



CENTRE DE RECERCA
EN ECONOMIA I SALUT·CRES
UNIVERSITAT POMPEU FABRA

DECISIÓN CLÍNICA: CÓMO ENTENDERLA Y MEJORARLA

Marisa Buglioli y Vicente Ortún

**Documento de Trabajo MSD-CRES/UPF
marzo 1999**

Nota sobre los autores y sobre el CRES, Centro que albergó este trabajo.

Marisa Buglioli Bonilla. Profesora Adjunta del Departamento de Medicina Preventiva y Social de la Facultad de Medicina de la Universidad de la República, Uruguay. Responsable de los módulos de Salud Pública, Economía Aplicada, y Gestión Clínica del Diploma de Especialización en Economía de la Salud del Centro Latinoamericano de Economía Humana-Instituto Universitario, Montevideo.

Es Doctora en Medicina por la Universidad de la República (1989), Uruguay, Médico Especialista en Administración de Servicios de Salud (1995), y Máster en Economía de la Salud y Gestión Sanitaria por las Universidades de Barcelona y Pompeu Fabra (1999).

Ha sido Gerente de Servicios Diagnósticos y Tratamientos Especializados en el Hospital Pediátrico-Centro Hospitalario Pereira Rosell-Ministerio de Salud Pública, Uruguay, asesora de racionalización de recursos médicos en la Cooperativa Médica Empleados Civiles, ha participado en cinco libros y ha realizado diversas pasantías y estancias científicas en EEUU, Brasil, Argentina, Cuba y España.

Vicente Ortún Rubio. Profesor Titular del Departamento de Economía y Empresa de la UPF y codirector del CRES. Director del Master de Economía de la Salud y Gestión Sanitaria de la UPF. Presidente de la Asociación Española de Economía de la Salud y Secretario de la Fundación Avedis Donabedian para la Mejora de la Calidad Asistencial.

Es Máster en Ciencias por la Universidad de Purdue (1970), Indiana, Licenciado y MBA por ESADE, Licenciado y Doctor en Ciencias Económicas por la Universidad de Barcelona y ha realizado estudios de Salud Pública en la Universidad de Johns Hopkins, Baltimore.

Ha sido Presidente de la Sociedad Española de Salud Pública y Administración Sanitaria y Secretario de la European Public Health Association. Cuenta con experiencia en gestión sanitaria (Subdirector Gral Ministerio Sanidad, Jefe Gabinete Técnico Generalidad de Cataluña, Administrador Hospital de Sta Cruz y San Pablo), en consultoría para organismos internacionales y en análisis financiero y valoración de empresas. Referencia de sus publicaciones recientes en <http://bonvent.upf.es/~ortun/>

El **Centro de Investigación en Economía y Salud, CRES**, es un centro especial de investigación de la Universidad Pompeu Fabra, Ramon Trias Fargas 25, 08005 Barcelona. El CRES nace en julio de 1996 con el propósito fundamental de contribuir, desde la óptica de la investigación universitaria, a los procesos de cambio y mejora en la formación y gestión de los servicios sanitarios. La fuerte vocación multidisciplinaria del CRES se ve reflejada en la composición de sus miembros, con dilatada experiencia investigadora y docente en Economía de la Salud y Gestión de Servicios Sanitarios. La mayoría de los integrantes del CRES pertenecen a grupos consolidados de investigación en España y participan activamente en redes europeas de excelencia. El CRES tiene convenios suscritos con el Ministerio de Sanidad de España, el Departamento de Sanidad de la Generalidad de Cataluña, la Fundación BBV, y Merck, Sharp & Dohme, S.A. Más información en <http://www.upf.es/cres/>

Los autores agradecen la colaboración que tuvieron en el presente trabajo de los miembros del CRES de la UPF, Joan Sánchez y Ramón Sabés.

ÍNDICE

Capítulo I

DECISIÓN CLÍNICA.....	7
I.1 La información y los incentivos como problema.....	7
I.2 Cambios en las Organizaciones Sanitarias	10
Bibliografía referenciada	20

Capítulo II

LA CALIDAD DE LA INFORMACIÓN SOBRE LA QUE SE BASAN LAS

DECISIONES.....	21
II.1 Introducción.....	21
II.2 Factores que limitan la calidad de la información en el proceso diagnóstico	22
II.3 Consecuencias en la toma de decisión: Desacuerdo clínico	28
Bibliografía referenciada	31

Capítulo III

FACTORES QUE LIMITAN LA RACIONALIDAD.....

III.1 Introducción.....	33
III.2 Sesgos psicológicos	33
III.3 El entorno	39
III.4 Fuentes de incertidumbre en Medicina.....	40
Bibliografía referenciada	43

Capítulo IV

HERRAMIENTAS ÚTILES PARA DECIDIR SOBRE LAS PRUEBAS

DIAGNÓSTICAS.....	45
IV.1 Introducción.....	45
IV.2 Decidir sobre la utilidad de una prueba diagnóstica.....	45
IV.3 Técnicas que ayudan a analizar la información que aporta una prueba diagnóstica.	47
IV.4 Situación real y perspectivas	74
Bibliografía referenciada	78

Capítulo V

HERRAMIENTAS ÚTILES PARA DECIDIR SOBRE LAS MEDIDAS

TERAPEUTICAS.....	81
V.1 Evidencia sobre la eficacia de los tratamientos.....	81
V.2 Forma de presentación de los resultados: RR, RA, RRR, NNT.....	85
V.3 Utilización de la evidencia: variabilidad de la práctica médica	90
V.4 Revisión de la utilización	93
V.5 A modo de síntesis.....	96
Bibliografía referenciada	97

Capítulo VI

TENDENCIAS EN MEDICINA.....

VI.1 La Medicina Basada en la Evidencia.....	101
VI.2 Orientación hacia las utilidades del paciente	104
VI.3 Interiorización del coste social de oportunidad.....	105
Bibliografía referenciada	107

Capítulo I

DECISIÓN CLÍNICA

I.1 La información y los incentivos como problema

La decisión clínica, diagnóstica o terapéutica, ocupa el centro del escenario sanitario. Cada día se asigna la mayor parte de los recursos sanitarios a través de los millares de decisiones clínicas que se toman en condiciones de incertidumbre. La decisión clínica, como cualquier decisión, puede ser descompuesta en cuatro fases diferentes: 1/ Inteligencia o recogida de información sobre las alternativas, 2/ Elección de alternativa, 3/ Implantación de la decisión, y 4/ Monitorización del impacto de la decisión. Habitualmente a las fases primera y tercera se las denomina gestión de la decisión y a las fases segunda y cuarta control de las decisiones. Siempre que quien decide no soporta directamente las consecuencias de sus decisiones la gestión de la decisión y el control de las decisiones recaen en dos personas distintas. Curiosamente, en la clínica y pese a que el médico no soporta directamente las consecuencias de sus decisiones (por ejemplo las consecuencias económicas) no se produce la citada división del proceso decisorio que está en la base del control interno de las organizaciones. Este reconocimiento organizativo del carácter profesional de la decisión clínica es el que confiere la centralidad a la gestión clínica en el conjunto de la gestión sanitaria.

Puede discutirse hasta qué punto resulta excesiva la palabra gestión para referirse al manejo clínico de los pacientes. Si se repasan las acepciones más habituales del término gestión se observara una gran coincidencia en su caracterización como actividades de coordinación y motivación destinadas a conseguir los objetivos de una organización. El clínico coordina recursos humanos y materiales (cada paciente es un 'proyecto' del cual es responsable) y ciertamente trata de motivar a quienes de él dependen así como al propio paciente.

A diferencia de los otros tipos de gestión sanitaria (la política sanitaria o gestión pública, la gestión sanitaria macro, y la gestión de centros, o gestión sanitaria meso), la gestión clínica, o gestión sanitaria micro, es específica del sector sanitario y tiene a las ciencias médicas como referente disciplinario fundamental.

La eficiencia clínica, marca del virtuosismo en Medicina, pasa por la maximización de la calidad de la atención y la satisfacción de los usuarios con los menores costes sociales posibles. La eficiencia tiene dimensión temporal, es dinámica: mejora en la medida que se innove en respuesta a los cambios demográficos, en morbilidad, en tecnología, en costumbres y preferencias, y en recursos disponibles.

Sin eficiencia clínica no puede haber eficiencia sanitaria. Un sistema sanitario funcionará bien si quienes asignan la mayor parte de sus recursos, los clínicos, tienen la información y los incentivos requeridos para tomar decisiones coste-efectivas.

Información equivale a reunir los conocimientos, habilidades y actitudes requeridos para el ejercicio profesional. La información en Medicina está sujeta a incertidumbre y, por otra parte, la capacidad humana para procesar información tiene unos límites claros. Cuando se adopta la decisión como objeto de estudio diversas disciplinas pueden ser relevantes y diversos los propósitos:

1. Descriptivo, estudiar cómo se efectúan las decisiones. La base para la finalidad descriptiva deberá buscarse en la psicología y en las ciencias del comportamiento en general. Su principal criterio de evaluación vendrá proporcionado por su validez empírica: el grado en que los modelos descriptivos se correspondan con las decisiones observadas.
2. Normativo, estudiar cómo deberían efectuarse las decisiones, establecer cuáles son los procedimientos de decisión lógicamente consistentes. La base para la finalidad normativa debe buscarse en la Estadística, las Matemáticas y la Economía. Su principal criterio de evaluación vendrá proporcionado por su adecuación teórica, esto es por el grado en que proporcione idealizaciones aceptables de la elección racional.
3. Prescriptivo, tratar de ayudar a la realización de buenas decisiones. La base para la finalidad prescriptiva debe buscarse en la investigación operativa y en las ciencias de la gestión. Su principal criterio de evaluación vendrá dado por su valor pragmático, esto es la capacidad de contribuir a la mejora efectiva de las decisiones.

Los sujetos decisores, en decisión clínica, son individuos. La Teoría de la Decisión se ocupa de las decisiones individuales en condiciones de incertidumbre. Cuando la incertidumbre proviene de un agente de intenciones neutras, referido genéricamente como naturaleza, se habla del núcleo de la teoría de la decisión individual en condiciones de incertidumbre: las decisiones frente a la naturaleza.

En otras ocasiones la incertidumbre proviene de no saber con certeza cuál va a ser el comportamiento de otro agente, generalmente no neutral sino antagónico, de cuya decisión dependen las consecuencias de las decisiones propias: teoría de juegos.

Muchas decisiones son fruto de hábitos o rutinas organizativas; están programadas. En el otro extremo, existen decisiones para las que resulta tan difícil establecer los conocimientos, habilidades y actitudes ideales de quienes las han de adoptar que se opta por el criterio de ‘decidirá bien quien ya ha venido decidiendo bien en el pasado’: es la estrategia de las empresas de cazatalentos. La Medicina está suficientemente definida como profesión para que sea posible establecer a priori el perfil del profesional idóneo. Ello no impide que diversas técnicas puedan ser útiles para analizar decisiones: No sólo la Teoría de la Decisión en condiciones de incertidumbre, a la que antes se ha hecho

referencia, sino también la Investigación Operativa (Simulaciones incluidas) y la Inteligencia Artificial.

Los **incentivos** (y la coordinación) dependen de las formas organizativas. Estos incentivos pueden ser implícitos o explícitos; los explícitos financieros o no financieros. No se entra aquí a valorar si el sistema de incentivos profesionales sirve a la salud de la población (ni la tendencia a la especialización excesiva, ni el premio al ‘hacer’ en lugar de al escuchar, ni la búsqueda del artículo en lugar de la atención al paciente pasarían el examen) sino simplemente a constatar su existencia para tenerla, consecuentemente, en cuenta. ¿Cómo? Reconociendo a la gestión clínica el protagonismo que tiene en el conjunto de la gestión sanitaria.

En lo que queda de este primer capítulo se abordan los cambios en organizaciones sanitarias que, a priori, pueden mejorar los incentivos y la coordinación: Gerencialismo (descentralización), organizaciones sin finalidad lucrativa, cooperativas autogestionadas, y organizaciones sanitarias integradas. Los capítulos 2, 3, 4 y 5 se dedican al análisis de la decisión clínica. El Análisis de Decisión, como disciplina, resulta apropiado siempre que: primero, exista cierta incertidumbre sobre la estrategia clínica a seguir ante un paciente con un determinado estado de salud (en muchas circunstancias está claro qué constituye la mejor actuación), y, segundo, que entre las alternativas que se consideran no exista ninguna que claramente domine a cada una de las otras en todas las dimensiones relevantes -desenlace clínico, utilidad del paciente, efectos secundarios, coste... La conveniencia de contemplar compensaciones -mayor efectividad por menor utilidad por ejemplo- entre las dimensiones relevantes constituye la segunda indicación para que el Análisis de Decisión sea aplicable a la clínica (Detsky et al, 1997).

Con las decisiones clínicas ocurre como con los pacientes: no hay dos iguales, pero si puede identificarse variables compartidas, entre pacientes o decisiones, que permitan pronosticar resultados de interés, tendrá sentido utilizar tanto sistemas de clasificación de pacientes como resultados de los análisis de decisión. Gran parte de lo que en estas páginas se trata resulta tan engorroso para una decisión clínica aislada como liviano y útil para los millares de decisiones clínicas suficientemente parecidas que cotidianamente se producen en condiciones de incertidumbre. Precisamente al ser decisiones parecidas, no idénticas, con frecuencia deberán adaptarse los resultados del análisis -una guía de buena práctica por ejemplo- a las circunstancias de cada paciente individual. Ahí se justifica el profesionalismo médico que permite ejercer una discrecionalidad responsable.

En el capítulo 2 se analizan los factores que limitan la calidad de la información sobre la que se basan las decisiones. En el capítulo 3 se estudian los factores que limitan la racionalidad para repasar en los capítulos 4 y 5 las herramientas útiles para decidir sobre las pruebas diagnósticas y terapéuticas, respectivamente. Se finaliza, en el capítulo 6, con un análisis de las actuales tendencias en Medicina.

I.2 Cambios en las Organizaciones Sanitarias

Los cambios en las organizaciones sanitarias que se contemplan son los derivados del gerencialismo, de la existencia de organizaciones sin finalidad lucrativa, de la presencia de cooperativas autogestionadas, y del redescubrimiento de las organizaciones sanitarias integradas.

El gerencialismo (desburocratización) implica una descentralización en las decisiones que continúan produciéndose en el dominio público. Las organizaciones sin finalidad lucrativa y las cooperativas constituyen formas intermedias entre los dominios público puro y privado puro. En el dominio público [privado] puro los derechos residuales de control -decidir sobre lo que no puede dirimirse según contrato, ley o costumbre- son públicos [privados] como la propiedad. Los derechos residuales de control y el derecho a la renta residual suelen ser complementarios por lo que en principio han de ir juntos (Hart 1995, pags 63-66). Ahora bien, en toda transacción en la que concurren beneficios colectivos y beneficios individuales tendría interés distribuir los derechos residuales de control (el médico sobre el tratamiento en función de lo que la profesión considera aceptable, el hospital no lucrativo sobre si recuperar unas fiestas imprevistas, el financiador público sobre si incorporar una nueva tecnología...).

Las organizaciones no lucrativas y las cooperativas permiten una distribución de derechos residuales de control que, como se verá, pueden ofrecer para ciertos servicios sanitarios unos menores costes de transacción.

I.2.1 Gerencialismo (desburocratización)¹

Las organizaciones burocráticas caracterizan el sector público aunque no sean exclusivas de éste. También en el sector privado abundan las burocracias (¿qué, si no, albergan los rascacielos?) pero en este último el problema del nivel óptimo de burocracia (relaciones públicas, control de gestión, planificación, asesoría jurídica, staffs de dirección...) se resuelve por comparación con otras empresas. Al fin y al cabo han de repercutir los costes de la burocracia en precios y los mercados obligan a curas de adelgazamiento.

Caracteriza a la organización burocrática la centralización de las decisiones, lo que conlleva estructuras jerárquicas, habitualmente acompañadas de otros rasgos como puedan ser: la especialización funcional, la existencia de mercados internos laborales, la reglamentación detallada de procedimientos, la presencia de funciones de producción ambiguas, la dificultad para medir el rendimiento, o el carácter funcional del vínculo que une a los empleados y la organización.

¹ Basado parcialmente en un trabajo previo de uno de los autores (VO) en Del Llano, Ortún V, Martín JM, Millán J, y Gené J. Gestión Sanitaria. Innovaciones y desafíos. Barcelona: Masson, 1998, págs 3-16.

Los cambios en la tecnología (vapor, telégrafo, ferrocarril a finales de siglo pasado) han estado detrás de los cambios organizativos en las empresas privadas que fueron adaptando su arquitectura interna para aprovechar las economías de escala y de gama que las nuevas tecnologías permitían. Unas organizaciones burocráticas que eran adecuadas para conseguir economías de escala bajo condiciones de producción en serie, se vuelven inadecuadas cuando el conocimiento gana importancia como factor productivo y la demanda se vuelve más sofisticada. En estas circunstancias el problema organizativo fundamental (coordinar y motivar) pasa a ser el de situar la capacidad decisoria allá donde está la información. Esto es: descentralizar. Los mercados -la organización virtual hayekiana- ofrecen el mejor ejemplo de descentralización; en ellos la asignación de derechos de propiedad, evaluación de la actuación y recompensa de la misma se producen de forma automática a partir de la información que contienen los precios. En las organizaciones, los derechos de propiedad deben ser asignados (quién hace qué), y hay que diseñar y utilizar sistemas de control e incentivos. Descentralizar implica no tan sólo transferir o delegar unos derechos de decisión sino también alterar en consonancia los sistemas de medir el desempeño e incentivar a los componentes de la organización. La arquitectura organizativa asentada en un trípode: asignación derechos de decisión, sistemas de medida del rendimiento, y sistemas de incentivos. Se trata de:

- Vincular autoridad y responsabilidad. Quien toma las decisiones ha de tener motivos (incentivos) para adoptar las decisiones que estime más correctas. Toda organización tiene su sistema de incentivos: explícito o implícito.
- Mantener la coordinación. A mayor descentralización, mayor necesidad de proporcionar información e incentivos a quienes deciden para que las diferentes actuaciones resulten consistentes entre sí. Las decisiones pueden no ser consistentes por falta de información o por comportamientos oportunistas.
- Disponer de personas formadas en las que pueda descentralizarse (el problema actual de la expansión en el extranjero de muchas empresas españolas).

Conforme se desplaza hacia abajo la decisión, la organización responde con mayor rapidez, hay menos necesidad de comunicación y menor pérdida de información, las decisiones están mejor informadas y los niveles superiores sufren menos sobrecarga administrativa. La alta especificidad de la información educativa, clínica, investigadora -de transmisión costosa- aconseja que las decisiones se tomen allá donde se posee esa información. Esto supone descentralización.

Descentralizado y burocrático son dos términos antitéticos. En una organización descentralizada, sus miembros poseen un alto grado de autonomía decisoria y dentro de él responden a los estímulos que emanan del sistema de incentivos implantado. La organización burocrática se regula por reglas preestablecidas de antemano, con procedimientos formales de decisión, con limitaciones en la discrecionalidad (menores en las burocracias profesionales) y, en suma, a través de rígidos sistemas de decisión que responden lenta e ineficazmente a las demandas de cambio y adaptación.

El gerencialismo (desburocratización o descentralización) se identifica (Mintzberg 1996) con los siguientes rasgos:

- Desagregación del sector público en unidades manejables;
- Medidas explícitas sobre objetivos, resultados, costes e impacto sobre el bienestar;
- Gestión profesional activa y discrecional, controles presupuestarios menos específicos y orientados a la evaluación ex-post y atribución clara de responsabilidades.

Se verá qué suponen estos rasgos y cómo el gerencialismo está limitado en el sector público.

Gerencialismo limitado por la dificultad de desagregar el sector público en unidades manejables

Para poder aislar actividades públicas de forma que hagan viable el gerencialismo, las políticas han de ser claras (no ambiguas), definibles y estables en el tiempo de forma que la decisión política y la ejecución de tal decisión queden claramente separadas.

Tal como señala Mintzberg (1996) ciertas actividades públicas como la de prisiones pueden ser aisladas horizontalmente (de las prestaciones sanitarias, por ejemplo) pero está menos claro que puedan aislarse verticalmente del proceso político. Y no tan sólo por la resistencia de los políticos a ceder control sino por el carácter iterativo de la formulación de políticas con lo que tiene de aprendizaje continuo y reelaboración permanente en un proceso de recogida de información, formulación de alternativas, toma de decisión y retroalimentación sobre los efectos observados.

Gerencialismo limitado por la dificultad de medir el desempeño

La dificultad en la medida del desempeño -segunda pata del trípode organizativo- impide alterar la tercera -la de los incentivos- en consonancia con el cambio habido en la primera pata al descentralizar derechos de decisión.

La intensidad (potencia, fuerza...) de los incentivos depende -inversamente- de la calidad de la medida del desempeño y de la aversión al riesgo del agente, y -directamente- de la contribución del desempeño del agente a los resultados de la organización y de la sensibilidad del desempeño a los incentivos. Sólo tendrá sentido ofrecer incentivos individuales a quien pueda controlar el ritmo y calidad de su trabajo, y con una intensidad proporcional a la sensibilidad del esfuerzo a los incentivos. Cuando el trabajo de una persona dependa de un equipo (producción de equipo mayor que la suma de las producciones de cada componente por separado) habrá que pensar en

ofrecer un incentivo basado en el desempeño conjunto del equipo.

El desempeño puede tener muchas dimensiones y cada una de ellas muchos modos de medida. Lo que se mida y recompense es lo que se pretenda mejorar. La potencia (fuerza) de los incentivos depende de lo bien que se mida la actuación. Incentivos fuertes, con efectividad, requieren una buena medida del desempeño.

Los servicios profesionales del Estado del Bienestar se caracterizan por producirse en condiciones de fuerte asimetría en la información y tener varias dimensiones y muchas no medibles. La calidad constituye una de estas dimensiones que, de entrada, admite muchas definiciones (clínica, técnica, social...). Estas características hacen coherentes a los incentivos de baja potencia.

Esta dificultad de medir el producto está en la base de la persistencia de organizaciones burocráticas, no descentralizadas. Resulta plenamente aplicable aquí la reflexión de Salas (1996) en el sentido de presentar la burocracia como la consecuencia, y no como la causa, de la mayor ineficiencia que se introduciría en el sector público si se pretendiera funcionar de forma descentralizada, incentivando a los agentes de manera poco relacionada con el esfuerzo que realizan (por problemas de información). Si se utiliza un criterio equivocado para medir el desempeño los resultados pueden ser peores que si no se establecieran incentivos explícitos. Peor todavía que tratar igual a los desiguales es tratar desigualmente sobre bases equivocadas.

Otros límites al gerencialismo relacionados con las dificultades en la medida del rendimiento se derivan de:

- El efecto trinquete, revisión de los objetivos a conseguir en base al mejor comportamiento observado lo cual, una vez anticipado, frena el entusiasmo productivo.
- Los problemas de linealidad: el avance se torna, con frecuencia, más difícil cuando el objetivo está próximo. Considerar el grado de proximidad al objetivo.
- Que los supervisores no soporten todas las consecuencias de sus decisiones, el intercambio será ineficiente para toda la organización (a diferencia de lo que ocurre cuando el supervisor es el propietario). La imposición de restricciones burocráticas, en estas circunstancias, constituye una profilaxis del amiguismo, del nepotismo, de la desviación de poder, de las actividades de influencia (malversación de recursos empleados en agrandar a quien manda), y de las perversiones que una medida errónea del desempeño clínico puede introducir.

I.2.2 Organizaciones sin finalidad lucrativa

Los beneficios tienen tres importantes funciones: retribución del capital (función del riesgo), premio a la innovación (en procesos y estrategias clínicas, por ejemplo) y recompensa a la eficiencia. Esto último requiere mercados competitivos e incluye la remuneración por el control eficiente del escaqueo (renta residual).

Otros inconvenientes de las ONL (además del que las define):

- El escaqueo gerencial puede, en principio, ser importante entre las ONL por menor mercado de gerentes, ausencia de control por la vía del mercado de capitales, y ausencia de control por la vía de las ofertas públicas de adquisición de acciones.
- Menor disciplina de mercado. Cabe existir aunque no se realice ninguna función valorada por el mercado
- Su éxito puede ser un artefacto derivado de la existencia de privilegios fiscales
- Puede permitir a los profesionales más poder del que tendrían en una organización con unos propietarios con derecho a la renta residual.

Ahora bien, las ONL también tienen ventajas derivadas de personas, gerentes y profesionales, que crean en una mejor manera de suministrar un servicio, que tengan una ideología que vaya más allá de la mera conveniencia pecuniaria:

- El control de calidad resulta más fácil donde los empleados están motivados por la misión. Ello resulta de especial importancia siempre que, como en sanidad, resulta difícil medir el producto y cuesta mucho, por tanto, establecer y controlar las características deseables de un servicio.
- Señal de confianza a usuarios desinformados. Los suministradores tienen menos motivos para aprovecharse de la asimetría informativa en servicios cuya calidad es muy difícil de medir (atención diligente y cariñosa a un anciano, cuidado de un niño...).
- Menor necesidad de supervisión pública por creación de reputación (menores incentivos a, por ejemplo, seleccionar riesgos).
- Las ONL no suben precios como respuesta a un exceso de demanda para sus servicios. Simplemente escogen (alumnos más brillantes, casos más interesantes...)
- Diferenciación de producto para atender a demandas heterogéneas.

I.2.3 Gestión por cuenta propia: la pertinencia de las cooperativas para los servicios sanitarios

Una cooperativa es la empresa gestionada por sus trabajadores. Se caracteriza por una ausencia de especialización: los trabajadores han de ser capitalistas y ejercer el control. Cuando estos trabajadores son profesionales aparecen las asociaciones de profesionales de la auditoría, el derecho, la banca, la arquitectura o la medicina.

La característica esencial de la empresa cooperativa es la de que los trabajadores ejerzan el control. Las formas jurídicas son varias: Desde las cooperativas laborales propiamente dichas a las sociedades laborales, o las sociedades civiles o mercantiles formadas por profesionales liberales.

La viabilidad de las asociaciones de profesionales dependerá de su eficiencia relativa: de que puedan funcionar con unos "costes de transacción" menores que los de formas organizativas alternativas.

En principio la asociación de profesionales reduce el conflicto entre discrecionalidad profesional y discrecionalidad gestora -al coincidir ambas en las mismas personas, da mayor control y satisfacción al profesional, e implica responsabilidad sobre los aspectos económicos de las decisiones profesionales. Esta responsabilidad - convenientemente modulada - lleva a la interiorización del coste social de oportunidad en las decisiones profesionales, y aunque complique tales decisiones puede constituir una mejor salvaguarda de la libertad profesional que la alternativa de una interferencia permanente en la misma.

Ahora bien, para que las asociaciones de profesionales puedan funcionar (funcionan en los hospitales japoneses y en la atención médica primaria del Reino Unido) se precisa:

- Tamaño de grupo pequeño que permita un proceso democrático, y no costoso, de decisión. Los tamaños excesivos complican los procesos decisorios, hacen la empresa menos manejable, y desincentivan el esfuerzo individual en la medida que se debilita la relación entre este esfuerzo y el resultado de la producción conjunta. Para que un tamaño reducido sea viable hace falta que no se produzcan economías de escala substanciales. Cuando una organización -caso de las cooperativas industriales de Mondragón- se desenvuelve en sectores con importantes economías de escala, el crecimiento constituye una estrategia obligada si quiere mantenerse la competitividad.
- Posibilidad de control mutuo de la actividad de cada socio. Requiere una función gestora no muy especializada, de forma que ésta puede ser ventajosamente substituida por el control mutuo, y tamaño de grupo pequeño. El inevitable control - de rendimiento, calidad- se encarece cuando debe realizarse sobre trabajo profesional ya que sólo lo pueden realizar otros profesionales; de ahí que convenga sea mutuo, entre pares. De nuevo, si el tamaño del grupo no es reducido, se desincentiva el

esfuerzo individual de control en la medida que se debilita la relación entre tal esfuerzo de control y el resultado de la producción conjunta.

- Que el capital humano, y no el físico, detente el mayor riesgo. Mayor control sobre el propio ejercicio lo cual puede motivar y estimular la cooperación en la medida en que se participa en resultados y se arriesga la reputación.

Dentro de los servicios sanitarios, estas condiciones se dan en la atención *médica primaria*, por lo cual se ejemplificarán en ella las cautelas precisas para cambiar la regulación. De nuevo, si unos mercados inexistentes no pueden sentenciar sobre la eficiencia relativa de las cooperativas de profesionales como forma organizativa, habrá que propiciar la diversidad mediante cambios en la regulación. Este fomento de la diversidad tiene un sentido de aprendizaje por ensayo y error en un entorno en el que quepa cierta competencia gestionada entre formas organizativas y donde otras alternativas de mejora de la gestión pública no sean viables: por falta de condiciones para la competencia (contratación externa) o por dificultades insuperables para avanzar en la vía del gerencialismo en las organizaciones públicas.

Cautelas ante las asociaciones de profesionales (ejemplificadas en la atención médica primaria):

- 1/ Peligro de selección de riesgos. El ciudadano tiene una razonable información para elegir médico de cabecera y este médico de cabecera tiene mayor información sobre el usuario que cualquier agencia compradora. Ello puede permitir al médico cribar - por tipo de servicios ofertados, accesibilidad, trato - a los "riesgos" que considere más interesantes. Estaríamos ante una selección de riesgos, conducta oportunista basada en la ventaja informativa. Existen dos formas de evitar la selección de riesgos: Restringir la libertad de elección del usuario (y asignar presupuestos en base poblacional con algunos ajustes por edad, sexo, mortalidad deprivación, y dispersión) o ajustar por riesgo el pago capitativo. Para realizar, aunque sea con escaso poder predictivo, este ajuste por riesgo individual se precisa un registro de utilización.
- 2/ Tamaño insuficiente para compensar riesgos. La población atendida por una asociación de médicos de cabecera -unos miles de personas- tiene un tamaño tan reducido que cualquier desviación en la incidencia esperada de enfermedades catastróficas desequilibra el presupuesto y provoca que el financiador público no pueda efectuar una transferencia efectiva de riesgo y responsabilidad a la asociación proveedora. Este problema puede resolverse de varias formas. Entre ellas: la delimitación de una cartera de servicios que excluya ciertas coberturas o el establecimiento de un techo anual de gasto sanitario por persona cubierta que limita la responsabilidad de la asociación de profesionales. Esta última ha sido la solución adoptada en el Reino Unido ("stop loss provision").

