

18



Evaluación de Acciones Formativas en Medicina basada en la Evidencia (MBE).

M^a Carmen Sánchez López
Juan A^o Sánchez Sánchez

1.INTRODUCCIÓN.

2. METODOLOGÍA.

3. RESULTADOS.

- 3.1. Revisiones sistemáticas.
- 3.2. Ensayos clínicos controlados aleatorizados (ecas) .
- 3.3. Ensayos clínicos controlados no aleatorizados.
- 3.4. Estudios pre y post intervención.
- 3.5. Herramientas validadas para evaluar acciones formativas.
- 3.6. Acciones formativas a distancia y “on-line”.

4.DISCUSIÓN.

5.CONCLUSIONES(PUNTOS CLAVE).

6.IMPLICACIONES PARA LA INVESTIGACIÓN FUTURA EN ESTE CAMPO.

7. TABLAS DE LOS ESTUDIOS REVISADOS.

8. BIBLIOGRAFÍA.

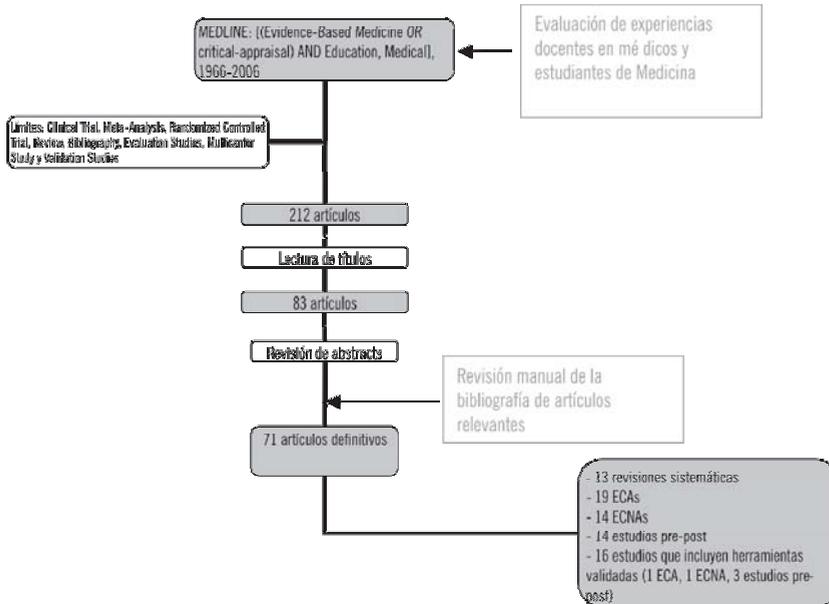
En un inquietante artículo de opinión¹ se comparaba la enseñanza de la Medicina Basada en la Evidencia (MBE) con el símil del “traje del emperador” y se reconocía el riesgo de que a los docentes de MBE se les pudiera señalar en un determinado momento por la multitud con el dedo, al creerse revestidos de ese invisible traje. Irónicamente si uno quiere desarrollar guías de cómo enseñar MBE, basada en resultados, deberán estar apoyadas en niveles bajos de evidencia, ya que la mayoría de los datos disponibles sobre la evaluación de la docencia proceden en general de estudios descriptivos observacionales².

Los resultados de la formación son difíciles de medir, el diseño de ensayos de intervención controlados es complejo y no ha sido tradicionalmente una prioridad de las agencias de investigación. Pero, aunque es pequeña la evidencia sobre la enseñanza de la cardiología en la competencia del profesional, no por ello se deja de enseñar. La experiencia ha mostrado que muchos médicos clínicos aprenden y utilizan en su práctica la MBE, luego enseñarla es posible. Parece que la cuestión no es si la MBE se debe enseñar, sino como enseñarla mejor. Para este propósito la evaluación de la formación es fundamental, para poder establecer mejoras en el diseño y en los objetivos de la misma.

Es necesaria una investigación diseñada de forma rigurosa en la efectividad de la educación para proporcionar resultados generalizables y para elevar el perfil de la investigación educativa en la profesión médica³. De hecho una intervención educativa puede ser estudiada a un nivel similar que otros ensayos clínicos. Lo que ocurre es que la formación puede estar influenciado por muchos factores, es una intervención compleja y a menudo muchos elementos actúan de forma sinérgica. El tamaño adecuado de la muestra, la posibilidad de contaminación del grupo control, los sesgos de selección, las pérdidas en el seguimiento son aspectos esenciales en el diseño y ejecución de un ensayo clínico y que cobran una especial dificultad al plantearse la evaluación de acciones formativas.

Con este capítulo pretendemos recopilar las experiencias docentes publicadas en MBE, que presenten algún tipo de evaluación, para extraer conclusiones y aprender de las mismas de cara a la planificación de actividades docentes y de investigación educativa. En concreto, sus objetivos son:

- Revisar experiencias evaluadas docentes en MBE dirigidas a médicos y a estudiantes de medicina.
- Conocer las herramientas validadas para medir competencias en MBE.
- Extraer unas recomendaciones sobre la planificación docente de acciones formativas en ASBE.



Se realizó una revisión bibliográfica en la base de datos MEDLINE utilizando como estrategia de búsqueda [(Evidence-Based Medicine OR critical-appraisal) AND Education, Medical] (con la selección en límites de Clinical Trial, Meta-Analysis, Randomized Controlled Trial, Review, Bibliography, Evaluation Studies, Multicenter Study and Validation Studies) desde 1966-2006 sobre evaluación de experiencias docentes en Medicina Basada en la Evidencia (MBE) dirigidas a médicos y estudiantes de medicina. Con la búsqueda se obtuvieron 212 artículos y, tras la lectura de los títulos, se seleccionaron 83 potencialmente útiles para nuestro estudio. Tras esta primera lectura se descartaron 8 artículos por presentar un idioma diferente al inglés, castellano, francés, italiano o portugués (los artículos tenían como idioma: 5 el japonés, 2 el alemán y 1 el polaco). También se excluyó dos artículos que no cumplían los criterios de inclusión en relación a la categoría profesional. Se revisaron después los resúmenes (“abstracts”) de los artículos seleccionados y tras su lectura fueron rechazados 20 nuevos artículos. Se completó la búsqueda con la revisión manual de la bibliografía de todos los artículos relevantes. Finalmente se obtuvieron un total de 71 artículos que fueron traducidos y analizados.

Tras esta primera búsqueda se realizó una complementaria [(Evidence-Based Medicine OR critical-appraisal) AND (Reproducibility of Results OR Validat*)], utilizando similar selección en el apartado de límites, para asegurarnos de la inclusión de todos los trabajos con cuestionarios validados presentes en la literatura. Esta búsqueda identificaba un total de 600 estudios, aunque tras la revisión de los títulos tan sólo se extrajeron 2 nuevos artículos potencialmente útiles. La lectura posterior de estos trabajos descartó su inclusión en el estudio.

3.1. REVISIONES SISTEMÁTICAS

Tabla 1

Las revisiones sistemáticas publicadas muestran como las intervenciones educativas analizadas mejoran, en general, el conocimiento sobre MBE y las actitudes hacia el uso de la literatura médica en la toma de decisiones aunque no hay evidencia de que esos cambios modifiquen el comportamiento de los profesionales o sus habilidades para aplicar la MBE en la práctica clínica. No obstante, estos hallazgos deben ser valorados con cautela al tratarse, en la gran mayoría de los casos, de trabajos de pobre calidad metodológica.

Por otro lado, las actividades docentes se suelen centrar en la lectura crítica, olvidando otras habilidades de la práctica de la MBE, y suelen impartirse de forma independiente a la clínica cuando revisiones más recientes, de Coomarasamy⁴⁹ y Werb⁵⁰, muestran como la integración de la formación con la clínica no sólo facilita la adopción de conocimientos de MBE sino que además mejora las actitudes, habilidades y comportamientos en relación a este paradigma.

Otra limitación apreciada a la hora de considerar los resultados es la escasez de trabajos que incluyan instrumentos de evaluación previamente validados.

Observamos también la heterogeneidad de los estudios en relación a objetivos de aprendizaje, contenidos curriculares, estrategias de enseñanza, intensidad de las intervenciones, diseño, evaluación de métodos y medida de resultados. Tal heterogeneidad complica la síntesis de la evidencia, por medio de la realización de análisis cuantitativos o meta-análisis, y suele motivar que los revisores establezcan decisiones subjetivas sobre cuales son los aspectos más importantes de la intervención. Sólo en la revisión de Norman y Shannon⁴² fue posible la síntesis cuantitativa de los hallazgos.

Otro aspecto importante a tener en cuenta es que el análisis de los resultados evaluados revela la carencia de trabajos que valoren los efectos de las experiencias docentes en resultados en salud y, en el caso de que este resultado se evalúe^{40,41}, nos encontramos ante la necesidad de estudios bien diseñados que permitan evidenciar cómo los cambios en el conocimiento y las actitudes de los alumnos se traducen en una mejora de resultados en salud. En relación a este último aspecto, Dobbie et al¹ encuentran, en un estudio cualitativo, resultados similares y se plantean la continuidad de estos programas docentes si no podemos encontrar evidencia de que la formación en MBE modifique la práctica clínica y los resultados en salud.

Sólo en dos de las revisiones^{41,45} se plantean el análisis de la efectividad de estrategias docentes específicas. Por una parte Davis et al⁴¹ comparan diferentes intervenciones educativas encontrando que las más efectivas para mejorar la práctica clínica y los resultados en salud incluían recordatorios, intervenciones basadas en el paciente, reuniones educacionales, opiniones de líderes y actividades

multidisciplinares; en cambio métodos de enseñanza tradicionales y ampliamente usados se muestran inefectivos. Por otra parte, Ebbert et al⁴⁵ se centran en la evaluación de las sesiones bibliográficas encontrando que dicha estrategia educacional puede mejorar hábitos de lectura, conocimientos en epidemiología clínica y bioestadística y comportamientos (uso de la literatura médica en la práctica clínica) en médicos postgraduados en formación, sin poder comprobar su papel en la mejoría de habilidades en lectura crítica.

Destacamos de forma aislada la revisión sistemática de Shaneyfelt et al⁵¹ por su reciente publicación y las diferencias que plantea, en cuanto a sus objetivos, con relación a revisiones previas. En ella se revisan instrumentos de evaluación y estrategias de formación en MBE, documentando su desarrollo, formato, nivel de aprendizaje, dominios de MBE evaluados, propiedades psicométricas y, de forma novedosa, factibilidad. Como vemos en la tabla, incluyeron en el estudio un total de 115 trabajos de los que se extrajeron 104 herramientas de evaluación. De forma similar a lo planteado en revisiones sistemáticas previas la mayoría de las intervenciones docentes se realizan en estudiantes y residentes. En cuanto a los instrumentos, en la mayor parte de los casos evalúan habilidades en MBE, seguido de conocimientos y comportamientos y, por último, actitudes. Dentro de las habilidades, la evaluación crítica de la evidencia fue incluida en la mayor proporción de instrumentos. Hallan sólo tres instrumentos^{4,5,6} que valoran los cuatro pasos de la MBE, de los cuales únicamente el test de Fresno y el cuestionario de Berlín están validados. De ellos, consideran el cuestionario de Berlín como el más factible de implantar por su formato de respuesta múltiple, aunque dicho formato restringe la evaluación de la aplicación de conocimientos en MBE. Cuando los instrumentos evalúan comportamientos, lo más frecuente es que recojan la auto-percepción de la ejecución de los pasos de la MBE, a pesar de que este enfoque puede estar muy sesgado pues los médicos tienden a subestimar sus necesidades de información y sobreestimar el grado de su actividad. En relación a actitudes, a pesar de que los ítems sobre actitudes hacia la MBE se incluyen en muchas de las herramientas, este dominio se suele valorar de forma superficial. En cuanto a las propiedades psicométricas de las estrategias de evaluación, los investigadores evaluaron la fiabilidad interobservador en 21 de los 51 instrumentos en los que era apropiado, usando fundamentalmente el estadístico K y los coeficientes de correlación; encuentran que al menos un tipo de validez se demuestra en el 53% de ellas pero sólo en el 10% se documenta más de un tipo de validez. Destacamos en este punto que los criterios de inclusión establecidos dejan fuera estudios que incluyen herramientas validadas que sí están presentes en revisiones sistemáticas previas. Finalmente, respecto a la factibilidad de la implantación, esta únicamente fue recogida en 19 de los 104 instrumentos.

