otorga un mayor poder estadístico (en potencia), y consecuentemente eficiencia del ensayo. Este es el caso del análisis "de tiempo para el evento", donde las combinaciones aumentan la tasa de eventos, y por lo tanto se puede aumentar la precisión del resultado. Haciendo una elección cuidadosa de los componentes a combinar, se podría llevar a cabo un estudio con un menor número de participantes y la evidencia de eficacia o no estaría disponible más rápido. La situación no es tan clara sin embargo cuando el evento tiene una distribución binomial. La integridad e interpretación de un ensayo es mucho más clara cuando existe una pre-especificación de los puntos finales primarios y secundarios con que se trabajará. Si el punto final primario es compuesto no se puede alegar como prueba de eficacia a los componentes que lo integran por separado. La presentación de todos los componentes ayuda a la interpretación. Los puntos finales compuestos están siendo favorecidos por las agencias reguladoras a fin de tomar las decisiones si/no al licenciar drogas y equipos. Sin embargo, las instituciones encargadas de analizar la inocuidad, calidad y eficacia de las tecnologías médicas lo ven como insuficiente en muchos sistemas de salud.

El área bajo la curva

Además de los conceptos vertidos al referirnos a los estudios diagnósticos, el estudio del ABC abarca asimismo otros aspectos. En muchas ocasiones la evaluación de los pacientes sometidos a distintos esquemas terapéuticos, impone la realización de controles periódicos de laboratorio a fin de determinar las variaciones de algunos marcadores sanguíneos vinculados al curso de la enfermedad, por ejemplo transaminasas en cuadros hepáticos. Concluido el estudio, disponemos de un buen número de determinaciones para cada paciente, tomadas en distintos momentos del mismo. Las maneras de analizar estas mediciones seriadas son variadas, cada uno con sus bondades y limitaciones. Si se trata de 2 grupos de pacientes, a priori, uno estaría tentado de calcular la *t* de Student, para cada momento en que se tomaron las muestras. Lamentablemente, lo más sencillo no siempre es posible ya que este tipo de enfoque tiene una serie de desventajas que lo invalidan (no contempla la variación global de los datos y múltiples valores de *p* no independientes, entre otros). A raíz de ello se debe recurrir a un tipo de estrategia denominada "mediciones sumarias o resumidas". Dentro de los cálculos que se pueden utilizar para tal fin se incluyen, la media de todas las evaluaciones (sin tener en cuenta el momento en que se efectuaron), el valor pico o el tiempo que demanda alcanzar dicho nivel, el tiempo requerido para que el parámetro se ubique por encima o por debajo de una determinada concentración, el tiempo que transcurre hasta que los valores descienden a los niveles basales y el ABC, entre los más importantes.

El ABC puede reflejar en cierta medida, la respuesta acumulativa a la intervención aplicada. Previo a su estimación es aconsejable graficar los datos, como se muestra en la Figura 4.4, puesto que de existir desemejanzas, las mismas se advierten fácilmente. Es un método muy adecuado para resumir la información que brindan las mediciones seriadas y se lo utiliza mucho en farmacología clínica. Se puede calcular el ABC aún

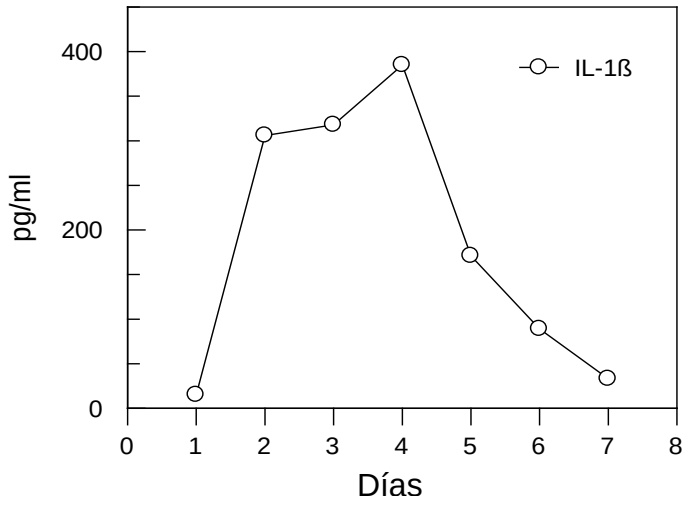
cuando algunos datos intermedios no están disponibles. El ejemplo que se indica en la figura está referido a los niveles plasmáticos de interleucocina 1β (una citocina involucrada en la reacción de fase aguda, en picogramos/ml) en distintos momentos –días- de una infección aguda experimental con un protozoario, más específicamente el *T. cruzi*. La tabla que se agrega al pie de la figura provee los valores obtenidos en cada momento de evaluación. El ABC se obtiene sumando las áreas para cada par de evaluaciones consecutivas. Vale decir con los valores y_1 e y_2 en los tiempos t_1 y t_2 se efectúa el siguiente cálculo $(t_2 - t_1)(y_1 + y_2)/2$. De acuerdo con los datos de la tabla se obtiene lo siguiente

ABC = 1x(15+306)/2 + 1x(306+318)/2 + 1x(318+385)/2 + 1x(385+171)/2 + 1x(171+89)/2 + 1x(89+33)/2

ABC = 1293

El valor que toma el ABC es el producto de las unidades utilizadas para las determinaciones (*y*) y el tiempo (*t*). En nuestro caso pg/ml y días.
Como ya hemos adelantado en el capítulo anterior (ensayos diagnósticos), el ABC es un método que también se puede aplicar cuando uno desea comparar distintas curvas ROC.

Figura 4.4 Mediciones seriadas de interleucocina-1β durante el curso de una infección con *T. cruzi*.



IL-1β	1	2	3	4	5	6	7
	15	306	318	385	171	89	33

4.5 Recomendaciones estadísticas para las distintas etapas del ensayo clínico

A. Previas a la realización del estudio

Componentes del estudio	Ítems a considerar
Grupo a evaluar	Descripción estadística no comparativa
Variables	Nominal, ordinal, continua
Potenciales sesgos	Análisis estratificado
Grupo control	Estadística comparativa (para la similitud)
Randomización	Identificar los potenciales sesgos
Número de personas	Errores tipo I y tipo II, poder del estudio

Ensayos multicéntricos Análisis interinos	Determinar si son combinables los datos Establecer las condiciones, corregir el nivel de significación

B. Durante el estudio

Adherir estrictamente al protocolo. En el caso de que existieran menos pacientes de lo calculado, o se presentaran casos que salen del estudio, tener cuidado que al tratar de compensar la muestra no se incluyan pacientes no representativos de la población blanco.

C. Análisis de los resultados

Describir a la población estudiada para determinar su representatividad	Graficar los datos (distribución de frecuencias), tener en cuenta el desvío y no el error estándar
Valores extremos	Analizar los datos para detectar aquellos fuera de rango ("*outlyers*") y decidir su inclusión u omisión
Sesgo	Análisis estratificado
Resultados negativos (no se puede rechazar la hipótesis nula)	Calcular cuál ha sido el poder del estudio que finalmente se llevó a cabo, estimar los intervalos de confianza (IC)
Eficacia	Calcular los IC
Efectos por estratos, los resultados son diferentes mujeres y varones	Utilizar métodos de ajuste

4.6 Los Sesgos

Quid natum totiens, crudelis tu quoque,
falsis Ludis imaginibus? Cur dextrae
jungere dextram Non datur, ac veras
audire et reddere voces?

¿Por qué a tu hijo –cruel tu también- tantas veces con falsas imágenes burlas?
¿Por qué a mi diestra unir a tu diestra no me es dado, y a voces oír y devolver verdaderas?

Eneida I (405-410)

La relación de los humanos con la deidad no ha sido fácil; en parte porque éstos no eran muy claros en sus mensajes y también debido al modo en que nosotros interrogábamos al oráculo. En esto de escudriñar en torno a los complejos procesos que llevan al desarrollo de una enfermedad, la relación del investigador con "*Matter Natura*" tampoco está exenta de infortunios. Uno de ellos, suerte de gnomos martirizantes, son los sesgos, al que hacemos votos para que nuestro intelecto nos preserve de aquellos errores abrasadores.

Una definición apropiada de sesgo sería una desviación de los resultados o inferencias acerca de la verdad del estudio, o los procesos que llevan a tal apartamiento. En parráfos anteriores se señaló que el sesgo tiene relación directa con el error sistemático el cual reduce la validez de lo que se estima, mientras que el error aleatorio disminuye la precisión de lo que se estima. A diferencia de este último, el sesgo no puede eliminarse o reducirse por un aumento del tamaño muestral.
El sesgo se produce a raíz de deficiencias surgidas en distintos momentos del proceso de investigación, tales como diseño del estudio, realización del estudio, recolección de los datos y análisis de los mismos, interpretación de los resultados, y finalmente publicación y revisión de la literatura. Además de lo explicitado, al referirnos a casos y controles, se describieron aquellos relacionados con los diferentes momentos de ese tipo de diseño.

Sesgos en el diseño
Como su nombre lo indica comprenden todas las situaciones vinculadas a un diseño defectuoso del estudio y su naturaleza es variada. Un punto crítico dentro de este contexto es el **sesgo de selección**. Se trata de una falla en la estimación del efecto debido al modo en que se seleccionaron los sujetos que participan del estudio. Las razones que pueden llevarnos a incurrir en este tipo de sesgo son las siguientes:
Defectos en la confección de la muestra como consecuencia que el contexto de base para la selección de los individuos no representa adecuadamente a la población blanco. Puede desagregarse en varios tipos. Por ejemplo, un *sesgo de valoración* puesto que el observador no efectúa una apropiada ponderación de la severidad de la enfermedad (leve, moderada o severa), o el proceso de diagnóstico de la patología no es el adecuado; o bien el *sesgo de atracción* donde el prestigio de determinados profesionales y/o las instituciones operan como elementos persuasivos para que pacientes con el trastorno específico, o expuestos al problema en cuestión concurran mayoritariamente a tales sitios. También pueden presentarse el *sesgo de estrictez diagnóstica,* vale decir cuando el criterio de selección es muy estricto y se excluyen cualquier tipo de co-morbilidad, que le restan representatividad a la muestra; o el *sesgo de acceso hospitalario* por el cual no se incluyen pacientes no muy comprometidos como para demandar cuidados hospitalarios, o imposibilitados de acudir a dichas instituciones.

Sesgo por muestreo no aleatorio como su nombre lo indica el muestreo se lleva a cabo en forma no aleatoria, y el proceso de selección se ve afectado consciente o involuntariamente por el investigador. Obviamente si nuestro estudio tiene que ver con un ensayo clínico es muy probable que recurramos a un muestreo no probabilístico pero esta es otra situación. Más allá del ejemplo típico que constituye el diseño de casos y controles, hay situaciones que devienen en errores de este tipo. Por ejemplo el *sesgo en el modo de diagnosticar*: el mismo se produce cuando el resultado de una prueba diagnóstica o un cribado afectan la decisión de solicitar el procedimiento de referencia ("*gold standard*") para el problema en cuestión. Supongamos que todos aquellos con un resultado negativo ante una prueba de barrido son excluidos sistemáticamente de la prueba de referencia. Existe asimismo el *sesgo por filtrado.* Esto es a medida que un grupo de pacientes detectados en servicios de atención primaria son referidos a centros de segundo y tercer nivel, aumenta la posibilidad que se vaya produciendo una concentración de casos raros, complicados o severos, o con múltiples patologías.

Sesgo por mala cobertura se presenta cuando algún segmento de la población no puede ser alcanzado o las personas se rehúsan a cooperar. Se puede generar bajo diversas circunstancias.
Sesgo vinculado al momento de inclusión: cuando se trata de un estudio cuyo período de reclutamiento es considerable prestar atención que los incluidos en el inicio no presentan un riesgo de exposición diferente a los incorporados tardíamente con las debidas consecuencias en términos del desenlace.
Sesgo por pérdidas en el seguimiento: está ocasionado por diferencias en las características de las personas que siguen integrando una cohorte y aquellos que han abandonado el estudio.
Sesgo entre los voluntarios y no respondedores: se origina en función de diferencias en las características de las personas que aceptan participar de un estudio respecto de aquellas que no acceden.
Sesgo por abandono: causado por desigualdades en el perfil de los sujetos que deciden retirarse del estudio respecto de quienes continúan dentro de la investigación.
Sesgo por no comparatividad: este sesgo se presenta si los grupos conformados no son comparables. Un caso es el *sesgo por tiempo de seguimiento*, es decir cuando el seguimiento de dos grupos no se inicia al mismo tiempo, por ejemplo un grupo de personas diagnosticadas precozmente de su enfermedad por haber sido sometidas a un cribado. Otra situación puede deberse al *sesgo por duración*: debido a la selección de un número excesivo de casos que sobreviven mucho tiempo en uno de los grupos. Es una disquisición a tener muy en cuenta en un estudio de casos y controles al momento de decidir si se incluyen casos prevalentes o incidentes. *Sesgo por tamaño muestral*: las muestras demasiado pequeñas pueden no revelar la presencia de un efecto, aún cuando éste exista. Por el contrario las muestras muy grandes suelen evidenciar efectos minúsculos de poca relevancia práctica.

Sesgo durante la realización del estudio
En este ítem se incluyen las fallas respecto al modo en que se aplica la intervención, se pondera la exposición a un determinado predictor o los *sesgos por observancia*. Este último se da en casos con problemas de adhesión al tratamiento, lo cual puede afectar el análisis de eficacia al no observarse las recomendaciones del estudio.

Sesgo en la recolección de los datos
También se los denomina sesgos de información u observación, errores de medición o clasificación. Tienen que ver con deficiencias al ponderar la exposición el predictor o el fenómeno de interés y que resultan en diferencias entre los grupos a comparar. El sesgo en la recolección de los datos puede deberse a (i) instrumentos de medición defectuosos, (ii) fuente de datos errónea, (iii) errores del observador, (iv) errores en los participantes del estudio y (v) errores durante el procesamiento de los datos. Los trataremos por separado.

Sesgos instrumentales
Dentro de la amplia gama de mediciones incorrectas, se incluyen aquellas vinculadas a problemas en análisis laboratoriales (calibración de los aparatos, instrumentos inadecuados, técnicas con una puesta a punto deficiente), o bien porque la técnica no tiene la suficiente sensibilidad como para detectar cambios menores. Ligado con esto puede darse un *sesgo en la definición de casos*: sea por los criterios que se toman en cuenta, duración de la patología, estadíos de la enfermedad, o terminología cambiante según criterios en boga.

Para el caso del empleo de cuestionarios también pueden ocurrir una serie de sesgos tales como el de *respuestas forzadas*: el caso típico en que la respuesta es Si o No, y no se contemplan formas intermedias; *sesgo en el armado*: vale decir la manera en la cual se van presentando las posibles respuestas a la preguntas; o *sesgo por preguntas sobre la privacidad* de las personas (ingreso salarial, orientación sexual, estado civil) o incomodan tienen más chance de recibir respuestas equívocas, o ser evadidas o aún rechazadas por el participante.

Sesgo en la fuente de datos

Se genera cuando los tipos o la fuente de datos plantean problemas de adecuación y exactitud. Habitualmente el deceso por enfermedades neoplásicas tiende a prevalecer en personas mayores, mientras que las debidas a patología infecciosa predomina en personas más jóvenes. En función de ello, el trabajo en áreas donde las enfermedades infecciosas son prevalentes, tiende a originar una subestimación de lo que ocurriría en relación con el cáncer debido a esta situación. Otro sesgo es por historia familiar, un dato de historia familiar positiva no es un indicador certero del grado de agregación familiar de una enfermedad y la influencia de los factores genéticos, dado que ello puede tener que ver con el número de parientes y la edad de los mismos. En investigaciones epidemiológicas el número de altas hospitalarias no refleja el número de admisiones dado que ello está afectado por la duración de la estadía, y por lo tanto no provee información adecuada para la enfermedad en estudio.

Sesgo del observador

Ello se debe a diferencias entre observadores (variación inter-observador) o a discrepancias en las lecturas efectuadas por el mismo observador en distintas ocasiones (variación intra-observador). En relación con este ítem se pueden mencionar: *sesgo por sospecha diagnóstica*: tiene que ver con el conocimiento previo sobre el tipo de predictor al que esta expuesto el sujeto lo cual puede influenciar tanto el modo como el resultado del diagnóstico; *sesgo por sospecha de exposición*: en este caso se está en conocimiento de la enfermedad del sujeto y ello puede influir sobre el modo en que se buscará el grado de exposición a la supuesta causa del padecimiento; *sesgo por la administración de tratamiento*: cuando la persona que administra el tratamiento no está cegada, el conocimiento acerca de la eficacia del producto podría influir tanto sobre el desenlace de la enfermedad como la forma en que se mide la evolución de la misma.

Sesgo en el observado está referido a la inexactitud de los datos provistos por el participante al momento de obtención de los mismos. Dentro del ítem se incluyen: *sesgo por aprensión o atención*: ciertas mediciones como la frecuencia cardíaca o la presión arterial pueden variar significativamente de sus valores habituales si el sujeto es aprensivo, o cambios automáticos en su comportamiento al saber que está siendo observado; *sesgo cultural*: las respuestas que brinda el participante pueden variar en función de las diferencias culturales, muchos individuos rehúsan referir públicamente problemas sobre diferencias matrimoniales y violencia familiar, entre otros; *sesgo por información familiar*: la historia familiar u otro tipo de información por el estilo puede variar sustancialmente si el miembro de la familia que provee la información es caso o control, como bien decía Sábato el hombre piensa contra lo que siente; *sesgo por sitio*: la respuesta del entrevistado puede variar de acuerdo al sitio en el cual se efectúa la entrevista (hogar, hospital, o lugar de trabajo); *sesgo por el informante próximo*: cuando se indagan acerca de personas fallecidas o que a pesar de haber sobrevivido tienen comprometida su capacidad de recordar detalles, la solicitud de información a las

personas en proximidad (esposa o miembro de la familia) puede llevar a una falta de coincidencia en los datos recabados; *sesgo de evocación*: esta ocasionado por diferencias en la fidelidad para recordar eventos pasados, por ejemplo madres cuyos hijos padecen una enfermedad en las primeras etapas de la vida son más proclives a recordar si estuvieron expuestas a alguna noxa durante el embarazo. Un caso también tiende a recordar o recabar más detalles de predictores que podrían haber influido sobre su enfermedad, no así los controles; *sesgo al relatar*: supresión selectiva de información sobre historias ocurridas o el haber presentado una enfermedad de transmisión sexual, suicidios, enfermedades mentales, alcoholismo, consumos de estupefacientes.

Sesgo en el procesamiento de los datos

Están relacionados al modo en que se ingresan los datos, sea por errores de códigos, en la entrada de la información a la base, o problemas en la transferencia a partir de bases pre-existentes. Otro problema que suele presentarse es al momento de efectuarse la unificación o creación de un archivo donde se fusiona información proveniente de distintos registros.

Sesgo durante el análisis

En esta instancia de la investigación los sesgos se originan por (i) a una falta de control para los FC, (ii) estrategias inapropiadas, y (iii) análisis *post hoc* de los resultados. El sesgo por FC ocurre cuando el efecto de un predictor está distorsionado por la existencia de un FC con efecto sobre el fenómeno de interés. No debe ser confundido con los eventos relacionados a la cadena causal de la enfermedad. Para citar algunos ejemplos podemos mencionar la existencia de exposiciones a múltiples predictores, muestra donde los participantes están agrupados por estratos y grupos heterogéneos de pacientes con diferentes proporciones respecto a la severidad o duración de la enfermedad.

Estrategias inapropiadas dentro de las mismas se incluyen,

Suposición errónea sobre la distribución de los datos: es un desliz que puede presentarse frecuentemente, por ejemplo las variables relacionadas con el tiempo y los porcentajes, se apartan de la distribución normal por lo que exigen otro tratamiento al momento de las comparaciones.
Sesgo del estimador: tiene que ver con la diferencia entre el valor esperado provisto por un estimador de determinado parámetro y el valor real de dicha medida. Un ejemplo típico es el OR el cual brinda una sobreestimación del RR.
Sesgo por datos faltantes: se debe poner especial atención al modo de tratamiento de los datos faltantes, no es lo mismo ausencia de información, que no respuesta.
Sesgo por el manejo de datos fuera de rango: se presenta cuando no descartamos a un valor inusual registrado en una muestra pequeña, o bien por la exclusión de un valor desacostumbrado que sí debería ser incluido. Vale decir la eliminación de datos fuera de rango o resultados desalineados para los cuales no existe justificación desde el punto de vista estadístico.
Sesgo por sobre emparejamiento: el emparejamiento para variables no confusoras que están relacionadas con el predictor pero no con la ocurrencia del evento (enfermedad) puede llevar a una estimación muy conservadora en los estudios de casos y controles pareados.