- 3/ Pérdida de equidad si los pacientes afiliados con asociaciones de profesionales resultaran mejor tratados. Problema evitable si se vigilan los presupuestos que se pactan con las asociaciones.
- 4/ Costos de administración y gestión que anulen las mejoras en eficiencia. La contratación de servicios de atención especializada no urgente, el manejo del presupuesto de farmacia, del personal y de las instalaciones supone costes de gestión, que pueden - no obstante - pactarse y ser objeto de monitorización.
- 5/ Densidad poblacional que permita comparaciones. La separación financiación/producción se basa en la idea de que los monopolios son perjudiciales porque permiten explotar a consumidores sin otra salida u opción. La explotación no necesariamente se presenta como precio excesivo sino que también puede presentarse -especialmente en trabajo profesional por cuenta ajena - bajo distintas formas de ineficiencia: retrasos, indolencia, derivación excesiva, prescripción inadecuada..."escaqueo" en general. Obviamente las motivaciones intrínsecas y sociales de los enseñantes o de los sanitarios pueden contrarrestar la invitación a la ineficiencia pero tampoco conviene esperar sistemáticamente todo de las mismas; especialmente si existe la posibilidad de traducir la ineficiencia - en la producción pública por cuenta ajena - en beneficio privado - por cuenta propia - a otras horas. Las bajas densidades poblacionales pueden impedir la elección del usuario y la comparación entre proveedores. Ahora bien, allá donde la densidad lo permita, las reglas del juego deberían ser claras y compartidas por todos: incompatibilidad, prohibido trabajar para una organización y para su competencia.

Garantía al usuario de que las motivaciones de lucro no interfieren en las decisiones clínicas. La autonomía profesional no puede subordinarse a intereses comerciales - propios o ajenos - sin erosionar el compromiso ético y social que sustenta la relación médico-paciente. Curiosamente esta cautela no se respetó en la modificación de la Ley de Ordenación Sanitaria de Cataluña (Ley 11/95 de 29 de setiembre) ni la contempla el documento de la Sociedad Española de Medicina Familiar y Comunitaria sobre autogestión (publicado en la revista Atención Primaria, de 30 de abril de 1998). Hay que compatibilizar la participación en los resultados de una asociación de profesionales con el correcto desempeño como agente del paciente. Una posible solución pasa por la definición de los resultados económicos y profesionales. Los beneficios económicos pueden utilizarse en equipo, instalaciones, formación, o contrataciones de personal, pero no son distribuibles como "dividendos". Los beneficios profesionales se derivan de una mayor satisfacción en el ejercicio -con capacidad para innovar y organizarse- y del desplazamiento de poder hacia la atención médica primaria al poder ejercer presión sobre los hospitales y otros niveles asistenciales con su correspondiente traducción en ganancias para el paciente en términos de menores esperas para atención especializada programable (cirugía por ejemplo) y mayor capacidad de resolución.

I.2.4 Organizaciones sanitarias integradas ²

La evolución del conocimiento y de la tecnología, la forma de organizar y financiar los servicios sanitarios y el poder de los distintos grupos profesionales dibujan los límites de la atención primaria, la atención especializada y la atención sociosanitaria. La tecnología influye tanto en la adecuación del lugar donde se presta la atención como en la combinación de recursos y habilidades que mejor resuelven un problema.

La Ley General de Sanidad establece dos niveles asistenciales interconectados entre sí: atención primaria y atención especializada. La atención primaria se constituye en la puerta de acceso al sistema sin que la intervención a este nivel esté limitada a la existencia de enfermedad. La asistencia especializada se concibe como apoyo y complemento de la atención primaria atendiendo aquellos procesos cuya complejidad exceda la capacidad de ésta. La labor de filtro aumenta la probabilidad de las enfermedades en el grupo de pacientes derivados y por ello mejora y justifica los métodos de diagnóstico y terapéutica de los especialistas. En la realidad la falta de continuidad entre una AP poco madura y una AE en ocasiones dividida entre ambulatoria y hospitalaria, sin criterios clínicos y organizativos comunes, constituye uno de los principales problemas del sistema sanitario en España.

Se concibe la integración vertical entre niveles asistenciales como una red coordinada de servicios que proporcionan la continuidad asistencial a una población determinada sobre cuya salud se tiene responsabilidad y cuyos consumos sanitarios implican un riesgo financiero para el proveedor integrado. Esta integración puede producirse mediante la propiedad conjunta (proceso de absorciones y fusiones) o a través de acuerdos contractuales. Una integración virtual, mediante contratos, puede evitar algunos de los inconvenientes organizativos ligados al mayor tamaño como son la atenuación de incentivos, la menor adaptabilidad ante los cambios y las mayores actividades de influencia.

El buen funcionamiento de un sistema sanitario va más allá del buen funcionamiento de los centros que lo integran. Una financiación capitativa ajustada por riesgo constituye una de las palancas regulatorias más potentes para fomentar la integración de proveedores y la cooperación entre clínicos. Habrá, por tanto, que optar entre proveedores integrados o no integrados (cuestión regulatoria). La forma de presupuestar -en la actualidad- o comprar -en el futuro- señalará la opción.

Si se continua presupuestando/comprando servicios sanitarios a proveedores no integrados -problema técnico: ajustar por aquellas variables que influyen en los productos intermedios y que no están bajo el control del proveedor- se favorece la competencia entre niveles, la duplicación de pruebas, el uso inadecuado y yatrogénico de tecnología y la diferenciación de servicios. Si, en cambio, se opta por

² Basado en un trabajo previo de uno de los autores (VO) en Del Llano J, Ortún V, Martín JM, Millán J y Gené J. Gestión Sanitaria. Innovaciones y desafíos. Barcelona: Masson, 1998, págs 349-357.

presupuestar/comprar atención sanitaria a una población -problema técnico: ajustar por riesgo el pago capitativo- se favorece la cooperación entre niveles ya que ni la retención de pacientes ni la actividad por la actividad favorece a ninguno de los niveles asistenciales en un juego de suma cero.

La integración, real o virtual, puede mitigar la actual falta de continuidad entre AP y AE pero, en última instancia, el que cada problema sea tratado en el lugar más adecuado -el más resolutivo- y que no se retengan casos ni se derive innecesariamente constituye un problema clínico. No habrá buena Medicina sin continuidad y eficiencia en la atención sanitaria, sin objetivos comunes, sin conocimiento mutuo AP/AE y sin el establecimiento conjunto de pautas de actuación.

Bibliografía referenciada

Del Llano J, Ortún V, Martín JM, Millán J, Gené J. Gestión Sanitaria. Innovaciones y desafíos. Barcelona: Masson, 1998.

Detsky A, Naglie G, Krahn M, Naimark D, Redelmeier D. Primer on Medical Decision Analysis: Part 1-Getting started. *Med Decis Making* 1997; 17:123-125.

Hart O. Firms, contracts and financial incentives. Oxford University Press, 1995.

Mintzberg H. Managing government, governing management. *Harvard Business Review* 1996; mayo-junio: 75-83.

Salas V. Gestión empresarial y sector público. *Ekonomiaz* 1996; nº 35: 146-163.

Capítulo II

LA CALIDAD DE LA INFORMACIÓN SOBRE LA QUE SE BASAN LAS DECISIONES

II.1 Introducción

“El examen clínico es mucho más poderoso que la evaluación de laboratorio para establecer el diagnóstico, el pronóstico y los planes terapéuticos para la mayoría de los pacientes, en la mayor parte de los sitios” (Sackett y cols, 1989).

La misma fuente cita trabajos en los que se demuestra que en medicina general el 73 y 88% de los diagnósticos fueron establecidos a través de esta herramienta; completando éste análisis los autores describen varios ejemplos frecuentes en la práctica clínica donde puede constatarse este hecho.

Como corolario de lo anterior, es necesario conocer cuales datos buscar en los exámenes clínicos, y una vez definido esto, conocer el cómo obtenerlos de manera que la información sea válida y precisa.

En el proceso del examen clínico puede ocurrir que algunos datos de la Historia Clínica o del Examen Físico sean poco reproducibles debido a que varían las respuestas de los pacientes, a que cada médico obtenga la información de manera diferente o a que varíe la forma de interpretar la misma (Goldman, 1991). Problemas similares pueden identificarse con las técnicas de laboratorio y demás exámenes paraclínicos.

A continuación se resumen algunas de las diversas fuentes de incertidumbre que pueden identificarse en la recolección de información y que repercuten en la elaboración de las hipótesis diagnósticas, haciendo más hincapié en el siguiente capítulo en los aspectos relativos a cómo se interpretan los datos a la hora de decidir sobre una terapéutica determinada. En este capítulo se consideran los sesgos que se producen al decidir, los problemas que la incertidumbre introduce en la información que se maneja así como las limitaciones a la racionalidad que pueden derivarse de la organización social y de la organización de los centros sanitarios. Los sesgos constituyen reglas sencillas que, sin ser racionales, proporcionan un atajo en el proceso decisorio al economizar esfuerzo mental. La utilización de sesgos y procedimientos heurísticos de decisión está ampliamente descrita y contrastada.

II.2 Factores que limitan la calidad de la información en el proceso diagnóstico

II.2.1 Relativos al paciente

Cuando se realiza una historia clínica, muchas veces existen problemas de *ignorancia* por parte del paciente de ciertos antecedentes personales o familiares, inclusive ambientales. En algunos casos el dato ignorado puede resultar esencial para formular o descartar ciertas hipótesis, o para las demás etapas del razonamiento clínico.

Otro aspecto a tener en cuenta es la forma en que un paciente *percibe los síntomas*, la cual puede variar según distintas características del individuo; se puede ver muy claro en las sensaciones como es el caso del dolor. Se recuerdan mejor los episodios agudos y más dolorosos, que los eventos vividos en períodos más prolongados pero de menor intensidad.

Se señala frecuentemente la distancia que existe entre la preocupación subjetiva y el nivel objetivo de peligro, o sea la dificultad para la percepción real del riesgo por parte del paciente. Existe una dificultad adicional y es que los elementos del interrogatorio no siempre pueden objetivarse a través ya sea del examen físico o de estudios complementarios.

Las diferentes percepciones se relacionan también con variables socio-culturales: clase social, continente-país de residencia, incluso dentro de distintas zonas de un mismo país, sin entrar a analizar el terreno de la accesibilidad y el tipo de servicios de salud disponibles. Esto puede determinar una consulta más o menos precoz, o incluso que el paciente no solicite atención para su “potencial” problema, con consecuencias evidentes en las posibilidades terapéuticas.

De todas formas debe tenerse cuidado con la interpretación de las diferencias entre poblaciones: Corin (1996), refiriéndose a afecciones psiquiátricas, plantea que resulta virtualmente imposible aclarar cuánto de la variación intercultural se debe a lo distinto de los procedimientos diagnósticos, cuánto a la diferencia de abordaje de los fenómenos de salud mental, y cuánto a diferencias en la propia realidad. Se reconoce, no obstante, que los síntomas comúnmente manifestados por personas afectadas por un mismo trastorno en distintas culturas presentan diferencias significativas; tales diferencias en el patrón sintomático prevalente estarían ligadas a otras de concepción cultural básica de la persona y del mundo. Se resalta por ejemplo, el papel en el contexto cultural y sociopolítico chino de los usos sociales de la somatización, en particular los beneficios que le puede reportar a las familias chinas enfatizar la dimensión física, ya que en esa sociedad los trastornos psicológicos se consideran como una desgracia, o algo sobre lo que se tiene la culpa.

La experiencia de un trastorno por tanto, no constituye un simple reflejo del proceso patológico orgánico, sino que incorpora expectativas y valores personales y colectivos; siendo los componentes sociales y psicológicos de dicha experiencia determinantes muy poderosos sobre la conducta de búsqueda de auxilio de las personas. (Corin, 1996).

Castellanos (1998) refiriéndose a la salud de la mujer, plantea que es necesario considerar que las diferencias culturales, educativas y laborales, así como su papel en la familia, han condicionado en muchos casos, patrones diferentes de percepción de las enfermedades y riesgos, que pueden traducirse en pautas distintas de demanda de atención médica; en algunos casos incluso se llegan a posponer necesidades propias de salud ante las de otros miembros de su grupo familiar.

De todas maneras no cabe generalizar. En la Encuesta de Salud de Cataluña, del año 1994, la mayor percepción de morbilidad se presenta en las mujeres, con cifras superiores a las manifestadas por hombres en todos los grupos etáreos.

Tabla 2.1 Población que declara restricción de sus actividades en los últimos 15 días por motivos de salud (en %) Cataluña 1994 (Plan de Salud 1996-1998):

<i>Grupo de edad en años</i>	<i>Hombres</i>	<i>Mujeres</i>
0-14	10.0	12.0
15-44	9.0	11.8
45-64	10.0	15.6
65 y más	10.8	19.7
Total por sexo	9.7	14.2

Estos datos son coherentes con la percepción de discapacidad que en el sexo masculino fue de 9.6% y en las mujeres del 14.4%, aumentando en estas últimas la declaración de distintos tipos de restricciones con la edad.

El *cómo se interpreta* la pregunta del médico depende también del nivel de educación, de conocimiento, y posiblemente de muchos factores más, como el siguiente ejemplo sugiere: cuatro médicos realizan entrevistas a un grupo de mineros en relación a varios síntomas comunes, encontrándose importantes diferencias en las respuestas al comparar los datos obtenidos por cada profesional. El porcentaje que declaró presentar tos varió entre un 23 y 40%, y ante la pregunta si presentaban expectoración el rango fue aun mayor: 13 a 42% (Eddy, 1984).

II.2.2 Relativos a los test diagnósticos: fiabilidad y validez

La indicación de pruebas complementarias sólo tiene una clara justificación si se dispone de una hipótesis, ya que de otra forma la recolección indiscriminada de datos puede suponer una interferencia más que una ayuda para el establecimiento del

diagnóstico: la frecuencia de resultados falsamente positivos se incrementa notablemente al aumentar el número de pruebas indicadas (Segura, 1998).

Por otra parte no siempre los estudios son fiables o válidos, es decir no registran en forma suficiente lo que se pretendía medir.

Los índices de validez de las pruebas diagnósticas más utilizados son la determinación de la sensibilidad y de la especificidad, que dependen de la misma prueba, y el cálculo de los valores predictivos positivos y negativos que dependen además de la prevalencia de la condición que se pretende diagnosticar (en el capítulo IV se abunda en estos conceptos).

Una fuente de desconcierto médico son los estudios donde el primer resultado es anormal y en una segunda indicación (sin mediar medidas terapéuticas) el resultado es normal. Un estudio sobre el nivel de bilirrubina en sangre mostró que en el 47% de los resultados anormales de un primer test, en el segundo test la prueba fue normal.

Es común que se repitan solicitudes que en muchos casos tampoco serían garantía de mayor certeza, ya que en general se resuelven con reglas informales: creer más en los positivos (frecuente en el caso por ejemplo de los cultivos bacteriológicos) o creer más en los valores negativos de los test (sería el caso de examen un funcional respiratorio) (Schwartz y Griffin, 1986)

II.2.3 Relativos al médico

La interpretación de un resultado anormal en una mamografía es mucho más común en los Estados Unidos que en otros países. Con igual sensibilidad y especificidad, una mamografía es interpretada como anormal en el 11% de los casos en éste país, comparado con un 2-5% en Suecia (Elmore y cols, 1998). Reconociendo que no se deban a factores subjetivos o personales del médico el total de diferencias percibidas³, la posibilidad de que los radiólogos de un lugar u otro interpreten de manera diferente los hallazgos patológicos en los pacientes existe, y algunos de los factores explicativos se pueden encontrar en los apartados siguientes.

1. Expectativas pasivas

La expectativas pasivas estarían determinadas por conductas y comportamientos que se perciben por largos períodos de tiempo en forma similar y que nos hace “tender a encontrar lo que esperamos o confiamos encontrar” (Sackett, 1989).

³ Se deberían valorar a que población van dirigidos los programas preventivos, y realizar diversos ajustes para poder comparar estos datos.

El ejemplo de la amigdalectomía es muy citado por lo elocuente: resultando que en alrededor del 45% de los *mismos* niños que los *mismos* médicos habían valorado previamente (en dos exámenes sucesivos) y descartado la necesidad de cirugía, se termina recomendando la amigdalectomía (Eddy, 1984).

Estas expectativas previas se describen como determinantes de percepciones -qué queremos escuchar o ver- de mayor importancia que las activas -que son las creadas para circunstancias específicas- (Schwartz y Griffin, 1986). La combinación de las expectativas pasivas con la incertidumbre de los procesos diagnósticos, es causa de variabilidad de la práctica médica.

En el caso del uso clínico de un fármaco, su efecto dependerá no sólo de su farmacología, composición química y dosis adecuada, sino también de las expectativas del médico y del paciente, del condicionamiento del paciente y de la enfermedad en si (Skrabanek y McCormick, 1992).

2. Percepción selectiva

Las expectativas pasivas se relacionan con otro concepto: la percepción selectiva. “La forma de buscar información en relación a nuestras hipótesis consiste, generalmente, en prestar atención a aquellos datos que las confirman y a pasar por alto aquellos que las rechazan” (León, 1994).

Es una característica de todas las profesiones, a la que no escapa la medicina. La mayor selección se relaciona con el conocimiento previo, experiencia, sistema de creencias y valores.

Frecuentemente la principal enfermedad y los datos que se obtienen en los primeros cinco minutos de interacción con el paciente, son suficientes para obtener un primer conjunto de hipótesis de trabajo. Este contenido de hipótesis es fundamental y se ve influido tanto por la experiencia como por las expectativas del médico. Además se tienden a enfatizar los hallazgos que se relacionan positivamente con nuestras hipótesis (información confirmatoria): se filtra y se selecciona información siendo los datos más recordados aquellos que concuerdan con las hipótesis generadas. Se describe como uno de los errores más frecuentes en medicina el que resulta de asignar más peso a la información confirmatoria de una hipótesis que a la que no lo es, existiendo evidencia de como incluso se llega a distorsionar información para ratificar opiniones previas (Elstein y Bordage, 1996).

3. Interpretación de números y palabras:

Rara vez se tiene certeza absoluta sobre un diagnóstico, en general se valora en término de probabilidades, y es común que éstas se expresen en palabras: “casi siempre”, “rara vez”, con interpretaciones que pueden llevar a no entenderse entre médicos o entre pacientes y médicos.

Algunas investigaciones muestran diferencias de interpretación entre los médicos de los términos que con “mayor frecuencia” se utilizan en medicina. A un grupo de médicos se les indicó que asignaran en una escala del 0 al 1, con marcas intermedias cada 0.05, su estimación de la probabilidad de presencia de enfermedad asociada con cada una de distintas expresiones o términos de uso común en los informes de radiología y laboratorio (probable, algunas veces, posible, etc). De las 30 expresiones utilizadas, solo 9, todas asociadas a expresiones de alta probabilidad, presentaron rangos menores a 0.5; algunas expresiones de baja probabilidad presentaron un rango de 0.9 (Bryant y Norman, 1980).

A continuación se presenta el resultado de otra investigación sobre las diferencias de interpretación de expresiones sobre probabilidad.

Tabla 2.2 *Interpretación de diferentes expresiones utilizadas para referirse a la probabilidad que suceda un evento: variación entre médicos (Schwartz y Griffin, 1986):*

<i>Palabras utilizadas</i>	<i>Rango de probabilidad</i>
Con frecuencia	0.4
Ocasional	0.3
Algunas veces	0.27
Infrecuente	0.25
Característico	0.20
Inusual	0.15
Raro	0.09

También se describen diferencias de interpretación entre médicos y no médicos (Schwartz y Griffin, 1986):

Incluso los médicos de mayor experiencia pueden ser incapaces de predecir con suficiente precisión la probabilidad de que ocurran ciertos estados patológicos. Los médicos presentan poca habilidad para cuantificar probabilidades que son altas o bajas: es el caso de la tendencia a sobrestimar la presencia de enfermedades relativamente infrecuentes (Goldman, 1991). Otros subestiman el riesgo de complicaciones de una intervención, por ejemplo en las endarterectomías, teniendo como consecuencia la

realización de ciertos procedimientos en los que la repercusión sobre el bienestar del paciente no es clara (McClellan y Brook, 1992).

Además no sólo se usa la misma palabra para expresar diferentes juicios sobre la probabilidad de un evento, sino que también es posible utilizar diferentes expresiones para realizar el mismo juicio sobre probabilidad.

4. Sobrecarga de información.

Es conocida la creencia del médico de que cuanto más información se obtiene es mejor para todos (incluyendo al paciente). Además está implícito el concepto “ante la duda hazlo⁴”, pues un error por comisión siempre se prefiere a uno por omisión (Eddy, 1984). Además distintos incentivos -que no analizaremos en esta publicación- pueden estar promoviendo esta actitud.

Los siguientes serían algunos de los factores que estarían influyendo sobre esa conducta: protección contra pleitos, costumbre y tradición de reunir todos los hechos así como el documentarlos en la historia clínica, formación recibida, y tipo de institución, pública o privada, donde se trabaja (Schwartz y Griffin, 1986).

El resultado de adquirir mucha más información de la que se puede manejar en la práctica, puede ser una decisión menos óptima: se llega a juicios menos precisos.

Una mejor decisión clínica se relaciona con una adecuada recolección de datos y su apropiada interpretación; esta última puede verse afectada negativamente por un exceso de información, al tener el clínico una capacidad limitada para manejar muchos datos (ver el siguiente apartado) y por otro lado se puede perjudicar la habilidad de seleccionar las variables más relevantes.

5. Límites en la capacidad de manejar información.

Está demostrada la existencia de límites en la cantidad de información que se puede manejar.

Describiendo las características del decisor humano, León (1994) habla de la memoria a corto plazo, o de trabajo, por la cual se llevan a cabo las operaciones de análisis de información, combinación de datos, y todas las actividades relacionadas con el proceso

⁴ El hazlo puede incluir la indicación de una prueba diagnóstica o una prescripción terapéutica.

de decisión. Esta memoria puede nutrirse (cargarse) del exterior o recuperar (recordar) de la memoria a largo plazo. La capacidad para memorizar elementos nuevos e independientes es de alrededor de siete, número que varía si se logra estructurar la información de manera de obtener nexos entre los distintos bloques de información (es el caso de las mnemotécnicas, muy utilizadas en medicina). El mismo autor cita un trabajo sobre decisión en el entorno médico donde se encontró que los clínicos raramente manejan más de cinco diagnósticos simultáneamente.

Elstein y Bordage (1996), describiendo la estrategia de generar y contrastar hipótesis por parte de los médicos, señalan que en general se basan en un número limitado de datos comparados con toda la información que recogen. El número de hipótesis generadas no excede el número de cuatro o cinco, no observándose cambios en el total aunque si pueden sustituirse unas por otras. Por otro lado, el número de problemas considerados raramente supera los seis o siete a la vez.

Otro de los aspectos que tiene repercusión en el proceso de decisión clínica, es la calidad de lo recuperado de la memoria a largo plazo: “estudiando el recuerdo se ha comprobado como, inconscientemente, al elaborar el relato, rellenamos huecos con datos de los que en realidad no disponíamos y tendemos a presentarlo de una forma argumentalmente coherente. Esta coherencia es tan personal, que, a veces, se convierte en una auténtica deformación fruto de nuestros prejuicios” (Leon, 1994). Este sesgo abarca la conducta no sólo de los médicos, sino también de muchos pacientes.

6. Factores biológicos.

Es indudable que la medicina es una de las pocas profesiones que debe utilizar prácticamente todos los sentidos: vista, oído, tacto, olfato (y el gusto al menos por practicarla) por lo cual las variaciones de percepción por parte de los numerosos clínicos es una fuente potencial de desacuerdo entre los mismos. Es interesante el resultado de estudios que muestran cambios en la capacidad perceptiva de los mismos médicos ante un estado de cansancio o de descanso suficiente (Sackett, 1989)

II.3 Consecuencias en la toma de decisión: Desacuerdo clínico

Se describen a continuación algunos ejemplos de desacuerdo clínico en relación a la evidencia clínica o paraclínica, dejando para el siguiente capítulo el análisis de lo que ocurre con las medidas terapéuticas.

Sackett (1989), describiendo el desacuerdo en la interpretación de la historia clínica, cita el caso de tres cardiólogos, que coincidieron en un 75% de los casos de las entrevistas realizadas a los mismos 57 pacientes, especificando además que cuando un cardiólogo

concluyó en un diagnóstico de angina de pecho, los otros dos coincidieron con él sólo en el 55% de los casos.

Eddy (1984) se refiere a un grupo de expertos que acuerdan lo siguiente en relación a 100 ECG: 50 con evidencias de IAM (infarto agudo de miocardio), 25 resultaron normales y los restantes con alguna otra alteración pero no indicativas de IAM. Cuando los mismos estudios se entregan para ser analizados por otro grupo de 10 cardiólogos, los distintos criterios utilizados hacen que si se tiene un IAM y se consulta a algún médico del primer grupo se tiene un 28% de probabilidad de que no lo diagnostique, y si no se tiene un IAM y se consulta al segundo grupo, la probabilidad que te digan que tienes un IAM es del 26%.

Algunos ejemplos más de desacuerdo clínico: grupo de médicos que al tomar el pulso a los mismos pacientes concuerdan en un 70% de los casos. Similares cifras se obtienen para el caso de interpretación de ECG, EEG, y estudios radiológicos (Eddy 1984, Schwartz y Griffin 1986).

También se describen desacuerdos con juicios propios previos, así como entre clínicos de gran experiencia (Sackett, 1989). Otro estudio aporta los siguientes datos para el caso de los anatómo-patólogos analizando biopsias de cuello de útero: el nivel de acuerdo intraobservador fue del 89% cuando se trataba de distinguir tejido normal de anormal, y comparando con un panel de expertos (acuerdo interobservador) de un 87 %; cuando se analizan los casos de pacientes con patología real el nivel de acuerdo intraobservador se reduce a un 68% y el interobservador al 55% (Eddy, 1984).

Se utilizará otro ejemplo de Sackett (1989) para describir como se puede medir el nivel de desacuerdo clínico.

Tabla 2.3 Cálculo del acuerdo observado entre dos clínicos en relación a 100 fotografías de fondos de ojo.

		Segundo clínico		
		<i>Poca o ninguna Retinopatía</i>	<i>Retinopatía moderada o grave</i>	
Primer clínico	<i>Poca o ninguna Retinopatía</i>	46	10	56
	<i>Retinopatía moderada o grave</i>	12	32	44
		58	42	100

El nivel de acuerdo observado sería: por un lado las 46 fotos diagnosticadas por ambos como de poca o ninguna retinopatía, más los 32 casos de retinopatía moderada o grave, en relación al total de casos estudiados ($46+32 / 100 = 78\%$). Pero una parte de éste

acuerdo se hubiera producido sin necesidad de estudiar cada uno de las pruebas, y es el llamado acuerdo debido al azar.

Tabla 2.4 Si el segundo clínico “tirara una moneda” se esperaría el siguiente resultado:

		Segundo clínico		
		<i>Poca o ninguna retinopatía</i>	<i>Retinopatía moderada o grave</i>	
Primer clínico	<i>Poca o ninguna Retinopatía</i>	(50 % de 56) 28	(50% de 56) 28	56
	<i>Retinopatía moderada o grave</i>	(50% de 44) 22	(50% de 44) 22	44
		50	50	100

Si el nivel de acuerdo producido por el azar fuera del 51%⁵, el acuerdo real más allá del azar sería: 78% - 51% = 27 %. La máxima posibilidad de acordar más allá del azar es del 49% en este caso (100% - 51% debido al azar).

El índice Kappa mide el nivel de desacuerdo clínico, y se calcula como acuerdo real producido más allá del azar en relación al acuerdo potencial. En éste caso la relación sería 27% / 49% = 0.55. Partiendo de una estimación inicial del nivel de acuerdo del 78%, se llega a este cálculo que con mayor objetividad mide las diferencias encontradas entre los clínicos. Cierta grado de coincidencia siempre aparece por casualidad (por azar), siendo el grado de coincidencia casual mayor cuando el hallazgo es muy común o muy raro (Goldman, 1991).

Cómo ya se vio, el desacuerdo se extiende también a los exámenes diagnósticos. En la tabla siguiente pueden verse algunos ejemplos de índice Kappa en esos casos.

Tabla 2.5 Desacuerdo clínico en la interpretación de resultados diagnósticos (Sackett, 1989):

<i>Situación planteada</i>	<i>Acuerdo observado</i>	<i>Esperado por azar</i>	<i>Índice Kappa</i>
2 radiólogos examinando 1.214 mamografías	97%	91%	0.67
2 cardiólogos examinando 38 ECG	57%	39%	0.30
1 cardiólogo examina 2 veces los mismos 38 ECG	74%	33%	0.61

⁵ El segundo clínico aprecia poca o ninguna retinopatía en el 58% de las ocasiones y retinopatía moderada o grave en el 42%. Aplicados estos porcentajes a los 56 y 44, respectivamente, que asigna el primer clínico a cada categoría se obtiene un acuerdo esperado por azar del 51%.

Bibliografía referenciada

Bryant GD; Norman GR. Expressions of probability: words and numbers. *N Engl J Med* 1980; 302: 411.

Castellanos PL. Los modelos explicativos del proceso salud-enfermedad: los determinantes sociales. En: Martínez Navarro et al. *Salud Pública*. Madrid: McGraw-Hill, 1998: 81-102.