En relación a las limitaciones que encuentran los autores que realizan revisiones sistemáticas sobre intervenciones en educación médica encontramos el trabajo de Reed et al⁷. Tal y como veíamos previamente, para este autor las principales dificultades que encuentran los revisores son la baja calidad de los estudios, con escaso número de ensayos clínicos aleatorizados controlados; la heterogeneidad de los trabajos, en cuanto a las estrategias de enseñanza descritas y la diferente intensidad de las intervenciones; y la carencia de métodos de evaluación de resultados educacionales que sean válidos y fiables. Al igual que en trabajos previos queda patente la falta de estudios que demuestren relación entre intervención educativa y resultados clínicos. Expone, finalmente, una serie de recomendaciones que mejorarían la validez interna de los estudios y permitirían la reproductibilidad de las intervenciones y generalización de los resultados.

3.2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADOS (ECAS)

Tabla 2

En los ECAs evaluados participan, en la mayoría de los casos, médicos fuera del periodo de formación. Observamos en estos estudios multitud de intervenciones, heterogéneas en cuanto a su intensidad docente, contenido de la intervención y forma de instauración. Son varios los trabajos que combinan diferentes estrategias educacionales. Entre las experiencias docentes evaluadas destacan aquellas que se centran en la evaluación crítica de artículos (6/19), aquellas en las que se instruye al alumno en la realización de una adecuada búsqueda de información (5/19) y los talleres o cursos de MBE (4/19). Otras intervenciones educativas que encontramos son sesiones bibliográficas, reuniones educacionales (visita individualizada de un educador a un profesional de la salud en su ámbito de trabajo), formación en el manejo de guías clínicas y de sistemas informáticos, formación en la construcción de preguntas clínicas e instrucción para la selección eficiente de literatura.

El principal instrumento de medida empleado es el cuestionario auto-contestable de respuesta múltiple que en 4 de los trabajos^{52,65,66,68} fue previamente validado. En dos de ellos, la herramienta utilizada fue la propuesta por Taylor et al^{66,68}. Con respecto al trabajo de Cheng⁶⁴, comenta el uso de un instrumento de evaluación validado sin describir, sin embargo, el proceso de validación.

Son múltiples los resultados evaluados y los métodos estadísticos manejados para su análisis. En la mayoría de los estudios se evalúan los cambios en las habilidades en lectura crítica y en la realización de búsquedas bibliográficas tras la experiencia docente. Otros resultados medidos son los hábitos de lectura de los participantes, las actitudes hacia la MBE, la aplicación de la evidencia a la práctica clínica y la utilización de fuentes electrónicas de información como apoyo a la toma de decisiones. En uno de los artículos⁶² se evalúan los cambios en los patrones de prescripción de fármacos tras una visita personalizada por un experto (reunión educacional).

En cuanto a los métodos estadísticos hay una amplia utilización de los tests tradicionales χ^2 , T de Student y U de Mann-Whitney. En relación a este tema Cheng⁶⁴ introduce de forma novedosa el uso de un modelo de regresión logística que permite calcular la OR y el NNT, procedimientos estadísticos más adecuados y acordes con la filosofía MBE.

Por último, en relación a los resultados obtenidos, los artículos muestran, en general, como las intervenciones educativas basadas o relacionadas con los principios de la MBE aumentan los conocimientos de los participantes, incrementan sus habilidades para la lectura crítica de un artículo y para llevar a cabo búsquedas bibliográficas y mejoran las actitudes hacia su uso. Los trabajos que analizan cambios en el comportamiento revelan como la formación en MBE y evaluación crítica puede modificar positivamente las decisiones clínicas, cambiar los patrones de prescripción y mejorar la aplicación de los principios de la MBE a la práctica clínica. En aquellos estudios en los que se evalúa se aprecia una mejoría en los hábitos de lectura tras la intervención docente.

Las estrategias docentes que se muestran poco efectivas, utilizadas de forma aislada, son la formación en el manejo de las guías clínicas, la opinión de líderes y la recepción de vía e-mail de resúmenes evaluados críticamente. En relación a las reuniones educacionales (visita de un educador a un profesional de la salud en su ámbito de trabajo en la que se suelen combinar técnicas de auditoría y retroalimentación) encontramos resultados enfrentados en dos artículos realizados en 2002^{62,8}. Con respecto a este tema, una revisión reciente de la librería Cochrane evalúa los efectos de este tipo de intervenciones sobre la práctica de los profesionales sanitarios y las medidas de resultado de los pacientes y concluye que la auditoría y la retroalimentación pueden ser eficaces para mejorar la práctica profesional, pero, cuando son eficaces, los efectos son generalmente de pequeños a moderados.

Se realiza un seguimiento del efecto de la intervención en los ensayos de Stevermer⁵⁷ y Cheng⁶⁴ (6 y 12 meses respectivamente); Stevermer muestra como la mejora de los hábitos de lectura, en relación al conocimiento de información relevante, persiste a los 6 meses de la intervención. Cheng, por su parte, deja patente que la ganancia en conocimientos y habilidades en búsquedas se erosionan con el tiempo aunque se mantienen los cambios en el uso preferente de fuentes electrónicas de información frente a fuentes tradicionales.

3.3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS

Tabla 3

En relación a los ensayos clínicos controlados pero no aleatorizados observamos como la gran mayoría utiliza como población a residentes de distintas especialidades (7/14): 4 estudios se centran en estudiantes de Medicina, uno en residentes y estudiantes y en otro participan estudiantes y médicos con actividad docente. Sólo en uno de los trabajos la población de estudio está compuesta por médicos fuera del periodo de formación. Como en el caso de los ECAs, son variadas las intervenciones educativas utilizadas, imperando, en este caso, las sesiones bibliográficas basadas en la MBE y los cursos o talleres en los que se trabajaban sus distintas habilidades, principalmente lectura crítica y habilidades de búsqueda. En uno de los trabajos⁸², la formación se complementa con la instrucción en el uso de una PDA (Personal Digital Assistant) y del programa InfoRetriever (sistema de apoyo para la toma de decisiones y recuperación de información clínica en bases de datos). Destacamos también, en relación a este tema, el ensayo de Bennet et al⁷² puesto que en él evalúan los cambios en la destreza para la evaluación crítica de estudiantes pero a través de la formación específica de sus tutores en tal habilidad.

Los instrumentos de medida utilizados (12/14) son cuestionarios auto-contestables que habían sido validados previamente en dos de los casos^{81,82}. Uno de los ensayos asociaba al test la evaluación de artículos en los que se habían introducido errores metodológicos y otro la grabación en audio de la interacción residente-preceptor. Otras herramientas de evaluación empleadas son la evaluación de la resolución de problemas clínicos y el análisis de la calidad de las búsquedas realizadas.

Estos instrumentos son utilizados para la valoración de múltiples resultados entre los que destacan habilidades en lectura crítica y realización de búsquedas, conocimientos en MBE y aplicación de sus principios en la práctica clínica, actitudes hacia la MBE o literatura médica, hábitos de lectura y auto-valoración de los participantes en relación a sus habilidades en MBE.



Prácticamente todos los estudios (9/14) emplean como métodos de análisis los tests X^2 y T de Student.

Para finalizar, en cuanto a los resultados obtenidos, destacamos que las sesiones bibliográficas se muestran como un método ineficaz para incrementar las habilidades en evaluación crítica, tal y como quedaba patente en la revisión sistemática de Ebbert⁴⁵. Langkamp⁷³ las considera también ineficaces para modificar los conocimientos de los profesionales sobre MBE. El uso de la PDA, con la herramienta InfoRetriever como sistema de apoyo, no parece aportar beneficios a los residentes. Respecto al resto de experiencias docentes se observan que consiguen incrementar las habilidades para realizar búsquedas bibliográficas, para evaluar críticamente un artículo, y para construir de forma más adecuada preguntas clínicas. Mejoran, además, los conocimientos sobre MBE y la utilización de fuentes de evidencia en la práctica clínica. Estos cambios positivos ocurren, sobre todo, cuando la enseñanza se integra en la clínica (basada en problemas). No encontramos estrategias docentes que modifiquen los hábitos de lectura.

3.4. ESTUDIOS PRE Y POST INTERVENCIÓN

En la revisión de la literatura se extrajeron 14 estudios antes-después^{10,25}.

Los participantes en los estudios son, en la mayor proporción, residentes de distintas especialidades (6 estudios) y médicos (3 estudios). En 4 trabajos participan médicos y residentes y en otros 3 la muestra la constituyen estudiantes. En un estudio, la población está constituida por médicos y enfermeros.

En todos los estudios se analizan los cambios que ocurren en los sujetos de la muestra tras intervenciones educativas de diversa índole: en 4 de los trabajos se describen talleres centrados en los conceptos de la MBE y su aplicación clínica, lectura crítica o búsquedas en MEDLINE; en otros 3 estudios la intervención consiste en la participación en sesiones bibliográficas reestructuradas en base a la MBE; en 5 trabajos los sujetos participan en cursos y seminarios de formación en MBE; en 2 estudios los participantes reciben en primer lugar sesiones didácticas sobre MBE y posteriormente intervienen en sesiones bibliográficas; y en un estudio la intervención educativa consiste en una rotación de 2 semanas durante la cual los residentes trabajan los conceptos y destrezas de la MBE. En el trabajo de Lucas et al²⁵, no se realiza una intervención educacional propiamente dicha sino que los médicos que participan en el estudio analizan los resultados de búsquedas bibliográficas realizadas por los investigadores en base a los problemas clínicos que presentan sus pacientes.

El instrumento de evaluación de resultados más frecuente fue el cuestionario, en 13 de los estudios, que combina preguntas de respuesta múltiple y preguntas abiertas y se aplica antes de la intervención educativa y después de esta. En el estudio de Cramer¹¹ y en el de Weberschock²², los cuestionarios utilizados fueron previamente validados, y en los trabajos de Thom²⁰ y Dinkevich²³ se utiliza una

versión adaptada del test validado de Fresno⁵. Straus¹⁶ realiza la evaluación por medio de la valoración de las modificaciones encontradas en los informes de alta de los pacientes ingresados más de 24 horas en el transcurso de 1 mes antes y 1 mes después de la intervención educativa. En el trabajo de Vogel et al¹⁵ se analiza la destreza que presentan los sujetos para realizar búsquedas bibliográficas antes y después de un taller centrado en este tema. Este estudio presenta además como interés el que, mediante el análisis de las búsquedas de los residentes hasta 11 meses después del taller, se pretende examinar la persistencia en el tiempo de los cambios. En esta línea, Schoenfeld¹⁸ proyecta el mismo objetivo para lo cual vuelve a aplicar el cuestionario 6 meses después de impartir su seminario. Por último, destacamos el estudio de Lucas et al²⁵, en el cual, los autores se proponen valorar el impacto de los resultados de búsquedas bibliográficas estandarizadas en la toma de decisiones de los médicos, formulando como hipótesis el que dichas búsquedas pueden cambiar el tratamiento elegido si revelan opciones terapéuticas no consideradas previamente o proporcionan datos que modifican la evaluación de las opciones anteriormente consideradas.

Los cuestionarios que hemos descrito miden conocimientos y habilidades en MBE (3 de ellos se centran en habilidades para la lectura crítica), conocimientos en investigación básica y bioestadística, auto-percepción de la habilidad para la práctica de la MBE, uso de fuentes informáticas y hábitos de lectura médica. El trabajo de Weberschock²² evalúa además la aplicación de la evidencia en la práctica, proponiendo para ello la resolución de un caso clínico de acuerdo a los principios de la MBE.