Sesgo por análisis *Post Hoc*

Un ejemplo típico de esto es el que se produce al buscar cualquier asociación posible entre los datos sin una hipótesis previa. Los resultados son únicamente aceptables como hechos generadores de futuras hipótesis, que en líneas generales conviene tener en el mazo. Otro sesgo es establecer un significado estadístico *post hoc*: vale decir cuando los niveles de decisión (una o dos colas) para los errores tipo I o tipo II se instauran luego de haberse examinado los datos.

Sesgo de interpretación

El sesgo de interpretación está ligado inexorablemente a fallas en cuanto a la inferencia y especulaciones que el investigador efectúa en función de los hechos recabados. Las variedades de sesgos relacionados con esta etapa del trabajo son las siguientes: s*esgo de presunción o conceptual*, se produce por una lógica defectiva o premisas erróneas por parte del investigador, que lo llevan a conclusiones totalmente desacertadas; *sesgo por correlación*, el querer establecer causalidad a partir de una correlación entre dos variables, confunde sea cual fuere el sentido conque se enfoca el problema; *sesgo de generalización o ausencia de validez externa*, la generalización de los resultados de un estudio más allá de las personas que participaron del mismo es un sesgo a todas luces; *sesgo por significado estadístico*, equiparar diferencias estadísticas con significación biológica, clínica o asistencial, puede llevarnos a conclusiones apartadas de lo realmente trascendente; *sesgo por sub-interpretación*, la incapacidad para extraer todas las posibles explicaciones a los datos obtenidos también puede llevar a interpretaciones erróneas.

Sesgo de publicación

Tiene que ver con la predilección a publicar determinados hallazgos, por ejemplo resultados positivos, lo cual puede distorsionar la idea general debido a una sumatoria de situaciones particulares no nedcesariamente representativa. Dentro de esta problemática también debe considerarse: *sesgo por literatura que no disiente*: algunas sociedades científicas o profesionales suelen publicar trabajos o editoriales donde se omiten o se toman muy poco en cuenta las controversias o resultados discordantes; *sesgo por resultados positivos*: los investigadores son más proclives a someter, y los editores a aceptar, resultados positivos; *sesgo por tópicos de moda*: cuando estamos ante una temática que podríamos llamar “la cresta de la ola”, tanto los investigadores como los editores están tentados a publicar resultados al respecto, sin que se tenga en cuenta si los mismos aún son preliminares o inmaduros.

Muy vinculado a estas cuestiones se hallan los **sesgos en la revisión de la literatura**. Un ejemplo típico tiene que ver con la supresión de trabajos que no están publicados en el idioma del país en el cual se lleva a cabo el estudio, sea un trabajo de investigación original, revisión o meta-análisis. Los que vivimos en esta parte del mundo conocemos algo de tal tipo de sesgo. Otro error en el que se puede incurrir es debido a una búsqueda incompleta por limitaciones en el acceso a la bibliografía, falta de destreza “informática” para recabar los estudios publicados, imposibilidad de obtener reportes aún no publicados o publicaciones de acceso bastante restringido.

Dentro de este ítem también se pueden incluir otros tipos de sesgos que denominaríamos menos “ingenuos” por ejemplo citar trabajos que están a favor de lo que el investigador sostiene o los recursos retóricos por el cual el sustento científico es reemplazado por un muy buen manejo de la redacción.

4.7 Criterios para valorar un ensayo clínico aleatorizado

"All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time"
Hill AB. Proc R Soc Med 1965, 58:295-300.

Una escueta mirada en las bases de datos sobre publicaciones científicas da cuenta de la gran variedad de ensayos clínicos que se publican a diario, lo que implica una necesidad cada vez más creciente de contar con pautas que nos permitan valorar los alcances de un estudio intervencionista de esta naturaleza. Dicho en otros términos a qué debemos atenernos para valorar si los resultados de este tipo de investigaciones se pueden trasladar a la práctica diaria.
Los lineamientos esenciales han sido elaborados hace varios años atrás y pueden resumirse en 3 aspectos fundamentales, esto es la confiabilidad, interpretación y aplicabilidad de los resultados provistos por el estudio. El primero de ellos tiene que ver con la fiabilidad y validez de los hallazgos, el segundo analiza el significado clínico/estadístico de los resultados en función del tamaño y la precisión del efecto; mientras que el último se refiere a la validez externa y los efectos adversos con el propósito de establecer cuáles serían los pacientes con más chance de beneficiarse con el tratamiento.
La literatura también provee de una serie de ítems a tener en cuenta cada vez que se desea efectuar un análisis crítico de un ECA (Consolidated Standards of Reporting Trials –CONSORT-, 2001, 2005 y 2010). Aún así se siguen observando deficiencias en el uso de CONSORT respecto a la calidad de reporte de inocuidad en los ECA.
Algunos hallazgos recientemente publicados indican que la confiabilidad de un estudio se la puede ponderar mucho mejor cuando uno dispone del protocolo original. Una evaluación realizada en Dinamarca (Chan y col. 2008) para analizar en qué medida se pre-especificaban los cálculos del tamaño muestral y los métodos de análisis estadístico en ECAs, comparó los protocolos y publicaciones de 70 ECAs aprobados entre 1994 y 1995 por comités ético-científicos. Entre los 62 estudios sólo 11 poseían cálculos del tamaño muestral tanto en el protocolo y la publicación. Respecto a los procedimientos para manejarse con los desvíos al protocolo, dicha información fue consignada en 37 protocolos y 43 publicaciones. La indicación de la prueba estadística para el análisis de los puntos finales primarios se especificó en 39/49 protocolos y en 42/43 publicaciones. Asimismo, se observaron discrepancias entre protocolos y publicaciones respecto al cálculo del tamaño muestral (18/34), datos faltantes (39/49), análisis de puntos finales primarios (25/42), análisis de subgrupos (25/25), y análisis ajustados (23/28).
Hace unos años Charles y col (2009) también señalan las incongruencias detectadas para el cálculo del tamaño muestral. Sobre un estudio de ECAs de superioridad con dos grupos en paralelo y un único punto final, publicados en 2005 y 2006 en 6 revistas médicas de alto impacto, 10/215 artículos seleccionados (5%) no especificaban el cálculo del tamaño muestral y 92 (43%) no brindaban todos los parámetros requeridos. Sólo en 73 ECAs (34%) se brindaban todos los datos requeridos para el cálculo.
Retornemos al tema que nos ocupa.

Confianza en los hallazgos

En líneas generales la evaluación del efecto de un tratamiento se realiza a través de la comparación entre dos grupos, uno con la intervención bajo análisis y el otro con el tratamiento estándar o placebo en algunos casos. La cuestión a esclarecer es si el desenlace/curso de la enfermedad es diferente en función de la terapéutica instituida.
El diseño de un ensayo clínico establece que los grupos a comparar deben ser similares entre sí respecto de factores capaces de influir sobre el curso de la enfermedad. Es indudable que la validez de las comparaciones efectuadas, requiere una similitud en cuanto a las características iniciales de ambos grupos de pacientes, tipo de cuidados administrados durante el estudio y métodos de evaluación a lo largo de su ejecución. Cualquier diferencia en este sentido introducirá un sesgo y la consecuente posibilidad de extraer conclusiones erróneas sobre la investigación. En la Tabla 4.16, se indican los procedimientos que nos ayudan a identificar tales sesgos potenciales.

Tabla 4.16 Sesgos que suelen surgir en la realización de un ECA y estrategias para neutralizarlos

Sesgos	Etapa	Estrategias para corregirlo
De selección	Asignación	Aleatorización
De ejecución	Conducción	Cegado del paciente y quien presta cuidados
De contracción	Seguimiento	Análisis de intención de tratar (ITT)
De detección	Desenlace	Cegado del evaluador del punto final

En primer término es necesario que la asignación de tratamiento se efectúe en forma aleatoria y no sea revelada. La aleatorización favorece que los grupos de pacientes a conformar sean comparables desde el inicio y así reducir el sesgo de selección. Para lograr una aleatorización apropiada se genera una secuencia de tratamientos y posteriormente se la implementa. En la actualidad la secuencia se prepara mediante programas de computación (generador de números aleatorios). Es fundamental que al momento de implementar la secuencia de aleatorización, la misma no sea conocida; ya que en su defecto el médico podría estar inclinado a asignar un determinado paciente a tal o cual tipo de tratamiento con el sesgo correspondiente. Hace unos años, Kunz y col. señalaron que un enmascaramiento inadecuado de la asignación de tratamientos habría ocasionado un incremento del 35–40% respecto de la estimación del efecto de tratamiento (2007). El procedimiento de ocultamiento no reviste mayores dificultades cuando los dos tipos de tratamientos son indistinguibles. De no ser así, el enmascaramiento se puede asegurar a través de una aleatorización centralizada. Una vez que el paciente firma el consentimiento, el investigador principal (IP) se comunica con un personal encargado de indicarle el grupo al que será asignado.
Como se señalara más arriba, si el estudio cuenta con una aleatorización correcta los grupos serán muy comparables al momento de la inclusión, pero no del todo. Si el trabajo muestra las características de los grupos, y no revela diferencias en cuanto a la distribución de factores pronósticos, estaremos más tranquilos.
Producida la incorporación del paciente, tanto los enfermos como quienes administran el tratamiento o evalúan el curso de la enfermedad no deben conocer el tipo de tratamiento administrado. Nobleza obliga que el doble ciego se haga extensible a todo el grupo; y el trabajo debe especificar quiénes estaban cegados a los tratamientos instituidos. Cuando los pacientes y las personas encargadas de brindar los cuidados desconocen el tratamiento que está siendo administrado el sesgo de ejecución se reduce ostensiblemente. Por otra parte, el cegado también nivela las diferencias relacionadas al seguimiento y atención para con el paciente como así también el efecto placebo.

Un sesgo al que se debe prestar especial importancia es el de detección, lo cual requiere que el personal médico encargado de evaluar el resultado de la intervención tampoco conozca el tratamiento. Este tipo de cegado adquiere más relevancia cuando las variables bajo análisis encierran un importante componente subjetivo, por ejemplo el prurito, y ni hablar si la persona que recaba la información está informada del tipo de tratamiento. Los puntos finales duros (por ejemplo infarto de miocardio, fractura de cuello de fémur) son menos vulnerables a este riesgo. Si se trata de dos tratamientos indistinguibles (similitud en cuanto a las características del medicamento, esquema de administración y dosis), el cegado de los pacientes, propiciadores de cuidados y evaluadores del punto final[8], es fácil de lograr. En caso que los tratamientos presenten diferencias se recurre a un procedimiento en el cual los pacientes del grupo A reciben el compuesto A más un placebo del compuesto B, mientras que los pacientes del grupo B reciben compuesto B y un placebo del compuesto A. No obstante, hay estudios donde el cegado no resulta posible (rehabilitación o psicoterapia) o si bien parece factible no dará resultado (taquicardia a raíz de la administración de broncodilatadores). En situaciones de este tipo la evaluación del punto final está reservada a un servicio específico para tal fin, obviamente cegado.

Para proseguir con un orden lógico, la pregunta subsiguiente será cuántos pacientes completaron el seguimiento y qué análisis se empleó en función de los grupos a los cuales fueron asignados los enfermos (ver tabla 4.17). En un ECA de superioridad, el método más vigoroso para prevenir el sesgo de contracción es el análisis de ITT. Como ya fuera señalado, la estrategia de ITT implica que los pacientes se analizan teniendo en cuenta el grupo al cual fueron inicialmente asignados, aún cuando haya existido un cruce hacia el otro brazo de tratamiento, abandono del mismo o falta de seguimiento. Este análisis es de particular importancia dado que los participantes sin cumplimentar el tratamiento al que fueron asignados en líneas generales no tienen los mismos factores pronósticos respecto de aquellos que si lo hicieron; pero en definitiva en la situación real esto suele ocurrir. Existe un método denominado análisis ITT modificado o *quasi* ITT el cual no toma en cuenta aquellos pacientes que a pesar de ser asignados a los dos tipos de tratamientos no recibieron medicación. Dado que los pacientes desconocen el tratamiento a administrar, la baja del estudio no puede ser atribuida a una decepción respecto del grupo que les correspondió. Por otro lado, si la contracción que se produce es baja y está uniformemente distribuida en los grupos, la posibilidad de sesgo debido a esta circunstancia es baja.

Asimismo se cuenta con el análisis por tratamiento recibido, en el cual los pacientes se estudian en función de lo realmente recibido, independientemente de que hayan suspendido la medicación o no completaran el seguimiento. Finalmente también se dispone del análisis por protocolo o por tratamiento, el mismo se restringe a quienes cumplieron con el tratamiento y la evaluación de los puntos finales. Vale la pena destacar que sólo el análisis ITT asegura una preservación del balance en los factores pronósticos surgido de la randomización y por ello debería ser el método de elección. La realización del análisis ITT implica usualmente elegir un método para manejarse con los datos faltantes. Dado que esto puede deberse a varias razones, incluido los efectos adversos relacionados con la terapia, el método empleado para la imputación de los datos debe ser conservador, es decir no favorecer al grupo de tratamiento.

Es necesario prestar atención a aquellos estudios donde se proporcionan datos de eficacia en pacientes que continuaron con el tratamiento bajo prueba una vez concluido el período de seguimiento. Tales resultados deben ser interpretados con cuidado puesto

[8] También se lo puede denominar variable de impacto, fenómeno de interés o desenlace

que pueden suscitarse cambios en la conducción posterior del estudio (cruce de tratamientos, falta de datos de seguimiento); ni el ITT o el análisis por protocolo son técnicas adecuadas para analizar este tipo de información.

Tabla 4.17 Alternativas para el análisis de los resultados de un ECA en razón de las limitaciones que podrían haberse suscitado durante su realización

Circunstancias durante el estudio	Análisis a utilizar
Pacientes asignados a uno u otro tratamiento	ITT
Tras excluir los pacientes que no recibieron tratamiento alguno	ITT modificado
Pacientes que ciertamente recibieron uno u otro tratamiento	Por tratamiento recibido
Pacientes tratados con seguimiento completo	Por protocolo

La cuestión siguiente está referida a cuán representativo es el punto final de lo que se pretende estudiar. Es un punto crítico pues en función de la información arrojada por esta variable estaremos en condiciones de definir si el tratamiento vale la pena o no. Es conveniente que la variable de impacto, para extraer conclusiones válidas en cuanto a la efectividad del tratamiento, sea un evento clínico muy conspicuo, por ejemplo conductas clínicas claves para el paciente, síntomas/signos referidos por el enfermo o una prueba complementaria. La mayoría de las variables de impacto se miden como variables dicotómicas (evento/no evento), continuas (mediciones de laboratorio) o tiempo para la ocurrencia del evento (muerte, aparición de metástasis o complicaciones). En cualquiera de los casos la variable por la que se opta debe poseer relevancia clínica, ser altamente confiable y reproducible, específica del fenómeno bajo estudio y al alcance de todos los pacientes.
Los lectores se tornan escépticos cuando el estudio establece puntos finales no convencionales y aún no consensuados.
La decisión de aplicar un determinado tratamiento a la práctica clínica requiere demostrar que la intervención contemplada en el ECA mejora el desenlace de la enfermedad. En general no es lo mismo un descenso en las cifras de tensión arterial que una reducción en el número de muertes por enfermedad vascular. Los puntos finales de relevancia son los decesos, días de internación y calidad de vida. Asimismo, para que el dato pueda ser recabado, el tiempo de seguimiento debe estar de acuerdo con el tipo de patología. Un seguimiento de 30 días es de poco valor para el caso de una enfermedad crónica.
En ciertos casos la medición del punto final se hace en base a la utilización de variables subrogadas o sustitutas, por ejemplo un marcador biológico o un estudio imagenológico, las cuales si bien suelen ser más fáciles de cuantificar constituyen mediciones indirectas del fenómeno clínico bajo análisis. Por ejemplo un cambio en determinados componentes del perfil lipídico como indicador de la enfermedad aterosclerosa. En casos de este tipo, debe existir una alta seguridad que tales cambios son representaciones fehacientes de las modificaciones en el curso de la enfermedad. En líneas generales, las variables sustitutas hacen que el período de seguimiento sea menor, y en muchas oportunidades sirven para orientarnos respecto a la marcha de un estudio en sus primeras etapas.
Respecto de la confiabilidad y reproducibilidad, una variable de impacto de valía debe ser relevante para el paciente y al mismo tiempo de fácil evaluación. Si la medición no es reproducible tampoco es confiable. Es recomendable que las personas a cargo de las

mediciones sean personal idóneo bien entrenado, preferiblemente dos observadores independientes bajo condiciones estandarizadas. Como ya se comentara el coeficiente kappa es una medición para evaluar reproducibilidad: un valor >0.6 implica un buen acuerdo y uno >0.8 muy buen acuerdo.

A raíz de la relevancia que posee el evento final (o variable de impacto primordial), debemos cerciorarnos que el mismo fue definido a priori, y así establecer el tamaño muestral capaz de detectar la diferencia; en definitiva el poder del estudio. Con ello se evita la hipertrofia de pruebas estadísticas que entrañan el riesgo de hallar resultados falsos positivos y por ende conclusiones erróneas.

El valor de alfa (o significado estadístico) es la chance de cometer el error tipo 1 y con ello declarar incorrectamente que la diferencia existe. Habitualmente se lo fija al 0.05, vale decir que no más de 1 en 20 pruebas estadísticas mostrarán un efecto cuando en realidad es inexistente. Al efectuarse muchas comparaciones estadísticas esta posibilidad se incrementa y podemos encontrar una diferencia significativa simplemente por chance. En ese caso concluiríamos incorrectamente que la diferencia es real. Para no caer en ello, es bueno que exista una limitación en el número de variables de impacto y las ocasiones en que serán analizadas. Si se pretende efectuar comparaciones de la variable de impacto en distintas oportunidades también habrá que definir previamente cuál es el momento de mayor interés. Para el caso que se hayan efectuado análisis interinos prestar atención a si los mismos estaban previstos a priori y la aplicación de métodos estadísticos "*ad-hoc*".

Hay estudios en los cuales se incluyen variables de impacto secundarias. También ellas deben estar pre-especificadas, y no mostrar únicamente las que arrojan diferencias significativas entre grupos.

Otro riesgo de incurrir en resultados falsos positivos es al efectuar un análisis de la misma variable en pacientes re-agrupados según distintas características de los mismos. A raíz de ello, el análisis por subgrupos no goza de mayor credibilidad (Rothwell, Lancet 2005). De toparnos con un trabajo que incluye análisis por subgrupos es recomendable verificar si el mismo fue planificado previamente y está reservado a un número reducido de cuestiones de relevancia clínica. Asimismo, el poder del estudio debe ser lo suficientemente capaz como para detectar la eventual existencia de interacciones entre efecto y subgrupo de pacientes. Si la interacción también se corrobora en otros estudios, la validez del hallazgo va adquiriendo consistencia. Aún cuando pueden mencionarse todos los análisis por subgrupos efectuados, se requiere mucha precaución al indicar los alcances del efecto del tratamiento por subgrupos puesto que se incrementa la proporción de resultados tanto falsos negativos como falsos positivos.

Qué dicen los resultados

El modo de informar los resultados del estudio depende de las características de la variable de impacto. Cuando se trata de una variable dicotómica generalmente se indica la proporción de personas que experimentaron el evento; mientras que para el caso de variables continuas se presenta la media ± DS de las mediciones practicadas. A pesar de ello, una apreciación más acabada del efecto del tratamiento requiere de un par de datos esenciales: el contraste entre los grupos, conocido como la estimación del efecto del tratamiento y la precisión de la estimación; la cual se obtiene a través del intervalo de confianza (IC) y/o el valor de *p*.

