

EVALUACION EPIDEMIOLOGICA DE TECNOLOGIAS DE SALUD

Agencia de Evaluación de Tecnologías Sanitarias

Organización Panamericana de la Salud
Oficina Sanitaria Panamericana
Oficina Regional de la Organización Mundial de la Salud

**Agencia de Evaluación de Tecnologías Sanitarias (AETS)
Instituto de Salud Carlos III
Ministerio de Sanidad y Consumo**

**c/ Sinesio Delgado nº 6 - Pabellón 3
28029 - MADRID (SPAIN)**

Tfno.: 91 - 387-78-40
91 - 387-78-00

Fax.: 91 - 387-78-41

Edita: INSTITUTO DE SALUD CARLOS SCIII
NIPO: 354-95-004-0
Imprime: MAP SERVICIOS GRÁFICOS SL
Fotocomposición: INDUSTRIA GRÁFICA SALJEN, SA

Este documento es un Informe Técnico de la Agencia de Evaluación de Tecnologías Sanitarias (AETS) del Instituto de Salud Carlos III del Ministerio de Sanidad y Consumo.

Dirección y coordinación:
José L. Conde Olasagasti

Elaboración y redacción:
Mónica Tafalla García

Colaboración y Asesoramiento:
En Página siguiente

Agradecimiento:

A Virginia García Mantilla por su colaboración en la edición de este documento.

Para citar este informe:

Ministerio de Sanidad y Consumo - Instituto de Salud "Carlos III"
Agencia de Evaluación de Tecnologías Sanitarias (AETS).
Evaluación Epidemiológica de Tecnologías de Salud
Madrid: AETS - Instituto de Salud Carlos III, Abril de 1995.

Este texto puede ser reproducido siempre que se cite su procedencia y no se vulnere su integridad.

AGRADECIMIENTOS

Al Dr. Eliseo Guallar, del Departamento de Epidemiología y Bioestadística de la Escuela Nacional de Sanidad, por la inestimable ayuda que me ha prestado siempre en la orientación de mi carrera profesional. Asimismo, quiero agradecerle la oportunidad que me ha brindado al ofrecerme la realización de este documento y las horas de trabajo que ha invertido en él.

A todos los componentes del Departamento de Epidemiología y Bioestadística de la Escuela Nacional de Sanidad, en especial al Dr. Fco. Javier Jiménez Jiménez, por su compañerismo y por aportar consejos y mejoras al documento que se presenta.

A los componentes de la Agencia de Evaluación de Tecnologías Sanitarias del Instituto de Salud Carlos III, en especial a su director, el Dr. J. Conde Olasagasti, que me ha dado la gran oportunidad de trabajar en esta institución permitiéndome, así, ahondar en el conocimiento de la evaluación de tecnologías y que ha aportado importantes matices a esta monografía.

Mónica Tafalla García

INDICE DE CONTENIDOS

1. INTRODUCCIÓN	11
2. CONCEPTOS BÁSICOS. DEFINICIONES	13
2.1. ¿Qué entendemos por tecnologías sanitarias?	13
2.2. Epidemiología	13
2.3. Eficacia, efectividad, eficiencia. Seguridad	14
3. ¿QUÉ, CUÁNDO Y CÓMO DEBE SER EVALUADO?	16
4. EVALUACIÓN A PARTIR DE LA EVIDENCIA DISPONIBLE	19
4.1. Especificación del propósito de la evaluación	19
4.2. Especificación de los criterios de inclusión/exclusión de los estudios	20
4.3. Técnicas de búsqueda bibliográfica	20
4.4. Interpretación e integración de la evidencia.	25
4.4.1. Revisiones narrativas de la literatura	25
4.4.2. Elección de un sólo estudio	25
4.4.3. Sistema de votación	26
4.4.4. Meta análisis	26
4.4.5. Juicios de expertos	30
4.5. Formulación de recomendaciones basadas en la calidad de la evidencia	32
5. GENERACIÓN DE DATOS NUEVOS	34
5.1. El papel de los estudios experimentales	35
5.1.1 Ensayo Clínico Aleatorio.	35
5.2. El papel de los estudios observacionales.	40
5.2.1. Estudios de cohortes	42
5.2.2. Estudios de casos y controles	47
6. EVALUACIÓN DE PRUEBAS DIAGNÓSTICAS.	51
7. EVALUACIÓN ECONÓMICA	58
8. DISCUSIÓN	63
9. BIBLIOGRAFÍA	65
10. ANEXO	66

INDICE DE TABLAS

Tabla I. Ventajas e inconvenientes de los Meta-análisis	30
Tabla II. Grados de calidad de la evidencia	33
Tabla III. Grados de seguridad de las recomendaciones en exámenes periódicos de salud	33
Tabla IV. Equivalencia entre calidad de la evidencia y seguridad de las recomendaciones sobre cribaje de cáncer de mama	34
Tabla V. Ventajas e inconvenientes de los Ensayos Clínico Aleatorios	40
Tabla VI. Ventajas e inconvenientes de los estudios de Cohortes	46
Tabla VII . Ventajas e inconvenientes de los estudios de Casos y Controles	51
Tabla VIII. Clasificación de los estudios de Evaluación Económica	60

INDICE DE FIGURAS

Figura 1. Estructura básica de un Ensayo Clínico Aleatorio	35
Figura 2. Estructura básica de un Estudio de Cohortes	42
Figura 3. Estructura básica de un Estudio de Casos y Controles	47
Figura 4. Curva ROC para el electrocardiograma de ejercicio como predictor de enfermedad coronaria	54

1. INTRODUCCIÓN

La tecnología médica juega un papel cada vez más importante en el cuidado de la salud de los pacientes y de las poblaciones. Los avances tecnológicos a los que hemos asistido en los últimos años han modificado por completo el panorama de la atención clínica y las posibilidades de intervención en salud pública. El rápido proceso de innovación tecnológica, sin embargo, hace necesaria la utilización de técnicas de evaluación que analicen, con rapidez y precisión, el impacto potencial de las nuevas tecnologías.

Al mismo tiempo, y motivado en parte por la incorporación de tecnología cada vez más sofisticada, se ha producido un encarecimiento progresivo de la atención sanitaria, que junto con el envejecimiento de las poblaciones y otros factores administrativos ha motivado un gran aumento del gasto sanitario global. Teniendo en cuenta que los recursos de los que dispone una sociedad para la atención en salud son limitados, la evaluación de tecnologías sanitarias no se puede limitar a medir la eficacia de las técnicas, sino que es necesario asegurar que se utilizan de forma óptima, es decir, produciendo el máximo beneficio posible. En estos términos se habla del coste de oportunidad, definido como el valor de la mejor alternativa a la que hay que renunciar, por la limitación de los recursos, al efectuar una elección, y que representa el beneficio no obtenido por emplear los recursos en la actividad elegida en lugar de en la mejor de las alternativas. Por ejemplo, ¿qué intervenciones sanitarias no podemos llevar a cabo por invertir una gran parte de los recursos en comprar aparatos de resonancia magnética? ¿Cuál es la mejor alternativa? ¿A qué se tiene que renunciar?

Aparte de estas cuestiones económicas, es necesario subrayar la existencia de una amplia variabilidad en la práctica médica que tiene su origen en las distintas tradiciones clínicas en las que se forman los médicos, tanto de atención primaria como especialistas. Aunque el grado de variabilidad de la práctica clínica difiere según las especialidades y las patologías asociadas, la implantación y evaluación de las tecnologías sanitarias se ve afectada por esta variabilidad, que frecuentemente resulta de difícil cuantificación.

Los tres elementos descritos, parámetros de eficacia, coste oportunidad y aplicación clínico/epidemiológica de las tecnologías deben ser considerados, por tanto, en la evaluación. En este sentido, los componentes más importantes de la evaluación de tecnologías sanitarias son los **resultados**, en términos de efectividad y seguridad, los **costes** que suponen a la sociedad y la **equidad**, que no se refiere a la distribución igualitaria de los recursos, sino a la mayor inversión en las poblaciones más necesitadas.

La evaluación de tecnologías tiene que ir encaminada a prestar un soporte a la toma de decisiones de los sistemas sanitarios. Debe ayudar a planificar el gasto y a decidir las prestaciones que cada sistema sanitario va a proporcionar a la sociedad, así como su óptima distribución geográfica, de forma que los recursos lleguen a aquellos que más los necesitan.

La epidemiología se ha consolidado progresivamente como ciencia del razonamiento objetivo en las ciencias de la salud, superando su inicial objetivo de evaluación del papel

causal de diversos agentes. El método epidemiológico, con sus aspectos tanto cualitativos como, sobre todo, cuantitativos permite obtener estimaciones de la efectividad y seguridad de las tecnologías que podrán ser de utilidad para decidir sobre la conveniencia de su utilización. Así, la adaptación de la epidemiología a la evaluación de las tecnologías constituye una herramienta básica en la toma de decisiones sanitarias, asignación de recursos y priorización de objetivos.

Con este documento se pretende introducir al lector en la evaluación epidemiológica de tecnologías sanitarias. Tras un primer capítulo de conceptos básicos y definiciones de términos fundamentales en evaluación de tecnologías, se revisarán las técnicas de evaluación a partir de la evidencia relevante disponible, por medio de su análisis crítico y síntesis. Este constituye el primer paso, y a veces el único, en toda evaluación y debe ser realizado de forma sistemática y rigurosa. Posteriormente se hará un breve repaso a los principales tipos de estudios epidemiológicos, tanto experimentales como observacionales, y se tratará como tema aparte, por sus características diferenciales, la evaluación de pruebas diagnósticas, haciendo especial énfasis en los conceptos de sensibilidad, especificidad y valores predictivos. Por último, se revisarán los principales conceptos de evaluación económica, un tema clave de la evaluación de tecnologías sanitarias.

2. CONCEPTOS BÁSICOS

2.1 ¿QUÉ ENTENDEMOS POR TECNOLOGÍAS SANITARIAS?

El término tecnología comprende todo lo relativo a instrumentos y maquinaria sea ésta del tipo que sea, aunque generalmente se suele identificar con la llamada “alta tecnología” o “nueva tecnología” y tiende a implicar elementos de sofisticación técnica, precisión en los resultados y coste elevado.

En la actualidad se ha implantado una definición mucho más amplia propuesta por la Office of Technology Assessment (OTA) de los EEUU en los años setenta, y que vamos a seguir en esta monografía. Según esta definición, el concepto de tecnología sanitaria engloba a “los **medicamentos, aparatos, procedimientos médicos y quirúrgicos** usados en atención médica, así como los **sistemas organizativos** con los que esta atención sanitaria se presta”, es decir, toda la práctica clínica y el modo en que se organiza.

La inclusión de los sistemas organizativos dentro de la definición de tecnología es consecuencia del reconocimiento de que los resultados de una intervención en salud dependen del contexto en que ésta se desarrolla y de que la organización de tecnologías individuales forma, a su vez, sistemas tecnológicos más complejos. Por ejemplo, una unidad de cuidados intensivos es una forma compleja de tecnología, con características propias, que surge, a su vez, de la organización de tecnologías aisladas. Asimismo, los factores de los que depende el éxito de un programa de cribaje de cáncer de cérvix incluyen desde el sistema de organización de atención primaria y de los servicios ginecológicos hasta la utilidad, en términos de sensibilidad y especificidad, de la propia prueba diagnóstica.

Igualmente, la OTA ha definido la evaluación de tecnologías como una "forma de investigación que examina **las consecuencias técnicas, sociales, económicas, éticas y legales** derivadas del uso de la tecnología, que se producen **a corto y largo plazo** y que pueden ser tanto directas como indirectas y tanto sobre los efectos deseados como sobre los indeseados".

Por tanto, la evaluación de tecnologías se ha convertido en una tarea ineludible para los sistemas sanitarios: es necesario conocer las repercusiones que los distintos aspectos de la práctica clínica van a tener sobre la sociedad. De hecho, en los últimos años hemos asistido a la creación y consolidación de diferentes organismos con funciones de evaluación de tecnologías en la mayor parte de los países desarrollados y cuya labor está teniendo un impacto progresivamente más importante en la toma de decisiones sanitarias.

La evaluación de tecnologías sanitarias, tal y como va a ser tratada en este documento, se define como un análisis sistemático y estructurado de la tecnología, fundamentalmente en términos de efectividad, seguridad y costes para el sistema y que tiene como función básica proveer de soporte técnico a las decisiones en política sanitaria.

2.2. Epidemiología

Si bien inicialmente la epidemiología estuvo directa y exclusivamente en las enfermedades infecto-contagiosas, en la actualidad ha extendido su campo de acción. Así, hoy en día se define como el “razonamiento y método propio del trabajo objetivo en medicina y otras ciencias de la salud, aplicados a la descripción de los fenómenos de salud, a la explicación de su etiología y a la búsqueda de los métodos de intervención más eficaces” (Jeniceck, 1987).

Algunas preguntas que se plantean en evaluación de tecnologías tienen una respuesta básicamente epidemiológica. Por ejemplo, ¿Cuál es el mejor tratamiento para la litiasis biliar sintomática?, ¿Qué sensibilidad y especificidad tiene una nueva prueba diagnóstica y a qué población se debe aplicar para obtener beneficio máximo?, ¿Cuál es la utilidad de la vacuna neumocócica?. El método epidemiológico aplicado a evaluación de tecnologías permite además responder otro tipo de preguntas, como ¿Dónde debe localizarse un aparato de resonancia magnética para que su aprovechamiento sea óptimo? o ¿Qué papel tienen actualmente las técnicas de radiocirugía?.

La aproximación epidemiológica en evaluación de tecnologías sanitarias permite obtener estimaciones de la utilidad, tanto en términos de eficacia y efectividad como de seguridad, de las tecnologías, al tiempo que puede valorar el impacto potencial de su aplicación sobre los niveles de salud de una determinada población. El método epidemiológico, por tanto, debe constituir uno de los pilares básicos en los que se fundamente la evaluación científica de las tecnologías sanitarias.

2.3. Eficacia, Efectividad y Eficiencia.

Los conceptos de eficacia, efectividad y eficiencia son fundamentales en el campo de la evaluación de tecnologías y van a aparecer frecuentemente en esta monografía, por lo que resulta conveniente definirlos con claridad.

La **eficacia** mide los beneficios obtenidos por una población cuando sobre ella se aplica, en condiciones ideales, una tecnología médica concreta. Se determina fundamentalmente de forma experimental y su validez es universal, siempre y cuando no se modifiquen las condiciones de aplicación de la intervención.

La **efectividad**, sin embargo, mide los beneficios obtenidos por una población cuando la tecnología se aplica en condiciones reales. Las medidas de efectividad no son tan universales como las de eficacia, sino que su generalizabilidad depende, en cierto grado, de la población a la que se aplica y de las condiciones de su aplicación. Como contrapartida, las medidas de efectividad pueden ser mucho más relevantes que las de eficacia al medir la utilidad de una tecnología en condiciones reales.

Es evidente que la diferencia entre eficacia y efectividad va a depender de forma directa de la diferencia que exista entre las condiciones ideales y las reales. Por ejemplo, la eficacia del tratamiento de la tuberculosis puede ser del 95% si se curan 95 de cada 100 pacientes tuberculosos que siguen correctamente este tratamiento. Pero si en una población sólo se diagnostica al 15% de los casos, se les prescribe el tratamiento correctamente, pero tan sólo el 30% tiene una cumplimentación correcta, la efectividad de la poliquimioterapia antituberculosa, en estas condiciones disminuirá al 4.3% ($0.95 \times 0.15 \times 0.30 = 0.043$), es decir, sólo el 4.3% de los enfermos de tuberculosis se curarán. En otra población en la que exista un programa de detección de casos nuevos de tuberculosis y se llegue a diagnosticar hasta el 50% de los casos reales, suponiendo que el cumplimiento del tratamiento sea igual que en la población anterior, la efectividad resultante sería del 14.25% ($0.95 \times 0.50 \times 0.30 = 0.1425$).