El análisis estadístico de los datos se realiza utilizando de forma mayoritaria el test T de Student. Otros test utilizados son el de Wilcoxon, U de Mann-Whitney, McNemar y ANOVA.

Los resultados de todos los estudios muestran un mejoría significativa de conocimientos y habilidades en MBE tras la intervención educativa y en uno de esos trabajos se observa como esta mejoría se mantiene pasados 6 meses. También se observa como, tras la intervención, se incrementa el uso de fuentes secundarias de información, el tiempo dedicado a la lectura de literatura biomédica y el uso de la informática. Mejora también la destreza en la realización de búsquedas bibliográficas y la auto-percepción de los sujetos en cuanto a sus habilidades. Straus¹⁶ muestra como la enseñanza modifica positivamente el modelo de práctica clínica con un mayor uso de tratamientos beneficiosos basados en ensayos clínicos aleatorizados y revisiones sistemáticas y Weberschock²² observa una adecuada transferencia del conocimiento en MBE a un escenario clínico. Lucas et al²⁵ establecen como los médicos, tras conocer los resultados de las búsquedas aportadas, modificaban, con la intención de mejorar, los tratamientos de muchos de los pacientes que atendían, incluso si previamente estos ya recibían un tratamiento basado en la evidencia.

Sin embargo hemos de ser cautelosos a la hora de interpretar los resultados que se desprenden de este tipo de estudios debido a sus limitaciones metodológicas. Destaca Bradley⁶⁸ en su trabajo la limitación debida al efecto aprendizaje, motivo por el cual los resultados de estudios donde se compara un pre y un post-test pueden resultar inválidos puesto que los participantes han aprendido como responder a los ítems a través de la exposición pre-intervención.

3.5.HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS

Tabla 4

Tras la revisión de la literatura obtenemos 16 estudios en los que se relata el desarrollo y validación de un cuestionario o test para evaluar actividades formativas en MBE a profesionales sanitarios. La construcción del cuestionario deriva de una revisión de lo publicado previamente en la literatura en 7 de los trabajos, y en dos de ellos se suman además las aportaciones de expertos en MBE. En los restantes 9 estudios los autores construyen los ítems de la herramienta con la ayuda, en la mitad de los casos, de la revisión o cooperación de expertos en MBE. El proceso de validación se describe de forma detallada y completa (entendiendo como tal los artículos en los que se describe al menos la evaluación de dos formas de validez y de la fiabilidad) en sólo 9 de los trabajos, careciendo el resto de un análisis adecuado de la validez del instrumento. Destacamos entre ellos el test de Hong Kong⁸⁹ por ser el único en el cuál se describe de forma completa la evaluación de los tres tipos de validez y de la fiabilidad.

En relación a los tipos de validez, los autores establecen de forma preferente la validez facial y de contenido, por medio de la realización de un pretest cognitivo y la evaluación del instrumento por un panel de expertos. En segundo lugar, encontramos la evaluación de la validez de constructo, fundamentalmente la validez discriminativa entendida como la capacidad de la herramienta para diferenciar entre participantes con distintos niveles de experiencia o competencia en MBE. En sólo 3 de los trabajos se analiza la validez de criterio^{85,89,90}. No se ha incluido en la tabla el test propuesto por Landry et al⁷⁵ al no hacer referencia el artículo el proceso de validación aunque se desprende de la lectura del artículo que dicho test muestra validez de constructo sensible, al ser capaz de detectar el impacto de la intervención educativa, y, al haber sido testado en un grupo de expertos en MBE, podemos presuponer que la validez de contenido también fue garantizada.

En la distintas fases de la validación participan estudiantes de Medicina, residentes o médicos en proporción similar. Uno de los cuestionarios utiliza como población a distintos profesionales de la salud⁸⁶.

Los cuestionarios, en la mayoría de los casos, están compuestos por preguntas de respuesta múltiple o con respuesta según escala tipo Likert. El test de Fresno⁵ presenta también preguntas abiertas que evalúan la capacidad del encuestado para aplicar los principios de la MBE. Todos los instrumentos, de una forma u otra, tienen como objetivo medir competencias en MBE, en términos de conocimientos, actitudes y habilidades. Algunos incluyen además la autovaloración de dichas competencias por parte de los encuestados y en otros casos también evalúan los hábitos de lectura médica, el uso de la MBE en la práctica clínica y el uso de las distintas fuentes de información. El cuestionario de MacRae⁹¹ consta de ítems con los que se evalúa la calidad de los artículos propuestos. El instrumento propuesto por Stern⁸⁴ incluye ítems que analizan la percepción de los encuestados sobre su capacidad para evaluar críticamente un artículo de tratamiento adjunto al test. Esta herramienta sirve de base para trabajos posteriores^{76,26}, entre los que destaca el de Thomas et al²⁶ que demuestra la generabilidad de este instrumento a otros marcos y a otro tipo de artículos.

La fiabilidad del instrumento se valora en 14 de los estudios. En la mayor parte de ellos, la valoración se realiza a través de la consistencia interna, medida a su vez por medio del coeficiente de Cronbach (9/14). Otros métodos empleados en la medición de la fiabilidad son la estabilidad del test, mediante el método test-retest, y la fiabilidad inter-ratio e inter-observador, utilizando coeficientes de correlación. Los test de Berlín⁴ y Fresno⁵ emplean además la correlación ítem-total.

3.6. ACCIONES FORMATIVAS A DISTANCIA Y ON-LINE

Son escasos los estudios encontrados sobre enseñanza on-line de aspectos relacionados con la MBE y, en general, son descripciones de experiencias docentes:

En un curso a distancia dirigido a médicos irlandeses sobre lectura crítica²⁷, los médicos respondieron que, tras la realización del curso, era más probable que hicieran referencia a información basada en la evidencia cuando toman decisiones clínicas. Concluye con la propuesta de que una mezcla de métodos de enseñanza, con inclusión de recursos interactivos y proyectos basados en la evidencia optimiza los beneficios de la educación a distancia.

Se ha publicado la experiencia de un curso on-line de MBE dirigidos a estudiantes de pre y postgrado en una Universidad Francesa²⁸; sin que se conste una evaluación formal de la misma.

Algunas experiencias apuntan a una efectividad de la enseñanza a través de internet, como el aprendizaje con casos clínicos que se ha mostrado, en algunos aspectos, con resultados superiores a la enseñanza en grupos²⁹. Un estudio muestra la efectividad parcial de la discusión por correo electrónico de casos clínicos, cuando se miden competencias a través de pacientes estructurados³⁰.

En un ensayo randomizado, en estudiantes de medicina en Noruega, el autoaprendizaje con un programa de ordenador no mostró diferencias estadísticamente significativas comparada con la enseñanza en pequeños grupos en la adquisición de conocimientos, habilidades y actitudes en MBE³¹.

18

→ 4 DISCUSIÓN

Como hemos comprobado, la gran mayoría de estudios identificados en esta revisión miden el nivel de aprendizaje de conocimientos y/o de habilidades, siendo muy pocos los que miden modificaciones en el comportamiento, al ser esta tarea mucho más compleja. Son prácticamente inexistentes los trabajos que evalúan modificaciones en el estado de salud provocadas por la actuación docente.



La evaluación educativa supone una valoración sistemática de la calidad de la enseñanza y aprendizaje³². A nivel de medición de resultados en evaluación educativa, se han definido cuatro niveles, que adquieren la forma de pirámide, en función del número de estudios realizados que se asignan a cada nivel³. En la base de la pirámide figuran resultados de la reacción a la actividad formativa, como sería la satisfacción con el curso de los alumnos. En el primer nivel se mide aprendizaje: medición de conocimientos y habilidades adquiridas. El segundo nivel mide modificaciones en el comportamiento, si hay una transferencia de aprendizaje que se refleje en el quehacer cotidiano y en el último nivel se miden resultados en términos de salud o de beneficio para la sociedad.

El abanico de competencias precisas para el uso adecuado de la MBE es muy amplio; aunque la mayoría de los estudios han valorado exclusivamente la adquisición de conocimientos y habilidades en lectura crítica. Se han señalado hasta 18 habilidades que debe tener un médico de familia para practicar la medicina basada en la evidencia³³. Estas van desde formular correctamente las preguntas de interés para clarificar una situación clínica o sanitaria, hasta conocer las diferentes estrategias de búsquedas electrónicas, valorar críticamente los diversos tipos de información científica o la efectividad desde la perspectiva de los pacientes. Recientemente se han definido unas herramientas de evaluación y de planificación de actividades docentes de MBE que diferenciando entre actitudes, habilidades, competencias, comportamiento y resultados clínicos las define para cada una de las cinco etapas de la práctica de la MBE³⁴. Este modelo diferencia entre varios perfiles en la relación del profesional con la evidencia:

1. El que “hace o fabrica evidencias” y que deberá realizar de forma autónoma las cinco etapas. Precisa de una formación y adiestramiento específico.
2. El “usuario de la evidencia”: formula adecuadamente las preguntas y conoce la estrategia de búsquedas, pero revisa información ya seleccionada por otros (por ejemplo revisiones sistemáticas, revistas secundarias o guías de práctica clínica). Reconoce si la información que está manejando es de calidad.
3. El que “replica la evidencia”: reconoce lagunas de conocimiento y formula adecuadamente las preguntas, pero revisa lo que otros han realizado y resumido (Ejemplo: usuario de un banco de preguntas sobre evidencias o de CATs).

Este enfoque asume que, aunque la mayoría de los médicos deben conocer a grandes rasgos los recursos de MBE, no es necesaria una formación reglada para poder utilizarla, aunque sí determinadas habilidades³⁵ y permite adaptar la formación a las necesidades reales de los alumnos.

Sobre el tipo de enseñanza, junto al método tradicional de la enseñanza presencial, surge con auge, en los últimos años la formación a distancia y la docencia virtual (“on-line”). El e-learning es definido como la utilización de las nuevas tecnologías multimedia y de internet para mejorar la calidad del aprendizaje, facilitando el acceso de recursos y servicios; así como los intercambios y la colaboración a distancia. Son muy escasas las experiencias encontradas en docencia en MBE en esta modalidad y la revisión no permite extraer conclusiones válidas sobre su efectividad comparada con la docencia presencial.

Es llamativo el esfuerzo por realizar estudios de síntesis (17 revisiones sistemáticas localizadas), que no se corresponde con un esfuerzo en realizar estudios primarios de calidad y con suficiente tamaño muestral que permita extraer conclusiones generalizables. Sin duda las dificultades metodológicas para llevar a la práctica este tipo de estudio justifica la ausencia de los mismos.

Hay dos aspectos a destacar de esta revisión, uno el efecto de las intervenciones educativas a largo plazo y otra, de gran interés, la integración en la práctica clínica de las habilidades entrenadas durante la formación. Son pocos los estudios que evalúan los cambios a largo plazo en las competencias en MBE. Wadland et al³⁶ muestran en su artículo la efectividad de un plan de estudios en lectura crítica en múltiples ámbitos, como el efecto de la instrucción continua después de la graduación y como los conocimientos adquiridos son considerados más valiosos y aplicables a la práctica clínica durante la residencia. Sobre la integración en la práctica clínica, se ha observado (Byrnes et al²¹) como el proporcionar acceso y formación individualizada en el uso de PubMed e Internet en los emplazamientos sanitarios produce cambios favorables en el comportamiento de los usuarios respecto al uso de la información (búsquedas en Internet y en Medline) y en la autovaloración de la confianza en su uso. También mejora tras esta intervención la percepción de la importancia de la MBE en la consecución de una atención de calidad a los pacientes.

La mayoría de los estudios proceden de un ámbito geográfico determinado (Inglaterra y USA) sin que se haya identificado ningún estudio realizado en España, ni publicado en castellano.