Corín E. La matriz social y cultural de la salud y la enfermedad. En: Evans RG; Barer ML; Marmor TR. *¿Por qué alguna gente está sana y otra no?* Madrid: Díaz de Santos, 1996: 103-46.

Eddy DM. Variations in physician practice: the role of uncertainty. *Health Affairs* 1984; 3: 74-89.

Elmore JG; Barton MB; Mocerri VM; Polk S; Arena PJ; Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998; 338(16): 1089-96.

Elstein A; Bordage G. Psychology of clinical reasoning. En: Dowie J; Elstein A. *Professional Judgment*. Cambridge University Press, 1996: 109-29.

Goldman L. Aspectos cuantitativos del razonamiento clínico. En: Harrison. *Principios de Medicina Interna*. Madrid: Mc-Graw-Hill, 1991: 6-13.

León O. *Análisis de decisiones*. Madrid: Mc-Graw-Hill, 1994.

McClellan M; Brook RH. Appropriateness of care. *Medical Care* 1992; 30(7): 565-86.

Plan de Salud de Cataluña 1996-1998. Barcelona: Departament de Sanitat i Seguretat Social de la Generalitat de Catalunya: 153-55.

Sackett DL; Haynes R; Tugwell P. *Epidemiología clínica*. Madrid: Díaz de Santos, 1989: 23-177.

Schwartz S; Griffin T. *Medical thinking. The psychology of medical judgment and decision making*. New York: Springer-Verlag, 1986.

Segura A. Validez de las pruebas diagnósticas. En: Segura A; Sánchez E. *Epidemiología. Máster en Economía de la Salud y Gestión Sanitaria*. Universidades de Barcelona y Pompeu Fabra. Barcelona, 1998.

Skrabanek P; McCormick J. *Sofismas y desatinos en Medicina*. Barcelona: Doyma, 1992.

Capítulo III

FACTORES QUE LIMITAN LA RACIONALIDAD

III.1 Introducción

Los datos de la Psicología apoyan la idea de que las capacidades humanas son limitadas y que las decisiones pueden tener sesgos cuando las actividades que se desarrollan son *complejas* o la información disponible tiene carácter *probabilístico*, circunstancias ambas intrínsecas a la decisión en medicina.

La capacidad de la mente humana para resolver problemas complejos es muy pequeña en relación al tipo de problemas que sería necesario resolver en el mundo real (Schwartz y Griffin, 1986). La capacidad para poder conservar información en nuestra memoria, recuperarla cuando sea necesario y “manipularla” cómo y cuando se requiera, no es ilimitada.

El objetivo de la presentación de los sesgos que se realiza a continuación, no es producir un estado de desconfianza sobre las capacidades clínicas que lleve a dudar de todo lo que se trate de resolver de ahora en adelante. El conocimiento de las limitaciones a la racionalidad constituye un punto de partida para su abordaje.

III.2 Sesgos psicológicos

III.2.1 Disponibilidad

Cuando se realizan estimaciones de probabilidad sobre diferentes hechos (padecer una enfermedad, respuesta a un tratamiento, asociaciones con factores de riesgo) existe fuerte influencia del contacto previo -acceso- que los profesionales tuvieron con el acontecimiento o resultado. Cuanto más disponible un evento más fácil será recuperarlo. Este fenómeno de la facilidad con la que unos casos vienen a nuestra consciencia se conoce como disponibilidad.

La probabilidad que se asigna a un evento, será juzgada con la facilidad con que dicho evento se recuerda (Sox y cols, 1988)

Los elementos que son más accesibles a nuestra memoria nos aparecen con una mayor probabilidad de ocurrencia. En general dominan nuestras mentes los casos más vistos, las causas de muerte más impresionantes, los temas más publicados en las revistas médicas, los ítems que aparecen en primer lugar en una descripción⁶.

Se presta más atención a las muertes producidas por ejemplo por homicidios, que las debidas a diabetes o asma. Aunque éstas últimas sean más frecuentes, es más fácil recuperar de la memoria a las que mayor impacto producen.

También es común que se sobrestimen como causa de mortalidad, aquellas enfermedades presentadas con más frecuencia en revistas médicas, las que por otra parte se relacionarían con una mayor indicación de test innecesarios (Schwartz y Griffin, 1986).

La mayor disponibilidad puede llevar a juicios equivocados en relación a la importancia de determinados eventos. Un médico sobrestimaré la probabilidad de que un paciente con diarrea tenga amebiasis, porque recientemente vio un caso, aunque dicha patología sea poco común en su entorno.

Puede relacionarse esta mayor disponibilidad con la aparición de “*correlaciones ilusorias*”, o sea la tendencia a percibir la asociación de ciertos hechos con mayor frecuencia de lo que realmente ocurre.

En una investigación se presenta a un grupo de enfermeras un hipotético caso, donde se relaciona un particular síntoma con la aparición de una enfermedad. La información debe presentarse en una tabla de dos por dos con las clásicas opciones: presencia de síntoma y enfermedad, presencia de síntoma ausencia de enfermedad, ausencia de síntoma presencia de enfermedad, y ausencia de ambos. En la realidad los 4 opciones ocurrieron con igual frecuencia por lo cual se concluye que no existe relación entre ambos eventos. Pero las enfermeras asignaron mayor relevancia a la información que apoyaba la hipótesis inicial en relación a la que la refutaba, señalando una correlación ilusoria entre el síntoma y la enfermedad (Schwartz y Griffin, 1986).

Además, la desagregación de hipótesis causales puede contribuir a que la memoria “recuerde” mejor casos relacionados con esas causas y se modifiquen las probabilidades previas asignadas a ciertos diagnósticos. Se recurrirá a un nuevo ejemplo para aclarar el tema.

⁶ Por ejemplo, en un listado de enfermedades, se consideran más frecuentes las que figuran en primer lugar en relación a las últimas descritas.

El caso se refiere a una consulta en urgencia por dolor en fosa ilíaca derecha de 12 horas de evolución, en una joven de 22 años, cuya última menstruación fue normal. Un grupo de médicos debe asignar la probabilidad de ocurrencia de dos diagnósticos principales - gastroenteritis y embarazo ectópico- y de una categoría residual que incluye el resto de posibles causas. Un segundo grupo debe realizar la misma actividad pero en este caso se presentan 5 diagnósticos -se agregan apendicitis, pielonefritis y enfermedad inflamatoria pélvica- más la categoría residual.

El resultado fue que el primer grupo asignó una probabilidad del 31% al diagnóstico de gastroenteritis mientras que el segundo grupo una probabilidad mucho menor (16%), concluyendo los investigadores que el desagregar la categoría residual -aportando mayor información sobre la misma- mejora la atención del profesional e incrementa su memoria sobre esos acontecimientos, modificando las probabilidades previas asignadas a los diagnósticos principales (Redelmeier y cols, 1995).

III.2.2 Representatividad

La probabilidad que un evento o resultado ocurra se suele asociar a la similitud que éste evento tenga con las características o lo que le sucede a la población general.

El efecto de este sesgo es que tendemos a evaluar el grado de pertenencia de un elemento a una clase por el grado de representatividad de ese elemento respecto a la clase. Un ejemplo: a usted le presentan, en un Congreso donde asisten vendedores (as) y bibliotecarios (as), una persona que se manifiesta como extremadamente tímida, y debe acertar su profesión. Probablemente dirá que es bibliotecario (a) pese a que la probabilidad mayor es que sea vendedor (a).

Profesión	Número	% de tímidos
Bibliotecario	10	40
Vendedor	800	2

Diferentes tipos de errores se identifican como desencadenantes del sesgo de representatividad. Enumeraremos algunos ejemplos.

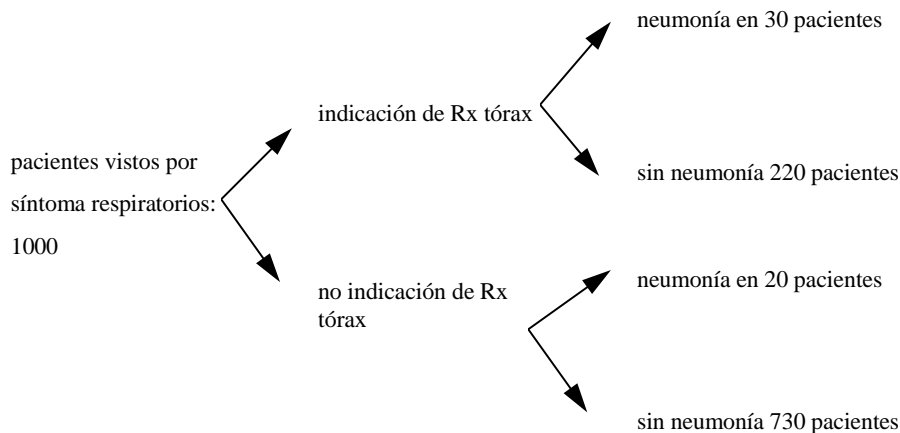
1. *Insensibilidad a las probabilidades previas*, sucede cuando se tiende a ignorar las probabilidades previas de presencia de la enfermedad en la población, en favor de los casos específicos estudiados.

2. *Insensibilidad con el tipo de muestra*, ley de los pequeños números: es pensar que la ley de los grandes números se aplica también a los números pequeños sin considerar el tamaño de la muestra. Un ejemplo: en un hospital público materno-infantil se asisten un número altísimo de embarazadas adolescentes procedentes de barrios

marginales de la ciudad. Puede suceder que el residente de guardia piense que la tasa de natalidad del país es altísima, debido a que ha leído que un indicador de dicha situación sería la frecuencia de embarazos en las adolescentes. Pero la realidad es que la natalidad en el país es baja, y sigue descendiendo en los últimos años. Resulta que en dicho hospital se asisten el 10% de los embarazos totales del país, y que las demás parturientas presentan una curva etárea que se desplaza lentamente hacia la derecha, con un pico máximo de embarazos en las mujeres que tienen entre 25 y 35 años.

3. *La idea errónea que se tiene del azar* se produce cuando existe la noción de que un fenómeno es menos probable que ocurra cuanto más regular sea el mismo. Por ejemplo, el sexo al nacer, ante estas dos opciones: fffffmmmmm // ffmfmmmfmf aparece en general como más probable la segunda (siendo m: masculino y f: femenino). Aunque desde el punto de vista matemático las probabilidades sean idénticas, la regularidad sería menos representativa que la no-regularidad.

4. *La ilusión de validez que se produce a raíz de un feed-back incompleto*, debido a que la estimación de la prevalencia puede distorsionarse. Por ejemplo, si se confirman un número determinado de neumopatías por radiografía de tórax siendo el criterio médico sólo indicarla en los casos de fuerte sospecha de tener la patología, la población a la que no se le indica el estudio pero tiene la enfermedad quedaría fuera de la estimación realizada (Schwartz y Griffin, 1986).



Un incompleto feedback produce una subvaloración de los casos de enfermedad: siendo la prevalencia real del 5%, y la conocida sólo del 3% (Schwartz y Griffin, 1986).

III.2.3 Punto de partida

La información que se recoge en un primer momento es la más fuerte, la que pesa más en el resto del proceso de decisión sobre el paciente, ya que a partir de esta información se van haciendo posteriormente sucesivos “ajustes”.

Por otro lado, no se realizan en general los ajustes suficientes a esta probabilidad inicial, cuando nueva información está disponible.

Experimentos realizados por psicólogos muestran que la estimación inicial de probabilidad de una enfermedad tiende a estar en alguno de los dos extremos: o muy cercana a 1 o a 0.

De estos dos comportamientos surge como consecuencia que si la estimación inicial es alta, la final también lo será, a la inversa si la inicial es cercana al 0 (Sox y cols, 1988).

III.2.4 Percepción selectiva: “ojo clínico”.

Si bien el concepto de percepción selectiva fue descrito en el capítulo anterior, nos interesa ahora subrayar el hecho que la convierte en limitadora de la racionalidad. León (1994) describe cómo los entrevistadores profesionales tienden a formular preguntas congruentes con sus primeras impresiones, tratando inconscientemente de encontrar datos que le permitan mantener sus hipótesis, “y su autoimagen de buen ojo clínico”. El problema que desencadena esta actitud hace que muchas veces queden sin explorar otras hipótesis, que de repente se ajustarían mejor a la situación real vivida por el paciente.

Resulta entonces que no sólo se percibe selectivamente sino que también se orienta de forma selectiva la búsqueda de información.

Se hace hincapié en los datos confirmatorios, cuando muchas veces sería mucho más eficiente encontrar sólo un dato que descartase la hipótesis de partida.

En Medicina es habitual tratar de contrastar hipótesis mediante la acumulación de datos confirmatorios, y cuantos más datos se recogen más seguros se está de la verdad de la hipótesis. La búsqueda del dato contrario sería mucho más efectiva, ya que encontrar un sólo dato que se oponga a la afirmación sería prácticamente suficiente para negarla; además la acumulación sucesiva de datos confirmatorios nunca daría la certeza total (León, 1998).

Es una característica del razonamiento científico -y la garantía de su avance- el sometimiento de las distintas hipótesis planteadas a métodos rigurosos que tratan de demostrar su falsedad.

“Leyendo con cuidado la mayoría de los artículos científicos es relativamente fácil averiguar la orientación que desean dar los autores a sus resultados, por lo que se ha de valorar siempre la posibilidad de que éstos hayan sido forzados a tomar una determinada dirección” (Skrabanek y McCormick, 1992). La utilización selectiva de evidencias así como de resultados, el omitir datos no coincidentes con las hipótesis de partida, la forma de expresar las conclusiones, son algunas de las herramientas que tienen fuerte discrecionalidad. Se retoma este tema en el capítulo V.

El “sofisma de la ley del silencio” es un agravante de ésta conducta. Como las consecuencias de un error en medicina pueden ser muy graves, existe una fuerte tendencia a negar la existencia de los mismos: un médico si es bueno, nunca se equivoca (Skrabanek y McCormick, 1992).

III.2.5 Sesgo retrospectivo

Se refiere a que una vez que sabemos lo que realmente ha ocurrido, tendemos a modificar el recuerdo de la opinión que teníamos previamente a que ocurrieran los hechos. Se describe que en general para los sucesos que realmente habían ocurrido, las probabilidades recordadas eran mayores que las realmente dadas; y para los sucesos que no habían ocurrido, las probabilidades recordadas eran menores que las que constaban en una primera respuesta (León, 1994).

Otros autores se refieren al tema de la siguiente manera: cuando valoramos hacia atrás la probabilidad de ocurrencia de un hecho, nuestro sentimiento es cada vez más a favor de que era inevitable.

Segundas opiniones, el asumir la atención de un paciente que previamente atendía otro colega o un estudiante, pueden aportarnos algunos ejemplos (Schwartz y Griffin, 1986).

III.2.6 Efecto Marco

Se refiere al efecto que produce en la elección entre alternativas la forma de presentación de la información: 80% de supervivencia y 20% de mortalidad, pese a expresar lo mismo, producen cambios de elección entre alternativas.

III.2.7 Placebos

La confianza en el tratamiento por parte del médico y del paciente ejerce un efecto de refuerzo mutuo cuyo resultado es la aparición de un poderoso remedio que mejora casi siempre al enfermo y que a veces lo cura. Se ha estimado que del 35-45% de las prescripciones cotidianas actuales es poco probable que ejerzan efecto específico alguno

sobre las afecciones para las que fueron recomendadas (Skrabanek y McCormick, 1992).

III.3 El entorno

El entorno institucional y organizativo puede limitar la racionalidad de las decisiones. Se comentan brevemente ambos entornos: el institucional –conjunto de restricciones, formales e informales, creadas por el hombre para regular su interacción social- y el organizativo –forma de distribuir derechos de decisión, evaluar comportamientos e incentivar individuos-.

Institucionalmente se requiere que sea atractivo para los individuos lo que resulta socialmente conveniente. Que resulte atractivo producir, innovar, esforzarse en lugar de litigar, copiar o parasitar esfuerzos ajenos.

Organizativamente, y como ya se ha señalado, hace falta que hospitales, centros de salud, mentales, sociosanitarios, u organizaciones integradas, asignen derechos de decisión de forma que los clínicos tengan la información y los incentivos requeridos para tomar decisiones coste-efectivas.

Grol (1997) describe el siguiente ejemplo: en un distrito X la tasa de cesáreas es particularmente elevada, por lo que se integró un comité de expertos para el desarrollo de planes que mejoren la atención obstétrica, tema que a todos preocupaba.

Los clínicos plantean que se organice un curso para incrementar el conocimiento y pericia clínica, los epidemiólogos proponen desarrollar un meta-análisis para obtener la mejor evidencia y elaborar guías para los obstetras, el académico recomienda la formación de pequeños grupos de profesionales para discutir los problemas identificados sobre la base de su propia práctica, el investigador en el área de servicios de salud solicita una auditoría multicéntrica para analizar la variabilidad entre los hospitales, el administrador señala la responsabilidad del sistema por lo que propone rediseñar el proceso de decisión y las estructuras relacionadas con el mismo, las autoridades locales de salud solicitan se asigne un límite en el número de cesáreas relacionado con un presupuesto razonable y se diseñen incentivos para reducir la tasa actual. El entorno incluye todos estos actores. Se retomará el tema en el capítulo sobre tendencias en medicina.

Algunos ejemplos más para considerar el rol de la autoridad y el papel del grupo que ocupan un destacado lugar en medicina. Skrabanek y Mc.Cormick (1992) se refieren en éstos términos al síndrome no-cardíaco: “...el jefe de servicio hace su ronda de rutina en la planta junto a cuatro médicos residentes y tres internos. El jefe ausculta al paciente y oye un soplo. Nadie más lo ausculta, pero el residente más antiguo afirma “yo sí lo

oigo”, con lo que sienta las bases para desatar una auténtica miniepidemia. Al cabo de un rato todos los miembros del grupo han adquirido la infección y auscultan el soplo...”

El respeto a la autoridad es un factor de importancia en gran parte de la educación médica; en ocasiones es prudente aceptar -al menos provisionalmente- lo que un experto tenga que decir, pero con la precaución de que “la duda es la que te ayuda a formarte”.

III.4 Fuentes de incertidumbre en Medicina

En el capítulo anterior se refirió a la importante variabilidad que se produce en la descripción de los síntomas y signos, así como en la interpretación de los resultados de las técnicas diagnósticas. Dentro de los problemas de consenso se comentan algunos ejemplos de desacuerdo producidos en el momento de utilizar la información para seleccionar ciertos procedimientos y medidas terapéuticas a ser aplicadas.

Los problemas detectados en el tratamiento de la enfermedad afectan tanto áreas de la medicina bien establecidas, como a las nuevas y relativamente poco sometidas a comprobación por el poco tiempo de aplicación. En un experimento se plantean una serie de casos hipotéticos describiendo pacientes potenciales candidatos a los siguientes procedimientos quirúrgicos: colecistectomía, herniorrafia inguinal, histerectomía, cesárea, amigdalectomía y colostomía. Ante la consulta a distintos médicos (en Canadá) se obtuvieron respuestas con un altísimo nivel de desacuerdo, por ejemplo las cesáreas fueron indicadas por el 65% de los médicos frente al 35% que no las realizaría (Roos y Roos, 1996).

En las patologías quirúrgicas es frecuente encontrar problemas de consenso sobre que tipo de cirugía utilizar o que técnica quirúrgica recomendar ante la misma patología. En esta falta de acuerdo no estaría influyendo demasiado los años de experiencia del profesional, aunque sería una variable de importancia el tipo de entrenamiento recibido, relacionado a qué tratamientos se recomiendan ante la presencia de diferentes riesgos (Timmermans y cols, 1997).

En el caso del Tromboembolismo Pulmonar (TEP) no existe un criterio diagnóstico estándar, y lo más útil para confirmar el diagnóstico -scanner positivo- no está presente en la mayor parte de los casos; además, los síntomas más frecuentes son inespecíficos. En una investigación donde se pide a diferentes grupos de médicos que ponderen 8 ítems relativos al diagnóstico de TEP se describe lo siguiente: en relación a la media por grupo fue similar la ponderación que se hizo del scanner y el examen de pantorrillas como los componentes de mayor relevancia. Cuando se analiza la estrategia individual, se detecta una altísima variabilidad inclusive entre los médicos con mayor nivel de entrenamiento, presentando todos los grupos analizados similar varianza intragrupo. Los autores describen como posibles causas de esta variabilidad lo confusa que es la literatura médica disponible, y la complejidad que implica ponderar múltiples variables,

que a su vez se ven influenciadas por los resultados de casos vistos en la práctica (Wigton y cols, 1996).

La incertidumbre sobre la efectividad de un procedimiento también puede ser importante entre los “expertos”. Otra investigación de Eddy (1984) describe que en una reunión de expertos en cáncer colo-rectal se pregunta cual sería el nivel esperado de reducción global de incidencia y mortalidad por ésta patología sometiendo a la población mayor de 50 años de ambos sexos al test para detectar sangre fecal oculta y una exploración con sigmoidoscopio flexible. Las respuestas variaron desde los que encontraron el procedimiento prácticamente inútil -sin ninguna influencia en la incidencia y mortalidad por cáncer de colon-, hasta los optimistas que pronosticaron la casi eliminación de la enfermedad. Cuando se muestran los resultados, la mayoría de los entrevistados no tenían conocimiento del nivel tan importante de desacuerdo sobre el tema.

Se han podido constatar diversas fuentes de incertidumbre en la práctica médica.

Eddy (1984) las resume de la siguiente manera: en el conocimiento de lo que el paciente tiene, en parte debido a que muchas condiciones no tienen una definición clara de enfermedad y esto conduce a diferentes opiniones en cuanto a la aplicación de criterios sobre que se entiende por la afección específica. Aunque todos los médicos estuvieran de acuerdo en la definición de enfermedad, pueden existir variaciones en: la habilidad para preguntar sobre síntomas asociados y en buscar los signos correspondientes (incertidumbre sobre si los síntomas y signos se corresponden con esa enfermedad), en seleccionar e interpretar resultados de estudios y en registrar las respuestas (sesgos sobre la información y los resultados), en el tipo de lenguaje utilizado, fallas en la comunicación y asignación de valores por parte del paciente.

La consecuencia lógica es la facilidad con que las “personas honestas” pueden llegar a diferentes conclusiones sobre un mismo proceso.

Se puede saber la dirección en la que potencialmente se movería la aplicación de diferentes alternativas, pero se desconoce que pasará en la vida real con el caso particular analizado, cual será su respuesta. El manejo de un paciente individual requiere de una compleja síntesis de información que como pudimos ver es “imperfecta”, pero de la que se espera suficiente precisión para alcanzar los mejores resultados posibles en las decisiones adoptadas.

Como ya se ha mencionado, el proceso se desencadena con la elaboración de ciertas hipótesis iniciales, con presencia de sesgos ya desde el punto de partida y en la etapa de selección (accesibilidad, representatividad), las que a su vez se confirman con una búsqueda selectiva de la información, caracterizada por la producción de errores en la interpretación de los datos recogidos a posteriori por profesionales que no escapan a los

diversos factores que limitan su racionalidad, y para complicar aun más en un entorno que puede ser un factor que distorsione más de lo que ayuda en la toma de decisión.

Ciertas investigaciones muestran que la mayor prevalencia de errores diagnósticos se deben a la generación y consideración de las hipótesis relevantes, las cuales estarían influidas por la experiencia y expectativas (Elstein y Bordage, 1996). Es lógico que se generen hipótesis tempranamente, pero al existir capacidad limitada en su manejo debería trabajarse con las más relevantes. Esto requeriría analizar pros y contras de cada alternativa posible, pero se vio que muchas veces se ignoran los aspectos negativos de las mismas, dándose más énfasis a los elementos confirmatorios, lo que se suma al hecho de que la información que se recoge en un primer momento es la más potente.

En los capítulos que siguen se presentan algunas herramientas que aparecen como útiles para ayudar al médico en algunas de las situaciones problema, si bien es el mismo profesional el que deberá identificar en su entorno que factores de los que producen información imperfecta son potencialmente modificables y cuales no, a su vez en los primeros en cuales sería más o menos fácil lograrlo.

Bibliografía referenciada

Elstein A, Bordage G. Psychology of clinical reasoning. En: Dowie J, Elstein A. Professional Judgment. Cambridge University Press, 1996: 109-29.

Grol R. Beliefs and evidence in changing clinical practice. Br Med J 1997; 315: 418-21.

León O. Análisis de decisiones. Madrid: Mc-Graw-Hill, 1994.

León O. Principios del análisis de decisiones en clínica. En: Del Llano J et al (dir). Gestión Sanitaria. Innovaciones y desafíos. Barcelona: Masson, 1998: 543-56.

Redelmeier D, Koehler D, Liberman V, Tversky A. Probability judgment in Medicine: discounting unspecified possibilities. Med Decis Making 1995; 15: 227-30.

Roos NP, Roos LL. Variaciones entre zonas pequeñas, estilo de práctica y calidad asistencial. En: Evans RG, Barer ML, Marmor TR. ¿Por qué alguna gente está sana y otra no? Madrid: Díaz de Santos, 1996: 253-76.

Schwartz S, Griffin T. Medical thinking. The psychology of medical judgment and decision making. New York: Springer-Verlag, 1986.

Skrabanek P; McCormick J. Sofismas y desatinos en Medicina. Barcelona: Doyma, 1992.

Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston: Butterworth-Heinemann, 1988: 27-64.

Timmermans D, Gooszen A, Geelkerken R, Tollenaar R, Gooszen H. Analysis of the variety in surgeons' decision strategies for the management of left colonic emergencies. Medical Care 1997; 35: 701-13.

Wigton RS, Hoellerich L, Patil KD. How physicians use clinical information in diagnosing pulmonary embolism: an application of conjoint analysis. En: Dowie J, Elstein A. Professional Judgment. Cambridge University Press, 1996: 130-49.

Capítulo IV

HERRAMIENTAS ÚTILES PARA DECIDIR SOBRE LAS PRUEBAS DIAGNÓSTICAS

IV.1 Introducción

En este capítulo se parte del concepto de que la probabilidad es una representación útil de la incertidumbre diagnóstica⁷. Los clínicos tienen que tomar sus decisiones sin contar con una información definitiva: no siempre se puede estar seguro de que un resultado negativo de un test es sinónimo de ausencia de enfermedad, como tampoco que uno positivo es indicador de certeza de su presencia. El razonamiento probabilístico es particularmente útil para sugerir las distintas hipótesis diagnósticas, y para asignar la importancia que tienen los hallazgos clínicos y los resultados de las pruebas diagnósticas (Kassirer, 1989). Por otro lado, Sox (1986) nos habla de que un test diagnóstico debería solicitarse solamente en los casos en que su resultado pueda alterar el manejo del paciente estudiado. Este será el principal criterio considerado en el presente trabajo, donde no se discutirán otras posibles razones para solicitar pruebas diagnósticas, tales como demandas por malpraxis, mejora de la satisfacción del paciente a través de posibles efectos placebos citados por algunos trabajos (mejora de la discapacidad y reducción de síntomas) (Sox y cols, 1981). Partiendo de esta premisa, se describen y analizan conceptos y herramientas que pueden contribuir a mejorar la racionalidad clínica con la que se indican las pruebas diagnósticas.

IV.2 Decidir sobre la utilidad de una prueba diagnóstica

El aporte que pueden realizar en el proceso de decisión clínica tanto los datos de la historia clínica como el practicar un adecuado examen físico es tema frecuentemente mencionado en las publicaciones médicas.

Los síntomas y signos producen, generalmente, pruebas de las hipótesis diagnósticas mucho más poderosas que las que podemos derivar jamás del laboratorio clínico. (Sackett y cols, 1989)

⁷ Siguiendo a Osler: la medicina es la ciencia de la incertidumbre y el arte de la probabilidad (Gerriti y DeVellis, 1990).

Aunque la mayor parte de la bibliografía y los ejemplos que se manejan se basan en el análisis de pruebas diagnósticas, el mismo razonamiento es aplicable a los datos de la historia clínica y del examen físico (Stoffers y cols, 1997).

En algunas situaciones las pruebas diagnósticas proveen una evidencia prácticamente definitiva (una biopsia cuando la anatomía patológica es patognomónica de la enfermedad); otras veces sirven como una contribución más a la evidencia diagnóstica.

Los criterios que permiten racionalizar la elección de una prueba diagnóstica tienen que ver con: la finalidad y capacidad de la prueba, el riesgo que supone y su coste.

Pueden distinguirse las pruebas por su distinto papel según sirvan para la detección, la exclusión o la confirmación diagnóstica.

La función detectora se refiere a la capacidad de “sospechar” la presencia de una enfermedad cuando todavía no se han producido manifestaciones clínicas; cuando este tipo de pruebas se aplica a la población general se habla de un “screening” o cribaje y cuando se aplican a los usuarios de los servicios médicos se habla de búsqueda oportunista de casos. La exclusión consiste en descartar la presencia de una enfermedad al realizar un diagnóstico diferencial; y la confirmación resulta de la constatación definitiva de una enfermedad. Se retoman estos aspectos más adelante.

La capacidad de la misma prueba se refiere a su validez para diagnosticar presencia o ausencia de la afección en estudio.

Las pruebas diagnósticas son imperfectas y se cometen errores al clasificar a una persona por su resultado.

La determinación de la sensibilidad y la especificidad (que dependen de la misma prueba) y el cálculo de los valores predictivos positivo y negativo (que dependen además de la probabilidad previa al test de padecer la condición que se pretende estudiar) son los criterios de validez que cuantifican la capacidad de una prueba para clasificar correcta o erróneamente a una persona según la presencia o ausencia de una exposición o una enfermedad.

También se debe considerar el riesgo que supone su realización o los efectos indeseables: iatrogenia por radiación, infección, intoxicación...

El coste de oportunidad de solicitar una prueba es aproximable en pesetas⁸.

⁸ Incluso los precios sirven en ciertos casos.