La amplia variabilidad de implantación de algunas tecnologías es consecuencia de la ausencia actual de consenso sobre la efectividad de muchas de ellas. Una importante contribución de la evaluación de tecnologías ha sido el énfasis puesto en la necesidad de valorar la efectividad global de la tecnología en cada una de las diferentes poblaciones a las que podría ser aplicada.

El concepto **eficiencia** incluye, además, los aspectos económicos asociados a la intervención. La eficiencia se define como la relación entre los beneficios que se obtienen al aplicar una tecnología y los costes que se han empleado para obtenerlos. Al ser un término relativo, la eficiencia de un procedimiento o intervención tiene que ser utilizada en comparación con la de otro. Por ejemplo, un determinado programa de lucha contra la malaria será más eficiente que otro si, a igual coste consigue mayores beneficios o si a igual beneficio le cuesta menos al sistema sanitario y, por tanto, a la sociedad.

Los estudios de eficiencia son un elemento fundamental en la evaluación de tecnologías siendo de suma importancia para la toma de decisiones en materia sanitaria. Ayudan tanto a priorizar los objetivos de los sistemas sanitarios como a la óptima asignación de recursos, completando la información obtenida por los estudios de efectividad y seguridad de las tecnologías.

El término **seguridad**, en evaluación de tecnologías, hace referencia a aquellos efectos indeseados, sean del tipo que sean, asociados a la tecnología evaluada. Toda evaluación deberá juzgar la aceptabilidad del riesgo inherente al uso de una determinada tecnología en una situación específica.

3. ¿QUÉ, CUÁNDO Y CÓMO DEBE SER EVALUADO?

¿Qué?

Siguiendo la definición de la OTA de tecnología sanitaria, pueden ser objeto de evaluación:

- 1.- Los medicamentos
- 2.- Los aparatos y dispositivos
- 3.- Los procedimientos médicos y quirúrgicos
- 4.- Los sistemas de organización

Las autoridades sanitarias de cada país son las encargadas de establecer los requisitos para la autorización de los **medicamentos**. Sin embargo, actualmente se está asistiendo a una armonización cada vez mayor de las regulaciones sobre investigación clínica en cuanto a las legislaciones nacionales en el área farmacéutica.

Asimismo, los **aparatos y dispositivos** están sujetos a normativas específicas que regulan su autorización, instalación y mantenimiento, para su aplicación en humanos. En estos casos, la evaluación de tecnologías es útil para determinar su efectividad y seguridad y ayudar a decidir si los proveedores de cuidados sanitarios deben ofrecerlos a los pacientes y en qué condiciones.

Los diferentes **procedimientos**, tanto médicos como quirúrgicos, empleados en la atención de la salud de la población son otro objetivo de la evaluación de tecnologías sanitarias. Se incluirán aquí los protocolos de diagnóstico y atención médica de una determinada enfermedad o las diferentes técnicas quirúrgicas que se utilizan hoy en día, entre otros.

Como ya se comentó en relación a la definición de tecnologías sanitarias, además de la evaluación de la efectividad y seguridad de los distintos medicamentos, aparatos y procedimientos utilizados en pacientes, la evaluación del **sistema de organización** de los mismos se ha convertido en una tarea esencial de los sistemas de salud y ha sido incorporada a aquellos organismos encargados de evaluar tecnologías.

Así, han sido objeto de evaluación por las diferentes agencias temas tan variados como los abordajes quirúrgicos (colecistectomía laparoscópica), técnicas de diagnóstico precoz de cáncer de cuello uterino, aparatos diagnósticos (resonancia magnética) o sistemas organizados (unidades de ictus).

¿Cuándo?

Toda tecnología atraviesa, a lo largo de su *vida*, varias fases de duración muy variable.

1. Experimental, en la que se establecen la eficacia y seguridad de la tecnología en las condiciones habitualmente ideales del laboratorio

2. Introducción y expansión en el sistema sanitario. La difusión dentro del sistema de una determinada tecnología responde a estímulos no sólo basados en la información científica que la acompaña. Muchas veces influyen otros factores difícilmente cuantificables como la presión social, las preferencias de los usuarios o determinados intereses comerciales.
3. Utilización generalizada, durante la cual es mayoritariamente aceptada la utilidad de la tecnología o bien no existe una alternativa de eficacia probada.
4. Declive o sustitución por otra tecnología nueva.

Cualquiera de estas fases puede y debe ser objeto de evaluación. Muchos procedimientos que se utilizan actualmente de forma generalizada nunca han sido evaluados o lo han sido de forma incompleta durante su fase experimental. Otros, en cambio, aún habiendo sido evaluados durante sus fases iniciales deben ser re-evaluados para establecer su efectividad, para estudiar sus efectos sobre grandes grupos de población o para completar el perfil de sus indicaciones y condiciones de uso.

Así, la evaluación debe ser planteada en todas las fases por las que atraviesa la tecnología. En la fase experimental deberá estudiarse fundamentalmente la eficacia y seguridad de la tecnología, en la fase de expansión la evaluación irá encaminada a determinar parámetros de efectividad y en la de utilización generalizada se podrán identificar posibles efectos adversos de escasa ocurrencia o nuevas indicaciones.

De hecho, entre los primeros informes de evaluación que publicó la OTA del congreso de los EE.UU., no sólo había tecnologías nuevas, sino que también se evaluaron otras que habían sido ampliamente utilizadas durante años, como el cribado de cáncer de cérvix, la diálisis o el cribado de cáncer de colon.

¿Cómo?

Los componentes básicos de la evaluación son la recogida o generación de información sobre la tecnología, el análisis crítico y síntesis de esa información y la presentación final de los resultados en un informe que contenga una serie de recomendaciones que ayuden al proceso de toma de decisiones de los responsables del sistema sanitario.

El proceso de evaluación es complejo y puede abordarse desde múltiples perspectivas. A la clásica visión clínica de efectividad y seguridad se han añadido otros puntos de vista más abiertos, como la satisfacción o preferencias de los usuarios. Así, la evaluación de tecnologías sanitarias se ha convertido en una tarea multidisciplinar en la que tienen cabida diversas ramas profesionales que han incorporado sus metodologías propias. Por tanto, no se puede hablar de un método propio de la evaluación de tecnologías sanitarias sino que, en función del aspecto de la tecnología que se quiere evaluar, se adoptará la metodología más apropiada.

La perspectiva epidemiológica de la evaluación de tecnologías dirige su actividad tanto a la evaluación de la efectividad y seguridad de los diferentes procedimientos como a la investigación del potencial impacto que puedan tener las intervenciones sobre la salud de las poblaciones. Además aporta las bases para el establecimiento de la utilidad de pruebas diagnósticas así como de la evaluación económica.

El arma fundamental de la evaluación epidemiológica es el ensayo clínico randomizado, al ser el tipo de estudio que conlleva conclusiones más sólidas. No obstante, por razones éticas, legales o logísticas no siempre se pueden realizar, por lo que la epidemiología ha buscado otra serie de instrumentos válidos de evaluación, como los estudios observacionales que serán descritos más adelante.

La evaluación de tecnologías debe partir de la información disponible a través de la búsqueda de los estudios epidemiológicos realizados y la síntesis de sus resultados. Sólo en caso de que esta información sea insuficiente se deberán diseñar y desarrollar estudios que generen datos nuevos, ya que el objetivo principal de la evaluación es obtener conclusiones o recomendaciones de la forma más precisa, barata y rápida posible y los estudios epidemiológicos son, en general, caros y de duración prolongada.

4. EVALUACIÓN A PARTIR DE LA EVIDENCIA DISPONIBLE

Aunque no exista un procedimiento único aplicable a todos los casos, en general, toda evaluación debe incluir los siguientes pasos:

1. Especificar el propósito de evaluación.
2. Especificar los criterios de selección de la evidencia relevante, incluyendo los criterios de inclusión/exclusión de los estudios individuales.
3. Formular un plan de búsqueda y recogida de la literatura relevante.
4. Interpretar e integrar la evidencia de los estudios individuales.
5. Formular recomendaciones basadas en la calidad de la evidencia.

4. 1. Especificación del propósito de la evaluación

Una formulación clara y explícita del propósito de la evaluación guiará la dirección y profundidad de la búsqueda de información relevante, la interpretación de la evidencia y el enfoque de las recomendaciones finales. Una agencia de evaluación se puede enfrentar, entre otros, a los siguientes problemas:

- aconsejar a una agencia gubernamental sobre la conveniencia de permitir el uso de una tecnología,
- aconsejar a médicos y pacientes sobre el uso adecuado de intervenciones en determinados problemas de salud,
- aconsejar a un hospital sobre la adquisición y precio de una tecnología,
- resaltar la necesidad y orientar futuras investigaciones sobre una tecnología.

Según el propósito de la evaluación, el tema deberá ser orientado desde alguna de las siguientes perspectivas:

1. Orientada hacia la tecnología propiamente dicha, evaluando la efectividad y seguridad de una tecnología.
2. Orientada hacia el problema de salud, evaluando todas las posibilidades preventivas, diagnósticas o terapéuticas de una enfermedad o condición clínica.
3. Orientada hacia programas de intervención, evaluando el impacto de programas concretos.

Así, en el caso de la hipertensión arterial (HTA), la primera perspectiva evaluaría la efectividad y seguridad de un fármaco antihipertensivo; desde la perspectiva del problema de salud se abordaría la evaluación de las medidas farmacológicas e higiénico-dietéticas que pretenden disminuir las consecuencias de la HTA sobre el organismo, así como sus formas de aplicación; y la perspectiva de los programas de salud enfocaría la evaluación en términos de la efectividad de un programa específico de prevención, diagnóstico precoz y tratamiento de esta enfermedad para una determinada área geográfica.

4.2. ESPECIFICACIÓN DE LOS CRITERIOS DE INCLUSIÓN / EXCLUSIÓN DE LOS ESTUDIOS

El segundo paso, tras delimitar el tema de estudio, consiste en buscar aquella información que más se ajuste a nuestras necesidades y mejor responda a la pregunta planteada en la evaluación.

Una búsqueda bibliográfica muy general nos puede proporcionar mucha información, pero de escasa relevancia para nuestros propósitos. Por tanto, se tendrán que establecer una serie de criterios, a partir de los objetivos de la evaluación, que limiten el campo de búsqueda ajustándolo a los propósitos concretos de la evaluación.

Por ejemplo, si se quiere evaluar la efectividad del uso de aspirina a dosis bajas en la prevención secundaria de enfermedad isquémica coronaria, los criterios de selección deberán basarse en el problema de salud de estudio (prevención secundaria de enfermedad isquémica coronaria), en la población a la que se aplica la tecnología (pacientes que han sufrido un infarto de miocardio), en el tipo de intervención (administración de dosis bajas de aspirina), en los resultados finales (mortalidad coronaria, mortalidad por todas las causas) y/o en el diseño de los estudios (ensayos clínicos controlados).

4.3. TÉCNICAS DE BÚSQUEDA BIBLIOGRÁFICA.

En evaluación de tecnologías, y con el objetivo de determinar el estado del conocimiento sobre el tema de estudio, se debe hacer una revisión sistemática y exhaustiva de la bibliografía existente. Hay diferentes métodos para la recogida sistemática de bibliografía, aunque ninguno de ellos es lo suficientemente exhaustivo como para poder ser utilizado de forma aislada. Se tendrá que recurrir, por tanto, a la combinación de diferentes estrategias de búsqueda alternativas.

En primer lugar, conviene conocer si alguno de los diferentes organismos encargados de evaluar tecnologías en otros países ya ha revisado el tema de nuestro interés. Para esto, resulta de gran ayuda ponerse en contacto con las diferentes agencias de evaluación, obtener sus catálogos de publicaciones y solicitar el informe que nos pueda interesar. Si el tema no ha sido estudiado o si el enfoque de la evaluación realizada es diferente del que nos interesa, se tendrá que obtener la información a través de otros métodos.

En general, el primer paso consiste en la obtención de un listado inicial de bibliografía relevante a través de repertorios bibliográficos o bases de datos informatizadas, y completar posteriormente la búsqueda con otros métodos que se comentarán al final de este apartado. Entre los repertorios bibliográficos más extensamente utilizados se encuentran el Index Medicus y el Excerpta Medica cuyas características se van a revisar brevemente a continuación.

Index Medicus.

Es un repertorio no selectivo de publicaciones biomédicas, publicado desde 1960 por la National Library of Medicine de los EE.UU. y contiene más de 3.000 revistas médicas de todo el mundo. Al no seleccionarse los artículos de las revistas que están incluidas, la presencia de un título en este repertorio no es indicativo de su calidad.

Se reparte en 12 volúmenes impresos mensuales que incluyen una sección por temas, una sección por autores y otra de artículos de revisión. Al final de cada año se publica el total refundido de los 12 números mensuales en 5 volúmenes titulados Cumulated Index Medicus.

La consulta del índice de materias requiere utilización de un diccionario de términos descriptores o palabras clave, Medical Subject Headings (MeSH), de estructura jerárquica y que aparece como suplemento al volumen de enero de cada año.

Excerpta Medica.

Se trata de otro repertorio impreso de publicaciones biomédicas que edita desde 1947 la Fundación Internacional Excerpta Medica de Amsterdam, con más de 4.500 revistas médicas. Como características que lo diferencian del Index Medicus hay que resaltar su carácter selectivo y analítico (ya que sólo aparecen artículos de un cierto nivel de calidad que se acompañan de un resumen elaborado por especialistas) y el hecho de que contiene además una selección de libros.

Se edita en series temáticas independientes cuya periodicidad depende de cada serie y contiene mayor número de revistas europeas que el Index Medicus.

A diferencia del Index Medicus, el índice de materias sigue un criterio abierto. Cada artículo aparecerá tantas veces como palabras clave lo identifiquen representado por el número de orden en que figure impreso su resumen.

Bases de datos bibliográficas informatizadas

El crecimiento casi exponencial de la difusión de literatura médica en los últimos años ha generado tal cantidad de artículos que las búsquedas manuales se han convertido en una tarea lenta y tediosa. Por esto se han creado diferentes bases de datos informatizadas que facilitan la labor de búsqueda y, al mismo tiempo, proporcionan una información más exhaustiva y muchas veces más actualizada.

Una de las más utilizadas es la base MEDLINE de la National Library of Medicine, que incorpora el Index Médico, el International Nursing Index y el Index to Dental Literature.

La base de datos del Excerpta Medica se denomina EMBASE y al igual que al MEDLINE se puede acceder a ella por línea telefónica (on line) o por CD-ROM (Compact Disk Read Only Memory), un sistema de almacenamiento óptico de datos que permite recoger una gran cantidad de información (hasta 250.000 páginas mecanografiadas) en un mínimo espacio.

Estas bases de datos informatizadas también necesitan la utilización de palabras clave o descriptores, pero permiten el uso de combinaciones lógicas tipo "AND", "OR" y "NOT" que facilitan mucho el acceso restringido a aquella literatura que se quiere revisar. Así, si nos interesara buscar artículos sobre las prótesis endovasculares, pero excluyendo su aplicación coronaria, una aproximación podría ser: "((intravascular OR endoluminal) AND stent) NOT coronary".

La información obtenida se puede grabar en un disquete para su posterior manejo con un procesador de textos, o bien imprimir directamente. El formato de las referencias por MEDLINE es:

MEDLINE (R) 1980 - 1981

1 of 33

Marked in Search: #5

TI: The results of cholecystectomy at a district general hospital. A reappraisal of operative cholangiography.