De todos los métodos encontrados de formación, es difícil extraer conclusiones sobre cuales son más efectivos para lograr resultados pues la heterogeneidad de los estudios impide extraer conclusiones. Los métodos integrados en la práctica, que combinan diferente metodología de enseñanza, parecen conseguir mejores resultados. Y hay otros que no se muestran efectivos, como la visita de líderes o la formación en el manejo de guías de práctica clínica.

En un artículo de revisión publicado recientemente³⁷, de forma similar a lo encontrado tras nuestra revisión, los autores concluyen que las estrategias docentes integradas en la práctica clínica mejoran conocimientos, habilidades, actitudes y comportamientos, mientras que si la enseñanza se desvincula de la clínica sólo se logran ganancias en conocimientos. También encuentran que frente a las actividades docentes didácticas (que por sí solas en poco probable que resulten en mejoras), las estrategias interactivas pueden mejorar los resultados educacionales (en términos de ganancias o logros de aprendizaje de los participantes) y los resultados en salud.

- Las revisiones sistemáticas publicadas muestran como las intervenciones educativas analizadas incrementan, en general, el conocimiento sobre MBE y las habilidades para su aplicación, fundamentalmente en relación a la evaluación crítica, y mejoran las actitudes hacia la MBE y el uso de la literatura médica en la toma de decisiones. Sin embargo, son escasos los estudios que analizan cambios en el comportamiento de los discentes y los que lo hacen no muestran resultados favorables. Se observa también una importante ausencia, que dificulta la obtención de conclusiones consistentes, de trabajos que evalúen si los cambios apreciados en las competencias en MBE se traducen posteriormente en una mejor atención clínica.
- Observamos una carencia de estudios con calidad metodológica adecuada, debido, en general, a las dificultades que surgen a la hora de evaluar la efectividad de actividades formativas.
- La mayoría de las experiencias docentes evaluadas se centran en la lectura crítica dejando a un lado el resto de competencias de la MBE.
- Las actividades docentes que, usadas de forma aislada, se presentan como poco eficaces para modificar competencias en MBE son formación en el manejo de guías clínica y transmisión de la opinión de líderes. En cuanto a las sesiones bibliográficas, estas se muestran ineficaces para la mejora de habilidades en evaluación crítica.
- La mayor efectividad (en términos de ganancia de conocimientos y mejoría de actitudes, habilidades y comportamientos) de las actividades docentes se observa cuando estas se integran en la práctica clínica y se combinan diversos métodos de enseñanza.
- Se ha detectado una carencia de herramientas fiables, válidas y reproducibles para medir el núcleo de competencias en MBE y su aplicación en la práctica clínica. No hay ninguna desarrollada en castellano.
- Las experiencias de formación on-line son escasas y no permiten extraer conclusiones sobre las ventajas sobre la docencia presencial.
- No hay resultados concluyentes en cuanto si la formación en MBE modifica los hábitos de lectura de los participantes.

18

→ 6

IMPLICACIONES PARA LA INVESTIGACIÓN FUTURA EN ESTE CAMPO

- Son muchos los instrumentos que contienen ítems sobre actitudes pero pocos aquellos que valoran este dominio en profundidad. Sin embargo, la evaluación de actitudes sobre MBE puede descubrir barreras ocultas pero potencialmente remediabiles en el desarrollo y ejecución de las habilidades en MBE⁵¹.

- Existe la necesidad de estudios que evalúen la competencia de los profesionales en la integración de la evidencia encontrada, las circunstancias clínicas y las preferencias de los pacientes para la resolución de un problema clínico concreto.

- Son necesarios estudios que determinen si las actuales herramientas de evaluación pueden ser usadas para evaluar un amplio rango de clínicos, tales como enfermero y otros profesionales de la salud relacionados.

- Siguiendo las recomendaciones propuestas por Reed et al⁷, al plantearnos la realización de futuros trabajos sobre evaluación de actividades docentes deberíamos:

- aumentar el rigor en cuanto a su diseño, análisis y recogida de datos;
- identificar claramente el propósito del estudio y utilizar un diseño adecuado para conseguirlo;
- describir claramente la población diana y demostrar como los grupos intervención y control son comparables.

Por otra parte, la intervención educativa debería ser detallada de forma completa (incluyendo contenido educacional, estrategias de enseñanza, ámbito de trabajo y fuentes requeridas) y adecuada, para permitir la reproducción del estudio, y habría de estar integrada en la práctica clínica. Los métodos de evaluación utilizados deberían ser objetivos y, a poder ser, estar validados, informando en ese caso el estudio sobre la fiabilidad y validez de dichos instrumentos.

Una conferencia de consenso realizada en el año 2005³⁸ entre expertos en docencia en MBE, recomienda, sobre la educación en MBE, que: i) los profesionales sanitarios deben incorporar los conocimientos, habilidades y actitudes necesarias en su formación académica; ii) que los currículos para alcanzar estas competencias deben estar basados en el modelo de cinco etapas de la MBE; y iii) que son necesarias más investigaciones en los métodos más efectivos y eficientes de enseñanza.

En definitiva es patente, fundamentalmente en nuestro entorno, la necesidad de más investigación sobre evaluación de las acciones formativas en Medicina Basada en la Evidencia y, sobre todo, de mejor calidad metodológica y enfocada a medir competencias, modificaciones del comportamiento y repercusión en la salud de los ciudadanos.

18

→ 7

TABLAS DE LOS ESTUDIOS REVISADOS

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
Audet et al, 1993 ³⁹ .	<ul style="list-style-type: none"> - Población: estudiantes de medicina en nivel pre o post graduado. - Intervención educativa dirigida a mejorar una o más de las siguientes áreas: conocimientos en epidemiología clínica o bioestadística, hábitos de lectura y habilidad para evaluar críticamente un artículo científico. 	10 artículos cumplieron los criterios de selección.	No realizado.	De los 10 artículos incluidos sólo 4 demostraban una mejora estadísticamente significativa en conocimientos de epidemiología y bioestadística, en hábitos de lectura y en la habilidad para evaluar críticamente un artículo. Además encontraron que la calidad metodológica de los artículos era generalmente pobre. Sólo uno de los estudios utilizaba un ensayo clínico controlado aleatorizado.	La efectividad de la enseñanza de lectura crítica es incierta.
Johnston et al, 1994 ⁴⁰ .	<ul style="list-style-type: none"> - Población : sanitarios en prácticas o formación. - Intervención: sistema informático de apoyo a las decisiones clínicas evaluado en un emplazamiento clínico. - Medidas de resultado: rendimiento del clínico o resultados en salud. - Estudios prospectivos con un grupo control. 	Se seleccionaron 28 ensayos clínicos controlados.	No realizado.	La mayoría de los estudios que evalúan sistemas informáticos de apoyo para el tratamiento, medidas preventivas y una actuación médica de calidad muestran mejora en el rendimiento clínico. Sólo en uno de los 5 estudios se observa un resultado favorable en el efecto de estos sistemas en el rendimiento diagnóstico. En 3 de los 10 estudios que valoran resultados en salud se encuentra una mejora significativa.	La revisión concluye que existen fuertes evidencias de que el uso de sistemas informáticos como ayuda para la toma de decisiones puede mejorar el rendimiento de médicos. Son necesarios nuevos estudios, bien diseñados, para valorar el efecto y coste-efectividad, especialmente en resultados en salud.
Davis et al, 1995 ⁴¹ .	<ul style="list-style-type: none"> - ECA controlados o ensayos en los que se usasen alternativas para aleatorización. - Intervenciones educacionales reproducibles dirigidas a modificar el comportamiento médico o los resultados en salud - Población: profesionales de la salud en los que la mayoría fuesen médicos o residentes - Medidas de resultado: cambios en el rendimiento de los médicos o en resultados en salud. 	99 artículos cumplieron los criterios de inclusión.	No realizado.	Los 99 ensayos incluían un total de 160 intervenciones educativas. De ellas 101 mostraban mejora en al menos una de las medidas de resultado: el 70% un cambio en el rendimiento médico y el 48% modificación positiva en resultados en salud. Las estrategias más efectivas incluían recordatorios, intervenciones basadas en el paciente, visitas individualizadas por personal especializado (reuniones educacionales), opiniones de líderes y actividades multidisciplinarias.	Estrategias ampliamente usadas (programas tradicionales de enseñanza didáctica como conferencias) tienen un pequeño impacto en la mejora de la práctica clínica. Métodos educacionales como estrategias basadas en la práctica y reuniones educacionales se muestran efectivos para mejorar el rendimiento de los médicos pero son escasamente utilizados.

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
Norman y Shamoun, 1998 ⁴² .	<ul style="list-style-type: none"> - Utilización de alguna forma de grupo control (no se requería una estricta aleatorización). - Medidas de resultado: conocimientos o destrezas en lectura crítica y epidemiología clínica; auto-percepción del uso de la literatura. 	<p>Se obtuvieron 10 trabajos pero finalmente 3 estudios fueron excluidos al ser sus resultados imposibles de interpretar.</p>	<p>De los 7 estudios se calculó la media de cambio en la puntuación de todas las medidas de resultado, la diferencia de la media y la desviación estándar. Se encuentra una considerable consistencia en los hallazgos de los estudios.</p>	<p>En estudiantes de medicina se aprecia cómo cursos de lectura crítica incrementan significativamente el nivel de conocimientos de epidemiología clínica pero la mejora en dicho conocimientos no garantiza su aplicación.</p> <p>Los residentes presentan un menor incremento de conocimientos, aunque estadísticamente significativo en 2 de los 3 estudios. Al igual que en los estudios con estudiantes, no se encuentra incremento en el uso de la literatura o mejora en la valoración crítica de artículos.</p>	<p>Los resultados muestran como la instrucción en evaluación crítica consigue mejorar el conocimiento en MBE en estudiantes con un efecto mucho menor entre residentes. No hay evidencias de que la ganancia en conocimientos se mantuviese en el tiempo, cambiase el comportamiento con respecto al uso crítico de la literatura o se pudiese trasladar a la mejora de los resultados en salud.</p>
Green, 1999 ⁴³	<ul style="list-style-type: none"> - Artículos que tratan sobre planes de estudio en relación a epidemiología clínica, lectura crítica y MBE en graduados médicos. - Evaluación del proceso de implantación del plan de estudio informando sobre la asistencia, duración o intensidad de la enseñanza, cargas administrativas y obstáculos o sostenibilidad. - Inclusión de análisis cuantitativo de la valoración de los alumnos de los aspectos del plan de estudios y del impacto del plan de estudios en las habilidades, conocimientos, comportamientos o actitudes de los alumnos. - Confirmación de la capacidad del instrumento de medida para discriminar entre diferentes niveles de habilidad. 	<p>18 artículos cumplan los criterios de inclusión.</p>	<p>La heterogeneidad de las intervenciones y de los resultados valorados impide un metaanálisis.</p>	<ul style="list-style-type: none"> - El objetivo más común de los planes de estudio fue mejorar las habilidades en evaluación crítica. - La intervención más común fue la sesión bibliográfica con pequeños grupos de residentes. - Siete de los artículos evaluaban la efectividad del plan de estudios y sólo 4 eran ensayos controlados. - El instrumento de evaluación de resultados más frecuente fue el examen de respuesta múltiple para medir los conocimientos en epidemiología clínica y lectura crítica. Sólo en un artículo se evaluaba la fiabilidad y validez del instrumento (Linzer et al) y en otro se evaluaba únicamente la fiabilidad (Green and Ellis). - El impacto en las habilidades en lectura crítica en los 4 ensayos controlados oscilaba entre el no efecto y el 23% de incremento absoluto en la puntuación del test. Sólo en 2 de estos artículos se recogían resultados conductuales. - Siete trabajos incluyen análisis de satisfacción y muestran alto nivel de agrado. 	<p>La mayoría de los estudios presentan descripciones incompletas e inadecuada evaluación de sus planes de estudio. Los planes de estudio suelen centrarse en la lectura crítica, excluyendo otras habilidades de la MBE, y tienen limitada efectividad.</p>