Para una variable dicotómica, la estimación de dicho efecto se puede obtener a través del RR, o diferencia de riesgo (reducción de riesgo absoluto –RRA-). Para datos referidos a la supervivencia el "*hazard ratio*" (cociente de riesgo) o la diferencia en el

tiempo de supervivencia media son adecuados. Cuando se trata de variables continuas se emplea la diferencia entre las medias. En todos los casos el valor de *p* no se tiene en cuenta para estimar el efecto de tratamiento.
En algunos cálculos se provee el coeficiente de disparidad (OR) en lugar del RR. El OR es la razón de eventos-no eventos en el grupo con la intervención respecto de la razón de eventos-no eventos del grupo control. Cuando la variable de impacto es rara el OR puede ser admitido como RR.
Para el caso que la variable de impacto es continua, la medida del efecto del tratamiento se logra a través de la diferencia en las medias del grupo tratado vs. el grupo control. El DS refleja la dispersión de los valores en torno a la media.
Los datos provistos por variables continuas son más trabajosos para la práctica clínica dado los planteamientos suscitados al momento de valorar la importancia clínica de tales resultados o en la relación riesgo/beneficio según las distintas modalidades terapéuticas. Al dividir la diferencia de medias por el DS se logra una comparación un tanto mejor. Dicha relación se la denomina tamaño del efecto (TE).

$$TE = (Media_{tratamiento} - Media_{control}) / DS_{global}$$

Si bien no existe consenso unánime, en líneas generales un valor de 0.2 indica un efecto reducido, 0.5 efecto mediano y 0.8 un efecto significativo.
Cuando resulta posible, algunos estudios convierten la medición continua en una escala dicotómica a fin de calcular el número necesario a tratar (NNT). Más allá de lo atractivo, no debemos dejar de lado que al dicotomizar una variable continua se corre el riesgo de perder potencia estadística y obviamente información respecto al mensaje del estudio.
Otro aspecto igualmente importante es la precisión de lo estimado. En términos reales, la medición exacta del efecto del tratamiento no es algo cognoscible. Lo que está a nuestro alcance es el guarismo surgido de los resultados del ECA, o sea la estimación puntual, que deriva de las observaciones efectuadas en la muestra. Habitualmente se utiliza el IC al 95% a fin de ponderar la "zona" dentro de la cual es factible que se presente el efecto real; es decir el rango que incluye el valor real del efecto en un 95% de las veces. Tener presente que a mayor tamaño muestral, mayor número de eventos estimables, y mayor confiabilidad de que el verdadero valor del efecto esté cerca del observado.
Aunque muchas revistas científicas recomiendan que los resultados muestren el IC y no sólo el valor de *p*, hay trabajos en que no se cuenta con el IC. A la hora de adjudicar peso recordemos que el valor de *p* nos indica cuán a menudo los resultados se podrían haber dado simplemente por chance de no existir la diferencia entre los grupos. En definitiva, el valor de *p* señala el riesgo de concluir erróneamente acerca de una diferencia que en realidad no existe. Al mismo tiempo no perdamos de vista que dicho refleja el significado estadístico de la diferencia pero no su tamaño. Una diferencia pequeña extraída de una muestra con muchos pacientes puede ser mucho más significativa, en términos estadísticos, que la alcanzada en un estudio con menor número de pacientes; pero el hecho biológico sigue siendo el mismo. El valor de *p* nos indica si la diferencia observada tiene posibilidades de ser cierta (< 0.05) o si sólo lo es por chance (> 0.05).
En definitiva podemos concluir que mientras el significado clínico pone énfasis en el tamaño del efecto el significado estadístico apunta a su credibilidad. Un hallazgo clínicamente significativo es beneficioso para el paciente. Diferencias clínicas de poca

monta pueden ser estadísticamente significativas si el tamaño muestral es grande, y a la inversa.

Retomando los conceptos del capítulo 1, el cálculo del tamaño muestral para variables de impacto dicotómicas requiere de 4 componentes: los límites asignados a las posibilidades de cometer el error tipo I (valor de α o significado estadístico), error tipo II (valor de β o poder del estudio), proporción de eventos en el grupo control y un efecto de tratamiento mínimo (digamos la tasa de eventos en el grupo experimental). Si uno no tiene en cuenta estas cuestiones es posible que el ECA no tenga el poder suficiente para detector la diferencia. Aún así, este tipo de estudios puede utilizarse cuando se desea efectuar un meta-análisis, a fin de disponer de toda la evidencia referida al tema.

Resta por comentar que los métodos para el análisis estadístico de una investigación requieren, en la mayoría de los casos, que los datos sean independientes. Si se trata de un ECA esta independencia exige que los datos a comparar provengan de una única observación por participante. Incluir varias observaciones de una misma persona como datos independientes es un error, cuanto menos muy grave. Cada participante cuenta una sola vez y para los casos en que nos confrontamos con mediciones sucesivas se debe recurrir a pruebas estadísticas diseñadas para muestras relacionadas.

Valor del ECA en cuanto a su aplicabilidad en el mundo real

Las cuestiones a las que nos hemos referido en los párrafos precedentes están relacionadas con la validez interna de un estudio. Si los datos mostrados por un ECA fueron obtenidos en condiciones donde las fuentes de error sistemático se controlaron adecuadamente, pasaremos a considerar la validez externa de la investigación. Específicamente, en qué medida la evidencia recogida del estudio puede trasladarse al cuidado de los pacientes a cargo del médico. Para posicionarnos correctamente en el tema, tengamos en claro de antemano que la validez externa *per se* no existe. En modo alguno, los resultados de un ECA pueden ser de relevancia para todos los pacientes y en cualquier contexto. No obstante, seguramente existirán situaciones donde los resultados son aplicables a una franja de pacientes, habida cuenta de las características del estudio.

El contexto clínico habitual, difícilmente se equipara al de un ECA, puesto que el paciente a menudo presenta características diferentes de aquellos reclutados en la investigación. La pregunta capital es si dichas diferencias son los suficientemente sustanciales como para afectar la respuesta al tratamiento y/o aumentar el riesgo de efectos adversos. El ejercicio es determinar hasta qué punto el paciente se aparta de los participantes del estudio y en función de ello barajar si debe aplicar o no los resultados del ECA.

Para allanar el camino, haremos mención a una serie de ítems que merecen ser tenidos en cuenta en la toma de esta decisión.

Contexto en el cual se llevó a cabo el estudio. En este punto podemos considerar la región geográfica de la cual provenían los pacientes y sus sistemas de salud; sitios de reclutamiento, por ejemplo centros de atención primaria, segundo o tercer nivel; como así también las pautas para selección de los centros y médicos tratantes en función de la experiencia y habilidades.

Selección de los pacientes. Aquí se tiene en cuenta la elegibilidad y criterios de exclusión de los pacientes; período durante el cual estuvieron sin tratamiento y relación entre pacientes aleatorizados/pacientes elegibles pero no aleatorizados.

Características de los pacientes aleatorizados. Es conveniente contar con datos sobre la historia natural de la enfermedad y severidad del cuadro, presencia de co-morbilidades, grupo racial y perfil clínico al momento de la inclusión.

Otro aspecto a considerar son las discrepancias entre el tratamiento contemplado en el protocolo y lo que se lleva a cabo en la práctica habitual, fundamentalmente la eficacia del tratamiento control. Asimismo hay que valorar si existieron intervenciones en simultáneo o prohibición de ciertos tratamientos que no correspondían al ensayo. Respecto a la variable de impacto y seguimiento volvemos a hacer hincapié sobre su relevancia clínica, la existencia de variables sustitutas o puntos finales compuestos; como así también la frecuencia y duración del seguimiento.
El manuscrito debe informar acabadamente sobre los efectos adversos de la intervención y la proporción de abandonos de tratamiento. En este aspecto es necesario insistir en efectuar un análisis exhaustivo en cuanto a las posibilidades que la intervención pueda entrañar algún daño para el paciente, el sacro principio de la no maleficencia. A menudo la comunicación de los efectos adversos recibe menos atención que la eficacia del tratamiento y puede ser inadecuada. Años atrás se agregó un anexo a la declaración de CONSORT en la cual se recomienda una comunicación más eficiente en cuanto a los efectos adversos ocurridos durante un ECA.
El cálculo del número necesario para dañar (NND) y su posterior comparación con el NNT permite sopesar los beneficios y riesgos de la intervención como para hacer una evaluación apropiada de sus bondades. El NND es el número promedio de sujetos receptores del tratamiento por el cual se produciría un evento adverso adicional respecto de aquellos que se les administra el tratamiento estándar. El cálculo de NND es como sigue:

NND = 1/(proporción de eventos adversos en el grupo de tratamiento/ proporción de eventos adversos en el grupo control).

La valoración de los riesgos de un tratamiento exige un análisis riguroso sobre la frecuencia de efectos adversos presentados en cada brazo de intervención. Lamentablemente, los estudios no tienen poder suficiente como para detectar una diferencia entre grupos estadísticamente significativa en función de la periodicidad con que usualmente se presentan las reacciones indeseables. En términos de inocuidad, los tamaños muestrales son generalmente reducidos como para que la chance de detección un efecto adverso no esperado sea aceptable. A raíz de ello, algunos autores plantean la confección de una canasta de variables de impacto compuesta, a fin de ser utilizadas como punto final primario de inocuidad en los ensayos de efectividad clínica (Tugwell y col. J Clin Epidemiol 2005).

Conflicto de intereses
Un aspecto igualmente importante es la fuente de financiación. Un trabajo financiado por una compañía farmacéutica podría contener mayor subjetividad respecto de un estudio subvencionado por fuentes oficiales.

4.8 Algunos conceptos sobre meta-análisis (MA)

Es una estrategia basada en el análisis estadístico combinado o integral de los resultados de varios ensayos clínicos independientes, que el investigador considera combinables. Sería una suerte de epidemiología de los resultados donde la unidad de análisis ya no es el participante de una investigación clínica sino el estudio en sí. El término fue propuesto por el Psicólogo Glass en 1976 y luego se incorporó a la investigación médica mayormente en el campo de los ensayos clínicos en cardiología y oncología, aunque también existen MA de estudios observacionales.

Con el objeto de promover este tipo de procedimiento, en 1992 se crea un Centro en la ciudad de Oxford que adopta el nombre de Cochrane, en homenaje a uno de los pioneros en el análisis de las intervenciones en el campo de la medicina. Esta institución con redes en distintos continentes e incluso en nuestro país tiene como objetivos estimular el conocimiento sobre los efectos que poseen las intervenciones aplicadas en el campo de la salud.

Existe acuerdo en que el MA es una herramienta de suma importancia para analizar, interpretar y dar a conocer los resultados de un conjunto de ensayos clínicos en torno a un tema determinado. Constituye una forma más objetiva de resolver zonas de incertidumbres donde no se consigue acuerdo sea a partir de los estudios originales, las opiniones de expertos o bien de las revisiones sistemáticas. La pregunta central que el MA intenta contestar es si del análisis de los estudios contemplados hay suficiente peso como para inferir que una determinada intervención modifica la variable de impacto, y siendo así cuál es la magnitud de ese efecto. Los MA que se efectúan con los datos obtenidos a través de los trabajos publicados se denominan MA de la literatura. Lo deseable sería poder trabajar con los datos individuales de los pacientes ya que al contar con esta información se consigue un perfil más acertado del marco en que se desarrollaron los estudios y permite a la vez confeccionar subgrupos de análisis de ser necesario.

Dentro de las variantes del MA, también pueden efectuarse análisis acumulados, vale decir repetir la práctica a medida que van apareciendo nuevos ensayos clínicos (publicados o no), lo cual permite consolidar una tendencia que al principio podría no haber sido tan clara. De ese modo es posible alcanzar un grado de evidencia tal, como para ser aplicada clínicamente sin necesidad de efectuar estudios posteriores. No obstante, y dado que no se trata de un método infalible es bueno ser cuidadoso al momento de efectuar recomendaciones ya que se han dado situaciones donde los estudios aleatorizados en gran escala no corroboraron los beneficios a que apuntaban los MA previos.

En consonancia con lo descrito para los estudios clínicos, la realización de un MA también requiere una tarea que partiendo de la pregunta lleva a confeccionar un protocolo de trabajo, en el que deberán consignarse los objetivos, la metodología y los criterios a utilizarse para la inclusión de datos. Al respecto de ello, es primordial tener en cuenta la calidad de los estudios, cuán combinables son los pacientes, los tipos de tratamientos aplicados, puntos finales y períodos de seguimientos. Del mismo modo en que el error sistemático resta validez a las conclusiones de un ensayo clínico, el MA también puede adolecer de un sesgo no desestimable. No puede desconocerse que ciertos estudios no llegan a publicarse, en parte debido a que el efecto de la intervención no es significativo. Para evitar eventuales distorsiones es fundamental que el MA no sólo incluya los trabajos publicados sino también aquellos que se hallan disponibles en bases de datos como "*The Cochrane Controlled Trials Register*" (www.cochrane.org), de lo contrario estaremos confeccionando un MA de la evidencia publicada. Opiniones muy autorizadas señalan que las revisiones Cochrane son superiores a las que aparecen publicadas en revistas de circulación periódica. Un trabajo muy reciente (Tricco et al. J Clin Epidemiol 2009) avala claramente esta presunción. En el mismo se determina qué factores predicen resultados favorables y conclusiones positivas en las revisiones sistemáticas –RS- a la vez que evalúa el nivel de acuerdo entre los resultados y conclusiones de las RS. Se tomaron 296 RS en inglés indexadas en MEDLINE (Noviembre, 2004). Dos investigadores categorizaron independientemente las características de los trabajos sus resultados y conclusiones. La información fue sometida a un análisis descriptivo y multivariado mientras que el nivel de concordancia se evaluó mediante el coeficiente kappa. En términos generales, el 36.5% de las RS tenían resultados favorables, con estimaciones del 57.7% para las Cochrane y 64.3% en el caso de las No

Cochrane en un MA de punto final primario. Las revisiones No Cochrane con un MA con punto final primario tenían un doble de conclusiones positivas respecto de los Cochrane (p <0.05). El coeficiente de acuerdo entre los resultados y las conclusiones de las RS fue 0.55, más baja para las Cochrane (0.41) respecto de las No Cochrane (0.67). Si esta tendencia se sostiene, la consulta a un registro de ensayos clínicos para equiparar el cuerpo de datos bajo análisis será esencial.

Esto que "*prima facie*" podría sonar como un excesivo prurito no lo es en modo alguno. Una investigación efectuada a principios de los años 90 puso en evidencia que la chance de publicar un trabajo es 3 veces superior cuando los resultados logran significación estadística. Seguramente el desinterés en publicar resultados negativos se debe a políticas tanto de la revista como de la industria farmacéutica. Un trabajo basado en datos correspondientes a 12 revisiones sobre estudios con antidepresivos en 12.564 pacientes compara la parte publicada con los resultados en poder de la FDA, primordialmente puntos finales y tamaño del efecto (Turner y col. New Engl J Med 2008). Sobre 74 estudios, un 31% del material, perteneciente a 3449 participantes, no estaba publicado. Un total de 37 estudios que la FDA consideraba con resultados positivos fueron publicados. Con excepción de 3 estudios, las investigaciones que la FDA agrupaba dentro de los resultados negativos o cuestionables no fueron publicadas (n=22) o bien se publicaron confiriendo un desenlace positivo (n=11). A partir de la evidencia publicada surgía que un 94% de los ensayos efectuados arrojaban resultados positivos, mientras que en la información de la FDA la proporción era del 51%. El trabajo no consigue determinar si la incongruencia se debió a que los trabajos no se enviaron a publicar, a decisiones de las revistas en no publicarlo, o ambos.

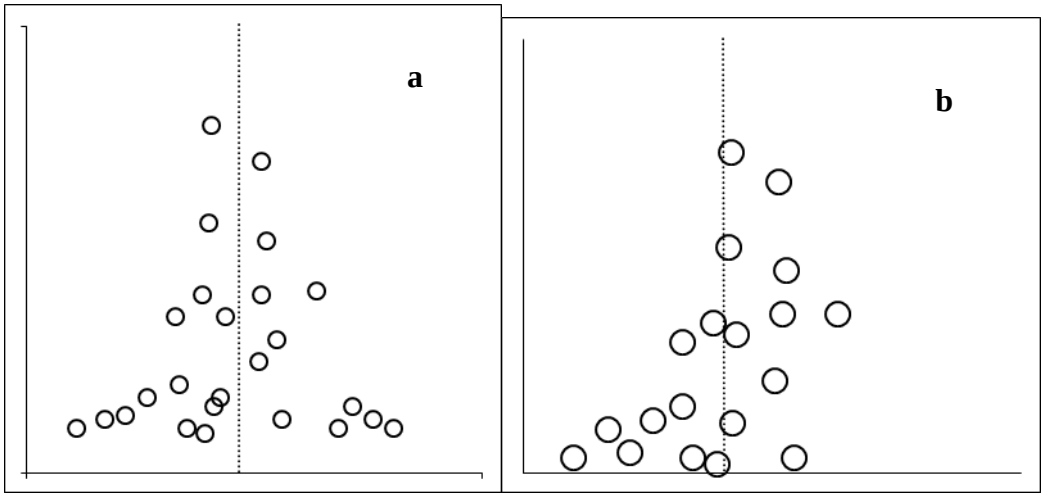
En consonancia con estos resultados un análisis de la literatura (PLoS Medicine 2008) analiza la proporción de ECAs, sobre nuevas drogas aprobadas por la FDA (1998-2000), publicados en revistas biomédicas a las que puede acceder el clínico, pacientes o hacedores de políticas; al momento de su aprobación o entre 2-5 años posteriores a la misma (PubMed y otras bases de datos). Se identificaron 909 ECAs relacionados con 90 drogas aprobadas, de los cuales el 43% (394/909) había sido publicado. En la subclase de ECAs descritos y clasificados como ensayos esenciales para el análisis (FDA-approved drug label), el 76% (257/340) estaba publicado. Mediante un análisis multivariado se observó que la probabilidad de publicación se correlacionaba con resultados estadísticamente significativos (OR 3.03, IC 95% 1.78–5.17); mayor tamaño muestral (OR 1.33 por cada duplicación en el aumento del tamaño, IC 95% 1.17–1.52); y de una calidad muy notable (OR 5.31, IC 95% 3.30–8.55). El hecho que más de la mitad de los ensayos que sustentan la aprobación de fármacos por la FDA siguen sin ser publicados tras 5 años de su aprobación, suscita preocupación.

Dentro del mismo contexto, hace unos años JAMA (2006) publica un trabajo donde se analiza si en los ECAs sobre patología cardiovascular existe asociación entre las fuentes de financiamiento y la probabilidad de comunicar resultados positivos. Se revisaron 324 ensayos de superioridad consecutivos publicados entre 01/012000 y el 30/07/2005, en *JAMA*, *The Lancet*, *New England Journal of Medicine*. En 21 ECAs no se menciona la fuente de financiación. De 104 ensayos subvencionados por organizaciones sin fines de lucro, 51 (49%) presentaron evidencia a favor de nuevos tratamientos mientras que otros 53 (51%) no iban en esa dirección. Entre los 137 ECAs subvencionados por organizaciones con fines de lucro, 92 (67.2%) favorecían nuevas aplicaciones. Respecto

de las que recibieron subvenciones mixtas, más de la mitad de los 62 ECAs ofrecían evidencia en el mismo sentido (n=35, 56.5%). Sobre un total de 205 ECAs que evaluaban fármacos, la proporción a favor de nuevos tratamientos fue 39.5% para las apoyadas por instituciones sin fines de lucro; 54.4%, las co-financiadas; y 65.5%, para las subvencionadas por organizaciones con fines de lucro (p = 0.002).

A la par de las dificultades comentadas, también se presentan los estudios interrumpidos a raíz de los resultados proporcionados por los análisis interinos. Ello plantea el interrogante en cuanto a la conveniencia de incluir estudios no finalizados, aunque en líneas generales no parece ser aconsejable porque introducen un sesgo. En contraposición al argumento de incluir los trabajos inéditos, otros autores sostienen que los mismos no deberían ser incorporados puesto que tales ensayos constituirían una muestra no representativa de los estudios no publicados y que los disponibles, lo son precisamente porque sus resultados concuerdan con lo "preestablecido". No hay dudas que el conocimiento científico es una actividad humana.
Otra fuente de error está dada por la barrera idiomática, ya que la franja mayoritaria de los MA se efectúa en base a trabajos publicados en inglés. Incluso, parece existir un desbalance en los estudios realizados en países de habla no inglesa donde aquellos que arrojan resultados positivos aparecen publicados en revistas internacionales cuyo idioma es obviamente el inglés. El factor de impacto de la revista donde fue publicado el artículo y la frecuencia con que el mismo es citado también aportan sus migajas a la hora de introducir sesgos. Esta suerte de "peso científico" puede asimismo favorecer una cierta subjetividad al momento de establecer los criterios de inclusión que se adoptarán para el MA.
Con el propósito de poder detectar sesgos en la selección de los estudios Egger y col (1997), proveyeron una herramienta muy útil en este sentido. Se trata de una representación gráfica "en cono" (*funnel plot*) en la cual se contrastan los efectos de cada uno de los estudios (eje de x) con su respectivo tamaño muestral (eje de y). Como se ilustra en la figura 4.5a, los resultados de los estudios con muestras pequeñas se esparcen en la parte inferior de la gráfica, mientras que los hallazgos surgidos de los ensayos a gran escala quedan agrupados en la parte superior, dando una imagen en forma de cono invertido. Cuando la silueta se aparta de esta configuración podemos estar en presencia de un sesgo (Figura 4.5b), dado que el vacío en la base del cono podría deberse a la no inclusión de estudios pequeños en los cuales no se observaron efectos beneficiosos. Para mejorar la capacidad de detección de sesgos de publicación y el efecto global de la intervención es conveniente utilizar el logaritmo natural (Ln) de los OR (ver párrafos subsiguientes).