AU: Reasbeck-PG

SO: Ann-R-Coll-Surg-Engl. 1981 Sep; 63(5): 359-62

ISSN: 0035-8843

PY: 1981

LA: ENGLISH

CP: ENGLAND

AB: A total of 487 cholecystectomies were performed at one hospital over a 44-month period by surgeons who differed considerably in the frequency with which they performed operative cholangiography. There were no differences in the frequency or severity of postoperative complications or recurrent symptoms between patients who did and those who did not undergo operative cholangiography; nor were any differences demonstrable between patients operated on by different surgeons. Although it is established that operative cholangiography will demonstrate otherwise unsuspected common bile duct stones in some patients, it has not yet been clearly proved that these stones would cause later complications if left undisturbed. The clinical advantages of operative cholangiography are not sufficiently clear to preclude their assessment by prospective controlled studies involving large numbers of patients.

MESH: Adult-; Common-Bile-Duct-Calculi-surgery; England-; Hospitals,-District; Hospitals,-General; Intraoperative-Period; Middle-Age; Postoperative-Complications

MESH: *Cholangiography-; *Cholecystectomy-; *Common-Bile-Duct-Calculi-radiography

TG: Comparative-Study; Human

PT: JOURNAL-ARTICLE

AN: 81280416

UD: 8112

En el que :

TI: título en inglés

TO: título en el idioma original

AU: autor

SO: fuente (revista año; volumen: páginas)

PY: año de publicación

LA: idioma

CP: país de publicación

AB: resumen

MESH: términos descriptores

PT: tipo de estudio

Las bases de datos proporcionan más información que los repertorios impresos, permiten el acceso a la información a través de más campos y resultan mucho más rápidas de revisar, ahorrando al investigador una gran cantidad de tiempo.

No obstante, su utilización requiere una cierta infraestructura, con ordenadores y lector de CD-rom y, para su manejo óptimo, es precisa una cierta formación o experiencia previa de la persona que realiza la búsqueda. Por tanto, si el investigador no está familiarizado con su manejo precisará de la colaboración de un documentalista experto en este sistema para obtener los mejores resultados posibles. En cualquier caso, la experiencia que proporciona su uso continuado es el factor que más ayudará a obtener el máximo partido de este método.

No obstante, a pesar de todas las ventajas de las bases de datos informatizadas, éstas no consiguen la localización de todos los trabajos publicados. Estudios diseñados para determinar el porcentaje de trabajos que identifica el MEDLINE lo han estimado en alrededor del 50% de los trabajos existentes. La denominada literatura gris, es decir, la contenida en las publicaciones de resúmenes de congresos, informes de organismos... etc., no aparece referenciada en repertorios bibliográficos de este tipo.

Además de todo lo señalado anteriormente, la introducción de los títulos en la base de datos y su posterior distribución requiere de un cierto tiempo, por lo que estos repertorios informatizados no están siempre actualizados.

¿Cómo completar una búsqueda bibliográfica?

Una vez conseguida la lista inicial de artículos de interés a través de los repertorios bibliográficos, será preciso una revisión exhaustiva de los listados de bibliografía que contienen cada uno de ellos. Una fuente especialmente rica en referencias bibliográficas son los artículos de revisión y meta-análisis, por lo que su revisión detenida será tarea ineludible para nuestro propósito.

Otro método, de completar la búsqueda será a través del contacto directo, cuando sea posible, con autores que están trabajando en el tema de interés, y que además podrán proporcionarnos información sobre trabajos que no han sido publicados o incluso trabajos que se están llevando a cabo en ese mismo momento.

Debido a que los repertorios no siempre están suficientemente actualizados, es muy conveniente realizar una búsqueda manual en los últimos números de aquellas revistas cuyos contenidos más se ajusten al tema de interés.

Es de suma importancia explotar al máximo todos estos sistemas para estar razonablemente seguros de que se ha obtenido toda la información disponible sobre el tema de estudio.

Conviene volver a resaltar la necesidad de la asistencia de un profesional de la documentación así como de expertos en el tema de interés, en este caso de la tecnología a evaluar, para conseguir toda la información relevante, ya que en la calidad de esta información se basará la calidad de la evaluación que se quiere realizar.

Sesgo de publicación.

Habitualmente no se va a poder localizar toda la información existente sobre el tema que se quiere evaluar. Por un lado, no todas las revistas que se publican en el mundo vienen recogidas en los repertorios y por otro, algunas publicaciones son extremadamente difíciles de obtener, bien por proceder de países con los que no resulta fácil contactar o por la propia limitación idiomática del investigador.

Además de esto, no todos los trabajos que se realizan llegan a ser finalmente publicados. Así, se define el **sesgo de publicación** como el “envío y publicación selectiva de trabajos con resultados positivos”. Esto es, un estudio con un resultado estadísticamente significativo tendría mayor probabilidad de ser publicado que otro que no lo es.

Las razones que llevan a los autores a no enviar y a los revisores a no admitir a publicación ciertos estudios son muchas y no siempre bien definidas. Además de la significación estadística, se han implicado otros factores en el sesgo de publicación, como la fuente de financiación del estudio (un laboratorio farmacéutico podría estar más motivado a publicar un trabajo que muestra un efecto positivo), el tamaño de la muestra utilizado, la presencia o ausencia de aleatorización, la metodología y hasta las modas que afectan a la medicina.

Por tanto, en el proceso de evaluación de tecnologías se tendrá que tener en cuenta este factor ya que la selección de trabajos puede modificar las conclusiones obtenidas. Se han propuesto diferentes opciones para evitar o disminuir, en lo posible, el sesgo de publicación. El contacto directo con todos los investigadores que se dedican a un mismo tema podrá ayudar a identificar estudios no publicados. Otra posible solución es utilizar, cuando sea posible, fuentes no sesgadas de información, como los registros de trabajos científicos que contienen datos de todos los estudios que se están llevando a cabo,

independientemente de su futura publicación. Hoy en día existen pocos registros disponibles y deberá realizarse un esfuerzo para aumentar su número y poder contar, aunque a largo plazo, con fuentes no sesgadas de las que extraer información.

4.4. INTERPRETACIÓN E INTEGRACIÓN DE LA EVIDENCIA

Una vez recogida la información relevante sobre la tecnología que se quiere evaluar, el siguiente paso consiste en la revisión y síntesis de los resultados que aportan los estudios individuales.

Si, por ejemplo, se ha realizado una búsqueda bibliográfica con el fin de conocer las diferencias en efectividad entre dos procedimientos quirúrgicos, la colecistectomía convencional y la colecistectomía laparoscópica, y se obtienen varios trabajos realizados con metodología diferente (ensayos clínicos, series de casos y comparaciones retrospectivas) cuyos resultados asimismo difieren, el trabajo del evaluador consistirá en revisar críticamente cada uno de los estudios individuales y tratar de sintetizar toda la información que contienen, teniendo en cuenta las limitaciones de diseño, de calidad o desarrollo que pudieran tener.

A continuación se van a describir diversos métodos de revisión y síntesis de la información disponible, así como algunas técnicas que utilizan los juicios consensuados de expertos que, aisladamente o en combinación con los diferentes métodos de revisión, buscan el mismo fin: obtener recomendaciones sobre la utilidad de una tecnología.

4.4.1. Revisión narrativa de la literatura

Es la forma tradicional de revisión donde un especialista en el tema de estudio elige y resume, según su propio criterio, la bibliografía relevante. Al carecer de una metodología sistemática en la búsqueda, extracción y evaluación de los resultados de los estudios originales, está sujeto a la subjetividad del revisor, en lo que radica su más importante debilidad como sistema de revisión.

Este método, no obstante, puede resultar útil cuando no existe controversia entre los resultados de los estudios; pero si éstos son claramente contradictorios, este tipo de aproximación es poco informativa y la conclusión a la que se llegue será muy subjetiva. Pueden ser, en cualquier caso, útil para la generación de nuevas hipótesis y para el estudio de las causas de divergencia entre los resultados de los estudios individuales.

4.4.2. Elección de un solo estudio

Cuando existe controversia entre los resultados de los diferentes estudios se puede escoger aquel cuya metodología es más correcta desestimando la información que proporcionan los de menor calidad. A pesar de que este sistema simplifica mucho la obtención de conclusiones su debilidad metodológica lo hace poco o nada recomendable.

4.4.3. Sistema de votación

Compara el número de estudios con resultados estadísticamente significativos positivos (a favor de la intervención) frente al número de negativos (a favor del tratamiento estándar) eliminando los nulos (estadísticamente no significativos). Es, como la anterior, una forma muy sencilla y rápida de sintetizar la información pero al no tener en cuenta la calidad de los estudios ni su tamaño, asignándoles a todos la misma relevancia, su aplicación es muy discutible metodológicamente.

4.4.4. El meta-análisis

Se define como un proceso de combinación **sistemática y cuantitativa** de resultados de estudios para extraer conclusiones. Sus dos pilares básicos son la recolección exhaustiva y sistemática de datos originales y la cuantificación, en la medida de lo posible, de los resultados obtenidos, es decir, la obtención de estimadores combinados de efecto a partir de los estimadores de efecto que aportan los estudios individuales.

A pesar de requerir más tiempo y mayores conocimientos metodológicos y estadísticos que cualquiera de los anteriores métodos descritos, es el sistema más válido y ajustado de integración de la evidencia. El diseño de todo meta-análisis, como el de cualquier otro estudio prospectivo, debe ir precedido por la formulación de una pregunta de investigación basada en una hipótesis de estudio, que en este caso es el propósito de la evaluación, y del desarrollo explícito un protocolo en el que se describirán cada uno de los pasos. Para ilustrar la explicación de los pasos de este método se va a seguir un trabajo recientemente publicado que evalúa la efectividad en adultos de la vacuna neumocócica por meta-análisis (Fine, 1994):

1.- **Identificación y localización de los estudios originales.** Ya se comentaron algunas de las diferentes fuentes de publicaciones disponibles, por lo que se remite al lector al apartado correspondiente.

En el ejemplo de la vacuna neumocócica los autores utilizan la combinación de tres distintas fuentes de información. La primera es una búsqueda en MEDLINE desde 1966 hasta 1991 con la estrategia “explore pneumococcal infections and explore immunization or explore vaccines”. Este sistema se complementa con la revisión de las referencias bibliográficas de los estudios localizados y con el contacto directo con los autores de los artículos y con los laboratorios farmacéuticos que producen la vacuna.

2.- **Selección de los estudios** que se van a incluir en el meta-análisis. Su objetivo es evitar la combinación de resultados de trabajos que no están midiendo lo mismo. Así, hay que valorar la homogeneidad respecto a diseño, utilización de grupo control, aleatorización, ajuste de variables de confusión y variables de efecto. Es necesario el establecimiento explícito “a priori” de los criterios de inclusión / exclusión para evitar en lo posible cualquier interferencia de subjetividad en la selección.

Normalmente, y con el fin de eliminar cualquier subjetividad del proceso de selección, dos investigadores del grupo de evaluación revisan independientemente los estudios y deciden cuáles se van a incluir en el meta-análisis. En caso de que existan artículos para los que no exista acuerdo, se puede añadir la opinión de un tercer investigador.

En el caso de la vacuna neumocócica, la revisión de los artículos identificados en la búsqueda fue realizada por dos investigadores independientemente para determinar si cumplían los siguientes criterios de inclusión:

- 1. Diseño: sólo ensayos clínicos aleatorizados.*
- 2. Población diana: adultos.*
- 3. Resultados: aquellos estudios que evalúan la aparición de una o más situaciones entre 10 posibilidades (infecciones o mortalidad) relacionadas con la infección neumocócica y que habían sido definidas de forma explícita antes del inicio del estudio.*

3.- Extracción de los datos relevantes de los estudios individuales. Una vez se ha decidido qué estudios van a formar parte del meta-análisis, se procede a la recogida de todos los datos que puedan influir en el resultado, así como las medidas de asociación de los efectos estudiados (estimadores de efecto) y algún indicador de variabilidad de los mismos (por ejemplo odds ratio -OR- y su intervalo de confianza). Resulta muy útil diseñar una hoja de recogida de datos específica que se utilizará para cada estudio, de modo que se sistematiza el proceso y se asegura la homogeneidad en la recogida de los diversos investigadores. También en este paso, la recogida por parte de dos o más investigadores independientemente aumenta la objetividad del proceso.

En el ejemplo que venimos siguiendo, la extracción de datos fue realizada de forma independiente por dos investigadores y contenía los siguientes apartados:

- 1. Descripción de la intervención en el grupo experimental y el de control, incluyendo datos sobre el número de serotipos de la vacuna.*
- 2. Datos demográficos de ambos grupos, incluyendo el número total de participantes, su edad y sexo.*
- 3. Las indicaciones para la vacunación, incluyendo el número de pacientes ancianos, sujetos con enfermedades crónicas, pacientes institucionalizados y adultos inmunocomprometidos.*
- 4. Tiempo de seguimiento.*
- 5. Variables de valoración de resultados*
- 6. Serotipos de las infecciones neumocócicas registradas*
- 7. Efectos adversos de la vacuna.*

Las discrepancias entre los dos investigadores fueron posteriormente resueltas con el consenso de ambos.

4. Revisión de la calidad de los estudios. Algunos meta-análisis incorporan al resultado final estimadores de la calidad de los estudios individuales. Se da un peso a cada uno de los estudios, ponderando, por tanto, el impacto que van a tener en el estimador combinado según su calidad. Si la escala de calidad va de 0,0 (calidad mínima) a 1,0 (máxima calidad) y se multiplica cada estimador individual de efecto por ese factor, la suma de estos productos dividida por el sumatorio de todos los pesos dará el estimador final ponderado por la calidad de los estudios individuales. El estudio de peor calidad tendrá una menor influencia en el resultado combinado, mientras que el de mayor calidad influirá más.

Otras veces, la evaluación de la calidad se utiliza para determinar qué estudios van a formar parte del meta-análisis, estableciendo un umbral de calidad aceptable, de forma que los estudios que no lo alcancen no serán incluidos.

En el ejemplo de la vacuna neumocócica el instrumento utilizado para medir la calidad de ensayos clínicos consta de 23 ítems referidos a 7 áreas diferentes: 1) selección de los sujetos del estudio, 2) procedimiento de aleatorización, 3) descripción de las pautas de tratamiento, 4) técnicas de enmascaramiento, 5) procedimiento de medida de los resultados, 6) manejo de las pérdidas y abandonos y 7) método estadístico utilizado.

5.- Evaluación del grado de heterogeneidad entre los estudios. Las diferencias entre los resultados de estudios que pretenden evaluar el mismo efecto, se miden a través de pruebas estadísticas de heterogeneidad. Estas pruebas, al ser de escasa potencia, deben ser complementadas con otros métodos, como la inspección visual de las discrepancias en un gráfico.

Las causas de heterogeneidad se deben a diferencias tanto en el diseño, como en la población de estudio o en los factores de confusión por los que se deberá ajustar el estimador final de efecto.

Existe una cierta controversia sobre la actitud correcta que se debe tomar ante un grupo de artículos con resultados heterogéneos. Algunos autores consideran que no debe intentarse la combinación de resultados heterogéneos y que el trabajo, en este caso, debe quedar reducido a la exposición de los resultados de los estudios individuales y a la búsqueda de explicaciones para esa heterogeneidad. Otros autores, sin embargo, admiten la combinación de resultados incorporando esa heterogeneidad a la precisión del estimador combinado de efecto.

6.- Obtención del estimador combinado de efecto. Las aproximaciones estadísticas para la obtención de un estimador combinado se basan en el cálculo de una media de los estimadores individuales de efecto, ponderada por la calidad de los estudios. La precisión del estimador individual de efecto es un indicador de la calidad del estudio, por lo que se puede utilizar la inversa del cuadrado del error estándar como medida de ponderación (peso = $1/ES^2$).