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
Taylor et al, 2000 ⁴⁴ .	<ul style="list-style-type: none"> - Exposición de una intervención educativa dirigida a mejorar las habilidades en lectura crítica - Presencia de un grupo de control que no reciba la intervención educativa o reciba una intervención placebo - Estudios cuya población sean profesionales de la salud - Objetivos: medir resultados educacionales o resultados en salud 	<p>14 estudios que cumplan los criterios de inclusión de los cuales solo se pudieron conseguir 10 a texto completo.</p>	<p>No se realizó un estudio estadístico de los resultados dada la heterogeneidad de los diseños de los estudios, métodos y resultados.</p>	<p>La calidad de los estudios es en general pobre (sólo uno presenta un diseño de ensayo clínico controlado y aleatorizado y únicamente dos de los artículos puntuaban más del 50% en la valoración de la calidad). Todos se realizaron en estudiantes de medicina o doctores en formación. Las intervenciones educativas usadas eran de diferente intensidad y analizaban distintos resultados (conocimientos de epidemiología y bioestadística, actitudes hacia la literatura médica, habilidades en lectura crítica y hábitos de lectura médica). Se recoge una mejora del 68% en conocimientos de epidemiología /bioestadística, resultados positivos en las actitudes hacia la evidencia y que los estudios cuyo grupo control no recibe una intervención educativa muestran una más alta proporción de resultados positivos (77%) que aquellos cuyo grupo control recibe una intervención educativa comparativa (44%).</p>	<p>La revisión muestra como la enseñanza de destrezas en lectura crítica mejora los conocimientos y las actitudes hacia la literatura médica de los participantes. La evidencia de una mejora en la habilidad para leer críticamente un artículo o para cambiar los hábitos de lectura fueron no convincentes.</p>
Ebbert et al, 2001 ⁴⁵ .	<ul style="list-style-type: none"> - Participantes: médicos postgraduados en formación - Intervención: journal club (sesión bibliográfica), definido como la reunión de un pequeño grupo para discutir uno o más artículos. - Diseño del estudio: ensayos aleatorios, estudios de cohortes, estudios antes-después, estudios transversales. - Resultados valorados: habilidades en lectura crítica, hábitos de lectura, conocimientos en epidemiología clínica y bioestadística, uso de la literatura médica en la práctica clínica o mejora en la atención al paciente. 	<p>De los 63 potenciales artículos solo 7 cumplieron todos los criterios de inclusión. Entre estos artículos uno era un ensayo aleatorizado, tres eran estudios de cohortes, uno era un estudio antes-después, y dos eran estudios transversales.</p>	<p>No realizado.</p>	<p>Sólo el ensayo controlado aleatorizado usó una herramienta de evaluación validada.</p> <ul style="list-style-type: none"> - Ensayo controlado aleatorizado: Muestra una mejora estadísticamente significativa en todos los resultado de interés (hábitos de lectura, conocimientos en epidemiología clínica y bioestadística, uso de la literatura médica en la práctica clínica) excepto en habilidades en lectura crítica. - Estudios de cohortes: Las sesiones bibliográficas no producen cambios estadísticamente significativos en los resultados evaluados. - Estudio antes-después: Muestra mejora en los conocimientos de los participantes sobre lectura crítica. Los resultados del estudio también muestran un incremento significativo en el tiempo dedicado a la lectura de literatura médica después del taller. - Estudios transversales: Sugieren que las sesiones bibliográficas permiten mejorar el entendimiento del diseño de la investigación y los hábitos de lectura. 	<p>Las sesiones bibliográficas son una intervención educativa que puede mejorar los hábitos de lectura, los conocimientos en epidemiología clínica y bioestadística y el uso de la literatura médica en la práctica clínica en médicos postgraduados en formación. No se ha comprobado su papel en la mejora de habilidades en lectura crítica.</p>

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
<p>Parkes et al (Revisión Cochrane traducida), 2001⁴⁶.</p>	<ul style="list-style-type: none"> - Tipo de estudio: ensayos aleatorios, ensayos clínicos controlados; estudios controlados del tipo antes y después y análisis de series. - Intervenciones educacionales sobre evaluación crítica. - Población: profesionales de la salud. - Resultados: proceso de atención; la mortalidad de los pacientes, la calidad de vida y la satisfacción; y el conocimiento/concienciación de los profesionales. - Uso de instrumentos estandarizados y fiables. 	<p>Tan sólo un ensayo controlado aleatorio del tipo antes y después cumplió con los criterios de inclusión..</p>	<p>No realizado.</p>	<p>La medida de resultado evaluada era el conocimiento de la evaluación crítica. No se evaluaron el proceso de atención, la salud de los pacientes o los resultados en actitud/concienciación. La enseñanza de evaluación crítica había resultado en una mejoría del 25% en el grupo intervención comparado con una mejoría del 6% en el grupo control (p = 0.02).</p>	<p>La enseñanza de la evaluación crítica presenta efectos positivos sobre el conocimiento de los participantes. Existen grandes vacíos en las pruebas sobre si repercute en el proceso de atención sanitaria o en la salud de los pacientes. También resulta incierto si el tamaño del beneficio observado es lo suficientemente grande como para tener significación práctica.</p>
<p>Garg y Turtlet, 2003⁴⁷.</p>	<ul style="list-style-type: none"> - Intervenciones educacionales de 1 a 8 horas sobre búsqueda de literatura on-line y recuperación de la información. - Resultados medidos: habilidades de los profesionales de la salud en búsqueda de literatura on-line y recuperación de la información. - Tipos de estudios: ensayos clínicos controlados, estudios controlados antes y después, estudios de cohortes controlados. - Participantes: profesionales de la salud cualificados y estudiantes del entorno de la salud. 	<p>Se encontraron 18 potencialmente útiles estudios pero sólo tres de ellos cumplieron los criterios de inclusión .</p>	<p>La síntesis cuantitativa no se aplicó puesto que los estudios tenían diferentes instrumentos de medida y distintos métodos de enseñanza.</p>	<p>De los 3 estudios analizados (dos ensayos clínicos controlados y un estudio controlado antes-después), uno de los ensayos clínicos y el estudio antes-después muestran mejoría significativa en la habilidad para realizar búsquedas en la literatura y en la calidad de dichas búsquedas tras la intervención educativa y en relación al grupo control.</p>	<p>Los estudios muestran mejoría en las habilidades para realizar búsquedas pero no hay clara evidencia sobre el nivel de efecto, si el efecto pasa y sobre los efectos en el conjunto de profesionales de la salud.</p>

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
Coomarasamy et al, 2003 ⁴⁸ .	<ul style="list-style-type: none"> - Participantes: postgraduados de Medicina. - Estudios que utilizaban un grupo control o cambios respecto a la situación antes de la enseñanza. - Medidas de resultado: conocimientos, habilidades, actitudes y comportamientos en relación a la IMBE; rendimiento de los médicos o ganancias en salud. 	<p>17 artículos cumplieron los criterios de inclusión (2 ECA, 6 ensayos clínicos sin aleatorización y 9 estudios antes-después).</p>	<p>No fue factible debido a la gran variación en la características de los estudios y en la calidad metodológica.</p>	<p>15 estudios valoran conocimientos y de ellos 10 encuentran resultados positivos (2 ECA, 3 ensayos clínicos no aleatorizados y 5 estudios antes-después). La habilidad en evaluación crítica se valoró en 6 estudios; de ellos dos encuentran mejoría tras la intervención educativa (1 ensayo clínico no aleatorizado y 1 estudio antes-después). Modificación en actitudes es analizada en 5 estudios y solo un estudio antes-después encuentra cambios positivos. En cuanto a cambios en el comportamiento 9 estudios los recogen como resultado y tan solo 2 de los ensayos clínicos no aleatorizados y un estudio antes-después encuentran mejoría.</p>	<p>La revisión de dichos estudios muestra una significativa mejoría en conocimientos pero no en actitudes, habilidades y comportamientos. Sin embargo, no pueden obtenerse inferencias conclusivas debido a el pobre diseño de muchos de los estudios, pobreza en el diseño de la instrucción y falta de desarrollo de instrumentos de valoración fiables y válidos.</p>
Coomarasamy y Khan, 2004 ⁴⁹ .	<ul style="list-style-type: none"> - Estudios que utilizasen un grupo control. - Participantes: postgraduados en Medicina. - Medidas de resultados: cambios en conocimientos en IMBE, habilidad en lectura crítica, actitudes y comportamiento hacia la IMBE y ganancia en salud de los pacientes. 	<p>Se seleccionaron 23 estudios (4 ensayos aleatorios, 7 estudios controlados no aleatorios y 12 estudios antes y después); 18 estudios evaluaban el método de enseñanza independiente y 5 el método de enseñanza integrada en la clínica.</p>	<p>No se realizó un meta-análisis debido a la heterogeneidad de los estudios incluidos.</p>	<p>La enseñanza independiente mejoraba conocimientos pero no habilidades, actitudes o comportamientos. La enseñanza integrada en la clínica mejoraba conocimientos, habilidades, actitudes y comportamientos (incluyen cambios en el hábito de lectura, en la elección de fuentes de información y cambios en el manejo de los pacientes y guías clínicas). Ningún estudio evaluaba resultados en salud.</p>	<p>La enseñanza debería trasladarse del aula a la práctica clínica para conseguir mejorar no solo los conocimientos y habilidades sino también las actitudes y el comportamiento. Cuando no sea posible la enseñanza adaptada en tiempo real a la clínica, los distintos métodos de enseñanza podrían basarse en problemas clínicos reales o de actualidad. Es más probable que mejoren las habilidades cuando la duración de la formación es al menos de 8 horas.</p>

TABLA 1. REVISIONES SISTEMÁTICAS.

Autores	Criterios de inclusión	Artículos incluidos	Meta-análisis	Resultados obtenidos	Conclusiones
Werb y Matear, 2004 ⁵⁰ .	<p>Las investigaciones originales se incluyeron si cumplían los cuatro criterios esenciales y dos deseables que aparecían en las guías del Evidence-Based Medicine Working Group, para asegurar que los estudios incluidos fueran al menos estudios casos-control con un buen diseño. Las revisiones sistemáticas fueron incluidas si cumplían los cinco criterios esenciales precisos para las revisiones sistemáticas.</p>	<p>Tras aplicar los criterios de inclusión sólo 12 estudios (9 investigaciones originales y 3 revisiones sistemáticas) fueron seleccionados y sometidos a análisis.</p>	<p>No realizado.</p>	<p>Los resultados de la revisión muestran que los estudiantes que han completado un curso didáctico en asistencia sanitaria basada en la evidencia y enseñanza basada en problemas muestran una mejoría en las siguientes áreas: conocimiento de estadística y terapias médicas, actitudes hacia la evaluación crítica, puntuación en programas de examen dental y percepción de uso de los principios de la MBE en la clínica. Sin embargo sólo 3 estudios analizan el uso de los principios de la MBE en la clínica.</p>	<p>La enseñanza basada en problemas y los cursos de asistencia sanitaria basada en la evidencia ha mostrado ser efectivos en la disseminación de los principios de la MBE y en provocar en los estudiantes actitudes positivas hacia su uso en la práctica clínica. La introducción de escenarios clínicos en la formación se muestra como la mejor forma para introducir los principios de la MBE en los estudiantes.</p>
Shaneyfelt et al, 2006 ⁵¹ .	<p>Se incluyeron estudios que:</p> <ul style="list-style-type: none"> - incluyan un instrumento o estrategia que evaluase conocimientos, habilidades, actitudes o comportamientos en PBE o resultados en salud - contuviese una suficiente descripción del instrumento o estrategia para permitir su análisis - presentase resultados de probar el funcionamiento /rendimiento del instrumento o estrategia. <p>Se limitó el análisis a estudios publicados en inglés y se excluyeron los estudios que recogían sólo satisfacción con un plan de estudios.</p>	<p>Cumplieron los criterios de inclusión 115 artículos que incluyeron 104 estrategias distintas de valoración.</p>	<p>No realizado.</p>	<p>1. Características de los instrumentos de evaluación en PBE: Empleados mayoritariamente en estudiantes y residentes. En la mayor parte de los casos, los instrumentos evalúan habilidades en MBE y centro de ellas habilidades en evaluación crítica. 2. Factibilidad y tests psicométricos: La factibilidad de la implantación fue escasamente recogida (18%). Los investigadores evaluaron la fiabilidad interobservador en el 41% de los instrumentos en los que era apropiado. Se testó al menos un tipo de validez en el 53% de los 104 instrumentos. 3. Categorización de la calidad de los instrumentos: - 11 artículos incluidos en el nivel 1; los instrumentos descritos presentan evaluación de fiabilidad interrato, medidas de resultado objetivas y evaluación de múltiples tipos de validez (incluyendo validez de constructo discriminativa). - 9 artículos se incluyen en el nivel 2; instrumentos en los que se evalúa la fiabilidad interrato y validez de constructo (sensibilidad), establecida por estudios con un ensayo controlado aleatorizado o un ensayo controlado pre-post, y una medida de resultados objetiva. - el nivel 3 lo constituyen 6 artículos: los instrumentos evalúan comportamientos en MBE.</p>	<p>Los investigadores encuentran que los instrumentos usados para evaluar la MBE son mas comunmente administrados a estudiantes de medicina y residentes y valoran habilidades en búsqueda y, sobre todo, en evaluación de la evidencia (lectura crítica). Son escasos las herramientas de evaluación que presentan una validación considerada completa o adecuada. Los instrumentos del nivel 1 son apropiados para evaluar la competencia individual pero dada la ausencia de criterios bien definidos para los diferentes niveles de estudiantes no deberian ser usados para otras evaluaciones como promoción académica o certificación. Los instrumentos del nivel 2 son adecuados para evaluar el impacto general, más que individual, de las intervenciones en MBE.</p>