La presencia o ausencia de enfermedad se determinan a partir de un criterio de referencia o "patrón oro"⁹ el cual sería siempre positivo en individuos enfermos y negativo en los que no presentan la afección que se estudia. El examen o patrón oro puede resultar un estudio simple, pero en la mayoría de los casos se trata de exámenes complejos, costosos y que implican riesgos. La realización de las pruebas más sencillas se presenta como sucesivas aproximaciones a los métodos más precisos, por tanto se debería conocer el riesgo de clasificación equivocada que se produce al utilizarlas.

IV.3 Técnicas que ayudan a analizar la información que aporta una prueba diagnóstica.

IV.3.1 Sensibilidad-Especificidad (Sox, 1986).

¿Cuál es la capacidad de la prueba de considerar como positivos a los individuos que realmente presentan la condición que se está investigando? A esta pregunta responde la sensibilidad de una prueba. En el lenguaje de probabilidad sería la probabilidad de ser diagnosticado como positivo a condición de padecer la enfermedad. Sus valores oscilan entre 0 y 1.

¿Cuál es la capacidad de la prueba de efectuar diagnósticos correctos de ausencia de enfermedad, cuando ésta está ausente? En este caso se habla de especificidad. Sería la probabilidad de ser diagnosticado como negativo cuando no se presenta la enfermedad (también sus valores oscilan entre 0 y 1).

		<i>Enfermedad</i>	
		<i>Presente</i>	<i>Ausente</i>
<i>Resultado de la Prueba</i>	+	Verdaderos + (a)	Falsos + ... (b)
	-	Falsos - (c)	Verdaderos - (d)

La prevalencia es la frecuencia en que el fenómeno que se estudia está presente en la población antes de la realización de una prueba, siendo sinónimos con la probabilidad de enfermedad pre-prueba para aquellos pacientes de los que no se tiene información previa alguna (de la historia clínica, el examen físico u otras pruebas).

$\text{Prevalencia o probabilidad previa} = \frac{a + c}{a + b + c + d}$
--

⁹ La evaluación de la exactitud de una prueba se basa en su relación con una forma de saber con elevada certidumbre si la enfermedad está presente o no: a este estándar de referencia se le denomina patrón oro.

$\text{Sensibilidad} = \frac{a}{a + c}$	$\text{Especificidad} = \frac{d}{b + d}$
---	--

La medida de la precisión global resultaría del siguiente cálculo:

$\text{Precisión global} = \frac{a + d}{a + b + c + d}$

Si se obtuvieran los siguientes datos de aplicar un test de screening, por ejemplo un Haemo-glucotest, a nivel de un grupo de individuos de la población general para detectar precozmente diabetes, y posteriormente confirmamos o descartamos la presencia de la misma:

		<i>Enfermedad¹⁰</i>		
		<i>Presente</i>	<i>Ausente</i>	<i>Total</i>
<i>Test</i>	+	24 (a)	42 (b)	66
	-	6 (c)	78 (d)	84
<i>Total</i>		30	120	150

Probabilidad previa (prevalencia) = $30/150 * 100 = 20\%$ Sensibilidad = $24/30 * 100 = 80\%$ Especificidad = $78/120 * 100 = 65\%$

Se puede resumir diciendo que la sensibilidad considera los aciertos positivos y la especificidad los aciertos negativos. Cuando más próximas se encuentren del 100% mayor es la capacidad de discriminar los que tienen la enfermedad de los que no la padecen. En la mayoría de las pruebas ambas son menores del 100% y surgen los errores: falsos positivos y falsos negativos; el margen de error que puede cometerse también debería conocerse. Con este objetivo se pueden calcular los valores predictivos.

¹⁰ Donde podríamos hacer la siguiente clasificación de individuos:

- (a) presentan la enfermedad y el test es positivo: verdaderos positivos
- (b) no presentan la enfermedad y el test es positivo: falsos positivos
- (c) presentan la enfermedad y el test es negativo: falsos negativos
- (d) no presentan la enfermedad y el test es negativo: verdaderos negativos

IV.3.2 Valor predictivo positivo (VPP) y negativo (VPN) de un resultado (Sox1986).

El VPP es la probabilidad de presentar la enfermedad ante la condición de un resultado positivo de la prueba, es decir la proporción de verdaderos positivos en relación a la población cuyo test dio positivo.

Se habla también de la probabilidad posprueba de enfermedad después de una prueba positiva.

El VPN es la probabilidad de no presentar la enfermedad ante un resultado negativo del test; es decir la proporción de verdaderos negativos en la población cuyo test dio negativo.

Se denomina también probabilidad posterior de no enfermedad después de una prueba negativa.

Esta es la situación a la que habitualmente se enfrenta el médico, tiene un resultado de una prueba diagnóstica -positivo o negativo- a partir del cual deberá concluir sobre si el paciente tiene o no la enfermedad.

$$\text{VPP} = \frac{a}{a + b} = \frac{\text{verdaderos +}}{\text{todos los resultados +}}$$

$$\text{VPN} = \frac{d}{c + d} = \frac{\text{verdaderos -}}{\text{todos los resultados -}}$$

$$\text{VPP} = 24/66 * 100 = 36 \%$$

En esta población ante la presencia de un resultado positivo tenemos un 36% de probabilidad de que se trate de un verdadero positivo (de que sean diabéticos).

$$\text{VPN} = 78/84 * 100 = 92 \%$$

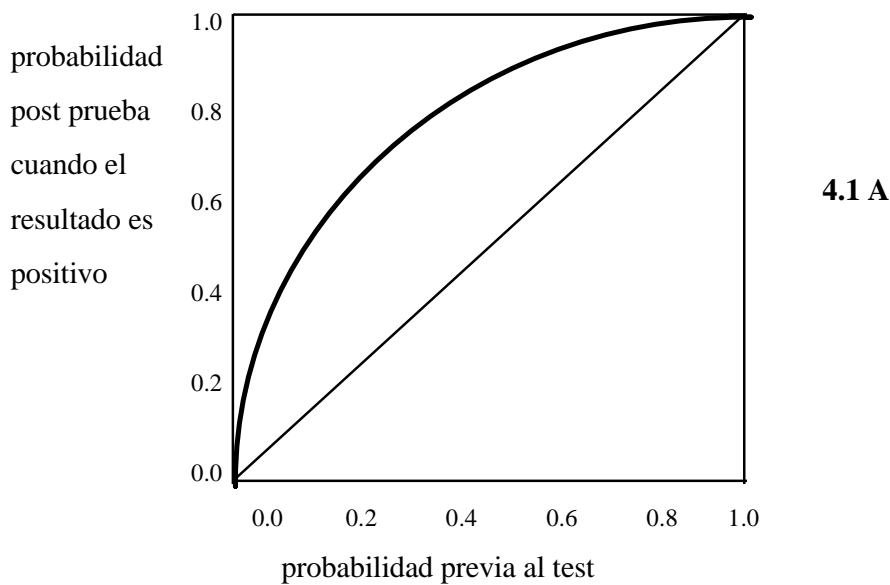
Cuando una prueba es muy sensible, un resultado negativo hace pensar en la ausencia de enfermedad, ya que el valor predictivo negativo es alto. En el caso de pruebas con

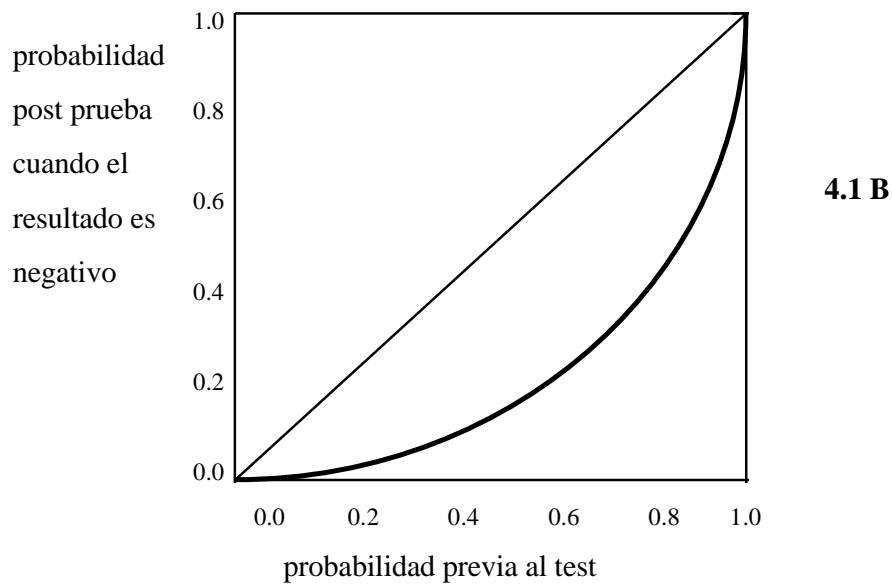
alta especificidad, lo que contribuye con más información es el valor predictivo positivo (los falsos positivos son pocos).

¿Podemos afirmar entonces que ante un resultado positivo de esta prueba diagnóstica la probabilidad de padecer la enfermedad es del 36% en todos los casos? **NO**. Los valores de predicción varían en relación a los pacientes que tienen la afección estudiada en relación a los sometidos a la evaluación diagnóstica.

La interpretación de los test depende de la prevalencia (o probabilidad previa) de la afección en la población estudiada: si el test es positivo, la probabilidad post-test de enfermedad crece en la medida que la probabilidad pre-test es mayor; en el caso de los resultados negativos, la probabilidad post-test decrece en la medida que la pre-test descende (Sox, 1986).

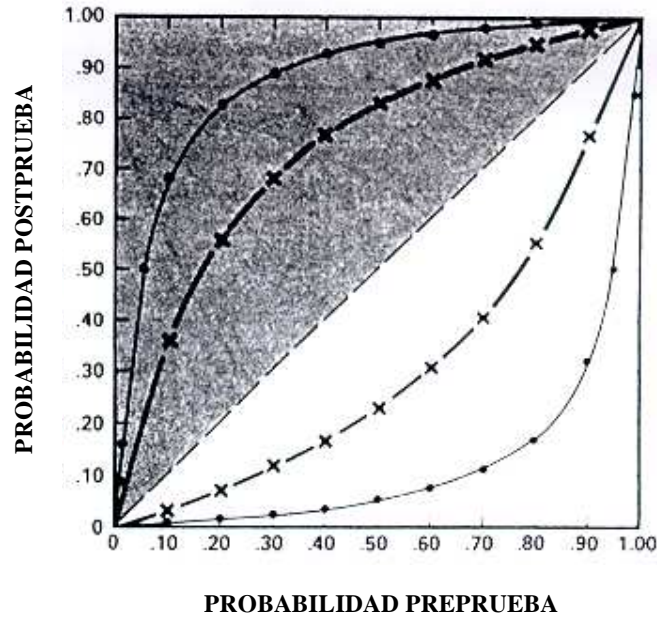
Fig 4.1 *Modificación de la probabilidad previa, según un resultado positivo o negativo del test diagnóstico.*





En estos gráficos puede evaluarse también la ganancia de información que puede obtenerse con una prueba diagnóstica: si las probabilidades pre y post prueba son similares, la curva coincidirá con la diagonal, la distancia vertical entre la diagonal y la curva es la diferencia entre las probabilidades pre y post prueba. La mayor ganancia en información global de un test se adquiere cuando se parte de una probabilidad intermedia entre el 40 y 60 %, un ejemplo puede verse en el uso de los marcadores tumorales en el cáncer de pulmón, donde el mejor resultado para tres distintos marcadores se obtiene con una probabilidad pre-test entre 30 y 40% (Pastor y cols, 1997).

Fig. 4.2 Probabilidad de padecer la enfermedad cuando el resultado de la prueba es positivo en función de la probabilidad preprueba. (Sackett y cols, 1989)



- Probabilidad postprueba cuando el resultado de la prueba es positivo
- Probabilidad postprueba cuando el resultado de la prueba es negativo
- Probabilidad de una prueba con sensibilidad 95 %, especificidad 95%
- x— Probabilidad de una prueba con sensibilidad 75 %, especificidad 85 %

Si este mismo screening se realiza a un grupo de población seleccionada según la presencia de ciertos factores de riesgo conocidos para la enfermedad investigada y tuviéramos por tanto una probabilidad pre-prueba mayor de enfermedad por ejemplo del 50 % -manteniendo constantes la sensibilidad y especificidad del test- obtendríamos los siguientes resultados¹¹:

		<i>Enfermedad</i>		
		<i>Presente</i>	<i>Ausente</i>	<i>Total</i>
<i>Resultado del test</i>	+	60	26	86
	-	15	49	64
<i>Total</i>		75	75	150

¹¹ Probabilidad previa (prevalencia) = $75/150 * 100 = 50\%$
 Sensibilidad = $60/75 * 100 = 80\%$
 Especificidad = $49/75 * 100 = 65\%$

$$VPP = 60/86 * 100 = 69.7\%$$

$$VPN = 49/64 * 100 = 76.7\%$$

Al aumentar la prevalencia o probabilidad pre-prueba (de 20% a 50%) incrementamos el valor predictivo de la prueba, pasando del 36% al 70%; simultáneamente se produce una disminución del valor predictivo de los resultados negativos.

Con otro ejemplo quedará más claro el concepto.

Tabla 4.1 Efecto de la prevalencia sobre el valor predictivo. Fosfatasa ácida prostática para detectar cáncer prostático -sensibilidad 70%, especificidad 90%- (Fletcher y cols, 1998).

Contexto	Prevalencia (casos/100.000)	Valor predictivo positivo (%)
Población general	35	0.4
Hombres, 75 o más años	500	5.6
Nódulo próstata Clínicamente sospechoso	50.000	93.0

A medida que la prevalencia de la enfermedad se acerca a 0, el valor predictivo de la prueba positiva también se acerca a 0. Cuando la prevalencia de la enfermedad en la población estudiada se acerca al 100%, el valor predictivo negativo se hace 0.

Las implicaciones de los conceptos anteriores, vistos en el caso de decisiones clínicas, se extienden también para considerar adecuadamente la efectividad de las acciones de salud pública, particularmente los cribajes poblacionales. Es el caso por ejemplo de la detección de infección asintomática por Chlamydia Trachomatis en mujeres donde se observa una disminución de casos conforme la actuación del programa va teniendo éxito, y al disminuir la prevalencia se incrementa la importancia relativa de los falsos positivos (Mariñas y cols, 1996).

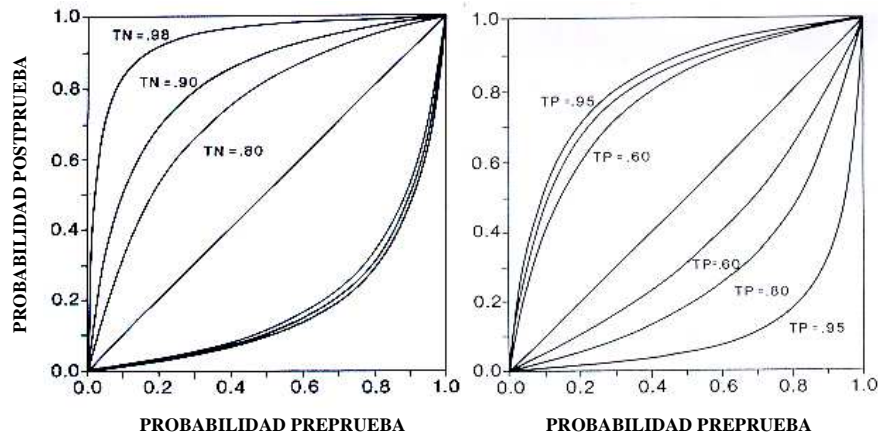
La prevalencia es el factor más determinante de los valores predictivos: cuando la prevalencia es alta un resultado positivo tiende a confirmar su presencia, pero si es negativo no permitirá excluirla (falsos negativos). Si es baja la prevalencia, un resultado negativo permitirá descartarla, aunque uno positivo no afirmará su presencia (los falsos positivos incrementan).

Véase algún ejemplo práctico de como una prueba diagnóstica con una sensibilidad y una especificidad determinada, tiene diferentes valores de predicción -positivo y negativo- cuando se utiliza en grupos con distinta prevalencia de la enfermedad. Un valor ligeramente alto de fosfatasa alcalina en un adulto joven que presenta un linfoma, sugiere un compromiso hepático por el tumor y por tanto es posible se trate de un verdadero positivo; mientras que el mismo valor de la fosfatasa alcalina en un individuo de la misma edad, sin síntomas, al que se le practicaron un conjunto de pruebas en

sangre dentro de una investigación sistemática, será más probable que se trate de un falso positivo (Goldman, 1991). Cuando se solicitan pruebas sin tomar en cuenta las probabilidades pre-prueba de la enfermedad se están incrementando los falsos positivos.

Por supuesto que el valor predictivo depende a su vez de la sensibilidad y especificidad del test: cuanto mayor sean la sensibilidad y especificidad mayor será el test predictivo (para una misma probabilidad pre-test de la enfermedad). Además la especificidad del test es un factor determinante de la probabilidad post-test después de un resultado positivo, se puede ver en el gráfico 4.3 como varía la ganancia de información al tener un test más específico (curva más arqueada), pero no altera demasiado la interpretación de un resultado negativo; en el caso de la sensibilidad es inverso: afecta fundamentalmente la probabilidad post-test ante un resultado negativo (Sox, 1986).

Fig. 4.3 Efectos de la sensibilidad y especificidad de un test en la probabilidad post-test (Sox, 1986).



Resumiendo: la ganancia de información de una prueba diagnóstica depende de la sensibilidad y especificidad de la prueba, y de la probabilidad previa de enfermedad.

También puede medirse la eficiencia de una prueba:

$$\text{Eficiencia} = \frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Total de resultados}}$$

Otra forma de expresar la eficiencia es a través del siguiente cálculo:

$$\text{Eficiencia} = \text{prevalencia} * \text{sensibilidad} + (1 - \text{prevalencia}) * \text{especificidad}$$

Si la prevalencia es alta, la eficiencia del test está más influenciada por la sensibilidad; mientras que ante bajas prevalencias la especificidad es particularmente importante (Bradley, 1993).

Otro aspecto importante a tener en cuenta es que el tipo de prueba a seleccionar depende de la situación que se trate, por ejemplo:

1. *Un test de screening*: también llamados pruebas de tamizaje o cribaje, se utiliza para la detección precoz de las enfermedades y permiten conocer en un momento dado la posible prevalencia de una enfermedad que afecta a una población determinada. En estos casos se utilizan técnicas de alta sensibilidad, es decir que capten el mayor número posible de enfermos, que sean fáciles de realizar, rápidas, no invasivas y de bajo costo. Debe tratarse de una intervención con muy bajo riesgo (la población no es la que toma la iniciativa de consulta, y la cantidad de verdaderos positivos siempre es pequeña en relación a todos los individuos estudiados). Dado que se aceptan cierto número de errores o falsos positivos, se requieren otras pruebas para la confirmación de la enfermedad. Las pruebas de confirmación tampoco deberían ser excesivamente peligrosas. Otro requisito es la necesaria garantía de que pueda modificarse favorablemente el pronóstico de los verdaderos positivos detectados, por lo que debe existir un instrumento terapéutico efectivo y fácilmente disponible.

2. *Una búsqueda oportunista de casos a nivel de la consulta médica*. Algunos ejemplos comunes: control de presión arterial, prescribir una colesterolemia, una glicemia, un PAP. En todos estos casos se debería asumir que el beneficio para el paciente en caso de estar afectado es mucho mayor que el perjuicio que le ocasionaría un error (falso positivo o falso negativo), así como los relacionados con el tiempo del procedimiento y con posibles efectos indeseables.

3. *Si se trata de orientar, confirmar o excluir un diagnóstico*. Se prescriben en personas con enfermedad y síntomas correspondientes a determinado cuadro mórbido, o con sospecha fundada de cierta patología (por ejemplo ante un test de screening positivo). Al tratarse de pruebas confirmatorias, requieren de alta especificidad, para evitar los falsos positivos.

Tabla 4.2 Diferencias entre un test de cribaje y un test diagnóstico

	<i>tamizaje</i>	<i>diagnóstico</i>
aplicación	enfermos ignorados y sanos	presuntos enfermos
practicado a:	grupos de población	indicación individual
preferencia	alta sensibilidad	alta especificidad
costo unitario	bajo	alto

Otras recomendaciones sobre las pruebas diagnósticas plantean que una prueba sensible debería utilizarse cuando las consecuencias de no detectar la enfermedad son importantes (procesos peligrosos pero tratables: tuberculosis, sífilis, o enfermedad de Hodgkin); también en las primeras etapas de un plan de estudio, cuando se plantean como probables muchas alternativas. Este tipo de prueba sería más útil cuando su resultado es negativo (Fletcher y cols, 1998). Las pruebas específicas se utiliza para

confirmar un resultado que ha sido sugerido por otras pruebas. Este tipo de pruebas da pocos resultados falsos positivos, y tienen gran utilidad en los casos en que un resultado falso positivo puede producir daños físicos, emocionales o afectivos.

IV.3.3 Curvas ROC

La precisión diagnóstica de un test se mide por su “habilidad” para discriminar entre diferentes estados de salud: presencia-ausencia de enfermedad, proceso benigno-maligno, respuesta o no a una terapia, indicar un buen o mal pronóstico. La capacidad propia de discriminar entre estos aspectos por parte de un test, está influida por factores externos al mismo que veremos a lo largo del capítulo, y que resultan imprescindibles evaluar a la hora de decidir sobre la utilidad de un test en una situación dada.

En este apartado se describe cómo medir la capacidad intrínseca de las pruebas para discriminar entre posibles estados clínicos, y su comparación con otras pruebas disponibles, requisito para luego poder decidir su contribución en el manejo del paciente particular.

En general se intenta buscar una prueba que tenga un valor de corte sensibilidad-especificidad que sirva de verdadera guía en la toma de decisiones, las pruebas más eficaces serían aquellas que logran obtener una alta sensibilidad y especificidad, pero lamentablemente no son las que se encuentran mayormente disponibles. ¿Qué criterio priorizar: una alta sensibilidad y definir como anormal un ECG durante el ejercicio si presenta un descenso ST igual o mayor a 0.5 mm, o priorizar la especificidad con una depresión ST igual o mayor a 2 mm? (Goldman, 1991).

Algunos ejemplos más: una prueba sensible es el Ecocardiograma en Modo M que permite excluir la estenosis aórtica grave en los adultos con una sensibilidad cercana al 100%, es decir un ecocardiograma normal de válvula aórtica prácticamente excluye la existencia de una estenosis aórtica grave en adultos. Pero es una prueba que resulta poco específica, y muchos pacientes que en el Ecocardiograma presentan alteraciones en la válvula aórtica, no tienen estenosis grave y requerirán otros estudios para confirmar o descartar el diagnóstico.

Un ECG que presente una elevación del segmento ST mayor a 1mm en dos o más ECG realizados en pacientes que consultan en urgencia de un hospital por dolor torácico puede considerarse altamente específico, pues es poco probable que se trate de un falso positivo, y la conducta terapéutica se basa en dicha premisa; pero el mismo estudio en las mismas condiciones es poco sensible, ya que el ingreso de pacientes con diagnóstico de IAM no se limita al hallazgo electrocardiográfico, debido a que las consecuencias serían graves (se omitiría el diagnóstico de casi la mitad de IAM) (Goldman, 1991).

La localización de un punto de corte aparece entonces como prioritaria. La relación entre ganancia en sensibilidad y pérdida en especificidad (y viceversa) puede observarse en la siguiente tabla:

Tabla 4.3 Relación sensibilidad y especificidad en el diagnóstico de diabetes (Fletcher y cols, 1998):

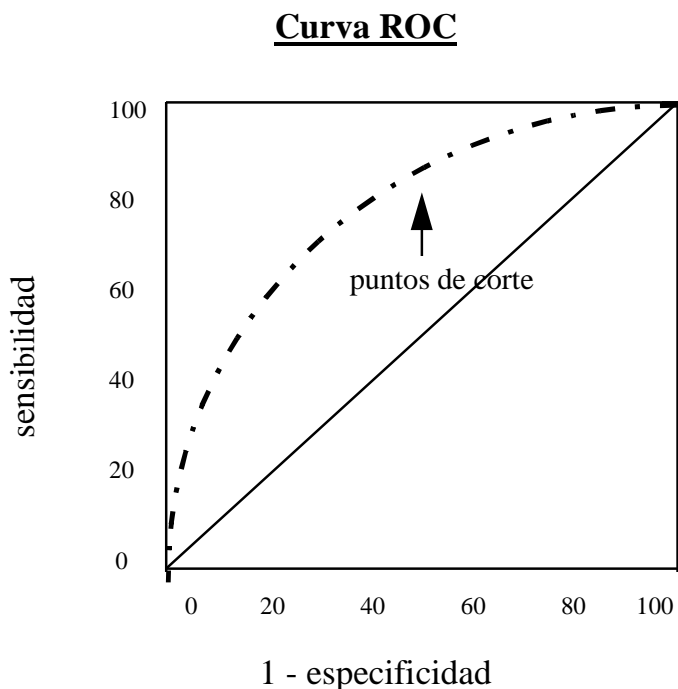
<i>Glicemia posprandial de 2 horas (mg/dl)</i>	<i>Sensibilidad (%)</i>	<i>Especificidad (%)</i>
70	98.6	8.8
80	97.1	25.5
90	94.3	47.6
100	88.6	69.8
110	85.7	84.1
120	71.4	92.5
130	64.3	96.9
140	57.1	99.4
150	50.0	99.6
160	47.1	99.8
170	42.9	100.0
180	38.6	100.0
190	34.3	100.0
200	27.1	100.0

Otra forma de expresar esta relación es a través de las llamadas curvas ROC (Receiver operator characteristic curve¹²). Se trata de una forma más global de evaluar la calidad de una prueba, a través de un gráfico en el que se observan todos los pares de sensibilidad/especificidad resultantes de la variación continua de puntos de corte en todo el rango de resultados observados. En el eje de las ordenadas se presenta la sensibilidad o tasa de verdaderos positivos, y en de las abscisas 1-especificidad o tasa de falsos positivos. Los valores de los ejes van de 0 a 1.

Cuando una prueba no discrimina coincidirá su curva ROC con la diagonal del gráfico formando un ángulo de 45 grados con los ejes, quedarían así representadas las pruebas que no adicionan información. En el otro extremo, si la discriminación fuera perfecta (sensibilidad y especificidad = 1) la curva pasaría por el extremo superior izquierdo.

¹² Tienen su origen en el año 1950, cuando son utilizadas para detectar problemas con las señales electrónicas de los radares.

Fig. 4.4



Para su trazado se requiere previamente calcular la sensibilidad y especificidad de todos los posibles puntos de corte de la prueba a evaluar¹³.

La precisión global de una prueba se describe como el área debajo de la curva, cuando mayor es esta área mejor discrimina la prueba. El máximo valor del área es de 1, para una sensibilidad y una especificidad del 100%.

El área bajo la curva ROC indica la probabilidad de clasificar correctamente un par de individuos sano y enfermo, seleccionados al azar de la población, mediante los resultados obtenidos al aplicarles la prueba diagnóstica.

El resultado de la prueba debería tener una mayor probabilidad de ser anormal en el paciente enfermo, que en el sano. Valores entre 0.5 y 0.7 indicarían baja exactitud o discriminación entre sanos y enfermos, siendo considerados útiles valores mayores de 0.7. Un valor de 0.5 sería equivalente a lanzar una moneda y catalogar como enfermo a quien tenga cruz.

Un ejemplo de aplicación de las curvas ROC es el siguiente: existen evidencias que las cifras de control de glicemia recomendadas presentan falsos positivos cuando son comparadas con la hemoglobina glicosilada (patrón oro). Al realizar el análisis de las

¹³ Para mayores detalles de su construcción puede consultarse: Clin Chem 1993; 39(4) 561-77; Med Clín (Barc) 1995; 104: 661-70.

curvas ROC se observa mayor exactitud diagnóstica con valores de glicemia algo más elevados que los recomendados. El valorar mal un diabético a través de la glicemia basal puede desencadenar riesgos importantes: en el caso de un falso positivo, una hipoglicemia iatrogénica; y si se trata de un falso negativo, un mal control. Si se adoptan valores muy bajos de glicemia basal ocurrirá lo primero, alta sensibilidad con baja especificidad; siendo lo contrario en caso de valores elevados (Orozco y cols, 1997). Un punto de corte en 100 de glucemia tendría un 11% de falsos negativos, y un 30% de falsos positivos; si el punto de corte lo moviéramos a 120, los falsos positivos disminuirían a menos del 10% pero a expensas de aumentar a un 30% las cifras de diabéticos no diagnosticados (Fletcher y cols, 1998).

Como criterio general, se plantea como mejor punto de corte el que quede más cercano al extremo superior izquierdo -valor máximo de (sensibilidad + especificidad)/2-, salvo situaciones en las que existan razones clínicas de peso que justificaran la necesidad de minimizar sólo los falsos positivos o sólo los falsos negativos.

Si se trata de una patología donde el establecer un diagnóstico y tratamiento precozmente influyen de manera determinante en la eficacia del resultado, deberían establecerse puntos de corte suficientemente bajos para incluir todos los posibles casos, a pesar de aumentar la cifra de falsos positivos. También conviene tomar en cuenta que el manejo de los falsos positivos no es un tema fácil, existiendo situaciones en las cuales el decir que se tiene la enfermedad puede ser prácticamente igual de invalidante que el tenerla en realidad: *“Lo que se percibe como cierto, aunque sea falso, tiene consecuencias reales”* (Renaud, 1996).

Se propone -para determinar las posibles mejores combinaciones de S/E- evaluar: los costes relativos de los resultados falsos -tanto positivos como negativos-, así como los beneficios de las clasificaciones correctas. Además, como se ha analizado, considerar la prevalencia del fenómeno de interés en la población.

Puede resumirse en la siguiente fórmula el cálculo de la tangente de corte sobre la curva (Zweig y Campbell, 1993):

Punto de corte =	$\frac{\text{Coste de los falsos positivos}}{\text{Coste de los falsos negativos}} * \frac{(1 - \text{Prevalencia})}{\text{Prevalencia}}$
------------------	---

Si el coste relativo de los falsos positivos es mayor que los falsos negativos, se obtendrá un punto de corte más cercano a la izquierda, priorizando la especificidad. Lo contrario en el caso que sea mayor el coste de los falsos negativos.