En este sentido, en el ejemplo de la vacuna neumocócica, se utiliza, para la combinación de los Odds Ratio (OR), el método de Mantel-Haenszel, que asume un modelo de efectos fijos, en el se presupone que los OR de los estudios individuales son constantes y que la variación observada entre ellos se debe al error de muestreo. Para la diferencia de riesgos (DR), sin embargo, se emplea el método de Dersimonian-Laird, que asume un modelo de efectos aleatorios, en el que se asume que existe una distribución de las DR; se utilizan las DR encontradas en los estudios individuales para estimar esta distribución. Así, la DR obtenida en cada estudio sería una combinación de la DR real en el estudio individual y del error del muestreo. Por tanto, la variabilidad observada se descompone en una variabilidad inter-estudio y una variabilidad intra-estudio. Con este método se puede incorporar la heterogeneidad al estimador de efecto final.

A grandes rasgos, se encontró que la vacuna neumocócica es un factor de protección frente a neumonía por todos los serotipos de la bacteria así como para los serotipos incluidos en la vacuna, aunque resultó poco eficaz en determinados grupos de alto riesgo.

7.- Finalmente, el último paso consiste en la **interpretación** de los resultados. Como en todo trabajo científico los resultados obtenidos deben ser interpretados en función de la validez del estudio, sus limitaciones y la plausibilidad biológica de las asociaciones encontradas. La **búsqueda de las posibles causas de heterogeneidad** entre los estudios individuales es una práctica necesaria y para muchos autores uno de los aspectos más atractivos del meta-análisis.

El meta-análisis es un técnica extremadamente útil en la evaluación de tecnologías ya que revisa de forma sistemática la evidencia relevante sobre efectividad y seguridad y permite su cuantificación y comparación con otras tecnologías. Sus principales ventajas e inconvenientes se resumen en la tabla siguiente.

Tabla I. Ventajas e inconvenientes de los meta-análisis

<p>VENTAJAS</p> <ul style="list-style-type: none">· Búsqueda sistemática de la información relevante· La cuantificación de los resultados consigue una mayor objetividad en las conclusiones· Aumento de la precisión de los resultados· Permite el estudio de subgrupos· Puede sugerir hipótesis explicativas de heterogeneidad· Permite una mayor generalizabilidad de los resultados· Bajo coste
<p>LIMITACIONES</p> <ul style="list-style-type: none">· Su calidad depende de la calidad de los estudios originales· Posible sesgo de publicación si no se obtienen los trabajos de registros no sesgados· En caso de heterogeneidad existe controversia acerca de la idoneidad de la combinación de resultados.

4.4.5. Juicios de expertos

La necesidad de tomar decisiones sobre temas para los que no existe respuesta cuantitativa posible ha conducido a la utilización cada vez más frecuente de aproximaciones cualitativas. En evaluación de tecnologías se utilizan diferentes métodos cualitativos cuyo objetivo final es conseguir recomendaciones basadas en el consenso de especialistas en el tema de estudio.

De una forma más o menos estructurada, las técnicas de juicios de expertos consisten en reunir a especialistas en una determinada materia a los que se proporciona una selección de la bibliografía más relevante para que, a través de una discusión, consigan elaborar un informe consensuado que responda a las preguntas determinadas previamente por los administradores sanitarios.

Las conferencias de consenso, uno de los sistemas de consenso más extendidos, fueron utilizadas por primera vez por los Institutos Nacionales de la Salud (NIH) de los EE.UU. en los años setenta. Constan de varias fases en su desarrollo:

- Fase preparatoria de **selección del tema** a tratar y de formulación de las preguntas concretas que se van a plantear en la conferencia. Los criterios que se utilizan para la elección de los temas tienen en cuenta el interés para el sistema sanitario, la proporción de la población afectada y la existencia de controversia en la comunidad médica sobre la aplicación de la tecnología a evaluar.
- **Conferencia** propiamente dicha, en la que un grupo de expertos expone la evidencia disponible sobre la tecnología a estudiar. A continuación otro grupo de expertos elabora un informe preliminar que pasa a ser discutido públicamente con participación de todos los estamentos de la sociedad implicados, hasta que se alcanza el consenso.
- **Difusión** del informe final.

Esta metodología ha sido ampliamente criticada en base a los siguientes razones:

- No siempre se hace una búsqueda sistemática y exhaustiva de la evidencia disponible, por lo que no se puede asegurar que los participantes tengan toda la información relevante sobre el tema.
- Esta información no es siempre revisada con el rigor deseable, por lo que las conclusiones obtenidas podrían estar sujetas a un cierto grado de subjetividad.
- Habitualmente, el tiempo de que se dispone en estas conferencias para revisar literatura es escaso.

En general, estas limitaciones hacen sospechar que la subjetividad pueda tener un importante papel en los resultados de las conferencias de consenso.

Método Delphi

Es una técnica interactiva de encuesta. Los participantes son encuestados sobre la pregunta de interés por medio de cuestionarios enviados por correo. Tras la primera ronda, el grupo investigador analiza los resultados y los remite de nuevo a los participantes, para que comparen sus opiniones con las del grupo. De esta manera y tras sucesivas rondas se obtiene el consenso con la agregación simple de las posturas individuales tras eliminar las más extremas.

Las cuestiones que pueden ser evaluadas con esta técnica son muy variadas. Así, se han utilizado para obtener estimadores de incidencia de enfermedad y reacciones adversas en el tratamiento preventivo de la tuberculosis cuando no existían datos empíricos, para identificar los problemas de salud más extendidos, para establecer criterios de buena práctica clínica o para determinar las implicaciones de los avances en tecnología biomédica.

La ventaja de este método sobre el anterior es que con unos recursos limitados se puede acceder a gran cantidad de expertos al tiempo que estos toman sus posturas independientemente, sin influencia de otros especialistas más o menos carismáticos. No obstante, este hecho siempre supone pérdida de la riqueza que surge de la discusión.

Entre las críticas que ha recibido el método Delphi destacan las que enfatizan que la eliminación de las posiciones extremas puede desestimar interesantes puntos de vista. Asimismo, se ha considerado al consenso obtenido de “artificial” o “forzado” al basarse en la agregación simple de las opiniones individuales en la última ronda.

En general, todos los métodos basados en la opinión de expertos son muy sensibles a la influencia de factores como:

- La selección de los participantes y la composición de los paneles. Por un lado, la credibilidad del informe final se va a basar, en gran parte, en la credibilidad individual de los expertos. Por otro, el enfoque del tema de discusión variará mucho según qué especialistas compongan el grupo de expertos. Así, si se pretenden evaluar las endoprótesis vasculares, es predecible que no llegue a las mismas conclusiones un panel formado por cirujanos vasculares que otro formado por radiólogos intervencionistas. Por esta razón es importante que los grupos de expertos estén formados por representantes de todos los estamentos que, de una u otra forma, puedan estar implicados en el tema.
- La eliminación de las opiniones más extremas lleva al consenso en ciertos “lugares comunes”, para los que existe un acuerdo previo mayoritario, haciendo que los informes resultantes a partir de éstas técnicas sean bastante conservadores, es decir, en contra de las innovaciones tecnológicas.

4.5. FORMULACIÓN DE RECOMENDACIONES BASADAS EN LA CALIDAD DE LA EVIDENCIA.

Una vez sintetizada la evidencia con cualquiera de los métodos descritos, el último paso en evaluación de tecnologías consiste en formular una serie de recomendaciones, apoyadas en esa evidencia, que sirvan a los administradores sanitarios en el proceso de toma de decisiones.

Basándose en el diseño de los estudios se han determinado, de mayor a menor, 5 grados de calidad de la evidencia:

Tabla II. Grados de calidad de la evidencia

- **Grado I:** Evidencia obtenida de al menos un ensayo clínico aleatorio controlado y correctamente diseñado.
- **Grado II-1:** Evidencia obtenida de ensayos clínicos sin randomización.
- **Grado II-2:** Evidencia obtenida de un estudio de cohortes o de casos y controles bien diseñados, preferiblemente de más de un centro o grupo de investigación.
- **Grado II-3:** Evidencia obtenida de múltiples series de casos con o sin intervención.
- **Grado III:** Opiniones de autoridades respetadas, basadas en experiencia clínica, estudios descriptivos o informes de comités de expertos.

Así, la Preventive Task Force de EE.UU. utiliza los siguientes niveles de seguridad, a partir de la calidad de la evidencia, en las recomendaciones formuladas para los exámenes periódicos de salud:

Tabla III. Niveles de seguridad de las recomendaciones en exámenes periódicos de salud

- A. Existe buena evidencia para apoyar la recomendación de que la intervención sea específicamente considerada en un examen periódico.
- B. Existe bastante evidencia que apoye la recomendación de que la intervención sea específicamente considerada en reconocimiento periódico.
- C. Existe poca evidencia para la inclusión de la intervención en un examen periódico pero se podría recomendar por otras razones.
- D. Existe bastante evidencia para apoyar la recomendación de que la intervención sea excluida del examen periódico.
- E. Existe buena evidencia para apoyar la recomendación de que la intervención sea excluida del examen periódico.

Estos niveles de seguridad de la recomendación se corresponden a grados de calidad de la evidencia. Así, en el caso de cribado de cáncer de mama, la Preventive Service Task Force establece los niveles de seguridad de sus recomendaciones y su relación con los grados de calidad.

Tabla IV. Equivalencia entre calidad de la evidencia y seguridad de las recomendaciones sobre cribaje de cáncer de mama.

Grupo de edad	Intervención preventiva	Calidad de la evidencia	Seguridad de la recomendación
40-49	Exploración clínica anual	III	C
50-59	Mamografía anual y exploración clínica	I	A
>60	Mamografía anual y exploración clínica	II-2	B
> 40	Aprendizaje de autoexploración	III	C

Esta tabla muestra cómo para las mujeres entre 40 y 49 años la evidencia acerca del beneficio de un examen clínico anual de mama es pobre (grado III), por lo que la seguridad de la recomendación realizada para este grupo de edad debe ser C; sin embargo, la existencia de un ensayo clínico randomizado y con grupo control, permite establecer que, en mujeres de 50 a 59 años, la recomendación de realizar anualmente una mamografía junto con el examen clínico de la mama sea de grado A, es decir, de máxima seguridad.

5. GENERACIÓN DE DATOS NUEVOS

En algunas ocasiones, la literatura existente aporta una información insuficiente para la consecución de recomendaciones utilizables en la toma de decisiones. En estos casos, las diferentes técnicas de síntesis, ya sea a través de la revisión del material publicado o de juicios de expertos, no van a poder responder a la pregunta de interés. En estos casos, se hace necesaria la generación de datos nuevos mediante la realización de estudios. La epidemiología aporta una serie de diseños válidos para este fin, que se dividen en:

1. Experimentales: cuando el investigador somete a una población a una determinada intervención controlada y evalúa su efecto.
2. Observacionales: en los que se observa el efecto de una tecnología sobre una población expuesta, pero sin que el equipo investigador actúe sobre la exposición. En estos casos, la exposición no es controlada, por lo que existen mayores fuentes de error potencial al medir el efecto de la tecnología.

5.1. EL PAPEL DE LOS ESTUDIOS EXPERIMENTALES.

El paradigma del método científico es el experimento, que en ciencias de la salud tiene unas características especiales:

1. El investigador no puede controlar completamente todas las variables que van a influir en el resultado.
2. Por motivos éticos, no siempre se puede experimentar sobre humanos, por lo que, en ocasiones, habrá que elegir otros diseños para llevar a cabo la investigación. Tampoco es infrecuente que un experimento tenga que ser interrumpido al sospecharse que la intervención es perjudicial, o bien porque el efecto beneficioso encontrado hace que no sea ético privar al grupo control de sus propiedades

El Ensayo Clínico Aleatorio

El objetivo principal de los estudios experimentales en medicina es **evaluar la eficacia o efectividad de intervenciones preventivas, curativas o rehabilitadoras**. El tipo de estudio experimental más importante es el ENSAYO CLÍNICO ALEATORIO (ECA) que ha sido definido como “toda evaluación experimental de una sustancia o medicamento, a través de su aplicación a seres humanos, orientada a alguno de los siguientes fines:

- a) Poner de manifiesto sus efectos farmacodinámicos o recoger datos referentes a su absorción, distribución, metabolismo y excreción en el organismo humano.
- b) Establecer su eficacia para una indicación terapéutica, profiláctica o diagnóstica determinada.
- c) Conocer el perfil de sus reacciones adversas y establecer su seguridad.

Aunque esta definición se refiere exclusivamente a ensayos clínicos con fármacos, los mismos principios son, en general, aplicables a otro tipo de tecnologías.

La estructura básica de un ECA se puede representar con el siguiente esquema:

Figura 1. Estructura básica de un Ensayo Clínico Aleatorio.

PARA VISUALIZAR ESTA FIGURA, VAYA A LA PENÚLTIMA PÁGINA

Un grupo de personas (muestra) es sometido a un proceso de randomización o aleatorización, por el que serán asignados al azar a recibir una intervención (grupo experimental) o no recibirla (grupo control). El grupo control puede recibir una intervención alternativa, un placebo o nada. Tras un tiempo variable de seguimiento se evaluarán las diferencias en una serie de parámetros para conocer qué efecto se ha conseguido con esa intervención.

El placebo es un fármaco con las mismas características fisicoquímicas e idéntica pauta de administración que el fármaco activo, pero sin su actividad farmacológica. El efecto placebo, a su vez, se define como el efecto fisiológico o psicológico de cualquier intervención independientemente de su efecto farmacológico.

Como en todo estudio científico, el primer paso en la realización de un ECA consiste en establecer una hipótesis, de la que se derivarán los objetivos específicos del estudio. Para la consecución de estos objetivos se diseñará un protocolo que describa con detalle los pasos a seguir, así como las incidencias o problemas previsibles y los mecanismos que se van a utilizar para su manejo.

A lo largo de la descripción de los diferentes pasos de un ECA, se va a utilizar como ejemplo un estudio que busca comparar los resultados obtenidos en el tratamiento de la litiasis biliar con dos procedimientos: la colecistectomía laparoscópica y la colecistectomía a través de minilaparotomía. La hipótesis de partida supone que la minilaparotomía tiene similares resultados que la técnica laparoscópica (McMahon, 1994).

A. Elección de la muestra

Una vez se han establecido los objetivos concretos se tendrá que definir la población sobre la que se van a comparar los procedimientos.

La **población diana** o de referencia es aquella de la que salen los individuos en los que se va a llevar a cabo la intervención.

La **población de estudio** es la que cumple los criterios de inclusión marcados a la población diana. En general, los criterios de selección se utilizan para conseguir una mayor homogeneidad en la muestra, ajustando las características de los participantes a los objetivos marcados. Habrá que excluir siempre a los individuos en los que la intervención esté contraindicada, a los que no vayan a seguir correctamente el tratamiento y a aquellos sin capacidad de dar consentimiento a los investigadores.

La población sobre la que finalmente se realiza la intervención es la **muestra**, cuyo tamaño se determina en función del efecto esperado y del error en la estimación que consideremos admisible. No obstante, muchas veces se decide según el número de participantes disponibles de acuerdo con los medios de que se cuentan.

La evaluación de las intervenciones debe realizarse siempre en poblaciones similares a aquellas en las que posteriormente se va a aplicar la intervención.

En el estudio de colecistectomías se consideró población de estudio a todos aquellos pacientes susceptibles de colecistectomía electiva, sin signos clínicos de coledocolitiasis y que no hubieran sido sometidos a cirugía en el abdomen superior. De estos, se escogió una muestra de 299 participantes.

Previamente al inicio del estudio se debe obtener el consentimiento de los participantes tras exponerles, de forma clara y evitando influir en su decisión, los objetivos que persigue el ECA, beneficios, incomodidades y posibles riesgos, alternativas, derechos y responsabilidades. La obtención del consentimiento informado, así como otros requerimientos éticos y legales de los ECA, deberán ajustarse a la legislación vigente en cada país.