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Linzer et al, 1988 ⁵²	44 internos de Medicina Interna.	Los participantes fueron asignados aleatoriamente a una sesión bibliográfica (grupo intervención) o a una conferencia estándar sobre medicina ambulatoria (grupo control). Dos internos abandonaron tras la aleatorización y no fueron incluidos en el análisis final.	Cada interno, dirigido por un facultativo, seleccionaba un artículo de interés clínico y metodológico y lo preparaba para presentarlo en la sesión bibliográfica. Se proporcionaban además guías sobre métodos de lectura crítica. En la sesión bibliográfica se discutían los métodos, resultados y conclusiones del artículo y se abría una discusión sobre la utilidad clínica del artículo revisado. El grupo control participaba en conferencias en la que no se trabajaba la evaluación crítica, epidemiología o bioestadística. La media de sesiones (de una hora de duración y realizadas semanalmente) por interno fue 5.	A todos los internos se les administró un cuestionario validado (sobre hábitos de lectura, conocimientos de epidemiología y bioestadística y habilidad para utilizar este conocimiento en la evaluación crítica) antes y después de las sesiones educativas. Estos tests fueron separados una media de 9.5 meses.	Los cambios entre grupos fueron comparados con el test χ^2 . El porcentaje medio de mejoría obtenido por cada grupo fue comparado con el test t de Student. Para valorar la ganancia de conocimiento y el número de sesiones a las que se asistieron se usó el análisis de regresión.	La intervención educativa descrita (sesión bibliográfica) mejora significativamente los hábitos de lectura, conocimientos en epidemiología y bioestadística y uso de la literatura médica en la práctica clínica. Sin embargo no hay una significativa mejoría en las habilidades en lectura crítica comparado con el grupo control.
Haynes et al, 1993 ⁶³	308 médicos y residentes de un hospital docente.	Todos los participantes recibieron 2 horas de formación básica en búsquedas en Medline y después fueron aleatorizados en un grupo control, que no recibía nueva formación, y un grupo intervención.	Los integrantes del grupo intervención elegían un tutor con experiencia en búsquedas en MEDLINE y recibía formación individualizada durante la realización de sus 10 primeras búsquedas.	Las búsquedas 1., 4., y 8., de los participantes en el estudio se puntuaban a partir de la comparación con otras búsquedas (para la misma pregunta clínica) realizadas por un clínico experto y un bibliotecario. Se calculaba la proporción de citas relevantes por medio del número de citas relevantes de cada búsqueda dividido por el número de citas relevantes de las tres búsquedas posibles (participante, clínico experto y bibliotecario) y la «precisión» calculada como el número de citas relevantes recuperadas en una búsqueda dividido por el total de citas recuperadas en esa búsqueda. Se valoró también el número total de búsquedas realizadas por los integrantes de ambos grupos control e intervención.	Para el análisis estadístico de los datos se utilizaron los test de Wilcoxon y χ^2 .	El grupo intervención no realizaba más búsquedas que el grupo control y no se encontraron diferencias significativas en la calidad de las búsquedas. Ambos grupos mejoraban en su elaboración de búsquedas con un incremento medio significativo estadísticamente del número de referencias relevantes recuperadas. Los expertos recuperaban de forma significativa más citas relevantes en la primera búsqueda pero no en la 4.ª o en la 8.ª. Por tanto esta intervención no aumenta la calidad o cantidad de las búsquedas. Si se observa como después de una introducción básica en técnicas de búsqueda mejora la habilidad de los clínicos de ambos grupos.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
MacAuley et al, 1998 ³⁴	343 médicos generales	La muestra fue distribuida de forma aleatoria (utilizando un sistema informático de asignación de números aleatoria) en un grupo intervención (n = 118) y un grupo control (n =125).	A los participantes de ambos grupos se les enviaron los mismos tres artículos (seleccionados de la sección de medicina general de BMJ en 1995). Al grupo intervención se le adjuntaba información escrita sobre el método READER (Relevancia, Educación, Aplicabilidad, Discriminación y Evaluación global) de lectura crítica y se les pedía que evaluaran los artículos en base a dicho método. Al grupo control se les pedía calificar los artículos en base a su calidad científica e importancia global.	Los investigadores recogían y analizaban la evaluación crítica que los participantes realizaban de los artículos evaluados. Se analizaron también las condiciones laborales de cada participante, su formación previa, actividades docentes y hábitos de lectura clínica.	Se utilizó el test U de Mann-Whitney para comparar las diferencias entre los dos grupos y el análisis de la varianza de una vía. Kruskal-Wallis, para comparar los 3 grupos (intervención, control y la evaluación de los artículos por un experto). El paquete estadístico utilizado fue SPSS para Windows. La regresión logística fue usada para examinar posibles factores asociados. La repetibilidad del estudio se evaluó mediante los test de Wilcoxon y McNemar.	Los participantes que usan el modelo READER dan una puntuación global más baja y puntúan más apropiadamente la metodología de los estudios que los que usan la valoración libre. El método resulta preciso y reproducible y no existen factores intrínsecos que influyan en la puntuación (el modelo es apropiado para el uso de todos los médicos de familia independientemente de su edad, ámbito de trabajo, experiencia en enseñanza o formación y número de artículos que regularmente leen).
Rosenberg et al, 1998 ⁵⁵	108 estudiantes de primer curso de Medicina.	Los estudiantes se aleatorizaron en (tabla estándar de número aleatorio grupo experimental (n = 54), que recibiría la formación, y control (n = 54), que no la recibiría durante el período de estudio. Los integrantes de cada grupo fueron después nuevamente aleatorizados y asignados a uno de los dos escenarios clínicos sobre los que se trabajaría.	Los estudiantes del grupo experimental debían realizar al principio del programa búsquedas en la base de datos Medline para encontrar evidencia relevante que contestase al problema planteado en el escenario clínico al que habían sido aleatorizados. Asistían después a un taller de 3 horas de duración sobre formulación de preguntas y habilidades de búsqueda impartido por dos bibliotecarios (a grupos de 4 a 7 estudiantes) a lo largo de 10 semanas. A los tres meses de completado el programa, los estudiantes realizaban una nueva búsqueda en relación al caso clínico sobre el que antes no habían investigado. Los estudiantes del grupo control realizaban una búsqueda libre en relación a su caso clínico a la vez que la segunda búsqueda del grupo intervención.	El método de la búsqueda era evaluado en base a una hoja de resultados que analizaba la presencia de sus diferentes componentes y asignaba a cada búsqueda una puntuación final. Se evaluaban también los títulos y abstracts recuperados y se puntuaban en base al grado de evidencia. Por último, la herramienta disponía de un ítem que valoraba la satisfacción del estudiante (opinión sobre la utilidad de la intervención educativa). Todos las búsquedas fueron evaluadas por tres de los autores cegados al grupo al que pertenecía el estudiante y al momento de la búsqueda.	El análisis se hizo según intención de tratar. Los análisis se hicieron por separado para cada problema clínico. Se usaron los tests U de Mann-Whitney y el de Wilcoxon para comparar las medianas de las puntuaciones obtenidas.	Este ensayo establece que las sesiones docentes son una efectiva forma de mejorar las habilidades de búsqueda de los estudiantes con ganancias estadísticamente significativas en la puntuación de las búsquedas tras la formación. En cuanto a la calidad de la evidencia recuperada también existen mejoras significativas aunque sólo en uno de los escenarios clínicos. La ausencia de evidencia de la difusión de información desde los estudiantes del grupo experimental a los del grupo control sugiere la necesidad de impartir la formación a todos los estudiantes como parte de su plan de estudios.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
MacAuley y McCrum, 1999:56.	Médicos de Atención Primaria asistentes a un taller anual para Formadores de Residentes y participaron en la parte de lectura crítica (n = 99).	Los participantes se dividieron en 4 grupos siguiendo una selección aleatoria en dos fases: en la primera fase se seleccionaron por la antigüedad (docentes 5 años o más, docentes durante menor tiempo); en la segunda fase, de los 4 grupos se seleccionaron 2 para constituir el grupo control y los otros dos el grupo intervención.	Los grupos intervención se sometieron a una sesión de 30 minutos de duración sobre el método READER de lectura crítica y, posteriormente, evaluaron uno de los artículos seleccionados. Los grupos control también llevaron a cabo la evaluación crítica de un artículo pero sin recibir formación previa en lectura crítica. Se analizaban también las características personales y de práctica de cada participante.	Los investigadores recogían y analizaban la evaluación crítica que los participantes realizaban de los artículos evaluados y la información sobre las condiciones de la práctica clínica de los participantes; formación previa, actividades docentes y hábitos de lectura científica.	El test 'X' fue usado para comparar las características personales entre ambos grupos; las diferencias en la puntuación se examinó usando el test U Mann-Whitney.	Aquellos que usan el método READER atribuyen a los artículos una puntuación mas alta pero son más críticos con la metodología que los que usan la evaluación libre.
Stevermer et al, 1999:57.	59 residentes de familia.	Los residentes fueron aleatorizados en bloques de seis, estratificados por lugar y año de residencia en un grupo intervención (n =29) y un grupo control (n = 30).	Los residentes del grupo intervención recibían sesiones personalizadas de 15 minutos durante 2 semanas. En ellas los investigadores enfatizaban sobre la cuidadosa selección de artículos para su posterior lectura. Les enseñaban como hojear las revistas para seleccionar artículos orientados a resolver problemas clínicos; discutían cuales eran los problemas comunes de la práctica de los residentes e informaban de recursos que podían cambiar su practica clínica. Animaban, pero no enseñaban, a la lectura crítica...	Todos los residentes rellenaron un pretest y, 6 meses después de la intervención, un posttest. En cada test, los residentes recogían si conocían ese artículo y, en el caso de que lo conociesen, si habían leído sólo el resumen o el artículo entero. Los residentes que sólo habían leído el resumen contestaban dos preguntas: la primera valoraba la habilidad del residente para extraer los principales resultados del artículo a través del abstract; la segunda cubría otra información del resumen. De igual forma, los que habían leído el artículo completo respondían dos preguntas basadas en la sección de métodos. Un cuestionario breve unido al posttest valoraba los cambios en el volumen de lectura de los residentes.	Se utilizó el programa estadístico SPSS para Windows para calcular los tests: T de Students, 'X' y tests no paramétricos.	En el grupo intervención, tras la estrategia educativa, eran más los residentes que conocían los artículos seleccionados y también eran más los que leían los artículos enteros. No hubo diferencias significativas entre el pre y posttest en el grupo control. En ninguno de los grupos se mostraban cambios en la cantidad de tiempo que dedicaban a leer. Por tanto, una breve intervención sobre como mejorar la selección de artículos por los residentes incrementa el conocimiento de artículos relevantes, persistiendo le efecto al menos 6 meses tras la intervención.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Neville et al, 2000 ³⁶	Dos grupos de 100 estudiantes de primer año de medicina (de 2 años consecutivos) Cada grupo a su vez constituido por 18 grupos tutoriales de 5 o seis alumnos cada uno.	Cada uno de los alumnos fue aleatorizado, bisemanalmente y utilizando una tabla de números aleatorios, para recibir el artículo real o el que contenía los errores.	Los estudiantes habían recibido formación en lectura crítica al comienzo del curso y el estudio se realizó durante el último trimestre de dicho curso. Se seleccionaron 6 artículos por expertos en gastroenterología, hematología y endocrinología y se reescribieron introduciendo en ellos uno, dos o tres errores metodológicos lo suficientemente notorios para rechazar las conclusiones del autor. Cada artículo se relacionaba con un caso clínico.	Los estudiantes recibían un caso clínico y mostraban su grado de acuerdo con la opción de manejo terapéutico o intervención recibida (mediante preguntas con respuesta según una escala de Likert). Tras trabajar un artículo, los estudiantes volvían a recibir el mismo caso clínico y las mismas preguntas y además se les pedía que identificasen posibles errores metodológicos en el artículo leído.	Se utilizó el test T de Students para comparar las puntuaciones medias obtenido por los estudiantes. El análisis de la varianza se empleó para evaluar si existía correlación entre el tiempo de contestación de la escala (pre o post) y si el artículo que habían leído era el bueno o el que contenía errores. Se analizó y se comparó la proporción de errores identificados en los artículos.	El artículo sugiere que los estudiantes si son capaces de modificar sus decisiones clínicas después de la lectura crítica de un artículo científico.