Las curvas ROC sirven también para la evaluación de pruebas alternativas para un mismo diagnóstico, a través de la comparación de las áreas debajo de la curva. En estos

casos se requiere considerar la forma de la curva, ya que con formas distintas pueden alcanzarse iguales áreas pero las conclusiones pueden ser erróneas (Moons y cols, 1997).

Las ventajas de las curvas ROC vienen dadas por su facilidad de interpretación visual de esta representación sobre la precisión y habilidad discriminativa de un test, siendo además independiente de la prevalencia. Un indicador de éstas ventajas sería su utilización en múltiples áreas¹⁴.

Algunos puntos negativos podrían relacionarse con su subutilización -a pesar de ser incluidas en publicaciones muchas veces no se sacan conclusiones sobre ellas-; además algunos problemas con el trazado de la curva (forma dentada) ante ciertos tipos de muestras; y una insuficiente disponibilidad de software.

Una consideración final, siempre deben evaluarse las implicaciones clínicas (riesgos y beneficios) de las decisiones diagnósticas y terapéuticas adoptadas una vez utilizado este instrumento ya que el mismo nos habla de la precisión de la prueba diagnóstica (eficacia) pero no de las consecuencias en el paciente particular estudiado (efectividad).

IV.3.4 Teorema de Bayes. Cociente de probabilidad.

En la mayoría de las situaciones clínicas, la información disponible no permite establecer un diagnóstico con certeza, y cada nueva evidencia clínica o paraclínica, promueve la revisión de las probabilidades de las varias alternativas planteadas.

El objetivo que se busca con la realización de las pruebas diagnósticas es descartar o confirmar la presencia de cierta enfermedad, en términos de probabilidades aproximar la probabilidad de una enfermedad hacia 0 o hacia 1 (Rué, 1998). Se trata de un proceso iterativo, donde pueden identificarse tres tipos de probabilidades: la probabilidad del diagnóstico antes de la presencia de una nueva evidencia se conoce como probabilidad previa; dada la supuesta evidencia (por ejemplo un nuevo síntoma o signo clínico o el resultado de un test) cual sería la frecuencia en que puede ser observada la enfermedad en estudio, es la probabilidad condicionada; cuando revisamos la probabilidad previa una vez presente este nuevo resultado diagnóstico, se habla de probabilidad posterior. Los términos previa y posterior quedan por tanto definidos en relación a cada nueva evidencia diagnóstica (Pauker y Kassirer, 1987).

Como se describe en los capítulos anteriores los datos clínicos pueden no resultar demasiados precisos; si se añade esto al hecho de que los resultados de las pruebas tampoco lo son, se plantea su integración como una alternativa a mejorar las predicciones diagnósticas en relación a si se consideran cada uno por separado.

¹⁴ En regresión logística, en la selección del umbral de decisión, comparación de tests diagnósticos.

La Teoría de las probabilidades del análisis de Bayes¹⁵ es una técnica matemática que se utiliza para integrar los datos que aporta el examen clínico con el resultado de las pruebas diagnósticas a través de la relación que establece entre las tres probabilidades antes mencionadas. Este enfoque puede concretarse siempre que los datos clínicos sirvan para estimar la probabilidad diagnóstica previa, y ésta se integre con el resultado y la sensibilidad y especificidad de una prueba.

Conociendo la probabilidad previa a la realización de la prueba y la sensibilidad y especificidad de la misma prueba, puede calcularse la probabilidad posterior, que es lo que nos interesa para la toma de decisiones.

Se trata de una vía alternativa para llegar al cálculo del valor predictivo a partir de la información que tienen los clínicos sobre sus pacientes.

Se ilustra a continuación la técnica con un ejemplo. Consideremos que la probabilidad previa de neumonía aportada por los datos clínicos y epidemiológicos es del 25%; para confirmar o descartar el diagnóstico se realiza una radiografía de tórax, de la que se obtiene una imagen compatible con el diagnóstico presuntivo (consolidación no segmentaria con broncograma aéreo) con una Sensibilidad = 90% y una Especificidad = 80%.

La fórmula que relaciona el valor predictivo de un resultado positivo con la sensibilidad, la especificidad y la prevalencia se calcula en acuerdo con el teorema de probabilidades condicionales de Bayes¹⁶:

$$P(E/\text{Prueba } +) = \frac{\text{sensibilidad} * \text{probabilidad previa enfermedad}}{\text{sensibilidad} * \text{prob. previa enferm} + (1 - \text{especificidad}) * \text{prob. no enferm}}$$

Con los datos del ejemplo:

$$P(E/\text{Prueba } +) = \frac{0.90 * 0.25}{0.90 * 0.25 + 0.20 * 0.75} = \frac{0.225}{0.225 + 0.15} = 0.6$$

¹⁵ Esta regla matemática se conoce desde hace más de dos siglos, pero sus aplicaciones en el razonamiento clínico comienzan en las últimas décadas. Ledley y Lusted fueron los primeros en aplicar esta regla a problemas médicos en 1959 (Bradley, 1993).

¹⁶ P (E/Prueba +) = probabilidad que ante un resultado positivo la enfermedad esté presente.
 Probabilidad previa enferm: es la estimada de acuerdo a la información disponible por el clínico.
 1 - Especificidad = índice de falsos positivos o probabilidad de un resultado + en ausencia de enfermedad (el b de la tabla 2 x 2).

Probabilidad de no enfermedad = 1 - probabilidad previa

¿Cómo se interpreta este resultado? La probabilidad de presentar neumonía pasó de un 25% previo a realizarse la prueba, a un 60% a partir de un resultado positivo de la misma.

Para que el Teorema de Bayes sea aplicable, la nueva información aportada por un test diagnóstico ha de ser independiente de la información previa que se disponga sobre el paciente.

Esto es, la probabilidad de un resultado positivo en la prueba no puede pronosticarse como diferente para ningún paciente. La independencia no ocurre normalmente en Medicina, pues cada individuo constituye una unidad fisiológica por lo que resulta improbable que en la mayoría de los casos las pruebas utilizadas tengan resultados totalmente independientes unos de otros.

Si la asunción de que las pruebas son totalmente independientes unas de otras es falsa el cálculo de probabilidades de la enfermedad a partir de las diversas pruebas tiende a sobrestimar el valor de las pruebas (Fletcher y cols, 1998). Si se combinan un ECG de esfuerzo y una angiocardiografía en un mismo paciente con sospecha de estenosis coronaria, se encontrarán muchos pacientes que siendo positivos en una prueba lo serán también en la otra; lo mismo en relación a los resultados negativos. Esto demuestra la dependencia de las pruebas entre si, y para poder aplicar el Teorema de Bayes se debe primero estimar el grado de concordancia entre ambas pruebas, para lo cual puede resultar útil una tabla con la siguiente información.

Tabla 4.4 Combinación de angiocardiografía y ECG de esfuerzo en el diagnóstico de estenosis coronaria (paciente con más del 70% de estenosis) (Sackett y cols, 1989)

		<i>Angiografía</i>		<i>Sensibilidad:</i>	
		<i>Positivo</i>	<i>Negativo</i>	<i>Angiografía</i>	
<i>ECG de</i>	<i>Positivo</i>	276	23	299	Con ECG + 276/299=92.3 %
<i>Esfuerzo</i>	<i>Negativo</i>	140	61	201	Con ECG - 140/201=69.7 %
		416	84	500	Total 416/500=83.2 %

Sensibilidad del ECG de esfuerzo

<i>Con Angiografía +</i>	<i>Con Angiografía -</i>	<i>Total</i>
276/416=66.3%	23/84=27.4%	299/500=59.8%

En el caso de la angiocardiografía, con una sensibilidad global del 83.2%, se produce un incremento de la misma si se consideran ambas pruebas con resultados positivos, pero un descenso de la sensibilidad al estudiar los pacientes con su ECG de esfuerzo negativo.

El uso de varias pruebas secuenciales en un mismo paciente es bastante frecuente, por lo que cabe aclarar que los términos probabilidad previa y posterior se definen en relación

a un test determinado, siendo la probabilidad posterior respecto a un test, previa en relación con el test siguiente. Si al paciente de nuestro ejemplo se le realizase otra prueba la probabilidad previa de enfermedad sería entonces del 60%.

Cuando se indican varias pruebas diagnósticas a un paciente, es posible realizarlas en serie o en paralelo.

en serie

prueba 1 ----- prueba 2 ----- prueba 3	disminuye sensibilidad aumenta especificidad
--	---

en paralelo

prueba 1 + prueba 2 + prueba 3	aumenta sensibilidad disminuye especificidad
--	---

Cuando se realizan en serie lleva al máximo la especificidad y el valor predictivo positivo, pero disminuye la sensibilidad y el valor predictivo negativo. En paralelo sin embargo, las consecuencias son inversas y se acumulan falsos positivos. Estas últimas son útiles en los casos en que sólo hay disponibles pruebas con poca sensibilidad, teniendo como efecto neto un incremento de la sensibilidad al utilizar las dos pruebas a la vez. El efecto negativo puede resultar de someter a tratamiento algunos pacientes sin la enfermedad.

La probabilidad de tener una prueba con un resultado anormal se incrementa con el aumento del número de pruebas.

Tabla 4.5 *Relación entre cantidad de pruebas realizadas y probabilidad de obtener resultados anormales*

<i>Número de pruebas</i>	<i>% de resultados anormales debidos al azar</i>
1	5 %
5	23 %
20	64 %
100	99 %

Este efecto puede verse también al practicar el mismo test en varias oportunidades a la misma persona, como sucede en los cribajes. Luego de realizar 10 mamografías la probabilidad de obtener al menos un falso positivo fue del 49% (Elmore y cols, 1998). En la misma fuente se obtuvieron los datos de la siguiente tabla, donde se muestra el

resultado a 10 años de la aplicación de un programa de detección precoz de cáncer de mama utilizando examen clínico, mamografía o ambos.

Tabla 4.6 Falsos positivos en cribaje de cáncer de mama luego de un período de 10 años (Elmore y cols, 1998).

<i>Variable</i>	<i>Mamografía</i>	<i>Examen clínico</i>	<i>Ambos test</i>
<i>Test de cribaje</i>	9762	10905	20667
<i>Falso positivos tests</i>	631 (6.5%)	402 (3.7%)	1033 (5.0%)
<i>Número de mujeres</i>	2227	2245	2312
<i>Mujeres con al menos un resultado falso +</i>	530 (23.8%)	300 (13.4%)	734 (31.7)

La estimación de la probabilidad también puede trabajarse en términos de "odds". Elegir uno u otro método es indistinto, depende de la practicidad que el cálculo tenga para uno u otro profesional y también del área que se estudie. En distintos trabajos se describen por un lado fundamentos a favor de utilizar el cociente de probabilidad (Gallegher, 1998) y por otro la aplicación de la ecuación de Bayes (Dominici, 1998).

La información que aportan la "odds" y la probabilidad es idéntica, lo que varía es la forma de expresarla: si la probabilidad de tener una afección es del 75%, la "Odds" es de 3:1.

Para convertir las probabilidades en odds se aplica la siguiente fórmula:

$$\text{Odds} = \text{probabilidad de estar enfermo} / (1 - \text{probabilidad de estar enfermo}).$$

Y para convertir "odds" en probabilidades:

$$\text{Probabilidad} = \text{Odds} / 1 + \text{Odds}$$

$$\text{Odds posterior} = \text{odds previa} * \text{cociente de probabilidad}$$

El cociente de probabilidad es la relación entre la sensibilidad (probabilidad de obtener un resultado positivo de la prueba en presencia de enfermedad) y la probabilidad de obtener un resultado positivo en ausencia de enfermedad (falsos positivos o complemento de la especificidad $1 - E$).

$$\text{Cociente de probabilidad} = \text{sensibilidad} / 1 - \text{especificidad}$$

Este cociente expresa las posibilidades que un nivel dado del resultado de una prueba diagnóstica cabría esperar en un paciente con la enfermedad en relación a otro sin la enfermedad.

El cociente de probabilidad o cociente de verosimilitud tiene, entre otras ventajas, que relaciona sensibilidad y especificidad en un solo índice, proporcionando una información de tipo horizontal como los valores predictivos, pero que no depende de la probabilidad pre-prueba.

Con el ejemplo anterior, tendríamos una razón de probabilidad de 4.5 ($0.9/0.2 = 4.5$); en cuanto a la interpretación, en el grupo de pacientes diagnosticados con neumonía la probabilidad de una imagen radiográfica típica es 4.5 veces mayor que en los que se ha descartado dicha enfermedad¹⁷.

Otra ventaja, es que pueden obtenerse valores para los diferentes niveles de un resultado (ver más adelante el ejemplo de la gammagrafía) y se ajustan bien a la descripción de la probabilidad global de enfermedad cuando se utiliza una serie de pruebas diagnósticas.

Entre las desventajas que se le atribuyen, por un lado mayor dificultad para razonar en términos de “odds” que de probabilidades; y por otro la necesidad de un cálculo para realizar la conversión.

Para facilitar el cálculo puede accederse a los llamados nomogramas, que permiten, partiendo de las probabilidades previa y el cociente de probabilidad, estimar la probabilidad posterior (Sackett y cols, 1989).

En los nomogramas se presentan 3 columnas: la primera con la probabilidad previa o preexamen, la segunda con el cociente de probabilidad y la tercera muestra la probabilidad postexamen. Ayudándonos de una regla se alinean los valores de las dos primeras columnas para obtener el valor de la tercera que es el que nos interesa conocer. Esta herramienta tiene la limitación que solo puede utilizarse en los casos en que se plantean dos opciones: enfermedad presente o ausente, y el test tiene un comportamiento binario: positivo-negativo.

Cuando se calcula la razón de probabilidad de una prueba negativa (RPN) se divide la especificidad y el complemento de la sensibilidad

¹⁷ Para el mismo ejemplo:

$$\text{Odds previa} = \frac{(\text{probabilidad previa})}{1 - (\text{probabilidad previa})} = \frac{0.25}{0.75} = 0.33 = 1/3$$

$$\text{"Odds" posterior} = 0.33 * 4.5 = 1.5$$

$$\text{Probabilidad posterior} = \text{odds} / (1 + \text{odds}) = 1.5 / 2.5 = 0.6$$

$$\text{RPN} = \frac{\text{especificidad}}{1 - \text{sensibilidad}}$$

Cada apartado de la historia clínica y del examen físico constituye un examen diagnóstico que modifica la probabilidad de un trastorno objetivo (la aumenta o la disminuye). Los cocientes de probabilidad indican hasta que punto un resultado determinado aumentará o disminuirá la probabilidad preexamen de un trastorno objetivo. Un valor de 1, significa que la probabilidad postexamen es exactamente la misma que la preexamen. Los cocientes superiores a 1 incrementan dicha probabilidad, y los inferiores a 1 la disminuyen. Una guía sobre la significación de los valores de estos cocientes sería la siguiente:

Tabla 4.7 Significado de distintos valores de los cocientes de probabilidad (Jaeschke y cols, 1997).

<i>Valores del cociente de probabilidad</i>	<i>Interpretación</i>
Superiores a 10 (o inferiores a 0.1)	Generan cambios importantes y frecuentemente concluyentes.
Entre 5 y 10 (o entre 0.1-0.2)	Implican cambios moderados.
Entre 2 y 5 (o 0.5 y 0.2)	Generan cambios pequeños (pueden ser importantes en ciertas situaciones).
Entre 1 y 2 (o 0.5 a 1)	son insignificantes.

Con el siguiente ejemplo practicaremos el empleo de esta herramienta

Tabla 4.8 Resultados de la aplicación de gammagrafía ventilación-perfusión en dos pacientes con embolia pulmonar (EP). (Jaeschke y cols, 1997):

<i>Probabilidad preexamen</i>	<i>Resultado gammagrafía (CP¹⁸)</i>	<i>Probabilidad postexamen</i>
Mujer de 78 años, con inicio	Súbito de disnea tras cirugía	Abdominal
70 %	Probabilidad elevada 18,3	97 %
70 %	Probabilidad intermedia 1,2	74 %
70 %	Probabilidad baja 0,36	46 %
70 %	Normal 0,1	19 %
Hombre de 28 años, con	Disnea y dolor torácico atípico	
20 %	Probabilidad elevada 18,3	82 %
20 %	Probabilidad intermedia 1,2	23 %
20 %	Probabilidad baja 0,36	8 %
20 %	Normal 0,1	2 %

¹⁸ CP: cociente de probabilidad.

La utilidad de un examen diagnóstico por tanto estará en añadir información más allá de la ya disponible, influyendo esto en la conducta terapéutica beneficiosa para el paciente. Nuevamente puede observarse en los datos de la tabla que partiendo de idénticos cocientes de probabilidad, la probabilidad posterior a la gammagrafía depende de la probabilidad previa al examen.

En cuanto a la interpretación de los coeficientes, podemos deducir que una paciente con una probabilidad previa del 70% de presentar EP tiene 18.3 veces más probabilidad de presentar una gammagrafía cuyo resultado sea “probabilidad elevada” que en una paciente sin EP. Otro componente importante de la tabla y que refleja situaciones frecuentes en la práctica clínica, es la presencia de una gama de resultados posibles, y no una conclusión dicotómica: presente-ausente, positivo-negativo.

El clínico para aplicar este razonamiento deberá estimar por un lado la probabilidad previa en su paciente y conocer la sensibilidad y especificidad del test (o el cociente de probabilidad). El primer dato es básico y depende directamente de su desempeño profesional, las diferentes alternativas para su obtención serán mencionadas más adelante.

En síntesis, el Teorema de Bayes es una herramienta que sirve como guía para interpretar en forma integrada la información clínica y el resultado de las pruebas diagnósticas.

Se describen a continuación una serie de aspectos para ser considerados y así evitar que se cometan errores en su aplicación:

1. Dificultades en utilizar las propiedades de una prueba –sensibilidad y especificidad-: es común que se evalúen individuos claramente enfermos en comparación con individuos que con bastante seguridad no presentan la enfermedad; además los pacientes pueden diferir en relación a su gravedad, estadio y duración (Fletcher y cols, 1998)

Cuando un test es introducido por primera vez, se utilizan personas voluntarias saludables, en las cuales la especificidad del test alcanza valores muy elevados (sería una sobrestimación de la especificidad). Cuando el test es utilizado con mayor frecuencia, los casos con resultados positivos son habitualmente referidos con el objetivo de un diagnóstico definitivo, no así los de resultados negativos; en esta situación la frecuencia de verdaderos negativos es subestimada, y la especificidad del test menor en relación a la población donde fue evaluado el test –sesgo de referencia- (Sox, 1986).

El sesgo de selección hace que en los pacientes donde primeramente se aplica el test (con mayor certidumbre de presencia de enfermedad, y muchas veces más graves) puedan elevar la sensibilidad del resultado del test. El criterio de referencia en estos

casos, hace que se subestimen los falsos negativos, sobrestimándose la sensibilidad del test (Sox, 1986).

2. Las afirmaciones sobre el valor predictivo en la bibliografía disponible, puede llevar a cometer errores si las pruebas se aplican en ámbitos con diferente prevalencia por tratarse de población que es referida a un especialista o a nivel de un hospital, por lo que se tiende a sobrestimar las prevalencias (Sox, 1986). Además es frecuente comparar el rendimiento de una prueba entre enfermos y una población igual en número de no enfermos, por lo que la prevalencia definida para el cálculo del valor predictivo es del 50% (Fletcher y cols, 1998). Por tanto el espectro de pacientes donde se evaluaron las pruebas puede diferir del espectro en el cual las pruebas serán utilizadas habitualmente.

3. Falta de información sobre resultados de pruebas en ausencia de enfermedad: el rendimiento de una prueba puede analizarse en forma errónea si el resultado de la misma solo se conoce en pacientes con la enfermedad: similares hallazgos se encontraron al realizar una resonancia magnética a pacientes con dolor lumbar y a otros sin sintomatología de columna, solo el 36% de los examinados presentaban una imagen normal de su columna (Jensen y cols, 1994).

4. Cuando se aplica el T de Bayes, se asumen la sensibilidad y especificidad como constantes, pero esto puede ser falso. Un test puede ser menos sensible en detectar una enfermedad si se aplica en etapas más tempranas, con una menor probabilidad pre-test, y mucho más en etapas avanzadas con un evidente cuadro clínico y con una mayor probabilidad pre-test. En estos casos deberían establecerse sub-grupos de acuerdo a que tipo de evidencia clínica se encuentra. Además, cada estadio puede tener situaciones de alta y baja prevalencia.

5. El otro problema con el T de Bayes ya fue referido: el asumir que la sensibilidad y especificidad de un test son independientes del resultado de otros tests. En estos casos la sensibilidad y especificidad deberán ser estimadas para dos situaciones: en pacientes con un resultado positivo en el primer test, y también en caso de resultado negativo. Si tanto la sensibilidad como la especificidad del segundo test son similares ante resultados diferentes del primer test, podrá asumirse la independencia. En la práctica, es más probable asumir independencia cuando los test miden diferentes aspectos de la fisiopatología.

6. Cuando las estimaciones de la probabilidad previa son comparadas en grupos de médicos, se detecta una elevada variabilidad en sus cálculos; y cuando las predicciones son comparadas con resultados conocidos, las correlaciones pueden ser muy pobres (Bradley, 1993).

En un primer momento las aplicaciones en el área de la decisión clínica de estos instrumentos relacionaban el valor predictivo de los datos clínicos y los resultados de las pruebas diagnósticas, con la probabilidad de ciertas enfermedades (Diamond y Forrester, 1979) área en la que se sigue utilizando hasta el momento actual. A

continuación, y solo en forma de breve descripción, se enumeran algunos ejemplos de publicaciones más recientes para conocer en que campos se están utilizando hoy en día.

1. Comparar la utilidad de distintas herramientas diagnósticas en una misma enfermedad. En el caso del asma, se comparan dos instrumentos: un cuestionario y el test de ejercicio (Demissie y cols, 1998).
2. Analizar el valor predictivo de los datos clínicos para ciertas afecciones. En los pacientes con sospecha de enfermedad colónica se propone un modelo predictivo para establecer diagnóstico a partir de ciertas variables clínicas, además de valorar la prioridad en la realización de la colonoscopia con el criterio de grupo de población seleccionado (que se describe más adelante) con la finalidad de tener una población con mayor probabilidad previa (Siles y cols, 1997).
3. En la predicción de los resultados después de cirugía mayor: supervivencia con o sin complicaciones, o probabilidad de morir (Millili y cols, 1998).
4. Comparar los resultados obtenidos al administrar diferentes terapéuticas, por ejemplo diferentes trombolíticos en el Infarto Agudo de Miocardio (Brophy y Joseph, 1995).

IV.3.5 Umbrales de decisión: diagnóstico y terapéutico.

Si tras la realización de un test diagnóstico obtenemos una probabilidad de enfermedad similar a la previa, la utilidad del mismo resulta dudosa. El modelo de los umbrales de decisión es una extensión de la idea de evaluar a un test según su efecto sobre la probabilidad de enfermedad.

A partir de la probabilidad previa Pauker y Kassirer (1980) describen tres situaciones posibles: a) si la probabilidad de padecer la enfermedad es muy baja, lo más racional sería no solicitar ninguna prueba en relación a dicha enfermedad y por tanto no trataríamos al paciente; con un resultado positivo de la prueba no obtendríamos una probabilidad suficiente como para decidir su tratamiento; b) lo mismo sucedería en el caso de una probabilidad previa muy alta: en este caso lo racional sería pasar directamente al tratamiento específico ya que un resultado positivo no mejoraría la información disponible; c) cuando tenemos una incertidumbre importante sobre la presencia de la enfermedad -probabilidades previas de alrededor del 50%- justificaría claramente la realización de la prueba. Es la situación en la cual ganaríamos mucha información para aumentar o disminuir la probabilidad de padecer la enfermedad, influyendo por tanto en la decisión clínica de tratar o no tratar.

Existen dos umbrales relevantes en la decisión clínica, el umbral diagnóstico y el umbral terapéutico. La información necesaria para su cálculo: probabilidad previa o

prevalencia de la enfermedad, los riesgos y beneficios de un tratamiento adecuado y los riesgos de la prueba diagnóstica y su precisión en términos de frecuencia de falsos positivos y falsos negativos (Rué, 1998).

$$\text{umbral diagnóstico}^{19} = \frac{\text{FP} * \text{Rtto} + \text{Rt}}{\text{FP} * \text{Rtto} + \text{VP} * \text{Btto}}$$

$$\text{umbral terapéutico} = \frac{\text{VN} * \text{Rtto} - \text{Rt}}{\text{VN} * \text{Rtto} + \text{FN} * \text{Btto}}$$

Donde:

VP: proporción de verdaderos positivos o sensibilidad

FN: proporción de falsos negativos o 1 - sensibilidad

FP: proporción de falsos positivos o 1 - especificidad

VN: proporción de verdaderos negativos o especificidad

Btto: beneficio neto del tratamiento o diferencia en utilidad para los individuos enfermos entre tratarse y no tratarse, o consecuencias positivas que como media se producen en los pacientes que tienen la enfermedad y son tratados.

Rtto: riesgo neto del tratamiento o diferencia en utilidad para los individuos sanos, entre no tratarse y tratarse. Mide las consecuencias negativas que como media se producen entre los pacientes que sin padecer la enfermedad son tratados por equivocación.

Rt: riesgo neto del test o diferencia de utilidad entre obtener un resultado con y sin exponer al paciente al riesgo del test.

La interpretación del cálculo sería: por debajo del umbral diagnóstico no está justificada acción ninguna; por encima del umbral terapéutico debe pasarse directamente al tratamiento; entre ambos umbrales lo indicado dependerá del resultado de las pruebas diagnósticas.

Se retomará un concepto que en forma rápida planteamos al principio del capítulo:

¹⁹ Umbral diagnóstico y terapéutico de realizar una gastroscopía para descartar un cáncer gástrico a un paciente de 60 años, hombre, con dolor epigástrico y hematemesis, y que presenta en el gastroduodeno una imagen de úlcus en curvatura mayor del estómago. Mortalidad por gastroscopía: 0.005%, Sensibilidad: 96%, Especificidad: 98% Los riesgos y beneficios del tratamiento (gastrectomía): mortalidad 2%, la ganancia en tasa de supervivencia con una cirugía temprana es del 33%.

$$\text{Umbral diagnóstico} = (0.02 * 2) + 0.005 / (0.02 * 2) + (0.96 * 33) = 0.0014$$

$$\text{Umbral terapéutico} = (0.98 * 2) - 0.005 / (0.98 * 2) + (0.04 * 33) = 0.60$$

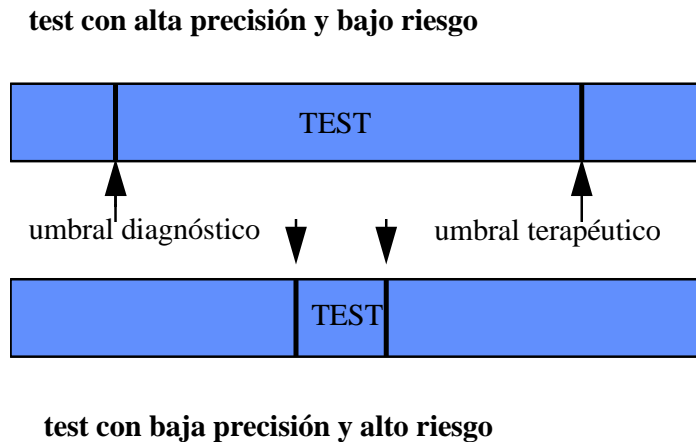
Probabilidad de cáncer menor a 0.0014%, ni gastroscopía ni cirugía son recomendables; si la probabilidad de cáncer es superior a 0.6 se debería operar sin previa gastroscopía; en el caso de una probabilidad previa intermedia, se debería realizar previamente la gastroscopía (de Pauker y Kassirer, 1980).

La racionalidad clínica de realizar un test diagnóstico radica en su capacidad de variar de la probabilidad de enfermedad y por tanto de modificar la decisión terapéutica una vez conocido su resultado.

El umbral diagnóstico aumenta en relación directa con el riesgo de la prueba, y disminuye cuando existe una alta tasa de verdaderos positivos y los beneficios del tratamiento son altos.

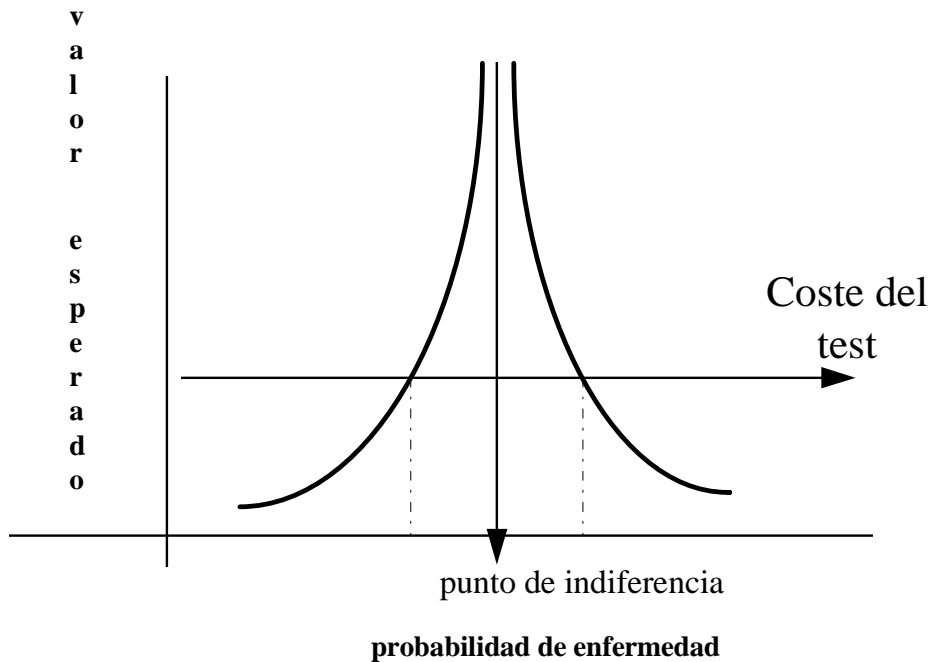
En las siguientes figuras puede observarse con bastante claridad el efecto del riesgo de la prueba diagnóstica y su precisión sobre el nivel de los umbrales (considerando que la línea horizontal representa un rango del 0 al 1).