B. Asignación aleatoria al grupo de tratamiento o al control.

Cada participante tendrá una probabilidad conocida y fijada de antemano de recibir la intervención o el placebo. Su función principal es conseguir la mayor imparcialidad posible al asignar la intervención y el control. La asignación aleatoria consigue que, con muestras suficientemente grandes, las características de ambos grupos tiendan a ser iguales, con lo que las variables que pudieran interferir (o confundir) en los resultados, conocidas o, fundamentalmente desconocidas, se distribuyan con la misma probabilidad en ambos grupos. De esta forma, los resultados que se encuentren tras la intervención se deberán a ésta y no a las diferencias que inicialmente existen entre los grupos. La randomización, por tanto, facilita la realización de inferencias causales en relación a las intervenciones estudiadas.

C. Intervención.

Los diferentes tipos de intervención que se pueden evaluar por medio de ECA incluyen:

- Intervenciones preventivas, que pretenden disminuir o eliminar el riesgo de enfermar de una población. Como ejemplos tenemos la evaluación del tratamiento con bajas dosis de aspirina en pacientes que han sufrido un infarto de miocardio (prevención secundaria) o los estudios de efectividad de la vacuna neumocócica (prevención primaria).

- Intervenciones diagnósticas que evalúan la utilidad de nuevos métodos para el diagnóstico. Así, se comparará la utilidad de una técnica de tinción de bacterias para el diagnóstico de infección urinaria con los métodos habituales o las diferentes técnicas de imagen para establecer el estadiaje de tumores.

- Intervenciones curativas, en las que se pretende evaluar qué aproximación terapéutica es más efectiva y segura en una enfermedad. Así, se puede evaluar la utilidad de un nuevo antibiótico para las infecciones del tracto urinario o la efectividad de la cirugía laparoscópica en comparación con otros abordajes quirúrgicos más convencionales.
- Intervenciones rehabilitadoras, que buscan disminuir las secuelas que producen determinadas enfermedades. Por ejemplo, la valoración de técnicas de rehabilitación motora tras accidentes cerebrovasculares.

Un problema especial en la realización de ECA en evaluación de tecnologías, lo constituye la evaluación de aquellos procedimientos que requieren un proceso de aprendizaje previo de los investigadores o personas encargadas de la intervención, como ocurre con los procedimientos quirúrgicos nuevos. En este caso, si la intervención es llevada a cabo por profesionales con diferente nivel de experiencia, los resultados pueden estar influidos por este factor. Al mismo tiempo, en la práctica real, no todos los profesionales tienen la misma experiencia y/o habilidad en la realización de determinados procedimientos. Por esto, un estudio que determine los resultados de una técnica quirúrgica realizada por cirujanos con una amplia variabilidad en experiencia obtendrá unos resultados más cercanos a lo que ocurriría en la práctica real, que si se elige a un grupo de expertos en la técnica, que obtendrán, previsiblemente, mejores resultados. En este último caso, al aplicar este procedimiento a la población general los resultados no serán tan favorables a la nueva técnica como los obtenidos durante el estudio.

El enmascaramiento

El enmascaramiento persigue que las personas involucradas en el estudio, tanto participantes como personal investigador, no sepan en qué grupo (intervención o control) está un paciente concreto. El objetivo es evitar que el conocimiento del tipo de intervención pueda interferir en el resultado final. Si, por ejemplo, el médico que atiende al individuo supiera cuál es la intervención concreta en ese caso, podría modificar su pauta de reconocimiento o de pruebas diagnósticas, alterando el resultado del estudio.

Existen varios tipos de enmascaramiento:

- **Simple ciego:** una de las partes involucradas, habitualmente el paciente, ignora el tratamiento que le ha sido asignado.
- **Doble ciego:** tanto los pacientes como los investigadores ignoran el tratamiento asignado
- **Evaluación ciega por terceros:** cuando otro investigador, sin intervención en el seguimiento de los pacientes y que desconoce la medicación administrada, realiza el análisis de los resultados. Es una alternativa útil cuando el tratamiento es difícil de enmascarar o cuando el enmascaramiento no se puede mantener por razones éticas.

Idealmente, el enmascaramiento debe afectar a todas las fases del estudio. En el caso de intervención con fármacos, el enmascaramiento no suele ser problemático aunque es necesario contar con la colaboración de las compañías que manufacturan el producto para la obtención de los placebos. Sin embargo, con otras tecnologías puede no ser posible, tal y como ocurre con técnicas quirúrgicas o con algunos aparatos diagnósticos.

El seguimiento de los participantes se tendrá que hacer desconociendo la asignación de tratamiento, ya que por muy objetivas que puedan ser las variables de estudio, existe la posibilidad de que los investigadores asociados al proyecto influyan en el resultado. Por último, el enmascaramiento tendrá que afectar también al análisis final de los resultados.

D. El seguimiento

La pauta de exploración a que son sometidos los pacientes debe ser independiente del grupo al que pertenezcan. Se utilizarán los mismos métodos en los individuos sometidos a la intervención y en los del grupo control. El tiempo de seguimiento será lo suficientemente largo como para que se puedan detectar los resultados que se quieren estudiar. Por ejemplo, será de unos días en la evaluación la efectividad de un tratamiento para infección del tracto urinario y de meses o incluso años en el del tratamiento con aspirina en pacientes de infarto. La exploración y pruebas que se realizan serán las más indicadas medir los parámetros que se quieren estudiar.

El seguimiento en el ejemplo de las colecistectomías consistió, durante su estancia en el hospital, en la medición del dolor postoperatorio, el consumo de opiáceos, la función pulmonar y la duración de la estancia hospitalaria. Posteriormente al alta se sometió a los participantes a dos reconocimientos físicos y se utilizaron cuestionarios enviados por correo para determinar la aparición de complicaciones y establecer el tiempo de convalecencia.

E. Análisis e interpretación de los resultados

El análisis supone la comparación de los resultados obtenidos en los grupos. Normalmente los ECA no permiten un tipo de análisis estándar aplicable a todos los casos, sino que dependerá de los parámetros estudiados (medias, proporciones).

Al interpretar los resultados obtenidos, debe valorarse si el protocolo ha sido fielmente seguido o si se ha modificado de tal manera que llegue a invalidar los resultados. Los resultados se deben comparar siempre con los de otros estudios y, en caso de discrepancia, se buscarán posibles explicaciones. Finalmente, se analizará la importancia clínica de la tecnología evaluada, teniendo siempre en cuenta las dificultades que pueden surgir al aplicarla a otras poblaciones cuyas características puedan ser diferentes de las de la población estudiada.

Tabla V. Ventajas e inconvenientes de los ECA

<p>VENTAJAS:</p> <ul style="list-style-type: none">· Permiten un mayor control de los factores de estudio· La asignación aleatoria tiende a controlar los factores de confusión conocidos y desconocidos, aislando el efecto de la intervención.· Permiten una mayor evidencia de la relación causa/efecto, lo que los convierte en el diseño que proporciona resultados más sólidos en evaluación de tecnología.
<p>INCONVENIENTES</p> <ul style="list-style-type: none">· Por razones éticas, existen muchas preguntas que no se pueden responder con este diseño.· Se suele realizar sobre muestras muy seleccionadas, lo que puede restar generalizabilidad a los resultados.· Suelen tener un elevado coste, dependiendo de la duración del estudio, del número de sujetos necesarios y de la modalidad de seguimiento.· Las intervenciones suelen ser estandarizadas y, en ocasiones, alejadas de la práctica clínica. Si en la población se aplica la intervención de forma diferente, no se repetirán los resultados obtenidos en el estudio.· Normalmente estudian una sola intervención y un sólo resultado.

5.2. EL PAPEL DE LOS ESTUDIOS OBSERVACIONALES.

Cuando no es posible, por razones éticas, logísticas o económicas, realizar un estudio experimental, se hace necesario buscar otras formas de evaluación.

En los estudios observacionales el investigador se limita a observar el curso natural de los acontecimientos. Estudia quiénes están expuestos a una tecnología determinada y quiénes no lo están, y dentro de estos dos grupos, quiénes desarrollan la condición de estudio y quiénes no.

El objetivo de los estudios observacionales es obtener estimadores no sesgados y precisos de la asociación entre la tecnología a que se exponen los individuos y un determinado estado de salud.

Los estudios observacionales se dividen en:

1. ESTUDIOS DESCRIPTIVOS: describen la ocurrencia de un hecho. Por tanto, se pueden utilizar para determinar la prevalencia de una condición clínica o enfermedad y para establecer la difusión de una tecnología. Como, por motivos logísticos, no se puede estudiar a toda la población, los estudios descriptivos utilizan muestras representativas de ésta para efectuar las mediciones, y posteriormente inferir (generalizar) los resultados obtenidos a la población diana.

El modelo de selección de la muestra más eficaz es **el muestreo probabilístico**, en el que cada unidad de muestreo tiene una probabilidad conocida de ser incluido en el estudio. Así se puede diferenciar entre:

- **Muestreo aleatorio simple:** otorga a todos los sujetos de una población la misma probabilidad de ser incluidos en la muestra. Normalmente se utilizan números aleatorios generados por ordenador para la obtención de la muestra, tras haber asignado un número a cada sujeto.
- **Muestreo aleatorio estratificado:** se divide la población en diferentes estratos (grupos homogéneos de sujetos) y se hace un muestreo aleatorio simple dentro de cada estrato, de forma que en la muestra queden representados todos los segmentos que nos interesen. Por ejemplo, si en una población muy heterogénea en cuanto a edad interesa que todos los grupos de edad estén representados, la solución es realizar un muestreo aleatorio simple dentro de cada grupo de edad.
- **Muestreo por conglomerados:** simplifica el muestreo cuando una población está agrupada en conglomerados (localidades, edificios, manzanas). Se puede elegir a la totalidad de los individuos de un conglomerado o a una muestra aleatoria simple del mismo.

Los datos en un estudio descriptivo se pueden obtener por medición directa o a través de cuestionarios. Así, las Encuestas de Salud describen, en la medida de lo posible, el estado de salud sentido por una población.

En este punto, se hace necesaria una referencia a los **registros** y las **bases de datos** como fuentes de información muy útiles en evaluación de tecnologías. En general, las bases de datos están más orientadas hacia una determinada enfermedad, tratamiento o procedimiento y recogen una gran cantidad de datos sobre un número reducido de pacientes, siendo su aplicación fundamentalmente clínica. Los registros, por el contrario, recogen un número mínimo de datos de grandes poblaciones, por lo que son especialmente útiles en la investigación de programas de salud, de la influencia de la

distribución geográfica de recursos y de las variaciones regionales en tasas de enfermedad y tratamientos.

2. ESTUDIOS ANALÍTICOS: pretenden comprobar una hipótesis formulada explícitamente sobre la asociación entre la exposición a una tecnología y el estado de salud de la población. Los diseños más importantes de estudios observacionales son:

2.1. Estudios de cohortes

2.2. Estudios de casos y controles.

5.2.1. Estudios de cohortes.

Los estudios de cohortes se definen como aquellos en que los individuos que componen los grupos de estudio son **seleccionados en función de que presenten una determinada característica o estén expuestos a un determinado factor**. Los grupos componen las cohortes (de expuestos y no expuestos) y son seguidos en el tiempo para medir en qué proporción desarrollan las variables de efecto que se desean estudiar.

Figura 2. Estructura básica de un estudio de cohorte.

PARA VISUALIZAR ESTA FIGURA, VAYA A LA PENÚLTIMA PÁGINA

El objetivo principal de los estudios de cohortes es estimar la incidencia de la enfermedad o condición de estudio en cada grupo y establecer comparaciones. A partir de las incidencias se obtiene el **riesgo relativo (RR)**, definido como la probabilidad de que una

enfermedad se desarrolle en el grupo de expuestos en relación a la probabilidad de los no expuestos. Esto es, cuánto más enferman los de un grupo que los de otro. Matemáticamente es la razón entre la incidencia de la enfermedad en expuestos (I_e) y en los no expuestos (I_o). Representándolo con una tabla de contingencia:

		Enfermedad		
		si	no	
Exposición	si	a	b	a+b
	no	c	d	c+d
		a+c	b+d	

- a:** número de expuestos que desarrollan la enfermedad
- b:** número de expuestos que no desarrollan la enfermedad
- c:** número de no expuestos que desarrollan la enfermedad.
- d:** número de no expuestos que no desarrollan la enfermedad.
- a+b:** total de expuestos.
- c+d:** total de no expuestos.
- a+c:** total de enfermos (personas que desarrollan la enfermedad).
- b+d:** total de sanos (personas que no desarrollan la enfermedad).

$$RR = I_e / I_o = a/(a+b) / c/(c+d)$$

El RR mide el grado de asociación entre una tecnología y el estado de salud de estudio. Tiene un rango de valores entre 0 e infinito. Así, si se mide la relación entre una tecnología y mortalidad, los posibles valores de los RR obtenidos se interpretarán de la siguiente forma:

- $RR < 1$. Menor probabilidad de morir en el grupo expuesto a la tecnología. En este caso la tecnología se considera un factor de protección.
- $RR = 1$. Existe la misma probabilidad de morir en ambos grupos. No existe asociación alguna entre la tecnología y la mortalidad.
- $RR > 1$. Mayor riesgo de morir en el grupo expuesto. La tecnología será considerada un factor de riesgo para la enfermedad de estudio.

Los estudios de cohortes también se utilizan para evaluar relaciones dosis-respuesta o para estimar la proporción de casos (enfermedad) atribuible a la exposición. De esta forma se puede evaluar el impacto potencial que tendría una intervención que eliminara la exposición a través del cálculo del **riesgo atribuible (RA)**, que se calcula:

$$\text{RA en expuestos} = (\text{Ie}-\text{Io})/\text{Ie}$$
$$\text{RA poblacional} = (\text{Ip}-\text{Io})/\text{Ip}^*$$

* Ip: incidencia total en la población

Para ilustrar la descripción de las fases de que consta un estudio de cohortes, se utilizará como ejemplo un estudio recientemente publicado cuyo objetivo es determinar si la exposición a aspirina previene del desarrollo de determinados tipos de cáncer (Schreinemachers, 1994).

A. Selección de la cohorte.

Según las condiciones en que se desarrolle el estudio, existen varias posibilidades:

- Cohorte poblacional. Cuando la exposición es frecuente, una población de un área geográfica o administrativa determinada contará con el suficiente número de individuos expuestos y no expuestos para establecer comparaciones de ocurrencia de enfermedad.
- Cohorte expuesta. Cuando la exposición es poco frecuente, el método anterior requerirá el seguimiento de una población extraordinariamente amplia para que entre ellos exista un número suficiente de expuestos y obtener suficientes enfermos al final del estudio. En este caso, suele ser más eficiente elegir una cohorte sólo de expuestos y compararla con otra cohorte de iguales características que la anterior, pero sin la exposición de interés.
- Cohortes de fácil seguimiento. Otras veces interesa más que se pueda recoger la información de forma segura y fácil. Se pueden buscar cohortes que por su estabilidad (instituciones) o su cooperabilidad (médicos, enfermeras) aseguren que el número de pérdidas sea pequeño.

En el ejemplo de la exposición a aspirina, al ser éste un hecho frecuente, se utilizó una muestra probabilística representativa de la población no institucionalizada de los EE.UU. recogida en una encuesta de salud realizada entre 1971 y 1975.

B. Clasificación de los participantes (expuestos / no-expuestos).

Primeramente se ha de tener la certeza de que los participantes no tienen la enfermedad de interés al inicio del estudio. Posteriormente, a través de diferentes métodos

(cuestionarios, exámenes físicos o pruebas de laboratorio) se recoge la información sobre si están o no expuestos a la tecnología que se quiere evaluar y se clasifica a los participantes en expuestos y no-expuestos.