Figura 4.5 Graficación para detectar si existen sesgos en los estudios que se incluirán en el meta-análisis. La línea vertical punteada indica la ausencia de efecto, OR = 1



Los resultados de un MA sobre estudios con un número reducido de pacientes suelen tener menos peso respecto de los ensayos en gran escala, aún cuando el efecto global sea significativo. La comunidad médica presta más atención a los resultados de grandes estudios por razones variadas. En parte porque el MA aún no constituye un recurso bien

incorporado para el ejercicio de la profesión y por otro lado debido a que el ensayo multicéntrico comparativo, controlado, aleatorizado y doble ciego tiene solidez y está más arraigado en el pensamiento médico. La recomendación que los expertos proponen es efectuar una revisión "*in extenso*" de toda la información existente y proceder a un análisis por separado donde se incluyan o no los resultados de los estudios no publicados. Respecto al recabado de los datos, es conveniente trabajar con una hoja de registro consensuada y que la recolección sea efectuada por más de un operador cegado a la fuente de información. Cuando se trata de una variable de impacto binaria, los ensayos clínicos aleatorizados con dos grupos de tratamiento proveen de cuatro valores esenciales el número de pacientes en cada grupo y la frecuencia con que los eventos se registraron en cada uno de los brazos. A partir de esos datos se pueden obtener los OR y tras ello extraer sus logaritmos, los cuales son buenos indicadores de la magnitud del efecto[9]. Supongamos que nuestro MA comprende una serie de 5 ensayos efectuados en distintos sitios. De ese modo obtendremos cinco ORs (OR_1, OR_2, OR_3, OR_4 y OR_5) más una estimación general que nos da el OR_{global} y su intervalo de confianza. El empleo de una escala logarítmica facilita mucho la representación de los resultados, ya que los OR superiores o inferiores a 1 ocuparán un espacio igualitario en la gráfica (un OR de 4 y otro de 0.25 serán equidistantes de 1 y existirá una simetría en los IC (ver Figura 4.6).

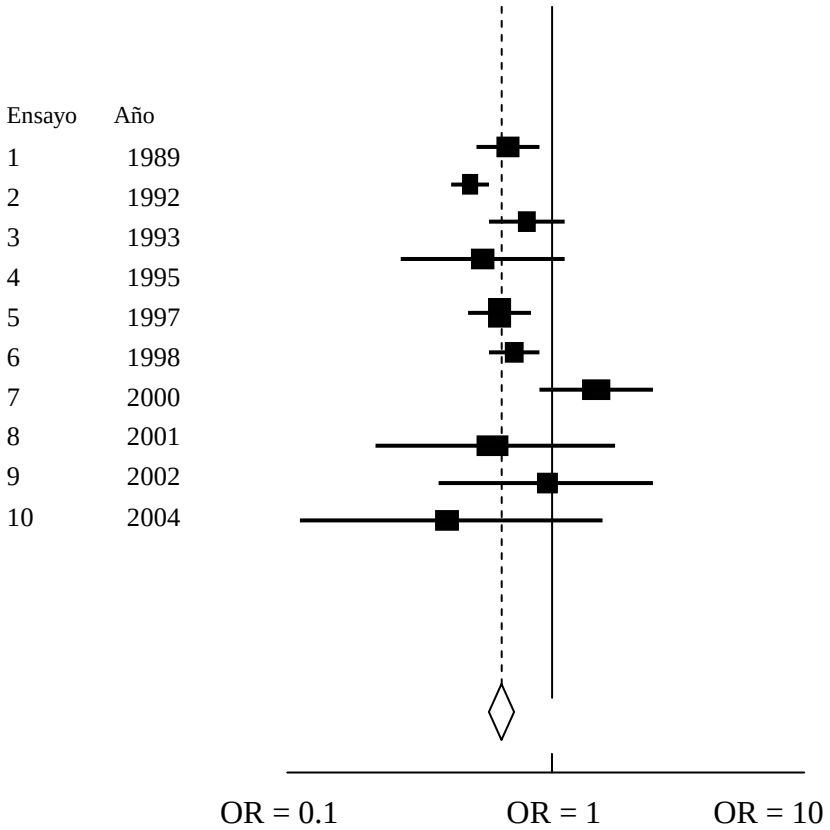


Figura 4.6 Representación de los resultados de un meta-análisis.
Supuesto meta-análisis acerca de 10 ECA sobre corticoides inhalatorios en Asma Bronquial. Cada recuadro con la línea horizontal indica la estimación obtenida del ensayo y el IC 95%, respectivamente. El tamaño del recuadro ilustra el valor de cada estudio. La línea vertical continua corresponde a un OR = 1, ausencia de efecto. El rombo representa el OR combinado (0.76) y la raya vertical punteada atraviesa casi todas las líneas horizontales, excepto dos estudios (adaptado de Egger M, et al. Meta-analysis: principles and procedures. BMJ 1997; 315: 1533-1537).

[9] En algunas investigaciones en lugar del OR se suele utilizar el RR para calcular el tamaño del efecto.

Si el fenómeno de interés es una variable cuantitativa, por ejemplo niveles de colesterol sanguíneo o presión arterial, lo que surge a primera a vista es analizar las diferencias que arroja cada uno de los estudios incluidos en el análisis. El problema es que este tipo de estimaciones está influenciado por las características de cada estudio en particular (tamaño muestral, tipo de pacientes reclutados). Para superar dicha limitación las diferencias se pueden indicar en función del desvío estándar que arroja cada estudio, como para brindar un cálculo ajustado.
Resulta claro que el sesgo es una suerte de pesadilla en investigación, y el MA tampoco escapa a ello. En los párrafos precedentes hemos señalado que en este caso el centro de la atención es el sesgo de publicación, vale decir la chance por la cual los estudios que refieren resultados poco atractivos o desfavorables tienden a ser menos publicados; y la utilidad del *funnel plot* para detectar esta situación. Como ya se explicitara la estimación del efecto obtenida a partir de los estudios pequeños suele ser más variable que la brindada por los estudios más grandes; las primeras se esparcen en la base del grafico creando en presencia de sesgo, una silueta de embudo asimétrica. Sin embargo, el sesgo de publicación puede no ser la única causa de asimetría. Cualquier factor asociado con el tamaño y efecto del estudio podría afectar la asociación real y causar asimetría. Así, de existir un defecto de diseño en los estudios pequeños como para ocasionar a una exageración sistemática del efecto, esto podría traducirse en una asimetría puesto que los estudios ubicados en la parte superior del gráfico (de gran escala y seguramente menos sesgados) ofrecerían una estimación diferente del efecto, respecto de los basales. Hace muy poco se introdujo un nuevo recurso para facilitar la interpretación de este gráfico en el sentido de distinguir la asimetría causada por el sesgo de publicación respecto del causado por otros factores. Los autores lo denominaron el "*funnel plot*" fortalecido por el contorno (*contour-enhanced funnel plot* J Clin Epidemiol 2008). Asumiendo que el efecto de tratamiento de cada estudio se distribuye normalmente, se procede a calcular su error estándar (ES) y el significado estadístico para cada caso. De este modo si los estudios parecen estar faltando en áreas donde no existe significado estadístico la presunción que la asimetría se debe a un sesgo de publicación cobra asidero. Por el contrario, si los supuestos estudios faltantes se ubican en zonas de significado estadístico, la causa de la asimetría se debería a otros factores por ejemplo calidad del estudio.
Recientemente se han desarrollado métodos no paramétricos (*trim & fill*) para investigar el número de estudios faltantes en un MA en base a la asimetría del *funnel plot* y de ese modo brindar un ajuste al efecto calculado (Stat Med 2007).

En continuidad con los procedimientos del MA, el paso siguiente será combinar los datos para el cálculo el efecto general. Obviamente que no se trata de extraer una simple media. Los estudios no tienen el mismo peso y seguramente los resultados de un ensayo de gran escala tendrán una mayor representación en la estimación general. La estadística que se utiliza para este procedimiento depende del modelo que hayamos empleado para la realización del MA, el modelo de los efectos iguales también denominado fijos (MEIoF) y el modelo de los efectos aleatorios (MEA). La diferencia entre los mismos reside en el modo de agrupar la magnitud de los efectos obtenidos de cada estudio para calcular el efecto global. El MEIoF propone que la magnitud del efecto de todos los ensayos incluidos es igual. La variación que pueda observarse se debe simplemente a cuestiones de muestreo o variabilidad dentro de cada estudio. De acuerdo con esto, la magnitud de los efectos individuales puede ser reemplazada por el efecto global. Este modelo goza de bastante beneplácito por cuanto asigna un mayor peso al efecto global y el IC que se obtiene es menos amplio. El MEA, establece por su parte, que la magnitud

del efecto de cada estudio es un valor aleatorio. Considera que los ensayos publicados son una muestra aleatoria de una población de ensayos "conceptuales" que presentan una variación más o menos pronunciada entre los ensayos. Los IC que se obtienen tras la aplicación de este modelo son más amplios puesto que permite asignar una mayor variación entre estudios.

La decisión sobre el modelo a elegir no es una cuestión simple y mucho menos aventurada. Para analizar el grado de "combinabilidad-inconsistencia" entre los estudios, habitualmente se utiliza la prueba de heterogeneidad (*Q* de Cochran), la cual parte de la hipótesis nula que todos los estudios incluidos en el MA analizan el mismo efecto. Dado que esta prueba no es lo suficientemente potente, un grupo de investigadores británicos ha provisto de en método alternativo (no necesariamente excluyente) para la evaluación de inconsistencias denominado I^2. El mismo indica qué proporción del porcentaje total de variación del MA se debe más bien a heterogeneidad que por chance. El cálculo se efectúa a partir de los mismos datos que provee el MA: $I^2 = 100\% \times (Q - \text{gl})/Q$, donde gl son los grados de libertad (número de estudios incluidos en el MA -1). De existir un resultado negativo al mismo se le asigna un valor de 0, con lo cual el I^2 se mueve entre 0% y 100%. Una estimación del 0% indica ausencia de heterogeneidad, mientras que valores del 25%, 50%, y 75%, estarían señalando, en líneas generales, inconsistencias bajas, moderadas y elevadas, respectivamente.

A partir de lo expuesto, cuando la prueba de heterogeneidad no es significativa y la inconsistencia de poca jerarquía, a menudo se aplica el MEIoF. Como fuera expresado, este modelo produce un IC del efecto global más estrecho y por lo tanto más atractivo. Para no dejar de insistir, debe tenerse en cuenta que una prueba de heterogeneidad negativa no implica necesariamente que las magnitudes de los efectos son iguales, puesto que la potencia de la prueba puede no detectar tal circunstancia.

Algunos expertos consideran que se debería proveer de un IC de la heterogeneidad, habida cuenta que los MA no documentan con frecuencia la incertidumbre en la heterogeneidad. Atribuir demasiada confiabilidad a la homogeneidad de los efectos puede ocasionar una falsa percepción. Una ausencia de evidencia de heterogeneidad no es evidencia de homogeneidad. En sentido inverso inclinarnos en demasía por la heterogeneidad de los efectos podría llevarnos a análisis exploratorios por subgrupos en buena medida espurios.

Así como en los ensayos clínicos uno podía establecer la conformación de algunos subgrupos de análisis para determinar si el efecto de la intervención mostraba cierta asociación con estas categorías, en el MA también se puede recurrir a esta posibilidad. Ello implica tomar debidos recaudos, puesto que los datos de las personas que participaron en cada uno de los ensayos no siempre están a nuestro alcance o porque existen diferencias importantes en las características generales de los estudios (enfermos con distintos grados de severidad, variantes del agente farmacológico utilizado, etc.). Es más, si los grupos de pacientes son disímiles en sus niveles de riesgo puede que no sea conveniente efectuar un MA, por más que existan métodos de ajuste para estos factores de confusión.

Los resultados de un MA también pueden plantear dificultades para el médico que en definitiva es el receptor final del mensaje. El dilema surge cuando el resultado de un ensayo clínico en particular se aparta del efecto global y se exacerba si el profesional se desempeña en la zona donde se llevó a cabo uno de los estudios contemplados en el MA, el cual arrojó un efecto global a favor de la intervención, no así en el ensayo de su región. No debe ser tarea fácil para el galeno decidir qué tratamiento asignar a sus pacientes. En la hipótesis que uno deba enfrentarse a una situación de esta naturaleza, es aconsejable tener en cuenta el modelo en que se basó el MA en cuestión. La elección

del MEIoF hace que los efectos individuales sean reemplazados por la estimación global con un menor IC y la consecuente "contracción" del cálculo. Mientras que el MEA es mas conservador.
En paralelo con este tipo de MA, el método también es aplicable a los estudios observacionales puros (cohortes y casos y controles mayoritariamente). Este tipo de estrategia persigue obtener una idea más acabada del papel que juegan los predictores en la causalidad de las enfermedades. En oposición a lo que uno podría esperar, los resultados de este tipo de MA suelen contradecirse con los hallazgos que proveen otros estudios. Quizás ello se deba a la presencia de FC capaces de originar asimetrías de mayor o menor cuantía. Muchas veces los FC son conocidos y en consecuencia es posible proceder al cálculo ajustado. El problema no siempre se supera por cuanto pueden existir FC desconocidos o bien no podemos estimarlo con suficiente precisión dejando espacio para un efecto de confusión residual. Este tipo de dificultades le han restado valor a los alcances del MA sobre estudios observacionales.
No obstante la tentación de proceder a efectuar los cálculos que el "*software*" pone a nuestro alcance, tengamos preparado de antemano un plan concienzudo sobre qué estudios analizaremos, la eventual heterogeneidad y los sesgos más arriba señalados para que la estimación del efecto general sea lo más genuina posible[10].
En los párrafos siguientes, se hace hincapié a las principales ventajas y desventajas del MA.

Fortalezas que se alegan

- Capacidad para resumir los efectos de las intervenciones (con su IC) a través de estudios realizados en un importante número de pacientes.
- Capacidad para poner de relieve el efecto moderador de los tipos de diseños sobre los resultados del MA.
- Capacidad para determinar si el efecto de la intervención es lo suficientemente importante en términos estadísticos y de aplicabilidad.
- Capacidad para reducir las incertidumbres interpretativas, si las mediciones con que se cuentan son los suficientemente válidas.
- Capacidad para lograr una evaluación más objetiva de la evidencia y de ese modo reducir el disenso.
- Capacidad para identificar resultados aparentemente falsos negativos y con ello promover la incorporación de tratamientos efectivos a la práctica clínica.
- Capacidad para reconocer si existen diferencias en cuanto al efecto del tratamiento por subgrupos de pacientes.
- Capacidad para clarificar si existe heterogeneidad entre los resultados del análisis.
- Elemento de ayuda para el cálculo del tamaño muestral en estudios a futuro.
- Capacidad para superar problemas habituales en la revisión de la literatura tales como, los tipos de estudios a examinar, visiones subjetivas en cuanto al peso de un estudio, características particulares de un estudio como responsable de la inconsistencias detectadas o análisis inapropiado de la participación de terceras variables que modifican la relación predictor-fenómeno de interés.

Debilidades que pueden presentarse

[10] También existen modelos Bayesianos de MA que pueden consultarse en la literatura especializada (Carlin 1992; Casella 1985; Gilks et al 1994; Louis 1991).

- Dependiendo de las características del diseño del MA, tipo de evidencia disponible (publicada y no publicada) y tamaño muestral de los estudios, el efecto promedio puede resultar erróneo.
- La falta de reparo en cuanto a las interacciones potenciales según las características de los pacientes, factores intercurrentes, o forma en que se controlaron las variables pueden limitar sus hallazgos.
- La inclusión de datos no publicados en revistas con comité de revisión puede ocasionar una distorsión del efecto.
- Dificultad para trasladar el resultado de un MA al acto médico individual.
- Dificultad para superar la subjetividad que se halla presente al momento de decidir el punto final y el peso que el mismo adoptará en el MA.
- Refutabilidad del tamaño del efecto (efecto final) aún siendo significativo, puesto que pueden existir variables no ponderadas.
- Afectación del grado de representatividad de los estudios cuando se incluyen investigaciones donde la información no es completa.
- Potencial sesgo de publicación puesto que al no disponerse de un registro cabal de todos lo ensayos que se llevan a cabo, los estudios con resultados positivos tienen una mayor chance de ser publicados.
- Posibilidad de enmascarar puntos de heterogeneidad en relación a las características de los participantes, modo en que se aplicaron las intervenciones y condujeron los estudios.
- Limitaciones estadísticas en cuanto a la potencia de las pruebas utilizadas para calcular heterogeneidad en los efectos de los estudios, como identificar diferencias reales entre ensayos. Una ausencia de significado estadístico no significa necesariamente que existe un grado suficiente de homogeneidad como para combinar los datos.

En conclusión, el sustento teórico para efectuar un MA sobre ensayos clínicos aleatorizados se asienta sobre el supuesto por el cual cada estudio brinda una estimación fidedigna del efecto de la intervención y que la diferencia entre los resultados de cada uno de los estudios se debe a una variación aleatoria. Si todo se cumple, el efecto global será un indicador fehaciente y más preciso de la acción terapéutica.
Habida cuenta que se publican 75 ECA y 11 revisiones sistemáticas por día resulta claro que la estrategia para la realización de MA es un área en plena ebullición en la cual los expertos intentan lograr un método cada vez más fidedigno en torno a la ponderación de los efectos provistos por distintos tratamientos. A la par de ello no es menos cierto que para satisfacer las necesidades de los pacientes, clínicos, y hacedores de políticas, no deberían llevarse a cabo ensayos innecesarios, y ocupar el espacio con revisiones sistemáticas valiosas y de libre acceso.

4.9 Indicadores utilizados en la Medicina basada en la evidencia (MBE)

Los propulsores de la MBE la definieron como el uso meticuloso, explicito, y razonable de la mejor evidencia actual al momento de tomar decisiones acerca del cuidado de un paciente. La misma definición implica de alguna manera que la práctica de la MBE es la integración de la experiencia clínica del médico (habilidades y juicio clínico surgido de la experiencia) con la evaluación crítica de la mejor evidencia disponible. Esta última tiene que ver con investigaciones médicas relevantes, de las ciencias básicas, pero fundamentalmente de los estudios clínicos en pacientes acerca de la exactitud y precisión de las pruebas diagnósticas, el poder de los factores pronósticos, y la eficacia e

inocuidad de los tratamientos, como así también pautas de prevención y terapias de rehabilitación.

La MBE parte por así decirlo de la Epidemiología clínica y va ocupando otros campos del sector Salud tales como cirugía basada en la evidencia, enfermería basada en la evidencia y hasta la misma Salud Pública basada en la evidencia donde el foco de atención ya no es el paciente sino la comunidad.

Con el transcurso del tiempo más y más centros han ido incorporando tales pautas a fin de que los clínicos (que ya tienen bastante con su profesión) puedan aplicarla. Tal estrategia no debe ser entendida como una receta de cocina, cada paciente es una situación particular y no se pueden aplicar "formulaciones" incuestionables. La evidencia nos puede informar pero nunca reemplazar a la experiencia clínica, y es eso lo que finalmente nos permite estimar si la información disponible es aplicable al paciente; sin perder de vista que el paciente tiene su singularidad.

Sin lugar a dudas la MBE es una disciplina joven y como tal irá evolucionando en función de las validaciones que vayan surgiendo. El proceso indudablemente se retroalimentará a medida que los programas de educación médica continua y la formación de pre y post-grado, la vayan adoptando y adaptando en función de las necesidades.