Fig. 4.5 *Relación entre el riesgo y precisión de la prueba diagnóstica, y el nivel de los umbrales (Pauker y Kassirer, 1980):*



La siguiente gráfica une el concepto de umbrales de decisión con el de ganancia de información antes mencionado.

Fig. 4.6 Valor esperado de la información aportada por un test diagnóstico (Phelps, 1997).



El mayor valor esperado de la información de una prueba diagnóstica se obtiene en el punto donde es indiferente tratar o no tratar al paciente.

No se está proponiendo estimar los umbrales ante cada paciente particular, ya que es evidente que no se trata de un cálculo simple; pero este enfoque contiene elementos que pueden ser muy útiles tanto en problemas clínicos genéricos o en casos específicos: en decisiones que se toman sobre pacientes ante situaciones comparables; como insumo para la elaboración de guías de práctica clínica, y ante procedimientos diagnósticos o terapéuticos de los que puedan resultar riesgos considerables para los individuos, potenciándose en los casos en que la precisión del test no sea muy elevada.

Algunos problemas de la técnica: se requiere para el cálculo reunir datos de diferentes fuentes, no siempre fácilmente disponibles para ciertas patologías; la asignación de los beneficios y riesgos no es simple en áreas donde entran a jugar varios factores que se relacionan con pérdida o ganancia de bienestar del paciente, además deben utilizarse unidades de medida comparables para lo que es necesario expresar todas las variables en términos homogéneos. Lo que se observa más frecuentemente son datos de mortalidad, de manera de simplificar las estimaciones, aunque el coste sea alejarse de representar fielmente la realidad.

IV.3.6 Probabilidad de enfermedad.

Se espera haber logrado demostrar la importancia de la estimación por parte del clínico de la probabilidad de que su paciente tenga o no el acontecimiento estudiado. La primera tarea cuando se utiliza el razonamiento probabilístico es estimar esta probabilidad previa: se parte de la premisa que una cuidadosa asignación de probabilidad es crítica, teniendo su valor la mayor influencia en la interpretación de los hallazgos clínicos y de los resultados de los test (Kassirer, 1989). Importa entonces, a la hora de mejorar la semiología, solicitar los test diagnósticos, así como la evaluación del tratamiento.

Existen diversas fuentes de información para realizar la estimación de probabilidad: la experiencia personal del médico, la literatura médica así como bases de datos del lugar, y las características peculiares del paciente.

La experiencia personal es el principal factor que influye en las estimaciones realizadas. En capítulos anteriores se han planteado algunos de los problemas más frecuentes: disponibilidad heurística, sobrestimación de los casos “raros”; determinación de la probabilidad inicial y los ajustes posteriores, etc.

También pueden encontrarse sesgos en los trabajos publicados. Es importante evaluar los potenciales sesgos de selección de pacientes, si se trata de casos vistos por especialistas, tenderá a elevarse la prevalencia de los pacientes más graves, en relación a la población que accede a un servicio de atención primaria.

Es muy útil considerar como punto de partida la información disponible sobre la prevalencia de la enfermedad en la población de donde procede el paciente, contrastada a su vez con la incidencia de la enfermedad en un período de tiempo determinado. Con este punto de partida, los médicos pueden ajustar la estimación a partir de su experiencia y de las características específicas que presente el paciente particular (Sox y cols, 1988).

La estrecha relación entre probabilidad previa y el valor predictivo de una prueba, hace evidente la utilidad de aplicar las pruebas diagnósticas teniendo en cuenta esta variable desde el punto de vista de la efectividad clínica.

Existen varias alternativas para incrementar la frecuencia de los casos en la población a la cual se le van a realizar los test; Fletcher propone las siguientes:

1. El *proceso de referencia* a especialistas y a hospitales desde atención primaria justifica la utilización en forma más intensiva de pruebas diagnósticas en estos entornos donde se incrementa la probabilidad pre-prueba. Un ejemplo puede verse en los casos de pacientes con dolor en el pecho, en los cuales según fueran grupos procedentes de

atención primaria o referidos para realizar coronariografía, aun presentando el mismo cuadro clínico (según puntuación pre-establecida) la presencia de enfermedad coronaria fue muy diferente (Sox y cols, 1990).

2. Los *grupos de población seleccionados* (las mujeres cuya madre o hermana han padecido cáncer de mama, tienen el doble de probabilidad de padecer este tipo de cáncer en relación al promedio en las mujeres).

3. Las *características específicas de la situación clínica* (signos, síntomas y factores de riesgo) que deberían tener la más fuerte influencia sobre la decisión de solicitud de las pruebas diagnósticas²⁰.

A modo de síntesis, la utilidad de una prueba diagnóstica se encuentra en la posibilidad de la misma en incorporar información en el proceso de elaboración diagnóstica a la que pueda obtenerse del examen clínico y de otras pruebas menos costosas, es decir la nueva probabilidad que surge a partir de la realización de la prueba deberá implicar un cambio en la conducta terapéutica o en la secuencia diagnóstica a seguir.

IV.4 Situación real y perspectivas

Algunas investigaciones muestran que de las diez pruebas diagnósticas habitualmente más indicadas en pacientes hospitalizados un 40 % son repetidas aunque no sean necesarias. En algunos casos estas cifras de repetición innecesaria alcanza el 70% (digoxinemias y nivel de gentamicina en sangre) (Browner, 1998).

La utilización con un criterio “defensivo” de las pruebas diagnósticas si bien mejora los resultados clínicos de algunos pacientes, puede empeorar la situación de otros, pero en conjunto la calidad de la atención médica se deteriora

Una actitud defensiva se produce cuando el médico es responsable ante los errores clínicos y espera reducir la probabilidad de cometer errores y demandas por malpraxis a través de la indicación de más técnicas diagnósticas, siendo en estos casos una conducta racional aunque no necesariamente apropiada. No solo se encarece la atención médica, sino que empeora la calidad de la misma: la presencia de falsos positivos hace que se requiera la indicación de otros estudios para confirmar o descartar el diagnóstico, lo que en general se logra con técnicas más invasivas, con mayor riesgo y morbilidad. Otro riesgo en que puede incurrirse con esta conducta es el tratar pacientes sin la enfermedad (Dekay y Asch, 1998).

²⁰ La utilización de programas informáticos facilita en la actualidad las estimaciones cuantitativas de la probabilidad de la enfermedad, dada ciertas combinaciones de hallazgos clínicos.

La responsabilidad de los errores por comisión (tratar un paciente sin la enfermedad en cuestión) y por omisión (no tratar a un enfermo) se pueden cuantificar a través de los criterios vistos al definir los umbrales y dependerá del valor del cociente falsos positivos/falsos negativos definidos.

En general depende que el médico sea más o menos averso al riesgo (cuanto más averso, más defensivo será en sus indicaciones). Además varía con el tipo de entrenamiento recibido, su experiencia, el tiempo de dedicación a cada paciente, la calidad de la relación médico-paciente, tipo de planes asistenciales (modelo capitativo, etc) (Pauker y Pauker, 1998). Las dificultades que los médicos presentan en trabajar con probabilidades, y en particular en estimar la probabilidad de presencia de la enfermedad²¹ también se ve afectada. Algunos errores descriptos: una actitud aversa al riesgo por parte de los médicos lleva en general a subestimar eventos comunes, mientras se sobrestiman las probabilidades de las condiciones clínicas más severas y complejas (Phelps, 1997).

Otro enfoque prioriza el problema de la incertidumbre. La expansión del conocimiento médico, como consecuencia de un mayor volumen de compleja información aportada a los médicos, en vez de disminuir ha producido un incremento de la incertidumbre (Eddy, 1984). Aunque la incertidumbre sea inevitable, no siempre se considera su verdadera importancia en medicina. Se plantea la necesidad de que el médico reconozca su presencia ante sí y ante sus pacientes, explicitarla, como un primer paso en la mejora de la calidad de la práctica médica (Logan y Scott, 1996)

También se ha investigado la incertidumbre como generadora de estrés en los profesionales. El nivel de estrés sería mayor en las mujeres profesionales y en los médicos con menos experiencia (Gerrity y DeVellis, 1990). La actitud ante la incertidumbre, repercute en el nivel de gasto incurrido en la atención de cada paciente: un médico con una mayor tolerancia ante la incertidumbre generaría menores costes por paciente (menos pruebas diagnósticas, menos prescripciones); los médicos que expresan mayor ansiedad ante la incertidumbre y con mayor rechazo a revelar la misma ante sus pacientes, presentan promedialmente mayores costes de atención (Allison y cols, 1998).

La necesidad de reducir el nivel de incertidumbre diagnóstica es una tarea que se vincula a la posibilidad de aplicar una óptima decisión terapéutica. Pero si bien realizar más tests no necesariamente producen mayor certidumbre (tabla 4.5) la disconformidad médica con la incertidumbre se expresa en una “cascada” interminable de pruebas diagnósticas (Thibault, 1994), potenciada por la formación recibida donde el dudar no se ve con buenos ojos (la incertidumbre es considerada sinónimo de ignorancia en la formación tradicional), el aval del paciente y en ciertos casos por la facilidad -desde un punto de vista institucional- de hacerlo. La escalada de costes en los servicios de salud ha sido uno de los factores principales en la limitación de su empleo, pero no siempre asegurando la reducción del uso inapropiado. Se han instrumentado diferentes

²¹ Una investigación refiere que sólo el 10% de un grupo de médicos estimó correctamente el valor predictivo de una prueba cuando la información era presentada como probabilidad, frente a un 46% si la forma de presentación era a través de frecuencias absolutas. Acad Med 1998; 73 (5): 538-40.

mecanismos para disminuir el número de exámenes solicitados, muchos de ellos basados en información a los médicos sobre los costes incurridos (Tierney y cols, 1990). Dichas medidas repercuten generalmente en el corto plazo, sin lograr un cambio en la conducta médica más allá del período de aplicación de las intervenciones.

Además, el control de costes en muchas instituciones repercute en un menor tiempo de intercambio directo médico-paciente, lo que reduce la información que el médico puede obtener en forma directa del mismo.

Con la cada vez mayor disponibilidad de pruebas diagnósticas en general se decide su solicitud aplicando un enfoque poco selectivo, teniendo como consecuencia inmediata una menor probabilidad previa de enfermedad y un bajo valor predictivo de las pruebas solicitadas. La magnitud de este efecto puede ser muy alta (Fletcher y cols, 1998).

Un agravante es la falta de conocimiento acerca del valor de ciertas pruebas y también de muchos tratamientos (Kassirer, 1993). En el caso de los ingresos hospitalarios, numerosos estudios demuestran la bajísima relación coste-beneficio que rinde la batería de test establecidos como rutina (Kassirer, 1989). También puede confirmarse la hipótesis de Kassirer (1989) de excesivo celo médico por la certeza diagnóstica al comprobarse en ciertas afecciones la indicación de tests que comprueban lo mismo, que son redundantes o agregan mínima información a la obtenida previamente.

El problema de la subutilización también existe. Como resultado de una investigación sobre pruebas de laboratorio complementarias, se detectó un 24% de subutilización en relación al total de pruebas realizadas (Castellví i Boada, 1996).

Por otro lado, los médicos raramente utilizan información sobre los test (sensibilidad-especificidad) y menos aplican razonamientos formales sobre la probabilidad en su práctica clínica. Una encuesta realizada a 300 médicos sobre la frecuencia de algunas de las técnicas antes descritas (Reid y cols, 1998) muestra los siguientes resultados: menos del 25% considera la información sobre sensibilidad y especificidad de las pruebas antes de indicarlas, 3% estima la probabilidad postprueba aplicando Bayes, solo 3 profesionales emplean las curvas ROC y 2 el cociente de probabilidad. Entre las causas de baja utilización de la sensibilidad, especificidad y cocientes de probabilidad, se mencionan: la no fácil accesibilidad a la información (ya sea porque el test no ha sido evaluado, o porque no se dispone de los datos); si bien el 70% refiere algún tipo de entrenamiento formal, el mismo sería insuficiente; también se nombran la diferencia de pacientes vistos en consulta con los utilizados para evaluar los test. La poca practicidad es mencionada como principal causa de no uso del Teorema de Bayes, también la “incomodidad” de realizar el cálculo, y el insuficiente entrenamiento.

Desde el punto de vista de facilitar el acceso a las herramientas descritas a lo largo del capítulo, las reglas de predicción clínica en las cuales se define a partir de qué conjunto de hallazgos clínicos y paraclínicos puede estimarse la probabilidad de enfermedad, el apoyo informático para realizar las estimaciones de probabilidad, así como algoritmos que orienten en la indicación de los test diagnósticos puede constituirse en una ayuda importante²².

Para otros el problema fundamental está en la estructura de la educación médica, siendo pobre el entrenamiento en estadística, epidemiología y otras disciplinas afines (Phelps, 1997), y donde se debería incluir la enseñanza de técnicas que permitan al médico combinar representaciones numéricas como probabilidades, riesgos y beneficios.

También se podrían mencionar los cambios institucionales necesarios.

Por otra parte se recomienda la aplicación de una mezcla de estrategias y de métodos, la cual sería mejor que la utilización de un mecanismo aislado.

Pero los distintos aspectos relacionados con el estilo de práctica médica se convierten en la principal dificultad para que se extienda el uso de estas herramientas: se trata de cambiar hábitos, modificar conductas establecidas y mantenidas por años. La misma práctica clínica nos enseña lo difícil que es hacerlo en los pacientes que requieren modificar su dieta, sus hábitos tóxicos, u otros componentes de sus estilos de vida. Se habla de la existencia de un requisito fundamental: “adherencia al tratamiento”, pues es el mismo médico el que durante los 365 días del año asumirá o no esta nueva conducta. Por tanto la necesidad de que el profesional esté envuelto en este proceso es básica.

Se requiere información adecuada, conciencia de la importancia de adoptar estas herramientas, incentivos para promover su uso, mejorar la relación con la academia donde recibe su formación y con los servicios de salud donde trabaja.

¿Cuales son las razones por las cuales los médicos deberían adherirse a la utilización en su práctica de una panoplia de instrumentos que promueva la mejora de la racionalidad clínica de sus indicaciones? El implicar al médico en el cambio deberá asegurar que como consecuencia: no se produce un descenso de su autoridad en la decisión clínica, no se reducen sus ingresos, y además repercute en beneficio de los pacientes (al menos no les produce daño) (Greco y Eisenberg, 1993).

Estos autores comparten el convencimiento de Williams (1990) de que las principales son razones éticas, la ética de fijarse como objetivos preservar la vida, aliviar el sufrimiento, minimizar el riesgo de daño, respetar los deseos del paciente, decir la verdad y asegurar que la distribución de la asistencia sea justa y equitativa.

²² Algunos ejemplos pueden verse en: Am J Emerg Med 1997; 15(7): 694-9 - J Med Syst 1997; 21(6): 429-44.

Bibliografía referenciada

Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Med Decis Making* 1998; 18(3): 320-9.

Bradley GW. *Disease, diagnosis & decisions*. New York: John Wiley & sons Ltd, 1993.

Brophy JM, Joseph L. Placing Trial in context using Bayesian analysis. *JAMA* 1995; 273: 871-75.

Browner WS. Diagnosis dilemma. *Am J Med* 1998; 104: 406-07.

Burgueño MJ, García-Bastos JL, Gonzalez JM. Las Curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clín (Barc)* 1995; 104: 661-70.

Castellví JM, Castells X. Appropriateness of physicians' request of laboratory examination in Primary Health Care: clinical over and under utilization study. *Clin Chem Lab Med* 1999; 37(1): 65-69.

Dekay ML, Asch DA. Is the defensive use of diagnostic test good for patients, or bad? *Med Decis Making* 1998; 18: 19-28.

Demissie K, White N, Joseph L, Ernst P. Bayesian estimation of asthma prevalence and comparison of exercise questionnaire diagnostics in the absence of a gold standard. *Ann Epidemiol* 1998; 8(3): 201-08.

Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Eng J Med* 1979; 300: 1350-58.

Dominici F. Testing simultaneous hypotheses in pharmaceutical trials: a Bayesian approach. *J Biopharm Stat* 1998; 8(2): 283-97.

Eddy DM. Variations in physicians practice: the role of uncertainty. *Health Affairs* 1984; 3: 74-89.

Elmore JG, Barton MB, Mocerri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998; 338: 1089-96.

Fletcher RH, Fletcher SW, Wagner EH. *Epidemiología Clínica*. Madrid: Masson-Williams & Wilkins, 1998: 43-75.

Gallagher EJ. Clinical utility of likelihood ratios. *Ann Emerg Med* 1998; 31(3): 391-7.

Gerrity MS; DeVellis RF. Physicians' reactions to uncertainty in patient care. *Medical Care* 1990; 28(8): 724-36.

Goldman L. Aspectos cuantitativos del razonamiento clínico. En: *Harrison, Principios de Medicina Interna*. Madrid: Interamer Mc-Graw-Hill, 1991: 6-13.

Greco PJ, Eisenberg JM. Changing physicians' practices. *N Engl J Med* 1993; 329: 1271-73.

Jaeschke R, Guyatt GH, Sackett DL. Cómo utilizar un artículo sobre un examen diagnóstico. En: *La Medicina Basada en la Evidencia*. JAMA (ed esp) 1997: 45-50.

Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic Resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994; 331: 69-73.

Kassirer JP. Our stubborn quest for diagnostic certainty: a cause of excessive testing. *N Engl J Med* 1989; 320: 1489-91.

Kassirer JP. Diagnostic reasoning. *Ann Intern Med* 1989; 110: 893-900.

Kassirer JP. The quality of care and the quality of measuring it. *N Engl J Med* 1993; 329: 1263-64.

Logan RL, Scott PJ. Uncertainty in clinical practice: implications for quality and costs of health care. *Lancet* 1996; 347: 595-98.

Mariñas JJ, Rego ME, Rosales M, Castro MI, Bouzas E. Evaluación de la efectividad de un programa de detección de infección asintomática por *Chlamydia Trachomatis* en mujeres. *Rev Esp Salud Pública* 1996; 71: 27-34.

Millili JJ, Philiponis VS, Nusbaum M. Predicting Surgical Outcome Using Bayesian Analysis. *J Surg Re* 1998; 77(1): 45-49.

Moons KG, Stijnen T, Michel BC, Büller HR, Van Es GA, Grobde DE, Habbema JD. Application to the comparison of areas under Receiver Operating Characteristic Curves. *Med Decis Making* 1997; 17: 447-54.

Orozco D, Gil VF, Pedrera V, Buigues F, Medina E, Merino J. Validez de la determinación de la glucemia basal en el control de los pacientes diabéticos no dependientes de insulina. *Med Clín (Barc)* 1997; 108: 325-29.

Pastor A, Menedez R, Cremades MJ, Pastor V, Llopis R, Aznar J. Diagnostic value of SCC, CEA and CYFRA 21.1 in lung cancer: a Bayesian analysis. *Eur Respir J* 1997; 10(3) 603-9.

Pauker SG, Kassirer JP. The Threshold Approach to Clinical Decision Making. *N Engl J Med* 1980; 302: 1109-16.

Pauker SG, Kassirer JP. Decision Analysis. *N Engl J Med* 1987; 316: 250-58.

Pauker SG, Pauker SP. Expected-utility perspectives on defensive testing. *Med Decis Making* 1998; 18: 29-31.

Phelps CE. Good technologies gone bad: how and why the cost-effectiveness of a medical intervention changes for different populations. *Med Decis Making* 1997; 17:107-17.

Reid MC, Lane DA, Feinstein AR. Academic Calculations versus Clinical Judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104: 374-80.

Renaud M. El futuro: ¿Hygeia versus Panakeia? En: Evans RG, Barer ML, Marmor TR. ¿Por qué alguna gente está sana y otra no? Madrid: Díaz de Santos, 1996: 347.

Rué M. Las técnicas de ayuda a la decisión en Salud Pública. En: Martínez Navarro F et al Salud Pública. Madrid: Interamer Mc-Graw-Hill, 1998: 288-301.

Sackett DL, Haynes R, Tugwell P. *Epidemiología Clínica*. Madrid: Díaz de Santos, 1989: 23-177.

Siles S, Garrigues V, Ponce J, Gálvez C, Berenguer J. Análisis del valor predictivo de los datos clínicos en pacientes con sospecha de enfermedad colónica. *Rev Esp Enf Digest* 1997; 89 (6): 445-50.

Sox HC, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. *Ann Intern Med* 1981; 95: 680-85.

Sox HC. Probability theory in the use of diagnostic tests. *Ann Intern Med* 1986; 104: 60-66.

Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Boston: Butterworth-Heinemann, 1988: 27-64.

Sox HC, Hickman DH, Marton KI, Moses L, Skeff KM, Sox CH, Neal A. Using the patient's History to estimate the probability of Coronary Artery Disease: A comparison of Primary Care and Referral Practices. *Am J Med* 1990; 89: 7-14.

Stoffers HE, Kester A, Kaiser V, Rinkens PE, Knottnerus JA. Diagnostic Value of signs and symptoms associated with Peripheral Arterial Occlusive Disease seen in General Practice. *Med Decis Making* 1997; 17: 61-70.

Thibault GE. The appropriate degree of diagnostic certainty. *N Engl J Med* 1994; 331: 1216-20.

Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic test. *N Engl J Med* 1990; 322: 1499-504.

Williams A. Incentivos, ética y libertad clínica. En: Reforma Sanitaria e incentivos. Barcelona: Asoc de Economía de la Salud, 1990: 55-66.

Zweig MH; Campbell G. Receiver-Operating Characteristic (ROC) Plots: A fundamental evaluation tool in Clinical Medicine. *Clin Chem* 1993; 39(4): 561-77.

Capítulo V

HERRAMIENTAS ÚTILES PARA DECIDIR SOBRE LAS MEDIDAS TERAPEUTICAS

V.1 Evidencia sobre la eficacia de los tratamientos

La eficacia mide la probabilidad de que un individuo, en una población definida, se beneficie de la aplicación de una tecnología médica en la resolución de un problema de salud determinado, bajo condiciones ideales de actuación.

Las diferentes alternativas de lo que podría resultar un tratamiento eficaz deberían someterse a rigurosas pruebas antes de su aceptación en la práctica clínica cotidiana. Los ensayos clínicos aleatorizados (ECA) son el patrón oro para las comparaciones de los efectos del tratamiento a lo largo del tiempo, y siempre que estén disponibles deberían tener la preferencia ante otro tipo de información. Aunque los ECA puedan resultar costosos y difíciles, la alternativa (tratamientos sin suficiente evidencia) puede serlo aun más (Fletcher y cols, 1998).

Para conocer en detalle que criterios utilizar en la valoración de un artículo sobre tratamiento existe literatura específica sobre el tema²³. Las recomendaciones que surgen de éstas lecturas parten de la base que la ayuda a evaluar la calidad de la evidencia redundará en una mejor asistencia de los pacientes. A continuación transcribimos algunas preguntas útiles para ayudar a evaluar la validez de las investigaciones clínicas y en particular de un ECA.

²³ Sackett y cols, 1989; JAMA 1997 (ed especial) La Medicina Basada en la Evidencia.

Directrices básicas para determinar la validez de los estudios clínicos (Fletcher y cols, 1998):

¿Qué tipo de pregunta clínica pretende contestar la investigación? El diseño de la investigación debe corresponderse con la pregunta a responder.

¿Qué tipo de pacientes, variables y resultados se estudiaron? Determina si un estudio puede generalizarse.

¿Cuál es la probabilidad de que los hallazgos sean consecuencia de un sesgo? Las diferencias sistemáticas de los grupos que se comparan (características de los pacientes, intervenciones, factores de riesgo) disminuyen la validez interna.

¿Cuál es la magnitud del efecto? Información fundamental para la toma de decisiones.

¿Cuál es la probabilidad de que los hallazgos se produjeran por azar? Obtener información sobre los límites de los valores dentro de los cuales es probable que se extienda el efecto real (intervalo de confianza).

Criterios para la valoración de un artículo sobre tratamiento (Guyatt y cols, 1997):

¿Son válidos los resultados del estudio?

Para responder a esta pregunta deberán considerarse los siguientes aspectos: se ha realizado de manera aleatoria la asignación de los tratamientos a los pacientes de manera de poder hacer comparaciones no sesgadas; se ha realizado un seguimiento completo y se han analizado los pacientes en los grupos en los que fueron asignados aleatoriamente; se ha mantenido un diseño “ciego” en relación al tratamiento aplicado, en cuanto a pacientes, clínicos y personal del estudio (contribuye a asegurar la imparcialidad de los participantes); eran similares los grupos al inicio del ensayo; aparte de la intervención experimental, se ha tratado a los grupos de la misma forma.

¿Cuales han sido los resultados del estudio?

Interesa considerar cuál ha sido la magnitud del efecto, y con que precisión se ha estimado; si fueron comunicados todos los resultados²⁴ relevantes; además de saber con qué fue comparado el resultado: placebo, otras terapias disponibles.

¿Son útiles los resultados para la asistencia a mis pacientes?

Para evaluar si pueden aplicarse los resultados obtenidos se deberá considerar si la población en estudio es similar a la de la zona donde se trabaja en cuanto a estado de salud, creencias, y otras variables de interés; si se cuenta con el potencial necesario para reproducir los servicios prestados en el estudio así como con las habilidades necesarias para brindar el servicio con la calidad adecuada. Por otra parte analizar si compensan los probables beneficios, los posibles efectos adversos y costes del mismo.

Las ventajas de un ECA son suficientes como para ser considerado el estudio de referencia. Obtiene una distinción especial en relación a otro tipo de diseño debido a que la intervención se asigna de manera aleatoria y por consiguiente sin sesgos.

²⁴ Los resultados puede considerarse cambios (tanto favorables como adversos) que se producen en la salud de las personas y que son atribuibles a la atención recibida.

Para aprovechar el potencial de diferentes ECA publicados sobre un tema, se realizan los llamados meta-análisis: su objetivo es el de resumir los resultados de toda la información procedente de estudios consistentes desde un punto de vista científico (ver en capítulo VI Medicina Basada en la Evidencia). Un requisito básico es que los estudios que incluya deberán ser homogéneos (similares en cuanto a características de los pacientes, intervenciones y medida de resultados).

Los ECA tienen también sus limitaciones. Enumeraremos algunas.

1. Puede no existir el número suficiente de pacientes con la enfermedad en cuestión para llevar a cabo el estudio.
2. El coste de realizarlos en general es muy elevado.
3. Pueden transcurrir muchos años antes de obtener resultados.
4. Nos informan sobre los beneficios de un tratamiento durante el período de tiempo de duración del ensayo, pero dichos beneficios (también las reacciones adversas) pueden variar a lo largo del tiempo (Laupacis y cols, 1988).
5. El conjunto de los resultados de los ECA puede que no siempre sean generalizables a los casos individuales, puede existir variabilidad en el efecto del tratamiento dependiendo -por ejemplo- del nivel de riesgo basal: pacientes donde puede resultar inefectivo y otros donde se obtiene un mayor beneficio (Rothwell, 1995). Los beneficios de una nueva medida terapéutica son en general demostrados en pacientes en situaciones extremas de la enfermedad, donde es también mayor el poder de la estadística para mostrarnos el efecto de las intervenciones.

Resulta entonces que para muchas decisiones sobre tratamientos no se dispone de evidencia procedente de ECA. En estos casos se utilizan otras fuentes de información por lo que se trata de un análisis secundario de datos (responder la pregunta sobre la eficacia del tratamiento no fue la causa que originó la recogida de los datos). Sus ventajas: puede contar con un número importante de casos (habitualmente se recurre a grandes bases de datos previstas para facturar los gastos de la asistencia recibida o para la administración), puede trabajarse con subgrupos de pacientes, los resultados pueden ser más generalizables que en los ECA al proceder la información de contextos naturales de la organización sanitaria, el coste es muy inferior al de un ECA, y las conclusiones pueden obtenerse en un período más breve (Fletcher y cols, 1998).

Los inconvenientes se relacionan con las ventajas de los ECA: es más común la presencia de diferentes sesgos que no permitan realizar comparaciones entre distintas alternativas; los datos no cuentan con la calidad de un diseño de ECA, pueden faltar variables que sea de interés analizar.

Las agencias de evaluación de tecnología médica realizan una valoración y clasificación de la evidencia sobre una tecnología determinada, utilizando distintas escalas basadas en

el rigor científico y la calidad de la evidencia disponible. A continuación se presenta un ejemplo de las escalas utilizadas²⁵

Tabla 5.1 Escala de Evaluación de la evidencia.²⁶

Niveles (de mayor a menor)	Calidad de la evidencia	Tipo de diseño del estudio
I	Buena	Meta-análisis de ensayos controlados y aleatorizados
II	Buena	Ensayos controlados y aleatorizados de muestra grande
III	Buena a regular	Ensayos controlados y aleatorizados de muestra pequeña
IV	Buena a regular	Ensayos prospectivos controlados no aleatorizados
V	Regular	Ensayos prospectivos controlados no aleatorizados
VI	Regular	Estudios de cohorte
VI		Estudios caso-control
VIII	Baja	Series clínicas no controladas Estudios descriptivos: seguimiento de la enfermedad, vigilancia epidemiológica, registros, bases de datos, Comités de expertos, conferencias de consenso.
IX		Anécdotas o casos.

En cuanto a la calidad de la evidencia su interpretación es la siguiente: *buena*, cuando la evidencia científica es adecuada, *regular* si existe cierta evidencia, y *baja*, la evidencia es insuficiente para recomendar o desaconsejar la adopción de la tecnología (Jovell y Navarro Rubio, 1995).