En el estudio de la aspirina la información sobre la exposición se obtuvo a través de los datos recogidos en la encuesta, en la que se preguntó a los participantes sobre la ingesta de aspirina en los 30 días previos.

C. Seguimiento de la cohorte y medición de los desenlaces. El seguimiento tiene varios objetivos: 1) Controlar modificaciones en el nivel de exposición (fumadores que han abandonado el hábito), 2) Registrar la aparición del desenlace (enfermedad, muerte, etc) 3) Mantener el contacto entre el grupo investigador y los participantes y mantener su nivel de cooperación.

Se debe hacer siempre una definición de desenlace rigurosa y aplicarla de igual forma a los dos grupos. También las exploraciones o pruebas a las que se somete a los participantes durante el tiempo de seguimiento, así como las fuentes que se van a utilizar para obtener la información sobre los desenlaces ocurridos, deben ser independientes de la exposición. Si la pauta de seguimiento se modifica en función del grupo al que pertenecen, la diferencia en la frecuencia de aparición de los desenlaces no se deberá, exclusivamente, a la exposición, sino que potencialmente podrá estar influida por esa modificación.

Por otro lado, el tiempo y el método de seguimiento dependerán, como en los estudios experimentales, de la enfermedad de estudio.

El trabajo que evalúa el papel de la aspirina en la prevención de cáncer, utilizó como definición de caso la contenida en la Clasificación Internacional de Enfermedades (CIE) en su 9ª versión. Para identificar los desenlaces, el seguimiento se hizo a través de la combinación de entrevistas, historiales médicos y certificados de defunción. Se obtuvieron además datos sobre edad, sexo, raza, estatus socioeconómico y enfermedades concurrentes que requerían el tratamiento con antiinflamatorios, es decir, todas aquellas variables que se presumía pudieran tener alguna influencia sobre la exposición o el resultado.

LOS SEGOS EN LOS ESTUDIOS DE COHORTES

Se define al sesgo como todo error sistemático que se comete en un estudio y que altera la validez de sus resultados.

1. **Sesgo de información:** aparece cuando el seguimiento es diferente en los expuestos que en los no expuestos. Si, por ejemplo, se está estudiando el efecto de los anticonceptivos orales en la aparición de trombosis de venas de miembros inferiores y el médico que realiza las exploraciones periódicas lo hace con mayor detalle a pacientes expuestas (al sospecharse que en este grupo será más

frecuente), es bastante probable que detecte mayor número de enfermas en este grupo; también es probable que tienda a dar por positivas más frecuentemente a las pacientes con signos dudosos. Por tanto, la frecuencia de signos de trombosis será mayor en el grupo expuesto y no se podrá determinar en qué proporción esto se debe al fármaco. Para eliminar en lo posible este sesgo, se deberá evitar que las personas que realizan el seguimiento tengan conocimiento del estatus de exposición de los participantes en el estudio.

2. **Sesgo de confusión:** se produce por una distribución diferencial de los factores de confusión, que son aquellos que actúan como factor de riesgo para la enfermedad y están asociados a la exposición que se quiere estudiar pero sin ser una variable intermedia en la secuencia causal. La solución a este problema pasa por el análisis estadístico de los datos (análisis estratificado o multivariante de los datos).

De una forma resumida se presentan a continuación las ventajas e inconvenientes de este diseño.

Tabla VI. Ventajas e inconvenientes de los estudios de cohortes

<p>VENTAJAS</p> <ul style="list-style-type: none"> · Permiten el cálculo de la incidencia, riesgo relativo y riesgos atribuibles. · Es una estrategia potente para investigar relaciones causales. · Permite medir las variables de exposición de forma exacta y conocer sus variaciones en el tiempo. · Permite estudiar varios efectos de una misma exposición.
<p>INCONVENIENTES</p> <ul style="list-style-type: none"> · No son eficientes en enfermedades con poca incidencia (necesitaría grandes cohortes y seguimientos muy prolongados). · Son caros.

Estudios de cohortes retrospectivos.

Se realizan en el momento en que el desenlace (enfermedad) ya ha ocurrido y se retrocede en la historia personal de los individuos para detectar la exposición. Así, se clasifica a los participantes en expuestos y no-expuestos y se calcula la frecuencia de enfermedad en ambos grupos. Al ser retrospectiva la medición de la exposición se pierde exactitud en comparación con el diseño prospectivo que mide la exposición directamente; pero la mayor ventaja de este diseño es que reduce enormemente los costes al no tener que realizar un seguimiento durante un largo período de tiempo.

Un problema que se puede presentar en este diseño retrospectivo, y que pondría en peligro la validez del estudio, es el **sesgo de selección** o distorsión en la estimación del efecto estudiado derivado de la incorrecta elección de los participantes si la inclusión en

alguno de los grupos depende de la presencia o ausencia de enfermedad. Para evitar este sesgo la recogida de datos sobre la exposición pasada debe ser ciega, esto es, realizada por personas que no conozcan el estado de “enfermo” o “no-enfermo” de los participantes. Otro problema surgirá con la dificultad de determinar exposiciones muy antiguas, ya sea por no existir registros de calidad, o por el efecto de la “memoria selectiva” (recuerdan las exposiciones previas más los que tienen una enfermedad potencialmente relacionada con ésta que los que no presentan ningún problema de salud)

5.2.2. Estudios de Casos y Controles

Se definen por **la selección de los participantes en función de la presencia (casos) y ausencia (controles) de la condición de estudio**. En evaluación de tecnologías la epidemiología estudia el nivel de exposición en el pasado a la tecnología sanitaria que se esté estudiando y se compara su frecuencia en ambos grupos.

Figura 3. Estructura básica de un estudio de casos y controles

PARA VISUALIZAR ESTA FIGURA, VAYA A LA PENÚLTIMA PÁGINA
--

El estimador de efecto en los estudios de casos y controles es la **Razón de Ventaja (Odds Ratio -OR-)**. Se calcula de la siguiente forma:

		Exposición		
		si	no	
Enfermedad	si	a	b	a+b
	no	c	d	c+d
		a+c	b+d	

a: número de enfermos (casos) que han sido expuestos

b: número de no enfermos (casos) que no han sido expuestos

c: número de no enfermos (controles) que han sido expuestos

d: número de no enfermos (controles) que no han sido expuestos

a+b: total de enfermos (casos)

c+d: total de no enfermos (controles)

a+c: total de expuestos

b+d: total de no expuestos

Ventaja de la exposición entre los casos: a/b

Ventaja de la exposición entre los controles: c/d

$$\text{Razón de ventaja (OR)} = (a/b)/(c/d) = ad/bc$$

Para ilustrar el desarrollo de este tipo de diseño se va a utilizar un estudio que se realizó con el objetivo de determinar si el cribado de cáncer de colon con sigmoidoscopio reducía la mortalidad por esta enfermedad (Selby, 1992).

Las fases de que consta un estudio de casos y controles son:

A. Definición de los casos

La enfermedad o condición que se va a estudiar debe definirse “a priori” con precisión, fijando criterios diagnósticos uniformes para ser aplicados por igual a todos los participantes.

Un tema de gran importancia es la diferencia que supone la elección de casos prevalentes (todos los casos) o sólo de casos incidentes (casos nuevos) de la condición de estudio. En el caso de que se pretenda estudiar la relación entre una enfermedad con algún fármaco, por ejemplo, la relación entre insuficiencia renal y el consumo prolongado de aspirina, la elección entre escoger casos incidentes o prevalentes puede tener muchas consecuencias en el resultado. Así, si se escogieran casos prevalentes se podría estar

midiendo la asociación entre el factor de exposición y una expresión menos grave de la enfermedad (los más graves habrían fallecido y no podrían entrar a formar parte del estudio). En este caso, de obtenerse en el estudio una asociación entre el consumo de aspirina con el desarrollo de insuficiencia renal, la enfermedad podría estar causada por otro factor y deberse esta asociación a que la aspirina proteja, de alguna forma, de fallecer por esta enfermedad.

En el ejemplo de la sigmoidoscopia para cribado de cáncer de colon, los casos se obtuvieron de entre las personas mayores de 45 años inscritas en un programa de cuidados de salud, diagnosticados de adenocarcinoma de colon y que habían fallecido antes de finales de 1988. Se revisaron las historias clínicas para verificar el diagnóstico histológico y la localización del cáncer y se excluyeron a aquellos que llevaban menos de un año adheridos al programa.

B. Definición de los controles.

Es probablemente el paso más problemático de este diseño ya que de su correcta selección dependerá la validez del estudio. El grupo control se utiliza para comparar la historia de exposición de los casos con la de los individuos que están libres de la enfermedad de estudio. Así, se pretende obtener de este grupo un estimador de la tasa de exposición esperada en los casos si la exposición y la enfermedad no estuvieran asociadas. Los controles deben, por tanto, ser representativos de la población de la que proceden los casos pero sin la enfermedad o condición de estudio. Deberían haber tenido la misma probabilidad de exposición a la tecnología estudiada que los casos. Este requerimiento dirigirá los criterios de inclusión.

Habitualmente, los controles se extraen por muestreo de dos diferentes ámbitos:

- Hospitalarios: más fáciles de obtener pero podrían ser substancialmente diferentes de la población de estudio.
- Poblacionales o comunitarios, que se obtienen de la población de la que han salido los casos.

En el caso que se está revisando los controles se eligieron, al igual que los casos, de entre los miembros del programa de cuidados de salud y se aparearon con los casos de acuerdo con la edad, el sexo y la fecha de entrada al programa. Se decidió escoger 4 controles por cada caso. Cada control tenía que estar vivo y no se excluyeron aquellos diagnosticados de cáncer o pólipos.

C. Recogida de la información sobre la / las exposiciones.

Una vez escogidos los participantes, el siguiente paso consiste en la búsqueda de información sobre la historia personal de exposición a la tecnología tanto de casos como de controles. La fuente de información debe ser la misma para ambos grupos. Dependiendo de la condición clínica que se estudie se utilizarán entrevistas, cuestionarios o la revisión de las historias clínicas de los participantes en el estudio.

En el caso que se está revisando, se quería obtener documentación sobre si a los participantes se les había realizado una sigmoidoscopia en los 10 años previos al inicio de los síntomas que habían llevado al diagnóstico. La fuente de información elegida fue la revisión de las historias clínicas de los participantes.

D. Análisis e interpretación de los resultados

Además de los puntos comunes a los otros diseños epidemiológicos descritos que este paso debe contener, en los estudios de casos y controles se tiene que hacer un especial esfuerzo en evaluar las posibles limitaciones del estudio, al ser diseños muy susceptibles de sufrir sesgos.

SESGOS EN LOS ESTUDIOS DE CASOS Y CONTROLES

- **Sesgo de selección:** se produce cuando existe una desigualdad en la selección de los casos y los controles basada en el estatus de expuesto o no expuesto o en el nivel de esa exposición. Algunos de los más frecuentemente apuntados son:

- Paradoja de Berkson: cuando se escogen controles hospitalarios que pueden diferir sistemáticamente de la población general. El sesgo aparece si la probabilidad de hospitalización es diferente en los casos que en los controles, o en los expuestos y no expuestos.

- Falacia de Neyman: la utilización de casos prevalentes cuando la exposición es un factor pronóstico.

- **Sesgos de información:** puede existir sesgo de memoria si las personas que sufren la enfermedad son más conscientes de las exposiciones pasadas (madres de recién nacidos con malformaciones congénitas pondrán más interés en recordar posibles exposiciones a fármacos durante el embarazo que madres de recién nacidos completamente sanos), sesgos introducidos por el entrevistador que encueste más exhaustivamente a los casos que a los controles o por realizar diferente pauta exploración en función del grupo a que pertenezcan.

Tabla VII. Ventajas e inconvenientes de los estudios de casos y controles

VENTAJAS <ul style="list-style-type: none">· Son eficientes en el estudio de enfermedades o condiciones raras o con largos periodos de latencia· Permiten evaluar el efecto de múltiples exposiciones sobre una enfermedad· Resultan relativamente rápidos y baratos
INCONVENIENTES <ul style="list-style-type: none">· Mayor susceptibilidad de sufrir sesgos· No estiman incidencias· Solo permiten el estudio de una enfermedad· Los casos (pacientes diagnosticados) no son una muestra representativa de todos los enfermos, quedando fuera, habitualmente, los casos fatales y los no diagnosticados.

6. EVALUACIÓN DE PRUEBAS DIAGNÓSTICAS

Un test diagnóstico es considerado de buena calidad si es capaz de clasificar correctamente a los miembros de la población a la que se aplica, dando resultados positivos en las personas enfermas y negativos en las carentes de la enfermedad de estudio. Es un grave error asumir que el resultado de una prueba diagnóstica es una representación fiel de la situación real de salud de un individuo. Por el contrario, el médico que reconoce las imperfecciones de un test se preguntará: ¿A la vista de este resultado, qué grado de seguridad tengo acerca del estado de salud de este paciente?.

La importancia de una prueba diagnóstica desde el punto de vista de evaluación de tecnologías se basa en varios factores. Al evaluar una prueba diagnóstica, no obstante, no debe tenerse en cuenta la calidad del test aisladamente, sino que se deben valorar también las acciones que se van a poner en práctica cuando el resultado sea anormal. Así, una prueba cuyo resultado no tenga influencia sobre el manejo del paciente ni sobre su estado de salud, tendrá poca importancia en salud pública, y su evaluación no será prioritaria. Otros factores que se deben valorar son el precio de la prueba así como su inocuidad, ya que pueden existir métodos menos invasivos o más económicos para obtener una información suficiente para nuestros propósitos.

La calidad de una prueba diagnóstica se mide en términos de validez y seguridad. La **validez de un test** es el grado en que mide lo que se supone debe medir y para determinarla se debe comparar con otro test de probada eficacia en la enfermedad de estudio o con el verdadero resultado en el caso de que pueda ser conocido. Estos

términos de comparación se denominan patrón de oro, “gold standard” o criterio de referencia. La validez se mide a través de la sensibilidad y la especificidad del test.

La **sensibilidad** de un test es la probabilidad de que un sujeto con la enfermedad sea clasificado correctamente como positivo.

La **especificidad** mide la probabilidad de que un sujeto sano tenga un resultado negativo en la prueba.

Así, se toma un grupo de personas con la enfermedad de estudio y un grupo de personas sanas y se les somete a la prueba diagnóstica que se quiere evaluar. Se tabulan los resultados en cuanto a positividad y negatividad de la prueba y se calcula la sensibilidad y la especificidad.

		Enfermedad		
		si	no	
TEST	+	a	b	a+b
	-	c	d	c+d
		a+c	b+d	

VERDADEROS POSITIVOS (**a**): número de individuos con la enfermedad en los que el resultado de la prueba es positivo.

FALSOS POSITIVOS (**b**): número de individuos sanos en los que el resultado de la prueba es positivo.

FALSOS NEGATIVOS (**c**): número de individuos enfermos en los que la prueba es negativa.

VERDADEROS NEGATIVOS (**d**): número de individuos sanos en los que la prueba es negativa.

a+b: total de positivos.

c+d: total de negativos.

a+c: total de enfermos.

b+d: total de sanos.

$$\text{Sensibilidad} = (a / a+c) \times 100$$

$$\text{Especificidad} = (d / b+d) \times 100$$

Las pruebas diagnósticas deben ser de muy alta sensibilidad cuando:

- La enfermedad sea grave
- No se pueda permitir la pérdida de ningún caso

- La enfermedad sea tratable
- Un resultado falso positivo no perjudique al paciente (tratamiento sin graves efectos secundarios).

Un ejemplo lo constituyen las enfermedades infecciosas curables.