A raíz de lo expresado y concentrándonos en el acto médico, es claro que la consulta con el galeno posee cuanto menos 3 componentes esenciales: el epistémico, donde se analiza la situación y se identifican las potenciales líneas de acción; el pragmático cuáles de ellas son posibles; y la cuestión ética, vale decir si las acciones que hemos escogido son moralmente aceptables o preferidas en función de los valores de la sociedad, el paciente, y el equipo de salud. Así como los dos últimos aspectos son capaces de suscitar cuestiones no fáciles de resolver, la elección de una conducta terapéutica también puede plantear problemas en torno al valor intrínseco de tal o cual medida. Para obrar con criterio, es que recurrimos a la MBE como para guiar las decisiones. Lo ideal sería contar con una suerte de manual-algoritmo "tranquilizador", pero no siempre es así.

Al consultar la literatura uno deberá sopesar los hallazgos que brinda un estudio en particular, en relación con el contexto de las diferentes investigaciones que se confrontan al efectuar un análisis de la evidencia científica, sea a partir de un MA o bien de una revisión sistemática. Por otro lado, establecer una jerarquía de la evidencia puede que no resulte tan simple dado los múltiples componentes que hacen a la calidad de una investigación (diseño, modo en que fue conducido, tamaño de la muestra, relevancia de los aspectos abordados). En un ensayo clínico nadie dudaría del peso que tiene la randomización de los tratamientos. Sin embargo hay otros aspectos muy significativos a considerar tales como la precisión y exactitud de las mediciones efectuadas, cegado de las personas que tuvieron a su cargo esta tarea, pérdidas en el seguimiento de los pacientes, IC de los efectos obtenidos. Ingredientes, todos ellos, no siempre fáciles de valorar cuando se trata de establecer las benditas jerarquías. Para reforzar o debilitar nuestras presunciones también conviene tener en cuenta los hallazgos que proveen las investigaciones mecanísticas o fisiopatológicas, ya que una concordancia entre lo clínico-epidemiológico y clínico-básico fortalecerá nuestras conclusiones.

Por su parte, si el interés del médico está centrado en lograr una idea bastante acabada de los efectos adversos del tratamiento en cuestión, es aconsejable efectuar una revisión de los reportes de casos disponibles en la literatura.

Un aporte muy importante de la medicina basada en la evidencia es otorgar una significación clínica de las diferencias que brinda un determinado tratamiento. En capítulos anteriores hemos mencionado al RR como un indicador de las modificaciones

que sufre el fenómeno de interés en función de la variable predictiva. Llevado al terreno de la terapéutica, la mejor resolución de la enfermedad según el tipo de tratamiento aplicado. Sin embargo, en términos de utilidad clínica esta estimación no es adecuada ya que no especifica la magnitud del riesgo absoluto (RA). Para ello conviene calcular dos estimadores, cuyos modos de obtención se representan en la tabla 4.16, la y el NNT. El primero es la diferencia en el riesgo absoluto entre el grupo experimental y el control. El segundo es el número de pacientes que debemos tratar para prevenir un fracaso terapéutico. Obviamente que al indicar el tratamiento estos guarismos deben ser confrontados respecto de los efectos colaterales de la medicación y la severidad de la enfermedad entre otros factores.

Tabla 4.16 Tabla de contingencia para el cálculo de los estimadores de riesgo

Grupos	Fenómeno de Interés		Total
	Si	No	
Control	a	b	a+b
Experimental	c	d	c+d

Para la obtención de dichos estimadores, en primer término se debe calcular:

- La proporción de eventos en Control (PEC) = riesgo en el grupo control a/(a+b)
- La proporción de eventos en Experimental (PEE) = riesgo en el grupo experimental c/(c+d)

Con estos datos estamos en condiciones de obtener:

- La reducción de riesgo relativo (RRR) = (PEC - PEE)/PEC
- La reducción de riesgo absoluto (RRA) = PEC - PEE

Con lo cual procedemos a estimar el NNT=1/RRA

Para ejemplificar estas cuestiones tomemos el caso del ECA con proteína C recombinante en sepsis (*N Engl J Med 344: 699, 2001*). En dicho estudio se registraron 210 fallecimientos en los 850 pacientes que recibieron este compuesto, mientras que en el grupo placebo el número de muertes fue de 259 entre 840 enfermos. El modo de expresar esta situación es a través del RR: vale decir el cociente entre el riesgo de eventos en los pacientes sometidos al nuevo tratamiento (0.247) respecto de los pacientes en el grupo control (0.308). Así obtenemos un RR de 0.8019, el cual indica que el tratamiento reduce la mortalidad en un 19.81%.
Sin embargo es la RRA lo que se considera como una mejor estimación del efecto de tratamiento dado que refleja el beneficio absoluto esperado, teniendo en cuenta el riesgo basal del paciente. El valor de la RRA (6.1%) nos permite extraer el NNT, que en nuestro ejemplo es igual a 16.
También se puede estimar el IC de la RRA y luego se calcula el valor recíproco. Para el caso de cálculos más sofisticados el lector puede recurrir a la bibliografía incluida al final del capítulo.
Hay circunstancias en que los resultados se prestan a interpretaciones diversas. Supongamos que hemos obtenido dos estimaciones de RRA una de 0.1 (IC 95% 0.05 a 02) y otra de 0.5 (IC 95% -0.01 a 0.8). Sólo el primero es estadísticamente significativo (porque excluye el 0). No obstante es probable que el significado clínico sea más

relevante en el segundo caso. La estadística Bayesiana ha ofrecido procedimientos alternativos a este respecto, aunque muchas veces difíciles de llevar a la práctica.
En analogía con el concepto del número de personas que deben ser tratadas a fin de prevenir un fracaso por el tratamiento instituido, también se ha hecho referencia al NND; un método que permite calcular la probabilidad de efectos adversos por el tratamiento instituido (el número de pacientes que debe tratarse para ocasionar un efecto indeseable más).
Los ensayos clínicos llevados a cabo por la industria proveen de hallazgos muy valiosos. No obstante ello, hay que tener en cuenta que no siempre estos estudios enfocan las cuestiones que son relevantes desde el punto de vista de la Salud Pública o de la situación clínica del paciente. El NNT pone un especial énfasis en cuanto se necesita para salvar una vida, pero este valor también pone de manifiesto cuántos son tratados sin experimentar beneficio alguno. A raíz de ello, una publicación aparecida en Lancet propone calcular el número de pacientes tratados innecesariamente (NTI = NNT-1), o bien el índice de impotencia terapéutica = NTI/NNT. De estar expuestos al canto de sirena, estos guarismos pueden preservarnos del aturdimiento.
A pesar de su amplia utilización, el NNT también tiene sus bemoles. En primer término, debemos considerar de qué enfermedad estamos hablando. No es lo mismo prevenir un accidente cerebrovascular que una cefalea. Un mismo NNT es diametralmente distinto según el caso. Asimismo, como el NNT no incorpora los efectos adversos del compuesto, se debe recurrir al NND el cual nos da una idea del riesgo del tratamiento. El NNT tampoco tiene en cuenta explícitamente la duración del tratamiento, por lo que en patologías crónicas es muy importante el momento en el cual se pondera el efecto beneficioso. Otra limitación del NNT es que sólo se restringe a variables dicotómicas. En contraste con las recomendaciones de la MBE acerca de que el NNT debe ser interpretado en el contexto de estas limitaciones y la situación del paciente, existe una suerte de umbral implícito, por el cual un valor entre 10–15 se lo considera atractivo independientemente de las diferencias clínicas, efectos colaterales, y duración de tratamiento. Algunos autores proponen modificaciones al NNT para mejorar la fidelidad de su estimación. Como fuera comentado en la sección de lectura crítica de un trabajo, la división del NNT por el NND también podría ser una alternativa o bien ajustar en función del tiempo, vale decir la duración del estudio.

4.10 Pautas para estimar la calidad de la evidencia

Si bien los ECAs proveen una evidencia más contundente que los estudios observacionales no aleatorizados, cuando no se dispone de los primeros recurriremos a los segundos. La mayoría de los expertos siguen sosteniendo que los efectos de tratamiento surgidos de ECAs están mejor protegidos de sesgos respecto de los no aleatorizados, y tienen más credibilidad en consecuencia. Algunos especialistas cuestionan, sin embargo, esta clara superioridad por cuanto las comparaciones sistemáticas entre ECAs y evidencia observacional (proveniente de estudios de gran calidad) sobre problemas clínicos similares demuestran los mismos efectos de tratamiento. En un estudio reciente se investigó el impacto de la calidad del diseño sobre el efecto de tratamiento según que se tratara de ECAs o estudios no aleatorizados a través de una muestra de 48 pares de ECA o no respecto de un mismo tipo de intervención y punto final (dolor lumbar) en contextos similares. No se observaron correlaciones respecto a la magnitud del efecto terapéutico en los pares de estudio de baja calidad. Se observó una correlación débil pero no significativa cuando por lo menos uno de los estudios del par era de alta calidad. En 20 pares considerados

similares respecto al contexto, pacientes, intervenciones, y resultados existió una correlación moderada, mientras que no se observó relación alguna en 28 pares donde la similitud de las variables clínicas bajo análisis era reducida. De ello se desprende que podemos esperar una buena correlación cuando ambos tipos de estudios son robustos en calidad y exploran aspectos de gran similitud. Las guías establecidas por GRADE (*Grading of Recommendations, Assessment, Development, and Evaluation*) reconocen la primacía del ECA y la necesidad de considerar los estudios observacionales. GRADE establece una serie de limitaciones que pueden reducir la calidad de la evidencia de un ECA (diseño inapropiado, imprecisiones en la realización, inconsistencia, y sesgo de información) y señala a su vez casos en que la alta calidad de los estudios observacionales generan datos confiables respecto a la magnitud del efecto.
Lo que en definitiva pretende el grupo de trabajo GRADE Working Group es determinar el grado de seguridad sobre cuán correcto es el efecto estimado, y el beneficio que acarreará adherir a tal recomendación sin desestimar el perjuicio.
Las recomendaciones efectuadas por el grupo se resumen a continuación:

➢En una primera etapa se debe establecer el proceso a seguir. Ello implica determinar cuáles son los problemas prioritarios, seleccionar el panel de expertos y que cada uno declare si tiene conflictos de interés, para finalmente alcanzar un acuerdo del grupo que desarrollará el trabajo.
➢Las dos actividades que se suceden a esta primer fase son: (1) la revisión sistemática de la mejor evidencia disponible respecto de un fenómeno relevante para su posterior evaluación crítica; (2) la preparación de un perfil de la evidencia para el evento bajo análisis, según cada grupo de riesgo, incluyendo la calidad de la evaluación y un resumen de los hallazgos
➢Tras ello se procede a graduar la calidad de la evidencia de lo cual se desprende el peso de las recomendaciones. Para ello se tienen en cuenta los siguientes aspectos:

- Calidad de la evidencia para cada resultado, la misma se pondera en función del resumen de la información que provee el perfil de la evidencia y criterios tales como diseño y calidad de los estudios, consistencia de los hallazgos, número de pacientes, magnitud del efecto e IC, entre otros.
- Importancia relativa de los resultados. En el perfil de la evidencia sólo se deben incluir los resultados relevantes, esto es aquellos críticos o substanciales.
- Calidad global de la evidencia, vale decir estimar cuán valiosos son los resultados obtenidos en los distintos estudios contemplados.
- Balance entre riesgo y beneficio. En este ítem se establecen cuatro categorías, beneficio evidente, riesgo-beneficio negociable, incertidumbre sobre la negociabilidad entre riesgo y beneficio, o no beneficio.
- Balance entre costo y beneficio. Lo central en este ítem es si el beneficio que confiere la intervención amerita su costo. Es un aspecto muy importante a la hora de efectuar recomendaciones.
- Peso de la recomendación. Las formulaciones al respecto deben basarse en el nivel de confiabilidad de que adherir a determinada pauta ocasionará más beneficios que desventajas.

➢Cumplidos estos requisitos, el paso siguiente es delinear estrategias dirigidas a lograr una implementación efectiva de las nuevas medidas, evaluar su aplicación y mantener al mismo tiempo una actualización de tales pautas.

A nivel del cuidado individual de los pacientes, el médico también puede recurrir a los lineamientos de PICO (del inglés, paciente o población, intervención, comparador o contexto, y punto final –outcome-). Las preguntas a formularnos son las siguientes,

Población

¿El paciente a su cargo es similar al participante del ensayo?

¿Los beneficios y perjuicios del tratamiento son similares a los del estudio?

Intervención

¿En cuánto se asimila el tratamiento que administrará respecto del empleado en el estudio?

¿El tratamiento es accesible?

¿Cuán probable es que el paciente lo cumpla?

¿Ha explorado las preferencias del paciente?

Contexto o tratamientos comparadores

¿El contexto específico del paciente es muy diferente al mostrado por el estudio y puede esa diferencia influir sobre el resultado que se busca?

¿Qué otras alternativas de tratamiento están siendo consideradas?

¿Tiene el paciente complicaciones o co-morbilidades que podrían afectar la probabilidad de beneficios o perjuicios?

¿Existen otros factores pronósticos que no se tuvieron en cuenta en el estudio?

¿Puede que el paciente mejore sin tratamiento?

Punto final

¿Lo contemplado en el estudio es similar a lo que Ud. considera importante para el paciente?

¿Ha establecido que es lo importante para el paciente?

La gradación de la calidad de la evidencia es un proceso dinámico y existen aspectos a lo cuales deberá otorgársele un peso a la hora de establecer pautas tales como plausibilidad biológica, reproducibilidad, generalización, temporalidad, consistencia y coherencia; habida cuenta que la representación de los estudios observacionales es insuficiente. Dentro de esta problemática, Atkins y col (BMJ 2004, 328:1490) publicaron una serie de lineamientos muy atingentes al respecto.

Hace muy poco un grupo multidisciplinario dio a conocer un método que puede ser muy útil para diferenciar cuán pragmático o explicativo es un ECA. El término pragmático está referido a ayudar al medico en opciones de tratamiento, mientras que es explicativo se relaciona con ensayos que analizan una hipótesis (tal intervención ocasiona un beneficio específico). Los primeros determinan los efectos de una intervención en las condiciones usuales que se aplicará, mientras que los explicativos determinan el efecto de la intervención en circunstancias ideales. El diseño de un estudio hace que un ECA sea más o menos pragmático que explicativo. Los autores parten de avances efectuados por expertos en terapéutica clínica, investigación, gasto en salud, especialistas en ECA y publicaciones. Así se estableció la herramienta PRECIS (*pragmatic-explanatory continuum indicator summary*) la cual utiliza los 10 componentes como para determinar cuan pragmático o explicador es el ensayo. Los elementos son: flexibilidad en la comparación de la intervención, flexibilidad de la intervención experimental, experiencia del médico en experimentación y comparación, fuerza del seguimiento, criterios de elegibilidad, análisis primarios, adherencia del médico, cumplimiento del paciente y puntos finales. Cada una de estas variables se la representa en una línea que representa una continua graduada en una escala de 1 al 10 (los rayos de una rueda). En

función de ello se marca cada una de las rectas en el punto correspondiente y luego se unen los mismos. Con ello se obtiene una imagen, la cual da una idea de cuán pragmático (figuras amplias más hacia la llanta) o explicativo (figuras comprimidas, cerca del eje) es el ensayo.

Tanto los ensayos explicativos como los pragmáticos arrojan datos útiles, pero ninguno puede brindarnos una medición exacta del efecto real de tratamiento, dado que el mismo varía según los escenarios por lo que visto desde este ángulo sería una concepción ilusoria.
Propuestas más recientes hablan de la necesidad de llevar a cabo ensayos clínicos del mundo real a fin de lograr mayor seguridad respecto a la validez externa de un estudio. Esta concepción tiene mucho de común con los ensayos pragmáticos pero habrá que ver si son factibles en términos de costo y financiación.

4.11 Medicina traslacional[11]

Es una corriente que en definitiva persigue concretar el omnipresente anhelo de los investigadores biomédicos en ayudar a los pacientes. A pesar del dinero invertido en I&D la provisión de nuevas medicinas viene disminuyendo ostensiblemente en los últimos años. La disciplina intenta llenar el espacio entre los nuevos desarrollos y lo que finalmente llega al paciente. Los experimentos en animales, la probeta y los ensayos en humanos iniciales no reflejan en todo la situación del paciente como para predecir con confiabilidad sobre la eficacia e inocuidad de un compuesto o biomarcador. La interdisciplina apunta a cristalizar esos objetivos.
La palabra está en boga en todas las instituciones de financiamiento científico, organismos reguladores, los mismos investigadores y el personal de salud. La transición entre los estadíos pre-clínicos y clínicos nunca ha sido fácil y la brecha es significativa.
Tanto la academia como la industria han entendido que la interfase es un punto central. Los problemas con los que se enfrenta el progreso de la medicina traslacional son científico, financieros, éticos, regulatorios, legislativos y prácticos. En Europa, los ensayos iniciados por la investigación académica son infrecuentes y con demasiados problemas tanto financieros como regulatorios para su implementación. La industria, que viene realizando ensayos clínicos y preclínicos mucho antes de la aparición de la medicina traslacional, enfrenta el desafío de cómo eslabonar las potencialmente nuevas herramientas en el proceso de descubrimiento/desarrollo. El problema es mucho más de interfase que una falta de acceso a los ensayos clínicos. La relación entre rédito económico, esencial para las compañías, no siempre está en sintonía con las sofisticaciones del compuesto, por lo que la ventana de oportunidades termina estrechándose. La traba para lograr un mayor espectro de posibilidades terapéuticas reside en que la mayoría de las actividades en la cartera de la medicina traslacional son mociones de anhelo que deberán “aprobar exámenes muy rigurosos para su graduación”. Cierto es que todos los compuestos disponibles en el mercado han experimentado un proceso exitoso en su historia.

Cómo desarrollar una ciencia traslacional para la medicina

[11] Para el lector interesado se puede consultar al siguiente sitio
http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm

Al igual que cualquier disciplina el logro de un objetivo exige un método, estrategias sistemáticas y herramientas apropiadas. La mayor dificultad que atraviesa la medicina traslacional es la carencia de una columna vertebral bien firme. En la academia, las iniciativas en este sentido son casi inexistentes, y los estudios fisiopatológicos (cuya relevancia no siempre es tan clara) están más bien fogoneados con un financiamiento cada vez más orientado a lo traslacional y la expectativa obvia para que los resultados vayan en la dirección esperada. En la industria, el logro de una acción traslacional más adecuada debería facilitarse más puesto que se cuenta con métodos o herramientas capaces de mejorar el proceso ya existente.

Si nos trasladamos a cuestiones no farmacológicas, aquí también se requieren de nuevos métodos indicadores como para que el proceso traslacional provea de biomarcadores capaces de predecir la eficacia e inocuidad de un producto surgido de la experimentación básica con miras al campo clínico. Los conceptos de precisión exactitud, reproducibilidad, valores predictivos, accesibilidad y experiencia en humanos, son esenciales en este caso. Otra área que también demanda un desarrollo acorde es la bioestadística. Cómo nos manejaremos en situaciones de multiplicidad de datos, análisis de expresión, "*single nucleotide permutation*" o paneles de marcadores séricos, como así también mejoras en el diseño de los ensayos clínicos iniciales tanto farmacológicos, como diagnósticos o pronósticos. El arte de diseñar un proyecto de medicina traslacional para el desarrollo de un fármaco difícilmente pueda ser asimilado por alguien que no acumule tanto la experiencia pre-clínica como clínica.

Bibliografía recomendada

Generalidades – Planificación y análisis

Abraha I, Montedori A. Modified intention to treat reporting in randomized controlled trials: systematic review. BMJ 2010;340:c2697 doi:10.1136/bmj.c2697

Altman DG, Matthews JNS. Interaction 1: heterogeneity of effects. BMJ 1996; 313: 486.

Altman DG. Practical statistics for medical research. Chapman & Hall. First CRC Press Reprint 1999. Boca Raton FL, EEUU.

Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003; 326: 219.

Altman DG, Bland JM. Standard deviations and standard errors. BMJ 2005; 331: 903.

Altman DG. Why we need confidence intervals. World J Surg 2005; 29: 554-556.

Assman SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355: 1064-1069.

Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995; 310: 170.

Bland JM, Altman DG. Bayesians and frequentists. BMJ 1998; 317: 1151-1160.

Brown RA, Swanson Beck J. Medical statistics on microcomputers. British Medical Journal, Londres, 1990

Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical significance: a simple Bayesian alternative to p values. J Epidemiol Community Health 1998; 52: 318-323.