Otras consideraciones sobre la eficacia de una terapéutica. Una parte del efecto que producen los fármacos puede ser en general atribuible al llamado efecto placebo. Puede distinguirse una acción mayormente específica (en antibióticos); una mixta (antidepresivos, narcóticos) y una mayormente inespecífica (antipruriginosos) (Fletcher y cols, 1998). Si bien lo que interesa desde el punto de vista del paciente es la mejoría que produce el tratamiento por encima de lo que sucedería en el curso natural de la enfermedad sin el tratamiento, debería saberse en cada caso cual parte del efecto puede atribuirse al tratamiento específico.

Si bien los ECA nos aportan la mejor evidencia sobre los beneficios y daños de un tratamiento, la aplicación de dichos resultados a los pacientes individuales puede tener algunas dificultades. Partiendo del supuesto que los ensayos clínicos fueron realizados

²⁵ Esta escala es utilizada por la Agencia de Evaluación de Tecnología Médica del Servei Català de la Salut. Departament de Sanitat y Seguretat Social.

²⁶ Esta escala incluye una especificación de las condiciones de rigor científico, pero su descripción excede el cometido de esta publicación.

cumpliendo con todos los requisitos necesarios, se hará referencia, a continuación, con mayor detalle a la presentación e interpretación de los resultados.

V.2 Forma de presentación de los resultados: RR, RA, RRR, NNT

La forma de presentación de los resultados de un tratamiento puede ser muy diferente, y existen diversas investigaciones que describen la existencia de una relación muy estrecha entre cual es la forma de presentación utilizada y cual la propensión de los clínicos a prescribir fármacos, o a utilizar programas preventivos o rehabilitadores. También influyen en las decisiones de los pacientes (Meneu y cols, 1998. Bobbio y cols, 1994).

La información usual sobre la eficacia de los diferentes tratamientos (fármacos, etc) en términos de *reducción del riesgo relativo* (RRR) aunque técnicamente correcta, tiende a operar en forma “subliminal” sobre el lector, magnificando el efecto de la intervención: no aporta información sobre la reducción absoluta del mismo (Bucher y cols, 1994).

Tabla 5.2 En el siguiente ejemplo se pretende evaluar la eficacia del uso de un antibiótico para evitar la aparición de complicaciones en la evolución de un cuadro infeccioso.

<i>Riesgo basal (sin tratamiento) (x)</i>	<i>Riesgo con el tratamiento (y)</i>	<i>Reducción absoluta del riesgo²⁷ (x - y)</i>	<i>Riesgo relativo: RR²⁸ (y / x)</i>	<i>Reducción del riesgo relativo: RRR (1 - RR) * 100</i>
0.02 o 2%	0.01 o 1%	0.02-0.01=0.01	0.01/0.02=0.5	(1-0.5)*100=50%

Si la RRR es de un 50% significa que el tratamiento aplicado reduce en esta cuantía el riesgo de complicaciones.

Cuanto mayor es el valor del RRR mayor eficacia tiene el tratamiento evaluado.

Pero este mismo RRR podría obtenerse partiendo de un riesgo basal del 20 %, o del 2 por mil, y no se vería reflejado en el indicador. La diferencia del impacto puede verse en la *reducción absoluta del riesgo*, que es de 0.01 en el primer caso, y de 0.001 para el segundo. Debido a que se trata de una fracción decimal, esta medida no es demasiado accesible para que el médico la recuerde e incorpore en su práctica clínica habitual.

²⁷ La reducción absoluta de riesgo es la diferencia de riesgo de padecer el problema que se quiere evitar entre el grupo sin tratamiento y el grupo al que se le aplica el tratamiento.

²⁸ El riesgo relativo se calcula a través del cociente entre el riesgo de padecer el problema que se quiere evitar entre el grupo sometido a tratamiento y el grupo control

Deberíamos analizar además si los beneficios del tratamiento compensan la utilización del mismo. Vamos a apoyarnos en el concepto de NNT²⁹ (*número que es necesario tratar*) para comprender mejor a que nos referimos. La utilidad de un tratamiento no sólo depende de la RRR, sino de cual es el riesgo basal de tener una evolución adversa sin la aplicación del tratamiento, que a su vez nos va a determinar cual es la reducción absoluta de riesgo que evitamos con el tratamiento (enfermedad, complicaciones, muerte). Este último nos permite deducir cual es el número de personas que deberíamos tratar para evitar un evento adverso (Laupacis y cols, 1988).

La ventaja de esta medida es que incorpora la información sobre el riesgo basal sin tratamiento y la reducción de riesgo con tratamiento.

En nuestro ejemplo el NNT con antibióticos sería $1 / 0.01 = 100$ pacientes para evitar la aparición de complicaciones en 1 caso. En los ejemplos pueden incluirse otras áreas como cirugía, programas preventivos y exposición a factores de riesgo: 6 pacientes con angina estable y estenosis de la rama principal de la arteria coronaria izquierda deben ser sometidos a bypass para lograr la supervivencia a 5 años de una persona; alrededor de 200 personas deben ser inmunizadas contra la hepatitis B para prevenir un caso de hepatitis en población de bajo riesgo de enfermedad, frente a solo 8 si se trata de población de alto riesgo; 1306 personas deben estar expuestas en forma pasiva a humo de 20 cigarrillos o más por día durante 14 años antes de que una persona muera por cáncer de pulmón (Laupacis y cols, 1988).

Para un mismo RRR, cuanto mayor es la probabilidad de un paciente de presentar una evolución adversa si no recibe tratamiento, menos casos serán necesarios tratar para prevenir un episodio no deseado.

Cuanto más bajo es el NNT, mayor es la eficiencia clínica.

En la siguiente tabla puede verse la influencia que tiene en el NNT la probabilidad basal o previa de padecer el efecto no deseado. Nuevamente presente la prevalencia.

Tabla 5.3 *Número necesario de pacientes a tratar en función de la probabilidad de padecer una condición y de la reducción porcentual del riesgo relativo que se consigue con el tratamiento (Sackett y cols, 1989).*

Probabilidad de Padecer la condición	Reducción del Riesgo Relativo		
	50 %	25 %	10 %
0.5	4	8	20
0.1	20	40	100
0.001	2000	4000	10.000

²⁹ Se trata de la inversa de la reducción del riesgo absoluto.

Otro ejemplo (Guyatt y cols, 1997): los bloqueadores beta tienen producen una RRR del 25% sobre la mortalidad después de un infarto agudo de miocardio (IAM). Si tenemos dos casos de IAM con diferente riesgo basal de muerte: caso 1, hombre de 40 años, con un IAM pequeño, con una capacidad de ejercicio normal y sin arritmia ventricular, quien está dispuesto a dejar de fumar, iniciar un programa de ejercicio físico, a perder peso y tomar ácido acetil salicílico diariamente. Por todo esto el riesgo de muerte al año se estima en 1%, con una disminución relativa del riesgo con beta bloqueantes del 25% y una reducción absoluta de 0.0025. La inversa de esta reducción absoluta del riesgo es el NNT, o la cantidad de pacientes con estas características que tendríamos que tratar con beta bloqueantes para evitar un caso de muerte tras un IAM: $1/0.0025=400$.

Si el caso 2 fuera un hombre, anciano, con capacidad limitada de ejercicio, con frecuentes extrasístoles ventriculares, que continua fumando, su riesgo basal sería mayor, por ejemplo un 10%, con una disminución absoluta de riesgo de 0.025, y una inversa de 40 (NNT).

Tabla 5.4 Eficacia de los bloqueadores beta sobre la mortalidad post IAM.

<i>Riesgo de muerte sin tratamiento</i>	<i>RRR</i>	<i>Riesgo de muerte con tratamiento</i>	<i>Reducción de riesgo absoluto</i>	<i>NNT</i>
0.01	25%	$0.01 \cdot 0.75 = 0.0075$	$0.01 - 0.0075 = 0.0025$	$1/0.0025 = 400$
0.1	25%	$0.1 \cdot 0.75 = 0.075$	$0.1 - 0.075 = 0.025$	$1/0.025 = 40$

Para la misma RRR, el paciente obtendrá mayor beneficio cuanto más alta sea su probabilidad de una evolución adversa si no recibe tratamiento.

Tabla 5.5 Veremos un último ejemplo sobre la indicación de antagonistas de los receptores H2 (Guyatt y cols, 1997):

	<i>Riesgo basal de hemorragia</i>	<i>Disminución del riesgo relativo</i>	<i>Riesgo de hemorragia con tratamiento</i>	<i>Disminución del riesgo absoluto</i>	<i>NNT</i>
<i>Paciente 1</i>	0.037	58%	0.0155	0.0215	45
<i>Paciente 2</i>	0.0014	58%	0.0006	0.0008	1.250

Paciente 1: en estado grave, con ventilación mecánica y/o presencia de coagulopatía

Paciente 2: en estado grave, que respira espontáneamente y sin coagulopatía

¿Se deben tratar ambos tipos de pacientes? Se requeriría la definición de un umbral de manera que si el riesgo de los pacientes sin tratamiento es elevado y el NNT fuera

inferior al umbral, se administraría el tratamiento; en caso contrario -riesgo bajo y un NNT por encima del umbral- no lo recibirían. En la definición del umbral deberían incluirse los costes: de tratar a los pacientes, que se ahorrarían al prevenir un acontecimiento objetivo, los que podrían surgir de prevenir dicho acontecimiento, y los de los efectos adversos (Guyatt y cols, 1997). La decisión sobre el NNT umbral puede variar en distintos contextos, ya que su valor depende de como se cuantifique cada componente.

Algunos problemas de medir a través de NNT:

1. Si tienen que tratarse, por ejemplo, 11 personas para evitar un evento adverso, nada se dice sobre el destino de los otros 10 en cuanto a si padecen o no el proceso adverso. Si el riesgo basal es de 0.9 y el RRR del 10%: el NNT es de 11, y 8 de los restantes pacientes sometidos al tratamiento tendrán el evento adverso. Si el riesgo basal es de 0.3 y el RRR un 30%, el NNT también será de 11, pero en este caso solo 2 de los otras 10 personas tratadas presentarán la enfermedad (Laupacis y cols, 1988).

2. Si quiere extrapolarse la medida a un caso clínico particular, una situación común es que la estimación que se hace del riesgo basal del paciente sea diferente a la del ECA. En estos casos puede realizarse un ajuste del cálculo: se considera la RRR constante (y publicada en el ECA), y el riesgo basal sin el tratamiento estimado en el paciente particular, se calcula el NNT (Cook y Sackett, 1995). También existen nomogramas, donde a partir de la información del riesgo basal y la RRR, se obtiene el NNT (Chatellier y cols, 1996).

3. Pero la premisa de RRR constante para todos los niveles de riesgo no es siempre cierta. La RRR atribuible a la cirugía con bypass coronario varia entre un 17 y 45% dependiendo del nivel de riesgo basal del paciente (Chatellier y cols, 1996). La terapia antihipertensiva puede contribuir también a ejemplificar este aspecto: 128 pacientes con hipertensión leve es necesario tratar para prevenir un resultado adverso (muerte, ataque cardíaco), siendo en el caso de hipertensión severa 9 el número de pacientes para evitar los mismos eventos (Deyo y Patrick, 1995). En el caso de los fármacos hipolipemiantes tendrían claros beneficios en pacientes de alto riesgo de enfermedad coronaria, y pocos beneficios (incluso posibles efectos adversos) en los de bajo riesgo cuando son utilizados a nivel de atención primaria (Smith y cols, 1993).

4. Otro factor que puede modificar el NNT, es el período de tiempo considerado en el ECA. Cuanto mayor es el tiempo de seguimiento del grupo control, se producirán un mayor número de eventos en dicho grupo, y con un similar RRR, será mayor la reducción absoluta de riesgo y por tanto el NNT disminuye. También pueden producirse modificaciones del RRR con mayores períodos de seguimiento (Chatellier y cols, 1996).

Otra forma alternativa de presentación de resultados es a través del incremento en el porcentaje de pacientes libres del episodio adverso, para el ejemplo de la tabla No. 5.1 pasa de un 98 a un 99%.

¿Cómo influyen las distintas medidas sobre resultados en la prescripción médica?

Los médicos valoran en forma más favorable un tratamiento cuyos resultados se presentan en términos relativos que en términos absolutos (Naylor y cols, 1992. Bobbio y cols, 1994).

La intención de prescribir es mayor cuando la medida utilizada es el RRR, en relación a la reducción de riesgo en términos absolutos, observándose una actitud más conservadora con información sobre el NNT (Bucher y cols, 1994).

En España también se describe que la intención de prescribir por parte de los médicos se magnifica cuando los resultados se presentan como RRR. Además la inclusión de la información sobre efectos negativos reduce en forma importante la intención de emplear el tratamiento (Meneu y cols, 1998).

Otro problema frecuente es la citación selectiva de los ECA, condenando en general al olvido los resultados opuestos a la hipótesis que se pretende defender. “Ravnskov concluye que la citación sesgada es el soporte de la evidencia sobre la eficacia de la prevención primaria de la cardiopatía isquémica a través del tratamiento de la hipercolesterolemia”; datos de 2 ECA muestran que el ensayo clínico con resultados favorables fue citado 109, 121, 202 y 180 veces en los años siguientes a su publicación, mientras el ensayo con resultados contrarios 6, 5, 3 y 1 vez solamente (Gérvas y Pérez F, 1997).

En resumen, si bien existen varias formas de presentar resultados sobre la eficacia de un tratamiento, puede decirse que la reducción absoluta de riesgo es superior a la RRR, porque incorpora la información tanto del riesgo basal como la magnitud de la reducción del riesgo. Por otro lado, su recíproca -el NNT- es una manera que puede ser comprendida fácilmente por los clínicos y además ser usada tanto para describir los beneficios como el daño que puede resultar de aplicar una terapia (Laupacis y cols, 1988).

En la medida que la forma de presentación de resultados sobre eficacia influye sobre la intención de prescribir, parece conveniente que en las publicaciones (reseñas, revisiones, libros de texto, CD-ROM con bases de datos) que difunden el conocimiento médico sobre efectividad se incluyan además de la RRR alguna de las medidas descriptas, especialmente el NNT, así como los resultados adversos. Smith y Egger (1994) recomiendan, tanto a los investigadores como a las empresas farmacéuticas el uso de medidas absolutas sobre resultados. La importancia de esto es aun mayor cuando la enfermedad tiene una baja incidencia (por ejemplo en prevención primaria) (Bucher y cols, 1994). Con estos datos se presentan en las revisiones sistemáticas sobre evidencia científica que realiza la biblioteca Cochrane. También importa que en las guías clínicas se relacionen los potenciales beneficios según el nivel de riesgo del paciente.

V.3 Utilización de la evidencia: variabilidad de la práctica médica

Las variaciones en la práctica médica (VPM) pueden definirse como las variaciones sistemáticas (es decir no-aleatorias) en las tasas estandarizadas para determinados tratamientos (o procedimientos diagnósticos) a un determinado nivel de agregación de la población (McPherson, 1995).

Se cita al trabajo de Glover en el año 1938 sobre la incidencia de tonsilectomía y la variabilidad debida a diferencias de opinión médica³⁰(Glover, 1992) como la primera investigación en el área. Pero sólo a partir de los años setenta se aplica el “bisturí a fondo” en el tema a partir de Wennberg y sus estudios en los Estados Unidos, y posteriormente otros investigadores en Canadá, Reino Unido, Noruega y Bélgica. La variabilidad se demuestra tanto en procedimientos médicos como quirúrgicos, en las distintas especialidades y subespecialidades, así como en pacientes internados o ambulatorios (Chassin y cols, 1986). Las diferencias de utilización, pueden reflejar sobreutilización en unas zonas (las de altas tasas), así como subutilización en otras (las de menores tasas), ya que la diferencia de morbilidad en poblaciones grandes tiende a desaparecer³¹. Se postula como variable explicativa de las importantes diferencias encontradas, el estilo de práctica médico.

Si bien parte de la variabilidad ha podido ser asociada a diferencias en la oferta de recursos, y a presencia de cuidados innecesarios, la hipótesis de mayor aceptación se relaciona con la discrecionalidad de las decisiones clínicas individuales o estilos de práctica médica (Peiró y Meneu, 1995).

McPherson (1995) plantea que desde el punto de vista epidemiológico pueden identificarse tres causas de variaciones sistemáticas: incertidumbre en casos de investigación insuficiente; ignorancia clínica, por insuficiente divulgación de la investigación o porque el resultado de la misma no se entiende o no se cree; o que se trate de la manifestación de preferencias informadas.

Se observa menor variabilidad cuando existe consenso entre los clínicos, y mayor en casos de alta incertidumbre³². Se describe alta variabilidad de la cirugía electiva, en particular histerectomías, amigdalectomías y prostatectomías en casos de hiperplasia benigna; siendo menor en la cirugía de urgencia, por ejemplo apendicectomías.

³⁰ “... existe una tendencia a practicarla como ritual profiláctico de rutina sin ninguna razón ni ningún resultado en particular ...”

³¹ Esto no significa que la diferencia en morbilidad no sea uno de los factores explicativos más importante cuando se analiza la utilización de servicios en casos individuales.

³² Esta constatación se relaciona con el análisis de Eddy D. sobre la dificultad de los médicos para combinar información de diferentes fuentes cuando tienen que resolver sobre un procedimiento particular, y por tanto las múltiples fuentes de incertidumbre (ver cap. II).

En los tratamientos médicos la variabilidad también es alta, las otitis medias agudas son tratadas con antibióticos en el 98% de los casos en Nueva Zelanda mientras solo el 31% de los niños ingleses los reciben para dicha afección. Varían los criterios diagnósticos (18 diferentes criterios son descriptos), el nivel de certidumbre con el que se inicia el tratamiento, así como el antibiótico seleccionado. Ante un diagnóstico más dudoso, mayor es la tasa de utilización de antibióticos (González-Vallejo y cols, 1998).

Algunos ejemplos en España: las intervenciones de cataratas en personas mayores de 65 año varían entre 40 intervenciones por cada 10.000 en Galicia, hasta 80 y 100 en Madrid, Cataluña, Extremadura y País Vasco (Peiró y cols, 1998). Otros ejemplos pueden verse en la siguiente tabla.

Tabla 5.6 Variabilidad en algunos tratamientos a nivel hospitalario en España.

<i>Tipo de patología</i>	<i>Coefficiente de variación</i> ³³
Apendicitis	1.5
Infarto agudo de miocardio	2.4
EPOC	2.5
Neumonía	2.9
Diabetes	3.3
Hipertrofia de próstata	3.6
Várices	4.8
Hipertensión arterial	4.9
Amigdalectomía	5.0

Fuente: Sarría y Rodríguez, 1998.

¿Cuál es el problema? La variabilidad estaría indicando que ante situaciones similares se adoptan estilos de práctica diferentes; esto genera dudas ante la indicación de ciertas intervenciones, los resultados obtenidos en los pacientes en función de la utilización de una u otra práctica, así como sobre la eficiencia del empleo de los recursos (Peiró y cols, 1998).

¿Cuál es la tasa de utilización correcta? Wennberg (1986) plantea que si los médicos no asumen la responsabilidad de responder a esta pregunta, la teoría dominante por defecto –en un contexto de contención de costes- será que el menor índice de utilización siempre será el mejor.

De la investigación se desprende que las VPM estarían en parte explicadas por las distintas opiniones de los médicos acerca de las diferentes alternativas diagnósticas y terapéuticas para un mismo problema de salud. Cuando no existe evidencia científica suficiente sobre los resultados de dichas alternativas (incertidumbre), o si existiendo

³³ El coeficiente de variación es el cociente entre la desviación estándar y la media, y se interpreta en términos de variación relativa.

evidencia científica el médico no las utiliza (ignorancia³⁴) se originan estas diferencias de opinión.

Peiró y Meneu (1997) describen dos estrategias de abordaje diferentes para cada situación. Cuando existe evidencia clara sobre la efectividad de un procedimiento pero la misma se ignora, es necesaria una mayor difusión de la información, programas de formación continuada, mecanismos de feed-back sobre los resultados obtenidos, instrumentar la revisión de utilización (ver siguiente punto) junto al desarrollo de incentivos para promover la utilización apropiada.

Ante situaciones de alta incertidumbre, promover la investigación sobre eficacia, efectividad y coste-efectividad de intervenciones alternativas, y mientras tanto aumentar la capacidad de elección por parte de los pacientes³⁵, así como la promoción de consensos de expertos hasta que la ciencia aporte evidencia sobre el procedimiento.

La *efectividad* al igual que la eficacia, también pretende medir cual es la probabilidad que tiene un individuo de beneficiarse con la aplicación de una tecnología médica, pero en las condiciones reales (no ideales) de aplicación, y por tanto podrá variar en distintos contextos organizativos y de actuación de la sociedad en cuestión. Una vez obtenida la información sobre la eficacia de una terapéutica, que el beneficio que obtenga un paciente particular de su aplicación sea el esperado dependerá de: un correcto diagnóstico por parte del médico; una indicación adecuada del tratamiento; que medie adhesión al mismo por parte del paciente, y que el nivel de cobertura de asistencia sea elevado³⁶.

Se concibe la *calidad* de la atención médica como la diferencia entre la eficacia y la efectividad que es atribuible al desempeño médico, tomando en cuenta el entorno donde se trabaja³⁷. Por tanto cuando se habla de mejora de la calidad implica que el proveedor del servicio adopte prácticas que lleven a reducir la distancia que separa lo que sucede en la realidad con un nivel de eficacia ideal.

El mejorar el conocimiento y la destreza de las técnicas cuantitativas, así como el promover su utilización en el terreno de las decisiones clínicas, puede por un lado delimitar el fuerte componente discrecional que distingue el ejercicio profesional en la práctica de la medicina, y por otro, contribuir con el control de algunos de los sesgos y de las fuentes de incertidumbre descriptas en los capítulos II y III.

³⁴ Algunos ejemplos de evidencia científica disponible pero no conocida o no utilizada por tiempo: efecto de los corticoesteroides en la prevención del parto prematuro; eficacia del empleo de trombolíticos en los IAM.

³⁵ Si no se sabe bien que hacer, que conducta seguir, mejor que lo decida el paciente.

³⁶ Aunque el tratamiento de la tuberculosis tenga una eficacia del 95%, si sólo se diagnostican el 10% de los casos nuevos, se tratan correctamente el 80% de los diagnosticados, con un cumplimiento del 50% de los pacientes, la efectividad será entonces de sólo 3.8% ($0.95 \cdot 0.10 \cdot 0.80 \cdot 0.50 = 0.038$).

³⁷ Es evidente que no toda la diferencia entre eficacia y efectividad es responsabilidad médica, pero si la parte que se relaciona con los temas de esta publicación.

Conocer qué se hace y cómo se hace, es el primer paso para poder mejorar el producto: atención al paciente. A continuación se describe una herramienta que puede resultar útil en esa dirección.

V.4 Revisión de la utilización

El marco conceptual de la revisión de la utilización está basado en establecer la relación entre la necesidad de la asistencia y los cuidados prestados. Su objetivo es identificar y reducir el uso innecesario o inapropiado, manteniendo la accesibilidad a los servicios necesarios (Meneu, 1996).

Puede definirse a la utilización inapropiada como aquellos casos de aplicación de tecnologías médicas ineficaces, o que teniendo una eficacia demostrada se utilizan en situaciones en que no son efectivas o existen alternativas más eficientes.

Se incluyen diversas variantes de servicios que si bien no necesariamente resultan perjudiciales a los pacientes, tienen en común que no le aportan ningún beneficio³⁸ (Peiró y Portella, 1993).

Pueden resumirse estas situaciones en la siguiente tabla.

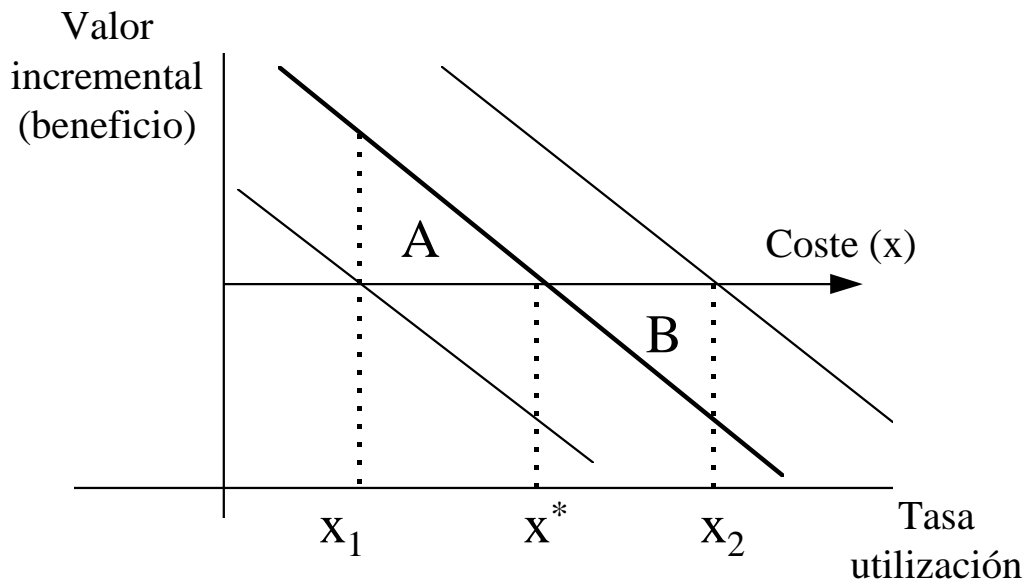
		<i>se suministra el tratamiento</i>	
		<i>Si</i>	<i>No</i>
<i>es necesario</i>	<i>si</i> <i>(dudoso)</i>	Utilización adecuada -----	Utilización insuficiente -----
	<i>no</i>	Utilización excesiva	Utilización adecuada

Si bien la subutilización es también un ejemplo de asistencia inapropiada, en general los programas de revisión se han dirigido a los problemas de sobreutilización (Meneu, 1996). En el mundo real no son tan nítidos los límites entre cada una de las posibilidades, por eso la zona de dudas. De éstos temas trata el campo de la “revisión de la utilización” ya sea de fármacos (Buetow y cols, 1996), pruebas diagnósticas (Walker y cols, 1994), camas hospitalarias (Oterino y cols, 1994), derivaciones a especialistas (Fertig y cols, 1993) entre otras áreas³⁹. El siguiente gráfico aborda el mismo tema pero a partir del concepto de pérdida de bienestar (Phelps y Parente, 1990):

³⁸ Indicación de antibióticos en casos de virosis, servicios de urgencia hospitalarios en enfermedades que no lo justifican, hospitalización de pacientes en situaciones que pueden resolverse en niveles de menor complejidad.

³⁹ Recomendamos la lectura del artículo “Revisión de la utilización. Definición, conceptos y métodos” Peiró S; Meneu R. Rev Calidad Asistencial 1997; 12: 122-136.

Gráfico 5.1 Relación entre la efectividad y la tasa de utilización de un procedimiento.



En general la mayor utilización de un procedimiento determina que el valor que aporta al bienestar de un paciente sea cada vez menor (cirugía en pacientes con cálculos biliares silenciosos, extensión de los screening a población de bajo riesgo, etc); en otras palabras la efectividad va disminuyendo a medida que aumenta la extensión de uso de un procedimiento diagnóstico o terapéutico. Por otro lado puede diferenciarse una efectividad objetiva o de referencia (x^*), y otra subjetiva que puede determinar una utilización superior (x_2) o inferior (x_1) a la objetiva. En ambos casos se producirá pérdida de bienestar: en la primera situación, la sobreutilización determina niveles de beneficios inferiores a los costes (la pérdida de bienestar queda representada por el área del triángulo B), y en la segunda, por subutilización se realizan menos procedimientos que los indicados de acuerdo a la efectividad de referencia (la pérdida de bienestar queda expresada por el área del triángulo A).

Algunas de las preguntas que pretende responder a través de las revisiones (Peiró y Meneu, 1997)

¿Qué asistencia se ha prestado? ¿Se hizo lo que el paciente necesitaba?

¿Cuánto se prestó de asistencia? ¿Fue en el momento adecuado?

¿Cuánta atención en términos de duración y frecuencia, fue prestada? Incluye duración de la estancia, medicación administrada, pruebas solicitadas, y sus pautas de aplicación.

¿Dónde se prestó la asistencia? ¿Fue en el nivel adecuado?

En el caso de los medicamentos (el recurso terapéutico más utilizado) el análisis de la adecuación incluye varios aspectos. Por una lado, la racionalidad farmacológica, que se relaciona con la eficacia y la efectividad del fármaco. Los problemas potenciales son: reacciones adversas, inapropiada indicación, interacciones peligrosas, fallos en la administración, selección inadecuada del fármaco, sobredosis o dosis subterapéutica. Por otro, la racionalidad económica o seleccionar el de menor coste para un mismo efecto, tomando en cuenta las preferencias del paciente en cuanto a vías de administración y otras condiciones del consumo (Saturno, 1996).

En ciertos contextos con abundante oferta de profesionales, presencia de distintos incentivos financieros, la mayor parte de reportes registran sobreutilización. Diversos estudios estiman las tasas de uso inapropiado de diferentes procedimientos – angiocardiografías, endarterectomias carótideas, bypass coronario- con un rango entre un 15 y 30%; otros datos sobre uso inapropiado: el 16% de las histerectomías en 7 health maintenance organizations, 24% de los días de hospitalización en hospitales pediátricos en Canadá (Phelps CE, 1993), el 27.6% de las admisiones y el 43.9% de las estancias se valoraron como innecesarias en España (Oterino y cols, 1994).

La utilización insuficiente se asociará a problemas de cobertura y retraso diagnóstico. Algunos ejemplos de reportes sobre subutilización se encuentran en la aplicación de los tratamientos para el dolor y la depresión (McDonald, 1996) y en el tratamiento del asma (Buetow y cols, 1996).