Sin embargo, se precisa una especificidad muy alta cuando:

- La enfermedad sea grave pero sin tratamiento disponible.
- Sea de gran interés la ausencia de enfermedad, por ser una enfermedad que pueda generar algún tipo de discriminación hacia el paciente.
- Un falso positivo sea perjudicial para el paciente.

Ejemplo: cáncer no tratable.

En los casos en que se puede modificar el punto a partir del cual se considera positiva una prueba, se ve que aumentos de la sensibilidad de un test se acompañan de disminuciones en su especificidad. Así, si se varía el número de milímetros de depresión del segmento ST que se consideran definitorios de resultado anormal al electrocardiograma de esfuerzo como predictor de enfermedad coronaria severa, se modificarán las características de validez del test. Si se fija la positividad a partir de 0.5 mm el test tendrá una gran sensibilidad al dar como positivos prácticamente a todos los pacientes que sufren esta condición, pero a costa de producir muchos falsos positivos. Si se aumenta a 2 mm el nivel de positividad, la especificidad será muy alta pero quedarán sin diagnosticar muchos pacientes con enfermedad coronaria, pero con depresión del segmento ST inferior a este punto.

Una excelente forma de ver este efecto es a través de la representación gráfica de la sensibilidad (en ordenadas) frente a la expresión {1-especificidad} (en abscisas), o lo que es igual, la proporción de verdaderos positivos frente a la de falsos positivos de los diferentes valores que tenga la prueba. Este tipo de gráficas se denominan curvas ROC (Receiver Operating Characteristics), desarrolladas en los años 50. Presentan todos los pares sensibilidad / especificidad resultantes de la variación continua de los puntos de corte en todo el rango de los resultados observados.

En el ángulo superior izquierdo se situaría la prueba diagnóstica perfecta por la que todos los individuos con la enfermedad son detectados (sensibilidad del 100%) y ninguno sin la enfermedad es falsamente positivo ({1-especificidad} del 0%). Por tanto, el punto de la curva en la que se representan los diferentes valores de la prueba diagnóstica, que más se acerque a este ángulo será el mejor punto de corte de la prueba. Las curvas ROC son, además, un instrumento muy útil en la comparación de dos o más pruebas diagnósticas.

En el gráfico (ver última pag.), cada punto representado es un par sensibilidad/(1-especificidad) correspondiente a los 5 niveles de decisión determinados (mm de depresión del segmento ST utilizados para diagnosticar enfermedad coronaria severa). La mayoría de las curvas ROC presentan una gran pendiente inicial en la que la sensibilidad aumenta mucho, mientras que la proporción de falsos positivos no se modifica, o lo hace

muy poco, hasta llegar a un punto a partir del cual se modifica muy poco la sensibilidad aumentando mucho los falsos positivos. La línea diagonal del gráfico es la representación de una prueba sin discriminación, con igual distribución de los resultados en los dos grupos.

Figura 4. Curva ROC para el electrocardiograma de ejercicio como predictor de enfermedad coronaria

PARA VISUALIZAR ESTA FIGURA, VAYA A LA ÚLTIMA PÁGINA

La **seguridad** de un test es la capacidad para predecir la presencia o ausencia de enfermedad y se mide en términos de los valores predictivos, positivo y negativo.

El **valor predictivo positivo** de un test es la probabilidad de que una persona con un resultado positivo esté realmente enferma.

El **valor predictivo negativo** es la probabilidad de que una persona con un resultado negativo del test esté realmente sana.

Para calcular estos valores, se necesita obtener una muestra representativa de la población a la que se va a aplicar posteriormente la prueba, y en la que la prevalencia de

enfermedad sea igual a la poblacional. Se someterá a la población a la prueba que se quiere evaluar y al patrón de referencia, tabulándose los resultados como se hizo previamente. Aplicando el mismo sistema de notación que se ha venido utilizando a lo largo de este texto, la expresión matemática de estos valores será:

$$\text{Valor predictivo positivo (VPP)} = (a / a+b) \times 100$$

$$\text{Valor predictivo negativo (VPN)} = (d / c+d) \times 100$$

Los valores predictivos dependen de la prevalencia de la enfermedad, de forma que cuanto menor sea ésta, menor será el valor predictivo de un resultado positivo, sea cual sea la sensibilidad y especificidad del test. Es decir, con prevalencias muy pequeñas es más improbable que alguien que presenta un resultado positivo tenga realmente la enfermedad, siendo más frecuente que el resultado represente un falso positivo.

Un error que se encuentra con cierta frecuencia en la literatura médica es el cálculo del valor predictivo a partir de un estudio que ha tomado el mismo número de sanos que de enfermos. En este caso se ha fijado artificialmente la prevalencia en el 50% y, al ser los valores predictivos dependientes de la prevalencia de la enfermedad, los resultados obtenidos de esta forma serán, obviamente, erróneos.

Un test de alto valor predictivo positivo debe ser utilizado cuando el tratamiento de falsos positivos pudiera tener graves consecuencias.

El **valor global** del test indica la proporción de resultados correctamente clasificados entre la totalidad de pruebas efectuadas y vendrá definido por la siguiente expresión matemática.

$$\text{Valor global} = (a+d / a+b+c+d) \times 100$$

Esta medida tiene muy poco valor en la práctica real ya que normalmente interesará conocer la sensibilidad y especificidad de la prueba por el diferente significado que tienen en cuanto a las características de gravedad de la condición clínica de estudio.

Ejemplo:

Con el objetivo de evaluar un método simple de tinción de bacterias con azul de metileno para diagnosticar infecciones del tracto urinario, se analizaron 1064 muestras de orina de un laboratorio. Como criterio de referencia se utilizó el cultivo en placas de agar de las mismas muestras, considerándose que existía infección si había un

crecimiento superior a las 10^5 UFC /ml de orina. Así, se determinaron 346 bacteriurias (32.5% del total de muestras analizadas)

		Enfermedad		
		si	no	
TEST	+	285	12	297
	-	61	706	767
		346	718	1064

A partir de los resultados obtenidos resumizados en la tabla, se calculan las siguientes medidas.

$$\text{Sensibilidad} = (285/346) \times 100 = 82.4\%$$

$$\text{Especificidad} = (706/718) \times 100 = 98.3\%$$

$$\text{VPP} = (285/297) \times 100 = 96\%$$

$$\text{VPN} = (706/767) \times 100 = 92\%$$

Si aplicáramos esta prueba (con esta sensibilidad y esta especificidad) a una muestra de 1.000 personas procedentes de una población en la que la prevalencia de infección urinaria fuera del 5%, la tabla quedaría:

		Enfermedad		
		si	no	
TEST	+	41	16	57
	-	9	934	943
		50	950	1000

A partir de los resultados obtenidos en la tabla se pueden calcular los valores predictivos:

$$\text{VPP} = (41/57) \times 100 = 71.9\%$$

$$\text{VPN} = (934/943) \times 100 = 99\%$$

Para mejorar la eficiencia en el diagnóstico de una enfermedad existen dos métodos de combinación de pruebas diagnósticas: en serie y en paralelo. La combinación en serie consiste en la aplicación consecutiva de la segunda prueba en caso de que la primera sea positiva. Así se mejora la especificidad ya que los falsos positivos de una prueba son comprobados por una segunda. En la combinación en paralelo se realizan las dos pruebas simultáneamente y sólo se considera negativo en caso de que ambas lo sean. Se aumenta,

por tanto, la sensibilidad global ya que los falsos negativos de una prueba son comprobados por la otra.

Etapas de un estudio de evaluación de pruebas diagnósticas.

Debido a que una misma prueba diagnóstica puede ser utilizada para el estudio de varias condiciones clínicas (por ejemplo, el diagnóstico por imagen con resonancia magnética puede aplicarse a gran número de patologías), se tendrá que decidir cuál se va a evaluar. Para guiar esta decisión se deben tener en cuenta los siguientes factores:

- Frecuencia de la enfermedad o condición. Se priorizarán aquellas aplicaciones que afecten a mayor número de personas.
- Coste de la tecnología: las tecnologías más caras deben ser evaluadas con el máximo rigor para evitar gastos innecesarios al sistema sanitario.
- Existencia de un criterio de referencia que permita la comparación
- Impacto potencial sobre el manejo del paciente: se debe hacer mayor esfuerzo en mejorar la calidad de las pruebas diagnóstica de las condiciones clínicas para las que hay tratamiento eficaz.
- Impacto potencial sobre el resultado clínico. Se dará prioridad a la evaluación de aquellas pruebas, que en definitiva, contribuyan a la mejora del estado de salud de los pacientes.
- Existencia de métodos diagnósticos inadecuados para la condición de estudio.

Una vez decidida la enfermedad o condición clínica que se va a estudiar, el diseño del estudio será similar al de cualquier otro estudio observacional y los pacientes elegidos no deberán diferir substancialmente de aquellos a los que posteriormente se aplicará la prueba. Se deben incluir, por tanto, pacientes con diversos estadios clínicos de la enfermedad, ya que si se escogieran sólo los más graves, por ejemplo, darían mayor número de positivos, aumentando de una forma ficticia la sensibilidad del test.

La elección del criterio de referencia resulta, en ocasiones, muy complejo, bien por ser costoso o potencialmente perjudicial para el individuo o porque simplemente no se ha desarrollado un mejor método de diagnóstico. Utilizar un criterio de referencia de baja calidad puede conducir, sin embargo, a que la nueva medida sea considerada falsamente de baja calidad. Por ejemplo, con un patrón de oro de baja sensibilidad muchos resultados positivos de la nueva prueba serán considerados erróneamente falsos positivos, disminuyendo el valor de la prueba.

7. EVALUACIÓN ECONÓMICA

La evaluación económica es un instrumento analítico de creciente utilización como soporte a la toma de decisiones relacionadas con la financiación y regulación de las tecnologías sanitarias. Es un instrumento básico en el establecimiento de prioridades dentro del sistema de salud para la asignación de recursos. La orientación que la evaluación económica puede aportar a las decisiones administrativas puede referirse a:

- Introducción o autorización de comercialización de una tecnología nueva o el mantenimiento de una ya existente.
- Fijación del precio de una tecnología.
- Priorización en la inversión en determinadas tecnologías.
- Establecimiento de recomendaciones o incentivos para la utilización y difusión de una tecnología.

La asignación de recursos a una alternativa concreta conlleva siempre el coste adicional de no poder lograr otros objetivos alternativos también deseables. En la vida diaria se toman continuamente este tipo de elecciones para emplear óptimamente los recursos económicos, comparando la satisfacción que generan los bienes o servicios obtenidos con los costes que supone su obtención.

Dentro de los servicios de salud la evaluación económica establece que los recursos deben ser aplicados a aquellas alternativas que generen mayor beneficio en relación con el coste ocasionado. Se puede referir a la compra de bienes, la prestación de determinados servicios o la distribución geográfica de éstos. Pero en ningún caso se debe olvidar que la medida de la eficiencia

estará siempre subordinada a la eficacia y efectividad de la tecnología, su aceptabilidad y su pertinencia. Esto es, debe estar condicionada a una evaluación clínica previa o simultánea, siendo los resultados obtenidos de los estudios epidemiológicos o de la síntesis de información existente los que van a estimar la efectividad de la tecnología.

La evaluación de la eficiencia de las tecnologías se basa en dos conceptos básicos: los costes y los beneficios. Estos conceptos variarán según el punto de vista que se adopte en el análisis. La evaluación económica se puede realizar desde la perspectiva del paciente individual, del médico, del hospital, del sistema de salud, del gobierno o de la sociedad. El punto de vista más global, que abarca más aspectos, será el de la sociedad, que incluirá todos los costes y consecuencias, sin tener en cuenta sobre qué o quién recaigan.

Normalmente, esta aproximación social de la eficiencia tiene que adoptar medidas y valores que no tienen precios de mercado establecidos. A esta dificultad se añade que además, determinados costes o ahorros no deben ser incluidos en el análisis porque son transferidos de un sector a otro en vez de recaer sobre la sociedad.

La primera cuestión a plantear en este punto es: ¿Qué costes se deben incluir en el análisis de evaluación de una tecnología? Desde la perspectiva más amplia son:

- Costes para el sistema sanitario. Gastos de personal sanitario, material médico empleado, servicio de hostelería, utilización de equipos, electricidad, calefacción...etc. Se pueden clasificar en fijos (que no dependen del nivel de actividad de la organización) y variables (relacionados con el nivel de actividad)
- Costes para el paciente y su familia: días en que no se acude al trabajo, transporte de familiares al hospital...etc.
- Costes externos: cuando personas no directamente relacionadas con un programa sufren un incremento de gastos, al repercutirse el coste a toda la sociedad. Por ejemplo, cuando un sistema especial de purificación del agua incrementa su precio para el consumidor.

Los costes, a su vez, se clasifican en:

1. Tangibles: que pueden ser valorados con precios de mercado.
2. Intangibles: a los que no se puede asignar un precio de mercado.

Las evaluaciones económicas se pueden clasificar en función de si examinan simultáneamente los costes y las consecuencias, y de si estudian una única alternativa o comparan dos o más.

Tabla VIII. Clasificación de los estudios de evaluación económica

¿SE EXAMINAN COSTES Y CONSECUENCIAS SIMULTÁNEAMENTE?

		NO		SI
		SOLO CONSECUENCIAS	SOLO COSTES	
¿SE COM- PARAN DOS O MÁS ALTER- NATIVAS ?	NO	Descripción de consecuencias (outputs)	Descripción de costes (inputs)	Descripción de costes y consecuencias
	SI	Evaluación de eficacia / efectividad	Análisis de costes	1. MINIMIZACIÓN DE COSTES 2. ANÁLISIS COSTE-EFECTIVIDAD 3. ANÁLISIS COSTE-UTILIDAD 4. ANÁLISIS COSTE-BENEFICIO

En este documento sólo se va a hacer una breve revisión de los 4 métodos más empleados en la comparación de dos o más alternativas, incluyendo análisis de costes y consecuencias simultáneamente.

Minimización de costes

El análisis de minimización de costes pretende identificar y cuantificar los costes de 2 o más procedimientos cuyas consecuencias son clínicamente similares y que se llevan a cabo en pacientes con las mismas condiciones basales, con el fin de elegir el procedimiento más barato. Un ejemplo lo podría constituir la comparación de dos procedimientos quirúrgicos para la apendectomía, que tuvieran la misma efectividad y seguridad. En este caso, se elegiría aquella opción que tuviera menor coste para el sistema sanitario.

Su máxima limitación es que no permite establecer comparaciones de proyectos de diferente naturaleza, siendo ésta la situación habitual a la que se enfrentan los administradores sanitarios.

Análisis coste-efectividad.

Se puede utilizar cuando las consecuencias de la aplicación de diferentes tecnologías se puedan expresar en medidas naturales de efectividad. Algunos ejemplos de estas medidas podrían ser: casos tratados adecuadamente, número de vidas salvadas, años de vida ganados, días libres de síntomas o dolor o casos diagnosticados con éxito.

La comparación entre 2 o más alternativas en este tipo de análisis se realiza en términos de costes por unidad de efecto (razón coste/efecto), como miles de pesetas invertidos por cada cáncer de cérvix diagnosticado precozmente a través de un programa de cribado o incluso por cada muerte evitada.

Su mayor inconveniente se debe a la dificultad para ajustar las medidas de efectividad derivadas de las distintas alternativas. Asimismo, no aporta información de si los costes superan a las consecuencias ni permite comparar proyectos con diferentes consecuencias.

Análisis coste utilidad.

Para algunos autores es una variedad del anterior. Sería una forma de evaluación económica en la que las consecuencias del procedimiento o programa se expresan en unidades basadas en la utilidad. Cuando se habla de utilidad se hace en referencia al estado subjetivo que experimentan las personas en diferentes estados de salud.