Cabell et al, 2001 ³⁷	48 residentes de Medicina Interna de 6 rotaciones continuas de entre 6 y 8 semanas.	Los residentes fueron aleatorizados en bloques (constituidos por los 8 residentes que realizaban cada rotación) en un grupo intervención (n = 24) y otro control (n = 24).	Los residentes de ambos grupos, durante el tiempo que duraba la rotación, realizaban reuniones con el jefe de residentes (que había recibido previamente formación en MBE) para discutir los pacientes ingresados en relación al diagnóstico y la decisión terapéutica. Participaban también en sesiones semanales con un bibliotecario en las que aprendían las diferentes fuentes disponibles para localizar información relevante y trabajaban sobre distintas técnicas útiles para responder preguntas clínicas. El grupo intervención además, durante la primera semana de la rotación, en una sesión didáctica de 1 hora de duración sobre los componentes clave de una correcta pregunta clínica. Finalmente se les animaba a usar unas tarjetas recordatorio para construir preguntas relevantes durante cada ingreso y que servían de apoyo en la elaboración posterior de sus estrategias de búsqueda.	Se utilizó un cuestionario auto-contestable para recoger datos demográficos, uso personal de MEDLINE, barreras percibidas para la búsqueda y recuperación de artículos y auto-valoración de habilidades de búsqueda. A partir de la revisión de un subconjunto de búsquedas realizadas por el grupo intervención se diseñó una herramienta de evaluación mediante la cual se analizaron las preguntas y las estrategias de búsqueda realizadas por los residentes.	Las diferencias entre el grupo control e intervención fueron evaluadas por medio del test exacto de Fishers, para variables categóricas, y el de Wilcoxon, para variables continuas.	La herramienta de evaluación empleada es capaz de mostrar el impacto de la estrategia educacional empleada y el estudio muestra como el número de conexiones a MEDLINE, el tiempo dedicado a las búsquedas, el volumen total de búsquedas, el número total de abstracts revisados y el número total de artículos completos revisados fue significativamente más elevado en el grupo intervención que en el control. No se encontraron diferencias significativas en las barreras percibidas por ambos grupos para la búsqueda y recuperación de artículos. Por tanto, la intervención educativa implantada es capaz de incrementar considerablemente la actividad de búsqueda de los residentes.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medición	Métodos estadísticos	Resultados principales
Villanueva et al, 2001 ⁶⁰ .	Los participantes fueron clínicos, usuarios por primera vez de un servicio informático creado para dar respuesta a preguntas formuladas por profesionales de la salud.	Utilizando números aleatorios generados por ordenador los participantes que aceptaban participar en el estudio (N = 52) se aleatorizaron en dos grupos: grupo control (n = 29), que recibía el formulario estándar para construir preguntas clínicas, y grupo intervención (n = 23), que recibía instrucción simple en formulación de preguntas clínicas.	El grupo intervención recibía un formulario revisado que contenía una breve explicación sobre la importancia de una apropiada formulación, algunas instrucciones escritas y un ejemplo de cómo los elementos de la pregunta deben ser ordenados. Cada grupo proporcionaba dos preguntas, una basal con la que habían contactado con el servicio informático y otra tras recibir uno u otro formulario.	Fue evaluado el cambio en la proporción de preguntas reformuladas que contenían las 4 dimensiones que debe incluir una pregunta clínica adecuadamente construida (PCO). Se analizó además el porcentaje de preguntas en las que cada dimensión estaba descrita de forma global o específica. Los dos evaluadores estaban cegados respecto al grupo de asignación. Estos asesores puntuaban si estaba presente o ausente cada una de las 4 dimensiones y, si estaba presente, si estaba definido de forma global o específica. Los desacuerdos fueron resueltos con la ayuda de un tercer asesor.	Las características de los participantes se compararon mediante el test exacto de Fisher. El grado de acuerdo entre los evaluadores se evaluó mediante el uso del estadístico Kappa. Las diferencias entre grupos fueron evaluadas usando una tabla de contingencia, el procedimiento binomial exacto y el test de Wilcoxon.	En las preguntas basales no se encontraron diferencias significativas entre ambos grupos y tampoco las hubo al analizar cada dominio individualmente. Al valorar las preguntas formuladas tras recibir la instrucción se observó una reducción significativa del número de preguntas consideradas no específicas en el grupo intervención en contraste con la ausencia de cambio en el grupo control. Por tanto, este ensayo demuestra el impacto positivo de proporcionar instrucción específica en la proporción de preguntas clínicas adecuadamente construidas. Quedaría pendiente el evaluar el impacto a largo plazo de estos cambios.
Bradley et al, 2002 ⁶¹ .	Participaron en el estudio 10 residentes de distintas especialidades médicas.	Los residentes fueron asignados arbitrariamente a 4 equipos que se aleatorizaron en un grupo control (n = 5 residentes) y un grupo intervención (n = 5 residentes).	Durante el periodo de estudio los residentes planteaban preguntas clínicas, en relación a sus pacientes, y realizaban búsquedas en MEDLINE para responderlas. Los componentes del grupo control recibían supervisión de las búsquedas por bibliotecarios pero sin recibir instrucción. Cuando el residente formaba parte del grupo intervención el bibliotecario proporcionaba formación activa en técnicas de búsquedas bibliográficas.	Los residentes completaban unos exámenes antes y después de la intervención para medir sus opiniones sobre la satisfacción sobre MEDLINE y las búsquedas. Además realizaron 3 búsquedas estandarizadas antes, después y 6 meses después de la intervención.	Los bibliotecarios evaluaban las búsquedas mediante el análisis de la estrategia de búsqueda: construcción adecuada de las preguntas, uso apropiado de términos MeSH, aplicación lógica de límites y búsqueda de palabras clave. El equipo (3 bibliotecarios) revisaba y puntuaba cada búsqueda 2 veces para asegurar la consistencia de la evaluación.	El grupo intervención, en el examen post-intervención, formula mejor las preguntas, usa los límites más efectivamente y muestra más seguridad en el uso de MEDLINE. Sus actitudes hacia las búsquedas también son más positivas.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Bernai-Delgado et al, 2002 ⁶² .	Participaron en el estudio 24 equipos de Atención Primaria con un total de 158 médicos de familia.	Usando una tabla de números aleatorios los equipos fueron asignados a uno de tres grupos: experimental (participación en visita externa educativa basada en la evidencia = 8 médicos), placebo (participación en una sesión educativa convencional = 8 equipos, 54 médicos) y control (sin intervención = 8 equipos, 56 médicos).	Los médicos del grupo experimental participan en reuniones educativas en las cuales se informaba, por medio de información oral y visual, de las conclusiones y recomendaciones extraídas de una revisión sistemática previa sobre el uso eficiente de tres AINEs (diclofenaco, piroxicam y tenoxicam) de amplio uso en AP y escasa prescripción indicada. Es esta visita se explicaban además la metodología seguida en la revisión, basada en la evidencia, y se daba un folleto recordatorio con las recomendaciones. La visita era realizada por el farmacéutico de AP que previamente había recibido formación sobre los principios de la MBE y las herramientas básicas de diseminación de la información. El grupo placebo participaba en una sesión educativa convencional sobre AINEs y no recibía material escrito de apoyo. También un farmacéutico conducía estas sesiones pero no recibía formación en MBE ni tenía conocimiento de los resultados de la revisión sistemática. El grupo control no recibía información durante el periodo del estudio (4 semanas).	Se evaluó el cambio en los patrones de prescripción de los fármacos por medio de la medición del número de paquetes prescritos en los 6 meses previos a la intervención y en los 6 meses posteriores. También se evaluó el coste promedio por paquete en los mismos periodos.	Para la comparación de las características basales de los tres grupos se usó el test X ² para variables cualitativas el test F de Snedecor para variables cuantitativas. Para el análisis de los resultados principales se usó la reducción relativa en el número de prescripción de cada tipo de AINEs. Se usó la curva S o de Gompertz para representar los cambios en el coste de las prescripciones.	El estudio muestra como los patrones de prescripción de AINEs cambian en el grupo experimental como resultado de la visita externa educativa. Este cambio fue significativo comparado con el grupo control pero no cuando se comparaba con el grupo placebo, que participaba en una sesión educativa convencional. El cambio no se acompañó de diferencias en el volumen total de AINEs prescritos ni en una significativa reducción del coste promedio por prescripción.
Verstappen et al, 2003 ⁶³ .	174 médicos de Atención Primaria.	El conjunto de médicos fue distribuido en 26 grupos que, a su vez se aleatorizaron en 2 ramas. Cada una de estas ramas actuaría como control de la otra.	Durante 6 meses los médicos recibían consecutivamente 3 correos sobre 3 diferentes problemas clínicos junto con información concisa sobre guías de práctica clínica en relación a esos problemas. Tras cada contacto se realizaba, unas 2 semanas más tarde, una reunión de 90 minutos, supervisada por un coordinador, en las que se discutía sobre las guías clínicas y las dificultades para su incorporación a la práctica clínica.	Se utilizó un cuestionario en el que se proponían distintas pruebas diagnósticas a realizar ante los diferentes problemas clínicos que se habían presentado. Se evaluó la disminución del número de repuestas total por médico y la disminución en el número de tests diagnósticos definidos en las guías clínicas como inapropiados.	Las diferencias en las características individuales de los médicos fue evaluada por medio del test X ² . Los efectos de la intervención fueron analizados con el análisis de covarianza usando el número de repuestas durante el periodo siguiente como variable dependientes y el número de repuestas en la línea de base como variable independiente.	El estudio muestra una modesta disminución en el número de repuestas globales y en la elección de tests inapropiados pero los cambios apreciados no son significativos.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Cheng et al, 2003 ⁶⁴	544 clínicos (médicos, enfermeros y profesionales relacionados) que mostraron su interés en participar en un taller de búsqueda.	Fueron aleatorizados de forma estratificada (médicos, enfermeros y profesionales relacionados) y usando números aleatorios en dos grupos: intervención (n = 257), que recibía formación, y control (n = 287), que no la recibía durante el periodo de estudio.	