La principal ventaja de esta metodología radica en centrarse en la utilización innecesaria, a diferencia de otras estrategias⁴⁰ que podrían tener como consecuencia una reducción de las prestaciones tanto inadecuadas como de las adecuadas. La limitación mayor está en la escasez de instrumentos válidos y fiables que puedan identificar los servicios que son innecesarios (Meneu, 1996).

⁴⁰ Diferentes tipos de co-pagos por parte de los pacientes, pago por proceso, incentivos económicos ligados a disminución de la utilización.

V.5 A modo de síntesis

Resumiendo los conceptos vistos hasta ahora, puede decirse que una decisión terapéutica requiere de la integración de varios elementos: solidez de las evidencias (ensayos clínicos controlados con asignación aleatoria y revisiones de conjunto), la magnitud del efecto de la intervención, y una utilización adecuada del procedimiento según la necesidad de salud.

Pero a este enfoque desde el punto de vista de un paciente individual, se debería incorporar una visión más colectiva: Maynard (1997) habla de maximizar las ganancias en términos de salud poblacional, a partir de un presupuesto dado. Para ello debe retomarse el concepto de ética social que mencionábamos en el capítulo anterior.

Si los costes no fueran un problema, debería buscarse la mayor efectividad posible en cada intervención individual, porque el óptimo coincidiría con el máximo; pero si se carecen de los recursos suficientes para otorgar la máxima calidad a todos los pacientes, en un contexto social, la eficiencia significará conseguir lo mejor para la mayoría de los enfermos, y no lo mejor para uno (Ortún, 1992).

Una regla racional es propuesta por Ortún y Rodríguez (1990) sobre la selección de alternativas tanto para diagnosticar como para tratar una enfermedad: de entre las más eficaces, inclinarse por aquellas que más beneficios aporten en la práctica real –las más efectivas–, y dentro de éstas las que más rindan en función de los costes (la más eficiente⁴¹).

Además, aunque para ciertas condiciones médicas concretas se investiguen beneficios y riesgos, y se mejore la adecuación de la práctica, otras muchas quedarán sin investigar y serán potencialmente influenciadas por el volumen de la oferta. Estos aspectos fundamentan el necesario establecimiento de prioridades sociales así como la regulación de la oferta de servicios a la hora de contribuir en forma eficiente al bienestar de la sociedad en su conjunto (Peiró y cols, 1998).

Por supuesto que estos aspectos trascienden la gestión clínica, pero tener presentes los temas de política sanitaria ayudará a promover su complementariedad con la práctica clínica. Se han propuesto una serie de vías que funcionarían de tránsito entre la efectividad clínica y la eficiencia social⁴², algunas de las cuales se explicitarán en el Cap.VI al referirnos a la Medicina Basada en la Evidencia, Análisis de Coste Efectividad y elaboración de las Guías de la Práctica Clínica.

⁴¹ La eficiencia relaciona los beneficios medidos a través de la efectividad, con los costes que supone obtenerlos.

⁴² Un análisis de las mismas puede obtenerse en Ortún V. La Economía en sanidad y medicina: instrumentos y limitaciones. Ed La Llar del Llibre. Barcelona, 1992.

Bibliografía referenciada

Bobio M; Demichelis B; Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; 343: 1209-11.

Bucher HC; Weinbacher M; Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *Br Med J* 1994; 309: 761-64.

Buetow SA; Sibbald B; Cantrill JA; Halliwell S. Prevalence of potentially inappropriate long term prescribing in general practice in the United Kingdom, 1980-95: systematic literature review. *Br Med J* 1996; 313: 1371-4.

Cook RJ; Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BrMed J* 1995; 310: 452-54.

Chassin MR; Brook RH; Park RE; Fink JK; Kosecoff J; Solomon DH y cols. Variations in the use of medical and surgical services by the medicare population. *N Engl J Med* 1986; 314: 285-90.

Chatellier G; Zapletal E; Lemaitre D; Menard J; Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *Br Med J* 1996; 312: 426-29.

Deyo RA; Patrick DL. The significance of treatment effects: the clinical perspective. *Medical Care* 1995; 33 (4): AS 286-AS291, supplement.

Fertig A; Roland M; King H; Moore T. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *Br Med J* 1993; 307: 1467-70.

Fletcher RH; Fletcher SW; Wagner EH. *Epidemiología Clínica*. Madrid: Masson-Williams & Wilkins, 2da edición, 1998: 141-70.

Gérvas J; Pérez Fernández M. Las hiperlipemias y la prevención primaria de la cardiopatía isquémica. *Med Clín (Barc)* 1997; 109: 549-52.

Glover JA. La incidencia de tonsilectomía en niños de edad escolar. En: Kerr L.White *Investigaciones sobre Servicios de Salud: una Antología*. Washington DC: OPS Publ Científica 534, 1992: 18-31.

González-Vallejo C; Sorum PC; Stewart TR; Chessare JB; Mumpower JL. Physicians' diagnostic judgments and treatment decisions for acute otitis media in children. *Med Decis Making* 1998; 18: 149-62.

Guyatt GH; Sackett DL; Cook DJ. Cómo utilizar un artículo sobre tratamiento o prevención. En: *La Medicina Basada en la Evidencia*. JAMA (ed esp) 1997; 30-40.

Jovell AJ; Navarro Rubio MD. Evaluación de la evidencia científica. *Med Clín (Barc)* 1995; 109: 740-43.

Laupacis A; Sackett D; Roberts R. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318: 1728-33.

Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. *Lancet* 1997; 349: 126-28.

McDonald CJ. Medical heuristics: the silent adjudicators of clinical practice. *Ann Intern Med* 1996; 124: 56-62.

McPherson K. Cómo debería modificarse la política sanitaria ante la evidencia de variaciones en la práctica médica. *Variaciones en la Práctica Médica* 1995; 7:9-17.

Meneu R. Gestión de la utilización. En: Meneu R; Ortún V. *Política y Gestión Sanitaria: la agenda explícita*. Barcelona: Asoc de Econ de la Salud, 1996: 219-38.

Meneu R; Peiró S; Márquez S. Influencia de la presentación de los resultados de los ensayos clínicos en la intención de prescribir: relativizando el riesgo relativo. *Aten Primaria* 1998; 21: 446-50.

Naylor CD; Chen E; Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992; 117: 916-21.

Ortún V; Rodríguez F. De la efectividad clínica a la eficiencia social. *Med Clin (Barc)* 1990; 95: 385-88.

Ortún V. *La Economía en sanidad y medicina: instrumentos y limitaciones*. Barcelona: Hogar del Libro, 1992.

Oterino D; Peiró S; Portella E; Marchan C; Aymerich S. Utilización innecesaria de la hospitalización: importancia de la gestión a nivel de servicio. *Rev Calidad Asistencial* 1994; 1: 8-16.

Peiró S; Portella E. La revisión de la utilización inapropiada de la hospitalización. *Variaciones en la Práctica Médica* 1993; 5: 1-2.

Peiró S; Meneu R. Revisión de la utilización. Definición, concepto y métodos. *Rev Calidad Asistencial* 1997; 12: 122-36.

Peiró S; Meneu R; Marqués Espí JA; Librero J; Ordiñana R. La variabilidad en la práctica médica: relevancia, estrategias de abordaje y política sanitaria. En: *Papeles de Economía Española*. Madrid, 1998 No 76: 165-75.

Phelps CE; Parente ST. Priority setting in medical technology and medical practice assessment. *Medical Care* 1990; 28(8): 703-23.

Phelps CE. The methodologic foundations of studies of the appropriateness of medical care. *N Engl J Med* 1993; 329: 1241-5.

Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995; 345: 1616-19.

Sackett DL; Haynes R; Tugwell P. *Epidemiología Clínica*. Madrid: Díaz de Santos, 1989: 221-57.

Sarría A; Rodríguez R. Variabilidad en la práctica clínica. En: Del Llano et al. *Gestión Sanitaria: Innovaciones y desafíos*. Barcelona: Masson, 1998: 513-28.

Saturno PJ. Monitorización del uso de medicamentos para mejorar la calidad de la prescripción. *Problemas, métodos e indicadores. Aten Primaria* 1996; 18(6): 331-38.

Smith GD; Song F; Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *Br Med J* 1993; 306: 1367-73.

Smith GD; Egger M. Who benefits from medical interventions? *Br Med J* 1994; 308: 72-74.

Walker D; Williams P; Tawn J. Audit of requests for preoperative chest radiography. *Br Med J* 1994; 309: 772-3.

Wennberg J. Wich rate is right? *N Engl J Med* 1986; 314: 310-11.

Capítulo VI

TENDENCIAS EN MEDICINA

El comportamiento profesional de los médicos, que determina fundamentalmente la asignación de recursos sanitarios, está sujeto a una ética -entendida como la serie de comportamientos que una profesión estima aceptables- con su correspondiente sistema de controles por colegas, y sus mecanismos sancionadores -recompensas y ostracismos. A todo médico le cabe la satisfacción de resolver problemas, curar a veces, paliar en ocasiones y consolar siempre. El contenido de su trabajo se enriquece con el progreso científico. Quienes alcanzan un cierto nivel de virtuosismo y los que publican, ganan reconocimiento profesional; quienes descubren o innovan tienen el reconocimiento intelectual de ser los primeros, primicia laureada a veces con eponimia, y más frecuentemente con distinciones y premios. Este comportamiento profesional vendrá afectado por los cambios tecnológicos, demográficos y epidemiológicos, así como por los cambios en las organizaciones sanitarias, vistos en el capítulo 1, que resulten.

Convendrá no perder de vista que las normas clínicas constituyen una auténtica institución social, con importancia comparable a la del mercado en la asignación de recursos sanitarios. Los cambios que afecten a esta institución deben sopesarse de forma adecuada.

Se finaliza este trabajo con un análisis de aquellas tendencias actuales en Medicina que parecen venir exigidas por los cambios de entorno y que supondrán una mejora de la decisión clínica. Tres son las tendencias fundamentales que actualmente se dan en Medicina: una primera de creciente apoyo de la práctica en la evidencia científica, que será referida como Medicina Basada en la Evidencia (MBE), una segunda de orientación hacia las utilidades del usuario, y una tercera hacia la interiorización del coste social de oportunidad en las decisiones clínicas. Así como la primera tendencia cuenta con una correlación de fuerzas sociales favorables la segunda y la tercera requieren de una actuación pública que las impulse.

VI.1 La Medicina Basada en la Evidencia⁴³

El desarrollo de la MBE viene motivado por cuatro evoluciones sociales distintas, aunque interrelacionadas: conciencia de amplias variaciones en la práctica médica (VPM), no explicables por diferencias en la morbilidad de las poblaciones; la

⁴³ Basado en partes de un trabajo más amplio realizado por uno de los autores (VO) con Salvador Peiró y Ricard Meneu y publicado en Meneu et al, 1998.

preocupación por la efectividad de las prácticas médicas, que aumenta en presencia de una fortísima innovación tecnológica (en productos, no en procesos) que, sin embargo, tiene un impacto decreciente sobre la salud; las políticas de contención de costes; y, cuarto, la mayor facilidad para acceder a la información.

Respecto al primero de estos factores, los estudios sobre VPM (Wennberg J, 1996) han mostrado la presencia de variaciones aparentemente arbitrarias en la forma de practicar la medicina entre países, regiones, áreas pequeñas y grupos poblacionales, propiciando la aparición de lo que Skrabanek denominó *escepticismo* y que, en lo esencial, es una quiebra en la creencia de que los sanitarios aplican de manera uniforme un tratamiento inequívocamente adecuado –científico– a cada problema de salud. En ausencia de explicaciones plausibles por el lado de la demanda (morbilidad, accesibilidad o preferencias de los pacientes), la constatación de VPM plantea importantes dudas sobre la indicación de determinadas intervenciones y sobre los resultados para los pacientes en función del uso de una u otra práctica posible, llevando directamente a preguntarse por la evidencia que sostiene cada estilo de práctica y, en situaciones de incertidumbre, a cuestionar la legitimidad de la aplicación de las valoraciones subjetivas de los profesionales a despecho de la evidencia disponible o las preferencias de los pacientes (Marión J et al, 1998). En nuestro país, diversos investigadores han empezado a documentar VPM utilizando, por ejemplo, la Encuesta de Morbilidad Hospitalaria (Sarria A, 1995; Compañ L et al, 1995) o los registros del Conjunto Mínimo Básico de Datos al Alta Hospitalaria (Jané E et al, 1996) con resultados muy ilustrativos. En uno de estos estudios (Marqués JA et al, 1998) se mostraba que la probabilidad de haber sido intervenido de cataratas a los 85 años se situaba en torno al 8% en un área de salud, mientras que en otra no alcanzaba el 1%. Las áreas con mayor oferta de recursos oftalmológicos, en relación a su población, mostraban tasas mayores de intervenciones, pero también mayor lista de espera por 10.000 habitantes, sugiriendo que donde hay mas recursos disponibles se indican mas intervenciones de cataratas. En otro reciente estudio sobre pruebas analíticas en Murcia, Huguet Ballester (1998) encuentra amplias variaciones en las tasas de peticiones de algunas pruebas entre zonas de salud: la VSG variaba entre 41 y 119 solicitudes por 1000 habitantes año (Razón de variación⁴⁴, RV: 2,9), los urocultivos de 19 a 91 (RV: 5), y algunas pruebas como la hemoglobina glucosilada (RV: 11), ASLO (RV: 70) o alfa-amilasa (RV: 92), aun mostraban mayores variaciones.

Aunque esta información de base poblacional tiene sumo interés, todavía resulta más propiciatoria de la *escepticismo* la constatación de variaciones en estudios de base individual abordados por los propios clínicos en respuesta a preguntas surgidas de la práctica cotidiana. Así, el porcentaje de partos que acaban en cesárea en hospitales públicos de la Comunidad Valenciana en 1994-95, variaba desde el 14% al 24%, pero para los partos de nalgas estos porcentajes oscilaban entre el 40% al 90% y, en presencia de sufrimiento fetal, entre el 7% al 97% (Librero J et al, 1997). En estos

⁴⁴La razón de variación (high-low ratio, extremal quotient, EQ) es la razón entre el más alto y el más bajo de los valores observados (valor máximo / valor mínimo). Pese a que apenas utiliza información, únicamente los dos valores más extremos, es una medida muy utilizada por su sencillez y su valor intuitivo para interpretar la variabilidad (una RV igual a 2 significa una variabilidad del doble entre el área con mayor y menor frecuencia del procedimiento). Debe ser interpretado con cautela cuando las áreas son pequeñas (menos de 10.000 habitantes) o la frecuencia del procedimiento es baja.

hospitales, y para 1995, la realización de una artroplastia de cadera tras fractura osteoporótica de cadera varió del 4% al 38% de los pacientes, y la duración de la estancia desde 9 a 19 días, sin que el ajuste por la gravedad de los pacientes atendidos consiguiera reducir estas diferencias (Peiró S, 1997). Estas variaciones se han relatado incluso en situaciones urgentes de riesgo vital, como el empleo de fármacos trombolíticos en el infarto agudo de miocardio (Agustí A et al, 1997).

La escepticismo conduce a preguntarse sobre la utilidad de algunas de las rutinas y procedimientos habituales y este tipo de preguntas lleva a preocuparse por la efectividad que, ahora, irá mas allá de las medidas clásicas de resultados (mortalidad, complicaciones, mejora de parámetros fisiológicos) para intentar valorar el impacto sobre la calidad de vida. La MBE utiliza los métodos habituales de investigación clínica (ensayos clínicos, estudios observacionales con o sin grupo control, etc.) para valorar la eficacia y efectividad de las prácticas médicas. Las revisiones sistemáticas de la literatura científica, la gradación de la confianza en los resultados en función de los diseños de investigación y facilitar su difusión e incorporación a la práctica habitual son las propuestas básicas de la MBE (Sackett D et al, 1996). La importancia de las revisiones sistemáticas se suele mostrar con el contundente ejemplo de los trombolíticos en el infarto de miocardio: mientras que los ensayos clínicos habían demostrado la eficacia de esta terapia desde mediados de los años 70, su aparición en los libros de texto y artículos de revisión no empieza hasta 10 años mas tarde, y su uso generalizado no se produjo hasta casi el final de los 90. En el interín, miles de personas fallecieron porque no se les aplicó una terapia cuya eficacia, aunque sólidamente evidenciada no estaba suficientemente difundida. (Antman et al, 1992).

Respecto a las políticas de contención de costes, las que depositan su fe en el mercado se han apuntado a la producción y difusión de investigación para que los pacientes y, sobre todo, los compradores influyan sobre los proveedores de servicios. En su vertiente positiva, el argumento que sustenta esta aproximación es que mediante guías de práctica basadas en la evidencia puede limitarse razonablemente la libertad clínica, restringiendo el uso de procedimientos inefectivos. En su vertiente más restrictiva, las guías de práctica se centran en la reducción de costes, afectando la autonomía profesional mas allá de lo que la evidencia sustenta. Así, mientras que en Europa la libertad clínica se mantiene relativamente estable, en Estados Unidos la autonomía profesional ha quedado seriamente afectada por la generalización de la "medicina gestionada" y un médico puede recibir varias guías clínicas diferentes sobre como tratar una patología –cuyo cumplimiento puede afectar sus honorarios- según quien sea el pagador/asegurador del paciente.

Respecto a la disponibilidad de información, los más de medio millón de ECA realizados hasta la actualidad, o los 2 millones de artículos que cada año aparecen en más de 20,000 revistas biomédicas, resultan ahora más accesibles, tanto por los procesos de revisión sistemática ya mencionados, como por la generalización y abaratamiento del acceso a bases de datos de documentación médica. Pese a la existencia de diversas limitaciones (selección, transmisión, preeminencia absoluta del inglés, con su correlato de dificultades para recuperar la información relevante), el acceso a la información aportada por la investigación ha experimentado recientemente

enormes avances. Un número creciente de personas están conectadas a Internet en España, red que permite la consulta inmediata -y gratuita- de bases de datos bibliográficas como Medline, el acceso a los principales fuentes de información, y la participación en foros de debate y listas de distribución donde compartir preocupaciones y resultados. Internet también permite una mayor diversidad de información que las bases bibliográficas convencionales, incluyendo textos completos, guías de práctica, revisiones sistemáticas y acceso a parte de la llamada literatura 'gris' y, mas allá de la fascinación por las novedades tecnológicas, se ha convertido en un instrumento importante para la aplicación de enfoques basados en la evidencia.

VI.2 Orientación hacia las utilidades del paciente

El clínico -en ejercicio de su libertad clínica- ha de aplicar las recomendaciones de buena práctica, las guías, a las circunstancias individuales de cada paciente. El enfoque clínico exige una adaptación de tal conocimiento a las circunstancias de cada individuo. La investigación probabilística predice resultados pero no determina acciones excepto cuando una intervención siempre es efectiva o nunca es efectiva.

Conforme una sociedad progresa el componente 'calidad de vida' en el producto de los servicios sanitarios aumenta. Una gran parte de las actuaciones sanitarias no tienen traducción en términos de mayor cantidad de vida, pero sí en mejor calidad de vida. Al ser la calidad subjetiva y multidimensional resulta inevitable el activismo del paciente para poder establecer el curso de acción más efectivo. Ello requiere proporcionar información pronóstica a los pacientes. Por ejemplo, la efectividad de la prostatectomía en el tratamiento de la hiperplasia benigna de próstata depende de cuáles son las dimensiones de calidad de vida apreciadas por los pacientes (incontinencia, impotencia). Un vídeo que muestra un par de médicos con prostatismo sintomático y en el que se aprecia la evolución del médico que optó ser intervenido quirúrgicamente y del que prefirió la estrategia de la vigilancia expectante, es una valiosa ayuda para que el paciente a tomar una decisión informada (siempre que la información utilizada para elaborar el vídeo sea válida); el que además haya desplazado las preferencias de los pacientes hacia menos cirugía resulta secundario (Wagner E et al, 1995). Igualmente, en Ottawa, un folleto y una grabación permiten a las mujeres que están considerando la terapia hormonal substitutoria, ponderar los beneficios y los riesgos que para cada una de ellas tendría el tratamiento en función de sus valores, factores de riesgo y estilos de vida (O'Connor A et al, 1996).

La orientación hacia las utilidades del paciente resulta generalmente exigible y se vuelve imprescindible para valorar la efectividad de los tratamientos con impacto importante en la calidad de vida y para realmente actuar como agente del paciente, haciendo lo mejor para él. La medida de la utilidad puede utilizarse en ocasiones para expresar la preferencia global de un paciente ante diversas alternativas. Esta utilidad viene muy afectada por la actitud ante el riesgo de los pacientes. La habitual aversión al riesgo hace que en bastantes ocasiones los pacientes no prefieran las alternativas más efectivas, sino aquellas a las que asocian la mayor utilidad. Por ejemplo, el tratamiento

más efectivo de la trombosis venosa profunda es una combinación de estreptokinasa y heparina, pero el pequeño riesgo de hemorragia intracraneal y muerte, asociado a este tratamiento más efectivo, hace que clínicos y pacientes prefieran el menos efectivo tratamiento con heparina sola, con su alto riesgo de pierna inflamada y dolorosa (Kassirer J, 1994). En bastantes decisiones diagnósticas (amniocentesis para el síndrome de Down) y terapéuticas (tratamiento médico o quirúrgico de la angina de pecho) la actitud del paciente ante el riesgo puede ser determinante de la elección.

La investigación existente sugiere que la utilización de la evidencia de modo compartido con los pacientes y estimulándoles a participar en las decisiones diagnósticas y terapéuticas produce mejores resultados sanitarios y a menor coste que las transacciones de arriba abajo (Hart JT, 1997). Al menos en los casos en los que la evidencia disponible no pueden traducirse en un resultado inequívoco deben desarrollarse estrategias de transferencia de información y elección informada. Estrategias que además deberían extenderse a toda intervención en que la dimensión relevante, o una de las dimensiones en conflicto fuese esencialmente subjetiva, como calidad de vida *versus* prolongación de supervivencia.

VI.3 Interiorización del coste social de oportunidad

La financiación pública de la sanidad, por razones de eficiencia y equidad, ha entronizado la responsabilidad clínica no ya ante un paciente sino ante el conjunto de todos ellos. Si hacer lo mejor posible por un paciente puede contradecir el conseguir lo mejor para el conjunto de los pacientes, la consideración del beneficio sanitario que deja de obtenerse en la mejor alternativa clínica disponible se vuelve inevitable. El coste de oportunidad -dimensión social- ha de interiorizarse en la decisión clínica -actuación individual.

Considerar el beneficio, diagnóstico o terapéutico, que deja de obtenerse en la mejor alternativa razonablemente disponible constituye una forma de asegurarse que se obtiene el máximo resultado, en términos de impacto en el bienestar, a partir de unos recursos determinados. El verdadero coste de la atención sanitaria no es dinero, ni los recursos que el dinero mide. Son los beneficios sanitarios -paliación de síntomas, recuperación funcional, mayor esperanza de vida- que podrían haberse conseguido si esas pesetas se hubieran utilizado en la mejor alternativa. El coste de oportunidad puede aproximarse a través de los precios y por otros mecanismos cuando éstos no existen. Mayor dificultad presenta la conceptualización y medición de los beneficios.

Convendrá distinguir entre beneficios en una decisión diagnóstica y beneficios en una decisión terapéutica. El beneficio en una decisión diagnóstica se mide en términos de reducción de incertidumbre. Un paciente tiene o no tiene una condición desde el inicio del proceso hipotético-deductivo de diagnóstico. Lo que va cambiando es la idea que médico tiene acerca de tal condición. Exploración, historia, pruebas diagnósticas sirven -en aquellas ocasiones donde observar y esperar no procede- para ir revisando esas ideas

hasta situar las probabilidades a una distancia próxima al cero o al uno. La proximidad exigida respecto al cero (descarte de una condición) o respecto al uno (confirmación de un diagnóstico) depende de múltiples factores: efectividad del tratamiento, iatrogenia del tratamiento en caso de error diagnóstico, sensibilidad, especificidad y riesgo de las pruebas que se estén considerando. El teorema de Bayes ayuda a relacionar las probabilidades posteriores de padecer una condición con posterioridad a la prueba diagnóstica con las probabilidades previas. Tendrá sentido continuar con un proceso diagnóstico hasta el punto en que los beneficios adicionales esperados -en términos de ganancias de información, de revisión de probabilidades- dejen de ser superiores a los costes de las pruebas diagnósticas. Y esos costes son los de oportunidad: los beneficios informativos que la prueba diagnóstica que se está considerando podrían reportar al paciente que, razonablemente, más podría beneficiarse de ella. En las decisiones terapéuticas, el beneficio se mide en términos de efectividad. Esta efectividad tiene, como mínimo, un par de dimensiones: cantidad y calidad de vida. La consideración del coste de oportunidad supone una reflexión acerca del beneficio del tratamiento que se está considerando en su mejor alternativa disponible.

Los incentivos para interiorizar el coste de oportunidad en las decisiones diagnósticas y terapéuticas y, en general, para estimular la eficiencia de la práctica clínica exigen un reconocimiento organizativo de la realidad de la gestión clínica. La asunción del protagonismo de la gestión clínica en los centros sanitarios supone reconocer como responsables de las diferentes líneas de producción sanitaria a aquellos que están al cargo del cuidado directo de los pacientes, los gestores clínicos que establecen los bienes y servicios que bajo las órdenes médicas se administrarán a cada paciente. Y esta asunción de la gestión clínica en los centros sanitarios admite dos escenarios diferentes: el del ejercicio por cuenta propia y el del ejercicio por cuenta ajena, con el problema fundamental que plantea la dificultad de medir el desempeño clínico (Ortún, V, 1998).

En todo caso, y para finalizar, la mayor elección por parte del usuario de aquellos servicios sanitarios que más se aproximen a un bien de experiencia o de información (elección de médico de cabecera, decisión sobre una intervención habitual de cirugía programable...) precisa de la recogida y elaboración pública de información sobre la efectividad de los servicios sanitarios ajustando debidamente por aquellas variables que explican resultados y que el proveedor no controla. También la mejora de los procesos colectivos para el establecimiento de prioridades sociales: democratización, intervención y responsabilidad de las autoridades locales en la gestión de servicios, explicitación de prioridades y -en general- mejora de la legitimidad del proceso que lleva a conseguir un consenso social en cuanto a objetivos colectivos.

Bibliografía referenciada

Agustí A, Mateu S, Molina LI, Arnau JM, Vidal X, Laporte JR. Utilización de fármacos trombolíticos en Barcelona. *Med Clin* 1997; 108: 167-170.

Antman E, Lau J, Kupelnick B, Mosteller F, Chalmers T. A Comparison of Results of Meta-analyses of Randomized Control Trials and Recommendations of Clinical Experts. Treatments for Myocardial Infarction. *JAMA* 1992; 268:240-48

Compañ L, Peiró S, Meneu R. Variaciones geográficas en hospitalizaciones quirúrgicas en ancianos (1985-1989). *Rev Gerontol* 1995; 5:166-170.

Hart JT. What evidence do we need for EBM?. *J Epidemiol Community Health* 1997; 51:623-629

Huguet Ballester J, Arccas García F, Benedito Rodríguez JE, Egea Caparrós JM, Gomez Gomez AE, Ortolá Devesa J. Variabilidad en la solicitud de pruebas analíticas en un área de salud. En: Variabilidad en la práctica clínica. Huesca: Asociación de Economía de la Salud, 1998 (en prensa).

Jané E, Barba G, Salvador X, Salas T, Sánchez E, Bustins M. Variaciones en la tasa de hospitalización por procedimientos quirúrgicos seleccionados. Aplicación del análisis de áreas pequeñas. *Gac Sanit* 1996; 10: 211-219.

Kassirer J. Incorporating patients' preferences into medical decisions. *New Engl J Med* 1994; 330: 1895-1896.

Librero J, Peiró S, Márquez S. Variabilidad en la práctica de cesáreas entre hospitales: ¿existe relación con las diferencias en población atendida? 1999, en revisión.

Marión J, Peiró S, Marquez S, Meneu R. Variaciones en la práctica médica. Importancia, causas e implicaciones. *Med Clin (Barc)* 1998; 110:382-390.

Marqués JA, Peiró S, Medrano J, Librero J, Perez-Vazquez MT, Aranz J et al. Variabilidad en las tasas de intervenciones de cirugía general por áreas de salud. *Cir Esp* 1998; 63:445-453.

Meneu R, Ortún V, Peiró S. Medicina basada en l'evidència. Possibilitats i limitacions per a la compra de serveis. En G López (dir): La contractació de serveis sanitaris. Barcelona: Departament de Sanitat de la Generalitat de Catalunya, 1998, págs 185-207.
O'Connor A, Tugwell P. Making choices: Hormones after menopause. Ottawa: University of Ottawa, 1996.

Ortún V, Del Llano J. Mercado y Estado en Sanidad. En Del Llano J, Ortún V, Martín JM^a, Millán J, Gené J. Gestión sanitaria: innovaciones y desafíos. Barcelona: Masson, 1998: 3-16.

Peiró S, Librero J, Ordiñana R. Perfiles de mortalidad hospitalaria: ¿una herramienta útil para la identificación de potenciales problemas de calidad? *Rev Calidad Asistencial* 1997; 12: 179-187.

Peiró S, Ortún V, Meneu R. Medicina basada en l'evidència. Possibilitats i limitacions per a la compra de serveis. En G López (dir): *La contratació de serveis sanitaris*. Barcelona: Departament de Sanitat de la Generalitat de Catalunya, 1998: 185-207. Existe versió castellana.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312:71-72

Sarria Santamera A, Sendra Gutierrez JM. Diferencias regionales en utilización hospitalaria. *Gac Sanit* 1993; 7:63-69.

Wagner E, Barrett P, Barry M, Barlow W, Fowler F. The effect of a shared decisionmaking program on rates of surgery for benign prostatic hyperplasia. *Med Care* 1995; 33: 765-770.

Wennberg JE. On the appropriateness of small-area analysis for cost-containment. *Health Aff* 1996; 15: 164-167.