Colton T. Statistics in Medicine. Boston, Little Brown, 1974.

Epi Info version 6.04. CDC-EEUU y OMS-Suiza, 1996

Epidat. Análisis epidemiológico de datos tabulados. Xunta de Galicia, distribuido por la OPS.

Feise RJ. Do multiple outcome measures require p-vale adjustment? BMC Medical Research Methodology 2002; 2:8.

Fisher RA, Yates F. Statistical tables for biological, agricultural and medical research. 6th edition. Oliver and Boyd, Edimburgo 1974

Fleming KK, Bovaird JA, Mosier MC, Emerson MR, LeVine SM, Marquis JG. Statistical analysis of data from studies on experimental autoimmune encephalomyelitis. J Neuroimmunol 2005; 170: 71-84.

Garcia-Berthou E, Alcaraz C. Incongruence between test statistic and *P* values in medical papers. BMC Medical Research Methodology 2004; 4: 13.

Goddman SN. Toward evidence-based medical statistics. II. The Bayes factor. Ann Intern Med 1999; 130: 1005-1013.

Gray R. Large scale randomised evidence: large, simple trials and overviews of trials. J Clin Epidemiol 1995; 48: 25-40.

Guyatt G, Jaenschke R, Heddle, N, Cook D, Shannon H, Walter S. Basic statistics for clinicians. 1. Hypothesis testing. Can Med Assoc J 1995; 152: 27-32.

Guyatt G, Jaenschke R, Heddle, N, Cook D, Shannon H, Walter S. Basic statistics for clinicians. 2. Interpreting study results: confidence intervals. Can Med Assoc J 1995; 152: 169-173.

Guyatt G, Walter S, Shannon H, Cook D, Jaenschke R, Heddle, N. Basic statistics for clinicians. 4. Correlation and regression. Can Med Assoc J 1995;152:497-504.

Hardy A, Magnello ME. Statistical methods in epidemiology: Karl Pearson, Ronald Ross, Major Greenwood and Austin Bradford Hill, 1900–1945. Soz Präventivmed 2002; 47:80-89.

Higgs E. Medical Statistics, patronage and the state: the development of the MRC statistical unit, 1911-1948. Medical History 2000; 44: 323-340.

Jaenschke R, Guyatt G, Shannon H, Walter S, Cook D, Heddle, N. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. Can Med Assoc J 1995; 152: 351-357.

Kilkenny C, Parsons N, Kadyszewski E, Festing MFV, Cuthill IC, Fry D, Hutton J, Altman DG. Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. PLoS One 2009; 4 (11) e7824

Matthews JNS, Altman DG. Interaction 2: compare effect sizes not P values. BMJ 1196; 313: 808.

Matthews JNS, Altman DG. Interaction 3: How to examine heterogeneity. BMJ 1996; 313: 862.

Miettinen O. Proportion of disease caused or prevented by a given exposure trait or intervention. Am J Epidemiol 1974; 99: 325-332.

Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypothesis. Philos Trans Roy Soc A 1933; 231: 289-337.

Perneger TV. What´s wrong with Bonferroni adjustments. BMJ 1998; 316: 1236-1238.

Pocock SJ. When to stop a clinical trial. BMJ 1992; 305: 235-240.

Pocock S. Clinical Trials. A practical approach. John Wiley & Sons, Londres, 1992.

Pocock SJ. The simplest statistical test: how to check for a difference between treatments. BMJ 2006;332:1256-1258.

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine, Boston Little Brown, 1991.

Siegel S. Estadística no paramétrica. Ed. Trillas, México 1979.

Sokal RR, James Rohlf F. Biometría. Principios y métodos estadísticos en la investigación biológica. H. Blume Ediciones, Madrid 1982.

Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to bayesian methods in health technology assessment. BMJ 1999; 319:508-512.

Sterne JAC, Smith GD. Sifting the evidence - what´s wrong with significance tests? BMJ 2001; 322: 226-231.

Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. BMJ 2010;340:c2073

Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976; 32: 829-849.

Woodeard M. Epidemiology. Study design and data analysis. Chapman & Hall, Londres 1999. Tablas científicas, 6ta edición. Ciba-Geigy, Basilea-Suiza, 1973.

Valier H, Timmermann. Clinical trials and the reorganization of medical research in post-second world war Britain. Medical History 2008; 52: 493-510.

Análisis ajustados – Estrategias Particulares

Besselinka MGH, van der Graafb Y, Gooszen HG. Interim analysis in randomized trials: DAMOCLES' sword? J Clin Epidemiol 2010; 63: 353-354

Bland JM, Altman DC. Survival probabilities. The Kaplan-Meier method. BMJ 1998; 317: 1572.

Bland JM, Altman DC. The logrank test. BMJ 2004; 328: 1073.

Bonifachich E, Chort M, Astigarraga A, Diaz N, Brunet B, Pezzotto SM, Bottasso O. Protective effect of Bacillus Calmette-Guerin (BCG) vaccination in children with extra-pulmonary tuberculosis, but not the pulmonary disease. A case-control study in Rosario, Argentina. Vaccine 2006; 24: 2894-2899.

Cox DR, Oakes D. Analysis of survival data. Chapman & Hall, Londres 1984

Freemantle N, Calvert M, Word J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003; 289: 2554-2559.

Freemantle N, Calvert M. Weighing the pros and cons for composite outcomes in clinical trials. J Clin Epidemiol 2007; 60: 658-659.

Fu P, Laughlin MJ, Zhang H. Comparison of survival times in a transplant study of hematologic disorders. Contemp Clin Trials 2006; 27: 174-182.

Geller NL, Pocock SJ. Interim analysis in randomized clinical trials: ramifications and guidelines for practitioners. Biometrics 1987; 43: 213-223.

Jódar L, Butler J, Carlone G, Dagan R, Frasch C, Goldblatt D, Kayhty H, Klugman K, Plikaytis B, Siber G, Kohberger R, Chang I, Cherian T. Serological criteria for evaluation and licensure of new pneumococcal conjugate vaccine formulations for use in infants. Vaccine 2003; 21:3265–3272.

KleinbaumDG, Kupper LL, Morgenstern H. Epidemiologic research. Van Nostrand Reinhold Company. New York, 1982

Knottnerus JA, Spigt MG. When should an interim analysis be unblinded to the data monitoring committee? J Clin Epidemiol 2010; 63: 350-352

Koch A. How much do we have to pay for how many interim analyses? Contemp Clin Trials 2005; 26: 113–116.

Lagakos SW. The Challenge of Subgroup Analyses - Reporting without Distorting. N Engl J Med 2006; 354: 1667-1669.

McPherson K. Sequential stopping rules in clinical trials. Stat Med 1990; 9: 595-600.

Pagano M, Gauvreau K. Fundamentos de Bioestadística 2da Ed. Thomson Learning. México, 2001.

Piantadosi S. Clinical trials: a methodologic perspective. New York, John Wiley ans Sons, 1997.

Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol 2007; 60: 663-669.

Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analysis. Lancet 2005; 365: 1657-1661.

Schoenbach VJ, Rosamond WD. Understanding the fundamentals of epidemiology, an evolving text. Fall 2000 Edition, NC, EEUU.

Sesgos

Choi BCK. Sensitivity and specificity of a single diagnostic test in the presence of work-up bias. J Clin Epidemiol 1992; 45: 581-586.

Choi BCK. A technique to re-assess epidemiologic evidence in light of the healthy worker effect: the case of fire fighting and heart disease. J Occupational Environm Med 2000; 42: 1021-1034.

Choi BCK, Noseworthy AL. Classification, direction, and prevention of bias in epidemiologic research. Journal of Occupational Medicine 1992; 34: 265-271.

Khoury MJ, Flanders WD. Bias in using family history as a risk factor in case-control studies of disease. Epidemiology 1995; 6: 511-519.

Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299: 926-930.

Sackett DL. Bias in analytic research. 1979. J Chron Dis 32, 51-63.

Valoración de un ensayo clínico aleatorizado

Altman DG, Bland JM. How to randomise BMJ 1999; 319:703-4.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134: 663-94.

Assman SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355:1064-9.

Bennett KJ, Sackett DL, Haynes RB, Neufeld VR. A controlled trial of teaching critical appraisal of the clinical literature to medical students. JAMA 1987; 257:2451-4.

Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337:a2299 doi: 10.1136/bmj.a2299

Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomized controlled trials: review. BMJ 2009;338: b1732 doi:10.1136/bmj.b1732

Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995; 48:167-71.

Evidence-Based Medicine Working Group. Evidence-based medicine: A new approach to teaching the practice of medicine. JAMA 1992; 268:2420-5.

Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1993; 270:2598-601.

Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based MedicineWorking Group. JAMA 1994; 271:59-63.

Greenhalgh T. Assessing the methodological quality of published papers. BMJ 1997; 315:305-8.

Haidich AB, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: Survey of major general medical journals. J Clin Epidemiol 2011; 64: 124-35.

Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999; 319:670-4.

Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D; CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004; 141:781-8.

Kaul S, Diamond GA. Trial and Error. How to Avoid Commonly Encountered Limitations of Published Clinical Trials. J Am Coll Cardiol 2010;55:415-27

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2007, Issue 2. Art No.:MR000012. doi:10.1002/14651858.MR000012.pub2.

Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987-91.

Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply? Lancet 2005; 365:82-93.

Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005; 365:176-86.

Sackett DL, Rosenberg WMC, Gray JAM, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312:71-2.

Schulza KF, Altman DG, Mohercc D, CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010; 63: 834-40

Tugwell P, Judd MG, Fries JF, Singh G, Wells GA. Powering our way to the elusive side effect: a composite outcome 'basket' of predefined designated endpoints in each organ system should be included in all controlled trials. J Clin Epidemiol 2005; 58:785-90.

Meta-análisis y Medicina basada en la evidencia

Altman DG. Confidence intervals for the number needed to treat. BMJ 1998; 317: 1309-1312.

Bagshaw SM, Bellomo R. The need to reform our assessment of evidence from clinical trials: A commentary. Philosophy, Ethics, and Humanities in Medicine 2008, 3:23 doi:10.1186/1747-5341-3-23

Bastian H, Glasziou P, Chalmers I. Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? PLoS Medicine 2010; 7: e1000326.

Bjerre LM, LeLorier J. Expressing the magnitude of adverse effects in case-control studies: "the number of patients needed to be treated for one additional patient to be harmed. BMJ 2000; 320: 503-506.

Bogaty P, Brophy J. Numbers needed to treat (needlessly?). Lancet 2005; 365: 1307-1308.

Borma GF, Lemmersa O, Fransen J, Donders R. The evidence provided by a single trial is less reliable than its statistical analysis suggests. J Clin Epidemiol 2009; 62: 711-715.

Carlin JB. Meta-analysis for 2 x 2 tables: a Bayesian approach. Stat Med 1992; 11: 141-158.

Casella G. An introduction to empirical Bayes data análisis. Am Statistician 1985; 39: 83-87.

Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995; 310: 452-54.

Cucherat M, Boissel JP, Leizorovicz A. EasyMA: a program for the meta-analysis of clinical trials. Computer Methods and Programs in Biomedicine 1997; 53: 187-190.

Davey Smith G, Egger M, Phillips AN. Meta-analysis: beyond the grand mean. BMJ 1997; 315: 1610-1614.

Davey Smith G, Egger M. Meta-analysis: unresolved issues and future developments. BMJ 1998; 316: 221-225.

DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986: 7: 177-188.

Dickersin K, Berlin JA. Meta-analysis: state-of-the-science. Epidemiol Rev 1992; 14: 154-176.

Dickersin K, Scherer R, LeFebvre C. Identifying relevant studies for systematic reviews. BMJ 1994; 309: 1286-1291.

Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315:629-634.

Egger M, Davey Smith G. Meta-analysis: potentials and promise. BMJ 1997; 315: 1371-1374.

Egger M, Davey Smith G, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997; 315: 1533-1537.

Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. BMJ 1998; 316: 61-66.

Egger M, Schneider M, Davey Smith G. Meta-analysis: spurious precision? Meta-analysis of observational studies. BMJ 1998; 316: 140-144.

Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modeling. The Statistician 1994; 43: 169-178.

GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004; 328: 1-8.

Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. JAMA 2000; 284:1290-1296.

Herbert RD, Bø K. Analysis of quality of interventions in systematic reviews. BMJ 2005; 331: 507-509.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-560.

Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965; 58: 295-300.

Ioannidis J, Patsopoulos K, Evangelou E. Uncertainty in heterogeneity estimates in meta-analysis. BMJ 2007; 335: 914-916.

Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998; 280: 278-280.

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054-1060.

Kendrick T, Hegarty K, Glasziou P. Interpreting research findings to guide treatment in practice. BMJ 2008;c338: 43-46.

Knottnerus JA, Tugwell P. Ongoing innovation in meta-analysis. J Clin Epidemiol 2011; 64: 117-8.

Kristiansen IS, Gyrd-Hansen D, Nexoe J, Nielsen JB. Number needed to treat: easily understood and intuitively meaningful? Theoretical considerations and a randomized trial. J Clin Epidemiol 2002;55:888-892.

Kunz R. Randomized trials and observational studies: still mostly similar results, still crucial differences. J Clin Epidemiol 2008; 61: 207-208.

Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Technol Assessment Health Care 1990; 6: 5-30

Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006; 333: 597-600.

Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988; 318: 1728-1733.

Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: A Literature Analysis. PLoS Medicine 2008, 5(9) e191

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326: 1167-1170.

Louis AT. Using empirical Bayes methods in biopharmaceutical research. Stat Med 1991; 10: 811-29.

McAlister FA, Straus SE, Guyatt GH, Haynes RB, Evidence-Based Medicine Working Group. Users' guides to the medical literature: XX. Integrating research evidence with the care of the individual patient. JAMA 2000;283:2829–36.

McGinn TH, Guyatt GH, Wyer PC, Naylo CD, Stiell IG, Richardson WS. User´s guide to the medical literature. XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2000; 284: 79-84.

Moher D,Cook DJ,Eastwood S,Olkin I,Rennie D,Stroup DF.Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement.Lancet 1999;354: 1896-900.

Moher D, Schulz KF, Altman DG.The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 2001;1: 2.

Moreno SG, Sutton AJ, Turner EH, Abrams KR, Cooper NJ, Palmer TM, Ades AE. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ 2009;339: b2981 doi:10.1136/bmj.b2981

NNT Tables [database on the Internet]. Toronto (ON): Centre for Evidence Based Medicine, University Health Network, 2003. Available at: http://www.cebm.utoronto.ca/glossary/nnts.htm#table.

Noble JH. Meta-analysis: methods, strenghts, weaknesses, and political uses. J Lab Clin Med 2006; 147: 7-20.

Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. JAMA 2002; 287: 2813-2814.

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006; 295: 676-680.

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med 2007; 26: 4544-62.

Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008; 61: 991-996.

Relly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. Ann Intern Med 2006; 144: 201-209.

Ridker PM, Torres J. Reported Outcomes in Major Cardiovascular Clinical Trials Funded by For-Profit and Not-for-Profit Organizations: 2000-2005. JAMA 2006; 295:2270-2274

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology-a basic science for clinical medicine. London: Little, Brown, 1991.

Sadler JZ, Hulgus YF. Clinical problem solving and the biopsychosocial model. Am J Psychiatry 1992; 149: 1315-1323.

Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011; 64: 163-71.

Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol 2005; 58: 894-901.

Thorpea KE, Zwarensteinb M, Oxmanc AD, Treweekd S, Furberge CD, Altman DG, Tunis S, Bergelh E, Harveyi I, Magidj DJ, Chalkidou K. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009; 62: 464-475.

Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D. Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. J Clin Epidemiol 2009; 62: 380-386.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252-260.

Vandenbroucke JP, de Craen AJ. Alternative medicine: a “mirror image” for scientific reasoning in conventional medicine. Ann Intern Med 2001; 135: 507-513.

Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trails. Prog Cardiovasc Dis 1985; 17: 335-371.

Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266: 93-98.

Capítulo 5. La preparación del manuscrito

Al escribir estos párrafos viene a nuestra memoria una recomendación que el becario recibe a muy poco de haberse incorporado a un laboratorio de investigaciones: "el trabajo sólo concluye cuando se ha preparado el manuscrito para publicar". Es una instancia en la cual debemos aplicar algunas estrategias de "*marketing*" para presentar los alcances de nuestra investigación a la comunidad científica. Como guía, conviene tener en cuenta las preguntas que habitualmente nos formulamos en nuestro accionar diario, por qué lo hicimos, cómo y con qué lo llevamos a cabo, qué es lo novedoso y cuál es la relevancia de lo observado. Cuando se dispone de la información según esta secuencia se terminan desarrollando las 4 secciones fundamentales del trabajo que se pretende publicar; esto es introducción, material y métodos, resultados y discusión.

5.1 Introducción

Al escribir la introducción, tener presente la necesidad de explicar/convencer al lector sobre el interés que suscita el tema abordado en el estudio. Por ejemplo, si lo sometemos a una revista de alguna especialidad conviene empezar que lo nuestro está cobrando cada vez más importancia en ciertas patologías relacionadas con la disciplina. Es obvio que los grupos abocados a los temas más candentes de la Medicina, por ejemplo cáncer y enfermedades cardiovasculares, no hallarán mayores dificultades para que los lectores adviertan el impacto potencial de sus estudios. Cuando la importancia del tema no es tan visible, generalmente se recurre a datos epidemiológicos que ponen de relieve su trascendencia. Para citar algunos ejemplos, (1) el notable incremento a partir de la última década, en la probabilidad de contraer determinada patología infecciosa, (2) enfermedades en las cuales la aparición de fracasos terapéuticos se ha vuelto muy frecuente, (3) necesidad de identificar los riesgos carcinogenéticos que implican la exposición cada vez más creciente a productos químicos utilizados por la industria o contaminantes medioambientales, (4) afecciones que si bien no condicionan una mayor chance de muerte determinan discapacidades con la consecuente repercusión socioeconómica. En definitiva, se trata de poner en un primer plano los aspectos que le otorgan jerarquía al problema investigado.

A continuación sobreviene lo que en líneas generales se denomina "estado de conocimiento del tema". En este punto el investigador debe explayarse demostrando su amplio manejo de la temática, que necesariamente se sustenta en la relevancia de las referencias bibliográficas incluidas. Conocer el problema en forma pormenorizada permite exponer los puntos no totalmente aclarados, ofreciendo un espacio para las cuestiones vinculadas con el objetivo a enunciar posteriormente. Asimismo, los investigadores con antecedentes en el tema hacen mención del grado de avance que significan sus propias investigaciones, lo cual contribuye a reforzar la imagen del grupo. Tras el desarrollo de estas ideas, resta la exposición de la o las hipótesis y los objetivos del estudio. Para ello uno debe conocer muy bien la biología y fisiopatología del problema, su epidemiología (determinantes, distribución, prevalencia, incidencia, y pronóstico), y limitaciones en el manejo clínico que llevan a resultados no satisfactorios. La afirmación de Ionesco "*no es la respuesta lo que ilumina sino la pregunta"*, también corre para nosotros y para ello hay que conocer muy bien el problema y la frontera entre lo conocido y desconocido. Si la cuestión es compleja pensar en profundidad y establecer qué direcciones habrá que tomar como para transitar en el borde del cráter sin caer en el. Generalmente en la investigación aplicada las preguntas son más fáciles de formular (habida cuenta de la información que provee la investigación básica) pero obtener la respuesta es más difícil. Las preguntas primarias van al inicio de una

investigación y las secundarias inmediatamente detrás y basadas en la hipótesis general del estudio. Estas preguntas adicionales nunca deben opacar el objetivo fundamental del estudio.

A diferencia de los párrafos previos donde el investigador suele explayarse en sus consideraciones, la enunciación de la hipótesis de trabajo y los objetivos debe redactarse del modo más claro y concreto posible. Por otro lado, la introducción del manuscrito necesita estar ajustada a "un hilo conductor", para establecer una secuencia ordenada y lógica de ideas que desemboque indefectiblemente en la propuesta de trabajo. En el área de la investigación clínica, lo habitual es que esta sección tenga una extensión no mayor a dos páginas, preferiblemente una. En consonancia con ciertas recomendaciones del "*marketing*" algunos autores agregan un párrafo final adelantando los principales hallazgos del estudio.

Como fuera comentado en el Capítulo 1, los estudios descriptivos no parten de un presupuesto, sino que sirven para generar futuras hipótesis de trabajo. En estos casos, la justificación del estudio se basa en la necesidad de conocer las características del fenómeno que nos ocupa.