La medida de utilidad más extendida son los “años de vida ajustados por calidad” o QALYs (Quality-Adjusted Life Years). Esta unidad combina una medida cualitativa especial de la calidad de vida con una medida cuantitativa de años de vida para obtener una única medida de tiempo de vida útil. La comparación se realiza en términos de coste por unidad de utilidad (razón coste/utilidad).

El mayor problema que presenta es medir la utilidad derivada de una intervención. La calidad de vida de un sujeto, al ser una medida de su sentir subjetivo, es difícilmente medible. Sin embargo, cada vez existe un mayor interés por programas que tienen como objetivo aumentar la calidad de vida de la población. Para evaluar estas intervenciones el análisis coste utilidad proporciona la metodología más ajustada.

Análisis coste-beneficio

Es la forma más amplia de evaluación económica. Ha sido utilizada como ayuda a la toma de decisiones en muchas áreas de las políticas económicas y sociales en las últimas décadas.

Su característica más importante es que asigna valor monetario tanto a los costes como a las consecuencias de la aplicación del programa o intervención que se desea evaluar, por lo que permite la comparación de proyectos de diferente naturaleza.

Entre sus desventajas destaca la orientación eficientista o economicista del método, lo que podría ir en detrimento de la equidad.

Todos estos métodos de evaluación económica son de un gran interés para la evaluación de tecnologías sanitarias al aportar información de la eficiencia de las diferentes alternativas sobre las que tienen que decidir los administradores sanitarios.

8. DISCUSIÓN

La evaluación de tecnologías es una técnica sistemática de generación de datos o recogida de la información disponible para que, a través de su análisis crítico y síntesis, se puedan obtener recomendaciones válidas para ayudar en la toma de decisiones sanitarias.

Las técnicas más utilizadas por los organismos que realizan evaluaciones de tecnologías sanitarias son las basadas en la revisión y síntesis de la evidencia relevante, así como las que se apoyan en el juicio de expertos. La evaluación de tecnologías precisa de respuestas rápidas a las preguntas que plantea, ya que estamos asistiendo a la creación casi constante de nuevas tecnologías que muchas veces se incorporan a la práctica clínica habitual sin haber demostrado inequívocamente su efectividad.

Una de las más importantes aportaciones de la evaluación de tecnologías a los sistemas sanitarios ha sido dar gran importancia a la calidad de la información y el énfasis puesto en mejorarla. Así, al trabajar con la información existente, la calidad del trabajo de evaluación dependerá, en gran medida, de la calidad de los estudios que se utilicen. Por otro lado, la información que se recoge no es, muchas veces, toda la que existe sobre el tema de evaluación. Para evitar en lo posible el sesgo de publicación se buscarán aquellas fuentes no susceptibles de estar sesgadas o se combinará el mayor número de técnicas de búsqueda bibliográfica a nuestro alcance, a fin de completar la información recogida.

También se debe poner el máximo interés en realizar de la forma más objetiva posible las revisiones y resúmenes de esta información. En este sentido, el meta-análisis constituye una herramienta de gran valor, al aportar sus dos pilares básicos: la recogida sistemática de la información y la obtención de un resultado cuantitativo. Debería ser utilizada siempre que sea posible, combinándola o no con los juicios de expertos.

Un tema de máximo interés en evaluación de tecnologías es la difusión de los informes realizados. A pesar de que no son muy abundantes los estudios que han investigado el impacto que los informes de evaluación de tecnología tienen sobre la práctica clínica, parecen apuntar al hecho de que esta influencia es todavía pequeña.

En contra de la innovación tecnológica están una serie de factores que deben ser tenidos en cuenta. Por ejemplo, las teorías prevalentes y explicaciones ampliamente aceptadas por la sociedad científica son, en ocasiones, difíciles de sustituir. Otros factores como la importancia de situación clínica a que se aplica la tecnología, los incentivos, incluso comerciales, y la presión social que acompañan, en ocasiones, a la innovación tecnológica o los canales de comunicación utilizados en la difusión de la información, juegan un papel crucial en la aceptabilidad científica y social de las tecnologías sanitarias. El papel que la evaluación juega entre todos estos factores resulta de difícil medición. Podría actuar directamente sobre la percepción de los médicos, influir a aquellos expertos con influencia, a su vez, en la opinión de los médicos o tener un papel directo en la toma de decisiones de los administradores sanitarios.

Para conseguir que la incorporación de nuevas tecnologías a los sistemas sanitarios se base cada vez más en datos concernientes a su efectividad, seguridad, eficiencia y

equidad, los canales de difusión de las recomendaciones formuladas por los equipos de evaluación deben ser lo más amplios y dirigidos posible. Así, además de implicar a los profesionales sanitarios, tanto clínico-asistenciales como a los encargados de la administración y planificación, el concepto de evaluación de tecnología, así como las recomendaciones que se vayan formulando, deben ser incorporadas a la formación pregrado de los futuros responsables de la salud.

Para finalizar, volvemos a destacar que la calidad de la información disponible es un elemento clave en el resultado de la evaluación de tecnologías. A pesar de esto, muchas veces nos vamos a encontrar con información insuficiente. Como soporte básico de la toma de decisiones sanitarias, la evaluación de tecnologías debe realizar todos los esfuerzos a su alcance para emitir recomendaciones sobre un determinado tema. Las decisiones han de ser tomadas continuamente y deben contar con un mínimo de información de apoyo. Así, la peor evaluación es siempre la que no se realiza.

BIBLIOGRAFÍA

Para profundizar en el conocimiento de temas relacionados con la Evaluación Epidemiológica de Tecnologías de Salud se pueden consultar, entre otros, los siguientes textos:

- **Bakke OM, Carné X, García Alonso F.** Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica. Ed Doyma, Barcelona. 1994
- **Dickersin K, Berlin JA.** Meta-analysis: state of the science. *Epidemiol. Rev* 1992;14:154-76
- **Hulley SB, Cummings SR.** Diseño de la investigación clínica. Un enfoque epidemiológico. Ed Doyma, Barcelona., 1993
- **Institute of Medicine.** Assessing medical technologies. National Academy Press. Washington DC, 1985
- **L'abbe KA, Detsky AS, O'Rourke K.** El meta-análisis en la investigación clínica. *Bol Of Sanit Panam* 1994;116:226-244.
- **Rothman KJ.** Epidemiología moderna. Ed Díaz de Santos, Madrid. 1987.
- **SBU.** Literature searching and Evidence Interpretation for Assessing Health Care Practice. SBU. Estocolmo,1993
- **Schlesselman JJ.** Case-Control Studies. Design, Conduct, Analysis. Oxford University Press, 1982.
- **Sox H, Stern S, Owens D, Abrams HL.** Assessment of Diagnostic Technology in Health Care Rationale, Methods, Problems and Directions. National Academy Press. Washington DC, 1989.

Bibliografía utilizada para los ejemplos:

- **Fine MJ, Smith MA, Carson CA et al.** Efficacy of Pneumococcal Vaccine in Adults. A Meta-analysis of Randomized Clinical Trials. *Arch Intern Med* 1994;154:2666-70.
- **McMahon AJ, Russell IV, et al.** Laparoscopic and minilaparotomy cholecystectomy. A randomized trial comparing postoperative pain and pulmonary function. *Surgery* 1994;115:533-9.
- **Schreinemacher DM, Everson RB.** Aspirin use and lung, colon, and breast cancer incidence in a prospective study. *Epidemiology* 1994;5:138-46.
- **Selby JV, Friedman GD, Quesenberry CP, et al.** A Case-Control Study of Screening Sigmoidoscopy and Mortality from Colorectal Cancer. *N Engl J Med* 1992; 326:653-7.

ANEXO

AL finalizar la lectura de este texto, las siguientes cuestiones deberían ser contestadas con facilidad:

1. Defina “Evaluación De Tecnologías Sanitarias”
2. ¿Cuál es el objetivo último de toda evaluación?
3. ¿Cuál sería el primer paso de un trabajo de evaluación de tecnología sanitaria?
4. ¿Qué pasos debe incluir una evaluación que parte de la evidencia disponible?
5. ¿Cómo definiría el meta-análisis y qué ventajas tiene sobre el resto de los tipos de evaluación a partir de la evidencia?
6. ¿En qué se diferencian los estudios observacionales de los experimentales?
¿Qué factores de deben tener en cuenta para elegir uno u otro tipo de diseños?
7. ¿Qué diseño epidemiológico sería el deseable en un proyecto de evaluación y por qué razón?
8. ¿Cuales son los problemas más importantes que pueden aparecer en los estudios observacionales?
9. ¿Qué estimadores de efecto se pueden obtener de los diseños epidemiológicos que se han revisado en el texto? Defínalo, escriba su expresión matemática e interprete los posibles valores que pueden resultar.
10. ¿Con qué parámetros se define la calidad de una prueba diagnóstica?. Defina cada uno de ellos y recuerde su expresión matemática
11. ¿Qué tipos de evaluación económica se pueden realizar? Comente las ventajas e inconvenientes de cada uno de ellos.

EJERCICIO

La concentración de hemoglobina en sangre es una variable continua. Si consideramos diagnóstica de anemia una concentración \leq de 12 g/100ml. Aplicamos esta prueba a una población de la que conocemos su situación en cuanto a anemia a través de un patrón de referencia y obtenemos los siguientes resultados:

		Anemia		
		si	no	
Hb \leq 12 g/100ml	+	19	10	29
	-	1	70	71
		20	80	100

Calcular la sensibilidad, especificidad y valores predictivos y comentar los resultados.

Si variamos el nivel de concentración de Hemoglobina considerado diagnóstico de anemia a \leq 10g/100ml, ¿Qué se puede predecir que ocurrirá con la sensibilidad y especificidad del test?

Los resultados aplicados a la misma población son:

		Anemia		
		si	no	
Hb \leq 10 g / 100ml	+	16	5	21
	-	4	75	79
		20	80	100

Calcular la sensibilidad, especificidad y valores predictivos y comentar los resultados.

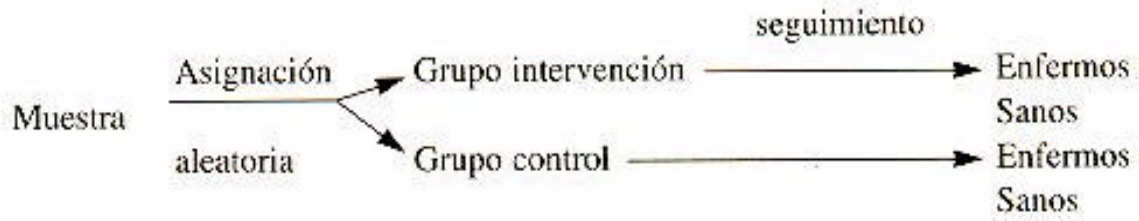


Fig. 1.- Estructura básica de un ensayo Clínico Aleatorio

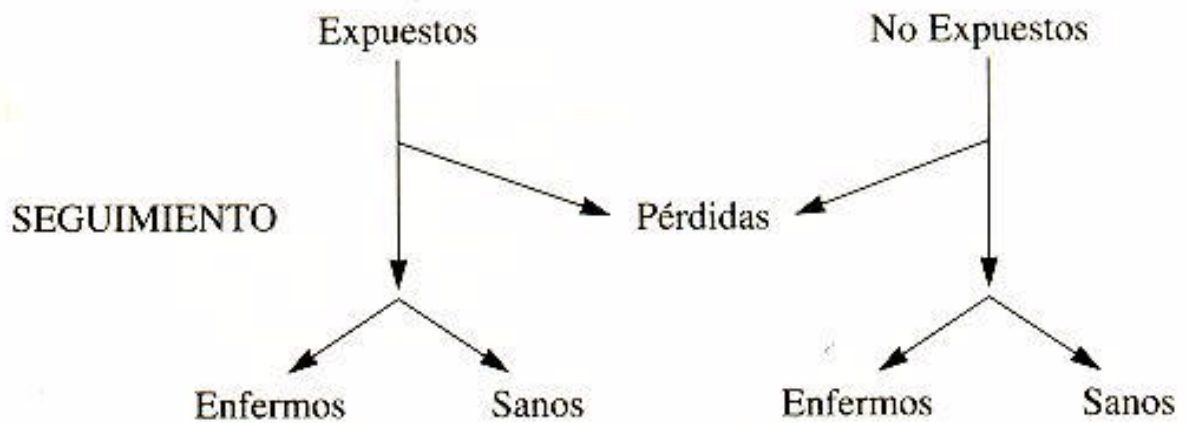


Fig. 2.- Estructura básica de un estudio de cohorte

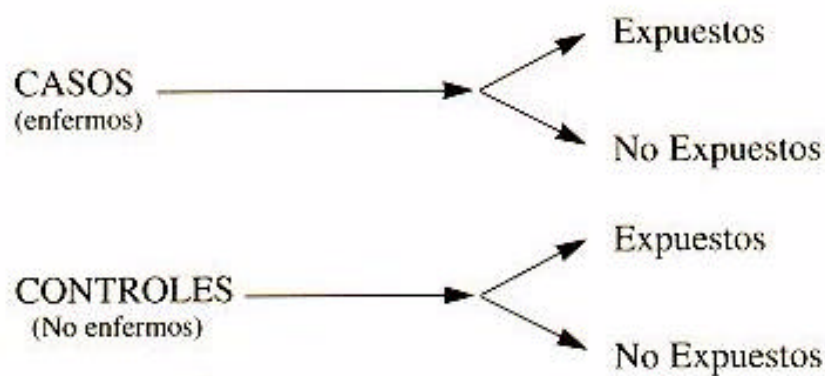


Fig. 3.- Estructura básica de un estudio de casos y controles

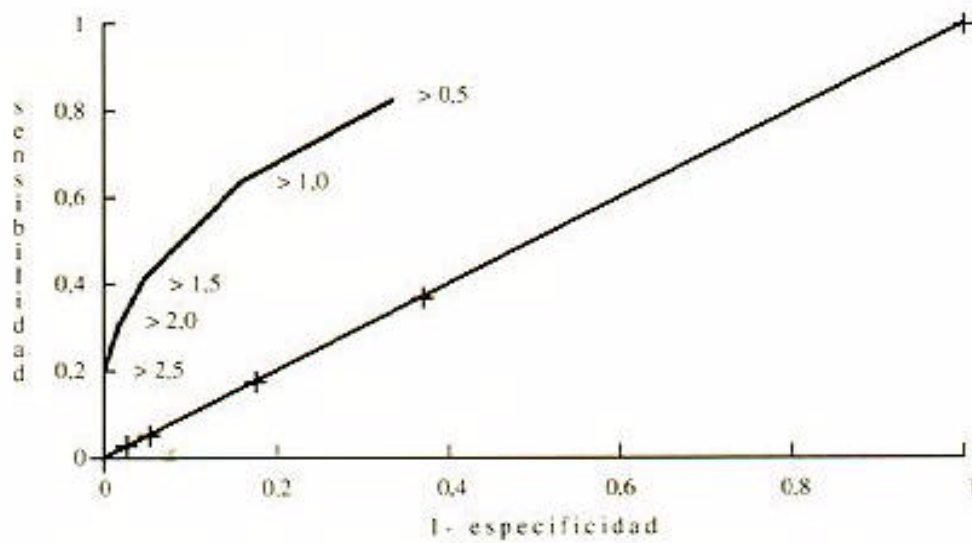


Fig. 4.- Curva ROC para el electrocardiograma de ejercicio como predictor de enfermedad coronaria.

NOTA: Rogamos a nuestros lectores que disculpen la baja calidad de estos gráficos, que esperamos poder sustituir más adelante por otros con mayor nitidez.

No obstante, le recordamos que , si solicita un ejemplar impreso, se les remitirá rápidamente.