5.2 Material y métodos

En casi todos los trabajos esta sección aparece desglosada en ítems referidos a los participantes, los estudios efectuados y el análisis de la información. Tratándose de investigaciones clínicas es lógico que el espacio destinado a describir la población en quienes se llevó a cabo el estudio merezca una atención especial. El documento debe brindar una información clara y concisa de los criterios de inclusión y exclusión que se utilizaron para incorporar a los candidatos, como así también de los procedimientos diagnósticos que permitieron identificar a los enfermos. Para el caso de los estudios observacionales (en particular de casos y controles) es aconsejable especificar debidamente las normas aplicadas para la conformación del grupo control. Según las características de la investigación puede que sea necesario consignar el período durante el cual se reclutaron los participantes, sobre todo en los casos donde el factor tiempo o las variaciones estacionales pueden ejercer alguna influencia sobre el fenómeno en observación. Por otra parte, es importante consignar la aprobación del protocolo por parte del Comité de Etica de la institución donde se radicó el estudio y que la inclusión de los participantes se efectuó con posterioridad a la obtención del consentimiento escrito informado.

Los párrafos siguientes están referidos a la manera en que se evaluaron las variables predictiva y de impacto. A modo de ejemplo, si el objetivo fue determinar el efecto del tratamiento con interferón alfa recombinante sobre el curso de la hepatitis C, deberá especificarse la dosis, duración y características del interferón empleado como así también de los criterios utilizados para definir los éxitos o fracasos terapéuticos. En general conviene adherir a los estándares internacionalmente aceptados ya que esto facilita la comparación con otros estudios. Cuando ello no es posible, deberán exponerse los fundamentos que llevaron a adoptar otros criterios. Según las características de la revista a la cual se somete el manuscrito, puede que sea necesario efectuar una descripción exhaustiva de los procedimientos empleados. La aplicación de técnicas sofisticadas y de última generación favorece, aunque no siempre, la publicación del trabajo en una revista de mayor prestigio.

Retomando los conceptos vertidos en el Capítulo 1 sobre precisión y exactitud, es aconsejable mencionar todas las medidas que se adoptaron para asegurar la calidad de las mediciones. Si se trata de una prueba de laboratorio suele indicarse que las mediciones para cada una de las muestras se efectuaron por duplicado o triplicado y se

consigna el promedio de las mismas. Algunos trabajos incluso proveen el coeficiente de variación para brindar una idea más acabada de la dispersión de los datos.
La sección de Material y Métodos generalmente concluye con un párrafo dedicado al análisis estadístico de los datos. En este ítem, es aconsejable especificar todos los procedimientos que se aplicaron para lograr un tratamiento estadísticamente adecuado de la información. Entre las cuestiones más salientes cabe indicar (1) si hubo necesidad de transformar alguna variable, (2) qué pruebas se emplearon para las comparaciones de los datos cuali y cuantitativos, (3) cuáles fueron las medidas de asociación entre las variables predictiva y de impacto (índices de disparidad, riesgo relativo), (5) qué técnicas se emplearon para evaluar la potencial influencia de algún FC sobre los resultados de las comparaciones (por ejemplo, regresión logística o regresión múltiple) y (6) a partir de qué valor de *p* se consideró estadísticamente significativa a la diferencia. Este comentario tiene que ver con lo señalado en capítulos anteriores respecto de estudios donde al efectuar muchas comparaciones estadísticas, generalmente el valor límite de *p* no se ubica al 0.05 sino por debajo de ese nivel a efectos de reducir la probabilidad de que alguna diferencia significativa lo sea simplemente por azar.
El análisis estadístico debe ser muy bien detallado si el trabajo se somete a una revista muy específica en el tema.
Tener siempre presente que nuestro trabajo debe proveer suficiente información como para que otro investigador puede reproducir el estudio.

5.3 Resultados

Si tuviéramos que resumir en pocas palabras el aspecto esencial de esta sección, coincidiríamos en que se trata de transmitir un mensaje claro e ilustrativo de los hallazgos obtenidos. A poco de concurrir a los seminarios de resultados, el becario aprende que las tablas y figuras deben ser autocontenidas. A través de los comentarios emitidos por quienes asisten a su presentación, nuestro joven entusiasta consigue entender, y a veces con dolor, que la audiencia se irrita cuando precisamente ellos no entienden. La buena predisposición del revisor de nuestro manuscrito también tiende a esfumarse si la forma en que hemos presentado los resultados es deficiente. Por lo tanto, mucha atención al momento de preparar las mostraciones.
La decisión de optar por tablas o figuras depende de varios factores, las exigencias de la revista, las características de los datos, y la magnitud de la información que se ha recogido, entre los más importantes. Si en las instrucciones para los autores se determina un número máximo de 4 presentaciones (entre tablas y figuras) y se han efectuado numerosas determinaciones de laboratorio, no cabe duda que la tabla es más adecuada para ese *quantum* de información. Aunque no conviene cargar con demasiadas Tablas, que dificultan el seguimiento del trabajo y hacen que el lector pierda interés en verlo

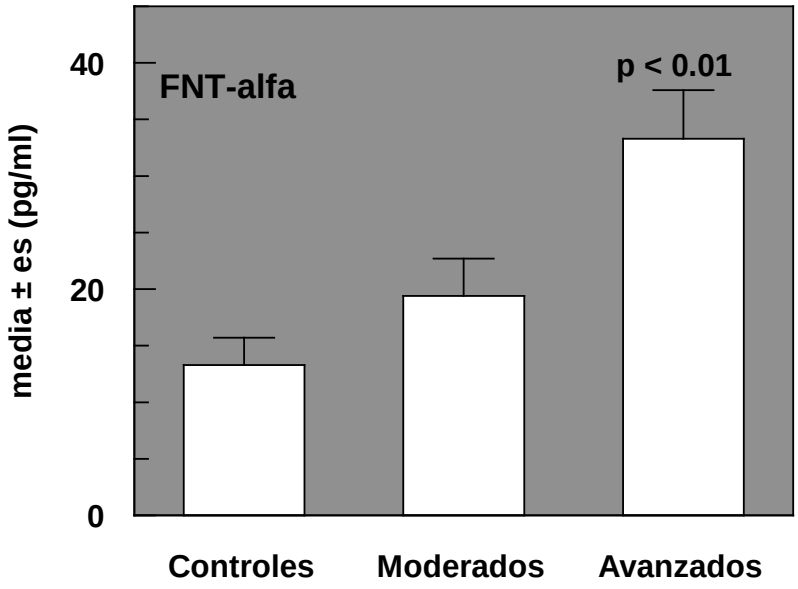
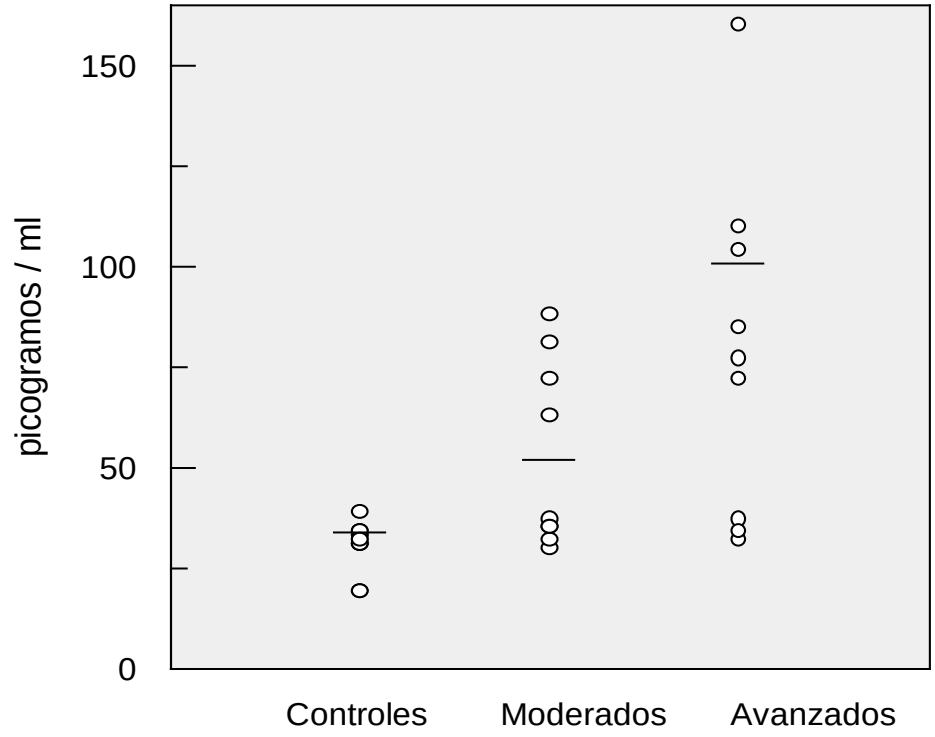


Figura 5.1 Niveles circulantes de factor de necrosis tumoral alfa (FNT-alfa) en enfermos tuberculosos (moderados y avanzados) y controles sanos.

La Figura, por su parte, confiere un efecto visual que es muy adecuado cuando se quiere poner énfasis en un resultado. Cuando se desea marcar los contrastes entre los resultados de un grupo respecto del otro el gráfico de barras es apropiado. A modo de ejemplo, en la gráfica 5.1 se hallan representadas la media más el error estándar de las concentraciones de factor de necrosis tumoral alfa (-FNT-alfa-, columna y el segmento de la recta, respectivamente) presentes en el suero de pacientes con Tuberculosis y controles sanos. La figura muestra con claridad que los niveles de esta citocina son significativamente superiores en los enfermos avanzados. Por su parte el gráfico de torta es útil para resumir los hallazgos especialmente si los valores que toma la variable no son demasiados.
En los casos donde a pesar de registrarse una diferencia estadísticamente significativa la variable muestra una dispersión bastante acentuada, es preferible mostrar los datos individuales. Un ejemplo de esta situación se muestra en la Figura 5.2 donde se representan las concentraciones de FNT-alfa que se midieron en los sobrenadantes de cultivo de las células periféricas mononucleares de enfermos tuberculosos. En este caso la línea que se traza para cada tipo de pacientes representa la media de las mediciones dentro de ese grupo. Las comparaciones estadísticas indican, nuevamente, que los valores de FNT-alfa son significativamente mayores en los enfermos avanzados.

Figura 5.2 Niveles de FNT-alfa en sobrenadantes de cultivo de células mononucleares de sangre periférica obtenidas de pacientes con Tuberculosis pulmonar de diferente severidad (moderados y avanzados) y controles sanos.



En las investigaciones donde los pacientes son estudiados en diferentes oportunidades puede ser útil la representación que se ofrece en la Figura 5.3. En ella se muestran las concentraciones sanguíneas de FNT-alfa en dos grupos de enfermos tuberculosos, los cuales fueron evaluados, previamente a la iniciación y tras 30 días de tratamiento con dos esquemas distintos de drogas antituberculosas (A y B).

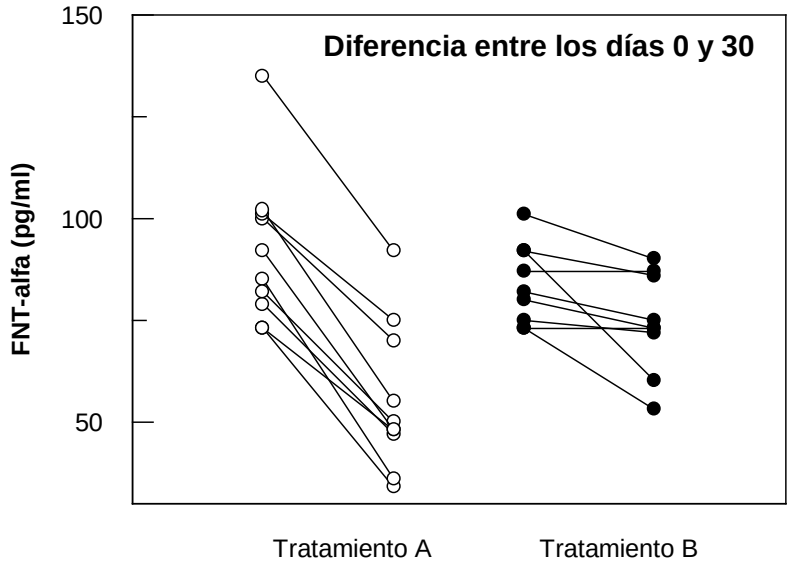
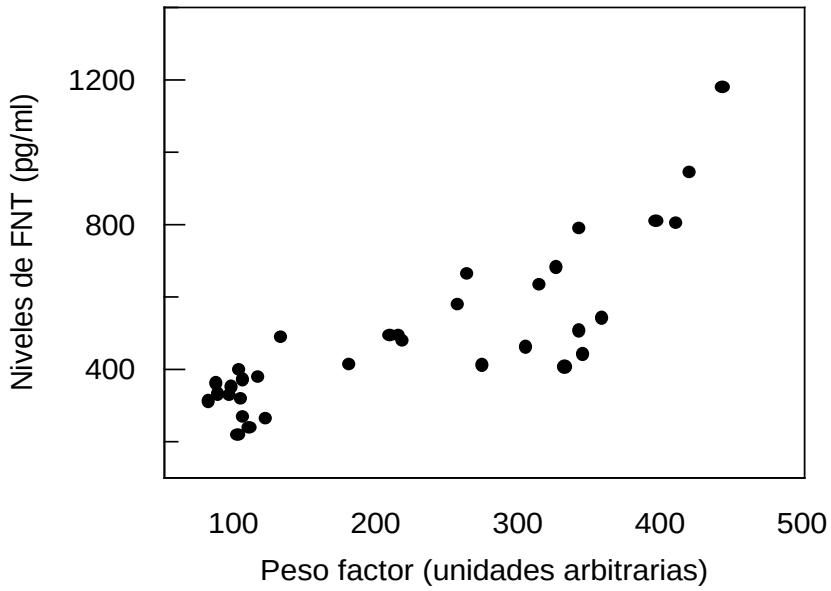


Figura 5.3 Concentraciones séricas de FNT-alfa en enfermos con Tuberculosis pulmonar previo y 1 mes después de iniciado el tratamiento específico.

La gráfica de la posible situación muestra que si bien se produce una caída en los valores de esta citocina con la administración de la terapia específica, el descenso es superior en los pacientes tratados con el esquema A.

Figura 5.4 Correlación entre los niveles de FNT-alfa y peso factor en pacientes con AR



La utilización de la figura es un recurso ineludible cuando se quiere representar la correlación entre dos variables. El caso que ilustra en la Figura 5.4, está referido a un ejemplo hipotético en pacientes con artritis reumatoidea donde se muestra que los niveles circulantes de FNT-alfa se asocian positivamente con el peso factor, un parámetro capaz de estimar la severidad de la enfermedad.
Para el caso en que tenemos múltiples observaciones de la variable a lo largo del estudio conviene emplear el gráfico de líneas que se muestra en la Figura 5.5

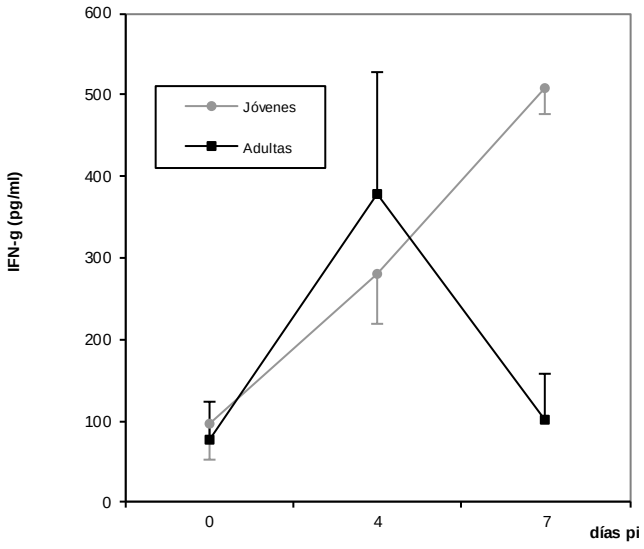


Figura 5.5 Niveles de interferón gamma (IFN-g, pg/ml) en animales infectados con *Trypanosoma cruzi* de distintas edades. El cuadrado representa la media y la barra el desvío estándar.

De constituir un verdadero escollo, la elaboración de los gráficos por medio de los diversos programas de computación disponibles en la actualidad ha pasado a ser una actividad *quasi* lúdica, en la que se advierte un número cada vez más creciente de apasionados.

5.4 Discusión

Para utilizar una expresión muy frecuentada en nuestros días, podemos afirmar que es la sección "más trabajada" del manuscrito. Ello no significa que deba primar la erudición, sobre todo en la actualidad, ya que los espacios asignados por las revistas se han reducido ostensiblemente. En líneas generales, tener presente que el mensaje debe ser claro, capaz de despertar la curiosidad del lector y con fuerza para resaltar la importancia de los resultados. Generalmente la discusión arranca con esas ideas claves de la introducción. Evitar los términos o expresiones innecesarias, para que el texto sea más fácil de leer y seguir; y tampoco abundar en adjetivos y adverbios. Un dato que de por sí es totalmente indicativo y preciso no necesita demasiados flecos. Las repeticiones disminuyen y no aumentan la fuerza del mensaje que se desea transmitir. Tratar, asimismo, que los párrafos tengan una extensión de una o dos frases, más de 5 frases no es recomendable.

Es difícil trazar una recomendación global para que los autores expresen todo lo que pretenden decir sin perder coherencia y unidad. A pesar de ello, se puede establecer una serie de prioridades que conviene tener en cuenta al momento de la redacción. Las mismas están referidas a:

- Señalar con claridad y contundencia cuál ha sido lo novedoso de nuestro estudio
- Establecer qué interpretación podemos darle a los hallazgos
- Comentar cómo se relacionan los resultados con los conocimientos actuales sobre el tema
- Indicar la potencial trascendencia o aplicabilidad clínicas que pudieran tener

En buena medida esto tiene que ver con la validez interna y externa de un estudio. La primera guarda relación con la confiabilidad del estudio para contestar la pregunta y depende de las características del diseño y de sus posibilidades. Por ejemplo cómo han sido las mediciones, definición de caso antes que nada y seguimiento de los pacientes. Generalmente conviene indicar que se trabajó con aquellos en quienes el seguimiento era posible, sin soslayar que ello podría entrañar un sesgo. Los estudios en patologías poco frecuentes plantean problemas adicionales puesto que requiere un esfuerzo mancomunado de distintos centros para lograr un número aceptable de pacientes.

La validez externa, por su parte, es la capacidad de generalizar los resultados del estudio a otras situaciones y pacientes. Un estudio puede tener un cierto grado de justificación en cuanto a validez externa si representa un paso adelante en la exploración de una nueva posibilidad. Mucho más si los pacientes que se han incluido representan a aquellos que se ven en la práctica diaria. Llevar adelante un estudio en circunstancias ideales generalmente cuesta menos que realizarlo en la situación habitual. Bueno sería plantear la pregunta que sea todo lo generalizable posible en la medida de nuestras posibilidades.

Es conveniente cerrar la discusión con un párrafo donde se sinteticen los puntos centrales de la investigación. En muchos trabajos se puede observar una concordancia entre los contenidos aquí expresados y el título del manuscrito que en definitiva es "el telegrama" sobre lo que se llevó a cabo o se descubrió. Los grupos con trayectoria en el tema suelen concluir con una frase donde adelantan las investigaciones que se han puesto en marcha para analizar los planteamientos surgidos del estudio actual. Como en

las jornadas de la tetralogía, ¡por qué no anunciar nuestra intención de hallar un héroe capaz de rescatar a Brunilda de su condenado sueño!

5.5 Referencias

Más que un capítulo de la metodología propiamente dicha, la estrategia para la obtención de información bibliográfica es una actividad perteneciente a la Bibliotecología. De modo que sólo nos referiremos a los procedimientos comúnmente más utilizados en el área biomédica. Previo a la aparición de las computadoras e Internet, los investigadores recurríamos a dos fuentes principales de información, *Index Medicus* y *Current Contents*. Quienes tenemos algunos años en el oficio recordamos perfectamente aquellos pesados volúmenes impresos en papel Biblia editados por la *National Library of Medicine*, cuyas búsquedas demandaban horas. Una vez obtenida la referencia el paso siguiente era ubicar la biblioteca donde se hallaba la revista. El *Current Contents* ofrecía la ventaja de consignar la dirección del autor a quien se le solicitaba un apartado de dicha publicación.

Felizmente en nuestros días el problema se ha simplificado notoriamente. No son pocas las bibliotecas que ofrecen la consulta de los *Medline* (versión electrónica de los *Index Medicus*) y lo que antes demandaba horas hoy se halla disponible, en cuestión de minutos, en la pantalla del monitor. Algo parecido ocurre con la versión digital de *Current Contents*. Con la lista de referencias y la posibilidad de acceder a Internet podemos visitar sitios de la red (por ejemplo, www.biomedcentral.com, www.pubmed.com y www.freemedicaljournals.com) donde se puede acceder a cientos de publicaciones periódicas, donde en muchos casos obtenemos la versión completa del artículo. Las universidades nacionales están relacionadas con un sitio del Ministerio de Ciencia y Técnica de la Nación por el cual se puede acceder a lista de publicaciones de Elsevier (*Science Direct*), *Springer Verlag* y *Wiley-Blackwell*, entre otros. Más allá de los trabajos que uno pueda acumular, ningún recurso informático podrá suplir el criterio profesional para decidir cuáles son los trabajos que ameritan ser citados en nuestro manuscrito.

Un punto que se torna cada vez más álgido está referido a la elección de la revista donde mandaremos a publicar el trabajo. Los más adentrados en estos quehaceres, conocimos una época hasta cierto punto "romántica", donde la jerarquía del trabajo importaba mucho más que la revista donde aparecía publicado. En la actualidad la cientimetría globalizadora aconseja, permítasenos el eufemismo, tener en cuenta el factor de impacto de la revista. Para quien desee conocer más al respecto, una publicación muy reciente (Biomedical Digital Libraries 2005, 2:7) brinda información en torno al modo en que se calcula dicho índice, como así también las bondades y desventajas del mismo.

Dependiendo de una serie de factores muchas veces imponderables, finalmente arriba una nota del Editor de la revista donde nos comunica que el trabajo ha sido aceptado para su publicación. Es un instante de gratificación y alivio que renueva el ánimo para seguir adelante.

La actividad del investigador está indefectiblemente ligada al conocimiento de las proposiciones temporalmente aceptadas ya que el acceso a la verdad es un anhelo poco probable, sino imposible, de cristalizar. Lo desconocido supera con creces a lo sabido, de modo que nuestras especulaciones e interpretaciones bien pueden ser erróneas. Manejar el método nos tiene que alertar que la afirmación tiene sus limitaciones. Seamos precavidos porque de lo contrario comienzan a operar los "*confounders*" de engreimiento para los cuales ¡no existen técnicas de ajuste!

Che onore? Che onor?
Che onor! Che ciancia! Che baia!
Può l´onore riempirvi la pancia? No
Può l´onore rimettervi uno stinco?
Non puo
Né un piede? No. Né un dito? No
Né un capello? No
L´onor non è chirurgo
Ch´ è dunque? Una parola
Che c´è in questa parola?
C´è dell´ aria che vuola
Bel costrutto!
L´onore lo può sentir chi è morto? No.
Vive sol coi vivi?
Neppure: perche a torto lo gonfiam le lusinghe,
lo corrompe l´orgoglio
l´ammorban le calumnie;
e per me non ne voglio!
No, no, no, no, no, no!
Ma, per tornare a voi furfanti,
Ho atesso troppo e vi discaccio.

(Acto I, final de la primera parte de la Opera Falstaff, Libreto de Arrigo Boito, sobre un texto de William Shakespeare)

DECLARACION DE HELSINKI DE LA ASOCIACION MÉDICA MUNDIAL

Principios éticos para las investigaciones médicas en seres humanos

Adoptada por la
18ª Asamblea Médica Mundial, Helsinki, Finlandia, junio 1964 y enmendada por la
29ª Asamblea Médica Mundial, Tokio, Japón, octubre 1975
35ª Asamblea Médica Mundial, Venecia, Italia, octubre 1983
41ª Asamblea Médica Mundial, Hong Kong, septiembre 1989
48ª Asamblea General Somerset West, Sudáfrica, octubre 1996
52ª Asamblea General, Edimburgo, Escocia, octubre 2000
Nota de Clarificación del Párrafo 29, agregada por la Asamblea General de la AMM, Washington 2002
Nota de Clarificación del Párrafo 30, agregada por la Asamblea General de la AMM, Tokio 2004
59ª Asamblea General, Seúl, Corea, octubre 2008

A. INTRODUCCION

1. La Asociación Médica Mundial (AMM) ha promulgado la Declaración de Helsinki como una propuesta de principios éticos para investigación médica en seres humanos, incluida la investigación del material humano y de información identificables. La Declaración debe ser considerada como un todo y un párrafo no debe ser aplicado sin considerar todos los otros párrafos pertinentes.
2. Aunque la Declaración está destinada principalmente a los médicos, la AMM insta a otros participantes en la investigación médica en seres humanos a adoptar estos principios.
3. El deber del médico es promover y velar por la salud de los pacientes, incluidos los que participan en investigación médica. Los conocimientos y la conciencia del médico han de subordinarse al cumplimiento de ese deber.
4. La Declaración de Ginebra de la Asociación Médica Mundial vincula al médico con la fórmula "velar solícitamente y ante todo por la salud de mi paciente", y el Código Internacional de Ética Médica afirma que: "El médico debe considerar lo mejor para el paciente cuando preste atención médica".
5. El progreso de la medicina se basa en la investigación que, en último término, debe incluir estudios en seres humanos. Las poblaciones que están subrepresentadas en la investigación médica deben tener un acceso apropiado a la participación en la investigación.
6. En investigación médica en seres humanos, el bienestar de la persona que participa en la investigación debe tener siempre primacía sobre todos los otros intereses.
7. El propósito principal de la investigación médica en seres humanos es comprender las causas, evolución y efectos de las enfermedades y mejorar las intervenciones preventivas, diagnósticas y terapéuticas (métodos, procedimientos y tratamientos). Incluso, las mejores intervenciones actuales deben ser evaluadas continuamente a través de la investigación para que sean seguras, eficaces, efectivas, accesibles y de calidad.
8. En la práctica de la medicina y de la investigación médica, la mayoría de las intervenciones implican algunos riesgos y costos.
9. La investigación médica está sujeta a normas éticas que sirven para promover el respeto a todos los seres humanos y para proteger su salud y sus derechos individuales. Algunas poblaciones sometidas a la investigación son particularmente vulnerables y necesitan protección especial. Estas incluyen a los que no pueden otorgar o rechazar el

consentimiento por sí mismos y a los que pueden ser vulnerables a coerción o influencia indebida.
10. Los médicos deben considerar las normas y estándares éticos, legales y jurídicos para la investigación en seres humanos en sus propios países, al igual que las normas y estándares internacionales vigentes. No se debe permitir que un requisito ético, legal o jurídico nacional o internacional disminuya o elimine cualquiera medida de protección para las personas que participan en la investigación establecida en esta Declaración.

B. PRINCIPIOS PARA TODA INVESTIGACION MÉDICA

11. En la investigación médica, es deber del médico proteger la vida, la salud, la dignidad, la integridad, el derecho a la autodeterminación, la intimidad y la confidencialidad de la información personal de las personas que participan en investigación.
12. La investigación médica en seres humanos debe conformarse con los principios científicos generalmente aceptados y debe apoyarse en un profundo conocimiento de la bibliografía científica, en otras fuentes de información pertinentes, así como en experimentos de laboratorio correctamente realizados y en animales, cuando sea oportuno. Se debe cuidar también del bienestar de los animales utilizados en los experimentos.
13. Al realizar una investigación médica, hay que prestar atención adecuada a los factores que puedan dañar el medio ambiente.
14. El proyecto y el método de todo estudio en seres humanos debe describirse claramente en un protocolo de investigación. Este debe hacer referencia siempre a las consideraciones éticas que fueran del caso y debe indicar cómo se han considerado los principios enunciados en esta Declaración. El protocolo debe incluir información sobre financiamiento, patrocinadores, afiliaciones institucionales, otros posibles conflictos de interés e incentivos para las personas del estudio y estipulaciones para tratar o compensar a las personas que han sufrido daños como consecuencia de su participación en la investigación. El protocolo debe describir los arreglos para el acceso después del ensayo a intervenciones identificadas como beneficiosas en el estudio o el acceso a otra atención o beneficios apropiados.
15. El protocolo de la investigación debe enviarse, para consideración, comentario, consejo y aprobación, a un comité de ética de investigación antes de comenzar el estudio. Este comité debe ser independiente del investigador, del patrocinador o de cualquier otro tipo de influencia indebida. El comité debe considerar las leyes y reglamentos vigentes en el país donde se realiza la investigación, como también las normas internacionales vigentes, pero no se debe permitir que éstas disminuyan o eliminen ninguna de las protecciones para las personas que participan en la investigación establecidas en esta Declaración. El comité tiene el derecho de controlar los ensayos en curso. El investigador tiene la obligación de proporcionar información del control al comité, en especial sobre todo incidente adverso grave**.** No se debe hacer ningún cambio en el protocolo sin la consideración y aprobación del comité.
16. La investigación médica en seres humanos debe ser llevada a cabo sólo por personas con la formación y calificaciones científicas apropiadas. La investigación en pacientes o voluntarios sanos necesita la supervisión de un médico u otro profesional de la salud competente y calificado apropiadamente. La responsabilidad de la protección de las personas que toman parte en la investigación debe recaer siempre en un médico u otro

profesional de la salud y nunca en los participantes en la investigación, aunque hayan otorgado su consentimiento.
17. La investigación médica en una población o comunidad con desventajas o vulnerable sólo se justifica si la investigación responde a las necesidades y prioridades de salud de esta población o comunidad y si existen posibilidades razonables de que la población o comunidad, sobre la que la investigación se realiza, podrá beneficiarse de sus resultados.
18. Todo proyecto de investigación médica en seres humanos debe ser precedido de una cuidadosa comparación de los riesgos y los costos para las personas y las comunidades que participan en la investigación, en comparación con los beneficios previsibles para ellos y para otras personas o comunidades afectadas por la enfermedad que se investiga.
19. Todo ensayo clínico debe ser inscrito en una base de datos disponible al público antes de aceptar a la primera persona.
20. Los médicos no deben participar en estudios de investigación en seres humanos a menos de que estén seguros de que los riesgos inherentes han sido adecuadamente evaluados y de que es posible hacerles frente de manera satisfactoria. Deben suspender inmediatamente el experimento en marcha si observan que los riesgos que implican son más importantes que los beneficios esperados o si existen pruebas concluyentes de resultados positivos o beneficiosos.
21. La investigación médica en seres humanos sólo debe realizarse cuando la importancia de su objetivo es mayor que el riesgo inherente y los costos para la persona que participa en la investigación.
22. La participación de personas competentes en la investigación médica debe ser voluntaria. Aunque puede ser apropiado consultar a familiares o líderes de la comunidad, ninguna persona competente debe ser incluida en un estudio, a menos que ella acepte libremente.
23. Deben tomarse toda clase de precauciones para resguardar la intimidad de la persona que participa en la investigación y la confidencialidad de su información personal y para reducir al mínimo las consecuencias de la investigación sobre su integridad física, mental y social.
24. En la investigación médica en seres humanos competentes, cada individuo potencial debe recibir información adecuada acerca de los objetivos, métodos, fuentes de financiamiento, posible conflictos de intereses, afiliaciones institucionales del investigador, beneficios calculados, riesgos previsibles e incomodidades derivadas del experimento y todo otro aspecto pertinente de la investigación. La persona potencial debe ser informada del derecho de participar o no en la investigación y de retirar su consentimiento en cualquier momento, sin exponerse a represalias. Se debe prestar especial atención a las necesidades específicas de información de cada individuo potencial, como también a los métodos utilizados para entregar la información. Después de asegurarse de que el individuo ha comprendido la información, el médico u otra persona calificada apropiadamente debe pedir entonces, preferiblemente por escrito, el consentimiento informado y voluntario de la persona. Si el consentimiento no se puede otorgar por escrito, el proceso para lograrlo debe ser documentado y atestiguado formalmente.
25. Para la investigación médica en que se utilice material o datos humanos identificables, el médico debe pedir normalmente el consentimiento para la recolección, análisis, almacenamiento y reutilización. Podrá haber situaciones en las que será imposible o impracticable obtener el consentimiento para dicha investigación o podría ser una amenaza para su validez. En esta situación, la investigación sólo puede ser

realizada después de ser considerada y aprobada por un comité de ética de investigación.
26. Al pedir el consentimiento informado para la participación en la investigación, el médico debe poner especial cuidado cuando el individuo potencial está vinculado con él por una relación de dependencia o si consiente bajo presión. En una situación así, el consentimiento informado debe ser pedido por una persona calificada adecuadamente y que nada tenga que ver con aquella relación.
27. Cuando el individuo potencial sea incapaz, el médico debe pedir el consentimiento informado del representante legal. Estas personas no deben ser incluidas en la investigación que no tenga posibilidades de beneficio para ellas, a menos que ésta tenga como objetivo promover la salud de la población representada por el individuo potencial y esta investigación no puede realizarse en personas competentes y la investigación implica sólo un riesgo y costo mínimos.
28. Si un individuo potencial que participa en la investigación considerado incompetente es capaz de dar su asentimiento a participar o no en la investigación, el médico debe pedirlo, además del consentimiento del representante legal. El desacuerdo del individuo potencial debe ser respetado.
29. La investigación en individuos que no son capaces física o mentalmente de otorgar consentimiento, por ejemplo los pacientes inconscientes, se puede realizar sólo si la condición física/mental que impide otorgar el consentimiento informado es una característica necesaria de la población investigada. En estas circunstancias, el médico debe pedir el consentimiento informado al representante legal. Si dicho representante no está disponible y si no se puede retrasar la investigación, el estudio puede llevarse a cabo sin consentimiento informado, siempre que las razones específicas para incluir a individuos con una enfermedad que no les permite otorgar consentimiento informado hayan sido estipuladas en el protocolo de la investigación y el estudio haya sido aprobado por un comité de ética de investigación. El consentimiento para mantenerse en la investigación debe obtenerse a la brevedad posible del individuo o de un representante legal.
30. Los autores, directores y editores todos tienen obligaciones éticas con respecto a la publicación de los resultados de su investigación. Los autores tienen el deber de tener a la disposición del público los resultados de su investigación en seres humanos y son responsables de la integridad y exactitud de sus informes. Deben aceptar las normas éticas de entrega de información. Se deben publicar tanto los resultados negativos e inconclusos como los positivos o de lo contrario deben estar a la disposición del público. En la publicación se debe citar la fuente de financiamiento, afiliaciones institucionales y conflictos de intereses. Los informes sobre investigaciones que no se ciñan a los principios descritos en esta Declaración no deben ser aceptados para su publicación.

C. PRINCIPIOS APLICABLES CUANDO LA INVESTIGACION MEDICA SE COMBINA CON LA ATENCION MEDICA

31. El médico puede combinar la investigación médica con la atención médica, sólo en la medida en que tal investigación acredite un justificado valor potencial preventivo, diagnóstico o terapéutico y si el médico tiene buenas razones para creer que la participación en el estudio no afectará de manera adversa la salud de los pacientes que toman parte en la investigación.

32. Los posibles beneficios, riesgos, costos y eficacia de toda intervención nueva deben ser evaluados mediante su comparación con la mejor intervención probada existente, excepto en las siguientes circunstancias:
- El uso de un placebo, o ningún tratamiento, es aceptable en estudios para los que no hay una intervención probada existente.
- Cuando por razones metodológicas, científicas y apremiantes, el uso de un placebo es necesario para determinar la eficacia y la seguridad de una intervención que no implique un riesgo, efectos adversos graves o daño irreversible para los pacientes que reciben el placebo o ningún tratamiento.
Se debe tener muchísimo cuidado para evitar abusar de esta opción.
33. Al final de la investigación**,** todos los pacientes que participan en el estudio tienen derecho a ser informados sobre sus resultados y compartir cualquier beneficio, por ejemplo, acceso a intervenciones identificadas como beneficiosas en el estudio o a otra atención apropiada o beneficios.
34. El médico debe informar cabalmente al paciente los aspectos de la atención que tienen relación con la investigación. La negativa del paciente a participar en una investigación o su decisión de retirarse nunca debe perturbar la relación médico-paciente.
35. Cuando en la atención de un enfermo las intervenciones probadas han resultado ineficaces o no existen, el médico, después de pedir consejo de experto, con el consentimiento informado del paciente o de un representante legal autorizado, puede permitirse usar intervenciones no comprobadas, si, a su juicio, ello da alguna esperanza de salvar la vida, restituir la salud o aliviar el sufrimiento. Siempre que sea posible, tales intervenciones deben ser investigadas a fin de evaluar su seguridad y eficacia. En todos los casos, esa información nueva debe ser registrada y, cuando sea oportuno, puesta a disposición del público.

Palabras Clave[12]

Análisis de la variancia
Análisis de la variancia de Friedman
Análisis de la variancia de Kruskall-Wallis
Análisis de los riesgos proporcionales (Regresión de Cox)
Análisis interinos
Análisis estratificado
Análisis multivariado
Area bajo la curva
Lineamientos éticos
Bayesianismo
Bonferroni corrección
Bradford Hill
Calidad de la evidencia
Casos y controles
Casos y controles anidados
Causalidad
Cochrane
Códigos
Coeficiente de desigualdad relativa (Odds Ratio)
Coeficiente de variación
Cohortes
Comité de Bioética
Consentimiento escrito informado
Correlación
Criterios de inclusión
Criterios de exclusión
Criterios de evaluación
Cuestionarios
Current Contents
Curva ROC
Datos secundarios
Declaración de Helsinki
Desvío estándar
Distribución normal
Drug Investigator Brochure
Efecto período
Efectos adversos
Eficacia
Ensayo cruzado
Ensayos clínicos
Ensayo clínico vs Cohorte
Entrevistas
Epi Info
Error aleatorio
Error sistemático o sesgo
Error estándar
Error tipo I (α)
Error tipo II (β)
Escalas
Especificidad
Estado de conocimiento del tema
Estudios clínicos, fases
Estudios descriptivos
Estudios pre-clínicos
Estudios prospectivos
Estudios retrospectivos
Estudios transversales
Etica
Exactitud
Exposición
Factibilidad
Factores de error
    Clasificación
    Implementación
    Medición
    Muestreo
    Sistemático
Factor de confusión (*confounder*)
Factor de riesgo
Falta de seguimiento
FDA
Figuras
Fracción etiológica
Fracción preventiva
Funnel plot
Hazard ratio
Hipótesis
Incidencia
Index Medicus
Inconsistencia
Indicadores de MBE
Indice de reclutamiento
Inocuidad
Intención de tratar
Interacciones
Intervalo o límite de confianza
Logaritmo OR
Magnitud del efecto
Manuscrito
Material y Métodos
Media
Mediana

[12] Su ubicación en el texto se puede lograr a través del comando especial del procesador

Medicina basada en la evidencia
Medicina traslacional
Mediciones
Medline
Meta-análisis
Método de Kaplan-Meier
Modificadores de efecto
Muestra
Muestreo
No participación
Número necesario a tratar
Observacional
Objetivos
OR diagnóstico
Placebo
Población
Poder del estudio
Precisión
Preguntas abiertas
Preguntas cerradas
Preguntas directas e indirectas
Prevalencia relativa
Probabilidad previa
Protocolo de ensayo clínico
Prueba de $\chi 2$
Prueba de Fisher
Prueba de la *Q* de Cochran
Prueba de la *t* de Student
Prueba de la U de Mann-Whitney
Prueba de Mantel-Haenszel
Prueba de McNemar
Prueba de Wilcoxon
Prueba del intervalo logarítmico (logrank test)
Prueba kappa
Pruebas estadísticas
    Paramétricas
    No paramétricas
Pruebas diagnósticas
    Fases
Puntajes
Punto o título de corte
Randomización
    Bloques permutados
    Desigual
    Estratificada
    Listas
    Minimización
Rango
Razón de probabilidad
Referencias
Reglas de la buena práctica clínica
Regresión de Cox
Regresión logística
Regresión múltiple
Reproducibilidad
Resultados
Riesgo atribuible
Riesgo relativo
Secuencia
Seguridad
Sensibilidad
Sesgo de selección
Significación estadística
Tabla de contingencia
Tablas
Tamaño muestral
Terceras variables
Tipos de mediciones
Validez del estudio
Validez interna
Valor delta
Valor predictivo
Variable de impacto o dependiente
Variable predictiva o independiente
Variancia explicada