La intervención educativa consistió en 20 sesiones docentes de 3 horas de duración impartidas a lo largo de un mes. En cada sesión, tras una parte teórica, se trabajaban 2 escenarios clínicos. Al finalizar el taller los participantes debían ser capaces de hacer una pregunta clínica contestable y transformarla en conceptos apropiados para la búsqueda, seleccionar la mejor evidencia desde diversas fuentes de información, entender los conceptos de búsqueda, aplicar filtros de búsqueda y conocer como «ponerse al día» de forma efectiva.	El instrumento, constituido por preguntas de respuesta múltiple y falso, fue diseñado para medir cambios en las habilidades y conocimientos en técnicas de búsqueda y en las actitudes antes y después de la intervención. El grupo control contestaba el test tras participar en el taller (post-test) y al grupo control se le ofrecía previamente a su participación (pre-test). La única diferencia entre el pre y post-test fue que en el pre-test fue omitida una pregunta que solicita exponer el grado de satisfacción con la intervención.	Para medir el efecto de la estrategia educativa se empleó un modelo de regresión logística que permite calcular la OR y el INNT. La OR fue usada para medir cuanto más se había beneficiado el grupo experimental, según su propia valoración y en cuanto a su uso de fuentes de información y bases datos, en relación al grupo control. El INNT aplicado a esta investigación se usó para predecir a cuantos clínicos era necesario formar para que obtuviese beneficio un clínico más.	La investigación muestran que la intervención es más efectiva que la no formación: - en mejorar las habilidades en formulación de respuestas clínicas; - en mejorar el conocimiento de la mejor evidencia; - en mejorar los conocimientos y habilidades en relación al proceso de búsqueda; - en la modificación de los comportamientos de búsqueda; - en mejorar los resultados de la búsqueda. El seguimiento (12 meses) de las búsquedas de información revela como las habilidades y conocimientos se van erosionando con el tiempo. Aumenta, sin embargo, el uso de fuentes electrónicas de información y disminuyen el uso de servicios tradicionales de biblioteca y el tiempo medio de sesión de búsqueda.
MacRae et al, 2004 ⁶⁵	Conjunto (N = 81) de cirujanos en activo de Canadá con acceso a Internet, e-mail y sin formación postgraduada en epidemiología clínica.	El conjunto final de participantes fue aleatorizado en un grupo control (n = 37) y otro intervención (n = 44). Se estratificó primero por el tipo de práctica (académica vs comunitaria) y después por el año de graduación.	A los sujetos del grupo intervención se les enviaba mensualmente, y durante un periodo de 8 meses, un paquete compuesto por un artículo clínico (de actualidad y de interés para la cirugía) y un artículo epidemiológico directamente relacionado. También se incluía una guía de lectura crítica. Tras un tiempo para trabajar el material, los sujetos participaban en una discusión sobre la metodología y recibían una revisión metodológica y clínica de cada artículo. El grupo control recibía los 8 artículos clínicos y la posibilidad de acceso on-line a la mayoría de las revistas médicas y quirúrgicas.	Al finalizar el periodo de estudio y tras la lectura de 2 artículos (similares en ambos grupos), todos los participantes completaron, un test, previamente validado y constituido por 54 ítems, sobre habilidades en lectura crítica	La puntuación media de los dos grupos fue comparada usando el test T de Student.	De los 55 cirujanos que completaron el examen, los sujetos del grupo intervención desarrollaron mejor el test de habilidades en lectura crítica que los del grupo control.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Acreditación	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Taylor et al, 2004 66.	De los 1305 profesionales de la salud invitados a intervenir en el estudio, aceptaron participar 145.	Un investigador independiente realizó la aleatorización, estratificada por la ocupación, en grupo control (n =72) y grupo intervención (n =73). Los investigadores que puntuaban los resultados del estudio estaban cegados respecto a la localización de los participantes.	La intervención educativa estaba basada en el Programa de Habilidades en Lectura Crítica (CASP). Consistió en un taller, didáctico y práctico, de 3 horas de duración y al menos una hora para preparar el material que se entregaba una o dos semanas previamente. El grupo control entraba en lista de espera para participar posteriormente en el taller.	Se utilizó un cuestionario validado (Taylor et al, 2001) para valorar: conocimiento de los principios de evidencia, actitudes hacia el uso de evidencia en la atención sanitaria, comportamiento respecto al uso de la evidencia, y percepción de la confianza en la evaluación de la evidencia. La habilidad para la evaluación crítica fue valorada mediante la valoración de la revisión sistemática.	Los test estadísticos usados fueron: - para variables continuas: modelo de regresión lineal multiple - resultados ordinales fueron comparados usando el test U de Mann-Whitney y test X ²	La formación en lectura crítica produce pequeña mejora en el conocimiento de los profesionales de los principios de la MBE y en algunos aspectos (valoración de resultados) de su habilidad para evaluar críticamente una revisión sistemática. No ocurren cambios en las actitudes o comportamientos de los participantes.
Johnston et al, 2004 67.	168 estudiantes de Medicina de 4º año	Los estudiantes fueron aleatorizados en 3 grupos. A lo largo de las tres rotaciones cada uno de los grupos utilizaba guías clínicas de bolsillo como apoyo a las decisiones clínicas, el software InfoRetriever para PDA o servía como control sin utilizar herramienta de apoyo.	Todos los estudiantes (en grupos de 20) intervenían en 2 sesiones de dos horas de duración al comienzo del estudio. En la primera sesión se revisaban los principios de la MBE y sobre cómo usarlos en la clínica. También recibían formación sobre el uso de la PDA y sus aplicaciones. En la segunda sesión se les pedía a los estudiantes aplicar las técnicas aprendidas mediante la construcción de preguntas clínicas, búsqueda y evaluación de la evidencia y aplicación de la información adquirida a la toma de decisiones sobre un caso clínico específico.	Se desarrolló un cuestionario de 41 ítems a partir de la revisión de la literatura. En el cuestionario se analizaba: la satisfacción con el uso de InfoRetriever; el grado de acuerdo con la funcionalidad de la PDA; la utilización de la PDA; la percepción de la utilidad de InfoRetriever; la percepción de las actitudes del tutor; la percepción de las habilidades informáticas; la estimación de la frecuencia de uso de InfoRetriever en diversas situaciones; la estimación del tiempo requerido para recuperar información clínica; la estimación de la proporción de búsquedas útiles. También se organizó un grupo focal de estudiantes para analizar sus sentimientos y experiencias con el uso de InfoRetriever.	La asociación entre dos variables fue analizada mediante los test X ² y ANOVA. La regresión logística se utilizó para evaluar la asociación entre múltiples variables.	El análisis cuantitativo y cualitativo muestra que los estudiantes encuentran la PDA y el software InfoRetriever como herramientas útiles. Sin embargo, están poco satisfechos con sus características funcionales y su utilización global es baja.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Bradley et al, 2005 ⁶⁸	3 cohortes (n = 175) de estudiantes de Medicina en su décimo semestre.	Cada cohorte fue aleatorizada en dos grupos (intervención directa, n = 90; intervención auto-dirigida, n = 85) mediante una tabla de número aleatorios. Los investigadores que puntuaban los resultados del estudio estaban cegados mediante el uso de códigos.	Consistió en 5 sesiones de media jornada durante las 2 primeras semanas del semestre. En estas sesiones se trabajaban los pasos básicos de la MBE. A los estudiantes de la intervención autodirigida se les animaba a ser autónomos en su aprendizaje y a elegir el tiempo, emplazamiento y velocidad del aprendizaje. Los estudiantes de la intervención directa seguían un modelo predeterminado de enseñanza basado en la teoría del aprendizaje social. Ambas intervenciones habían sido pilotadas previamente en una cohorte de estudiantes en 2001.	El cuestionario utilizado para medir conocimientos y evaluar actitudes hacia la MBE, se basó en el de Taylor et al sometido a doble traducción y revalidación. Para evaluar las habilidades en lectura crítica, cada uno de los alumnos valoraba críticamente un artículo durante un tiempo asignado de 60 minutos, y realizaba anotaciones en relación a diseño del estudio, metodología, resultados y generalización de los resultados a un escenario específico. Estas evaluaciones eran puntuadas por dos de los investigadores en base a unas respuestas previamente desarrolladas. La medición de los resultados se llevó a cabo a final de cada semestre.	Los resultados fueron comparados usando un análisis según intención de tratar. Se utilizaron los test T de studente, χ^2 y de Wilcoxon para datos no distribuidos normalmente. También se calculó el intervalo de confianza del 95% para la medias de los resultados y efectos de la intervención.	No se encontraron diferencias significativas entre los dos tipos de intervención en los tres resultados evaluados
Mukohara y Schwartz, 2005 ⁶⁹	Participaron en el estudio 107 internistas.	Usando un programa generador de números aleatorios, los participantes del estudio se dividieron en dos grupos: intervención (n = 54) y control (n = 53).	El grupo intervención recibía, vía e-mail semanal durante 3 meses, resúmenes evaluados críticamente de uno o dos artículos de la semana previa publicados en alguna de las 5 principales revistas médicas; el grupo control recibía, también por medio de e-mail semanales, una web de noticias sobre salud y comentarios de expertos.	Los participantes en el estudio completaban, antes y después de la intervención, un cuestionario a través de la web que medía su uso de la evidencia en la práctica, sus actitudes hacia el rol de la evidencia en la práctica, sus habilidades en evaluación crítica y sus hábitos de lectura. La parte del cuestionario que medía actitudes y auto-percepción de habilidades estaba constituido por ítems con respuesta según escala tipo Likert y se mostraba una fiabilidad adecuada a partir del cálculo de α de Cronbach. En el post-test los encuestados también evaluaban la utilidad de la intervención y grado de satisfacción en relación a ella.	Las características basales se compararon mediante los test T de Student y χ^2 . Estos test también se utilizaron para analizar las diferencias en la puntuación media del pre y post-test de ambos grupos.	Tras los 3 meses del periodo de estudio, los dos grupos no mostraban diferencias en el uso de la evidencia en la práctica. Los cambios en las actitudes hacia el rol de la evidencia en la práctica clínica y en la auto-percepción de su habilidad en la evaluación clínica fueron similares. Los hábitos de lectura mejoraron significativamente en el grupo control mientras que disminuyó el tiempo semanal dedicado a la lectura en el grupo intervención. La mayoría de los integrantes del grupo intervención la consideró útil y el 90% solicitó continuar recibiendo los e-mail. Sólo el 9% de los integrantes del grupo control estaba satisfechos con la intervención recibida.

TABLA 2. ENSAYOS CLÍNICOS CONTROLADOS ALEATORIZADO.

Autor	Población	Aleatorización	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Schilling et al, 200670.	238 estudiantes de 3 ^{er} curso de Medicina.	Los participantes se aleatorizaron en dos grupos: intervención (n = 134), que participaba en un taller que incluía formación on-line sobre MBE; control (n = 104), que participaba en un taller tradicional sin formación en MBE.	<p>Consistió en un curso interactivo on-line sobre MBE dentro de un taller de Medicina de Familia que tenía tres objetivos fundamentales:</p> <ul style="list-style-type: none"> - promover una efectiva recuperación de literatura biomédica. - incrementar el conocimiento y uso de relevantes guías clínicas - enseñar el cálculo e interpretación de medidas fundamentales dentro de la MBE: NNNT. <p>Todo este proceso se desarrollaba a lo largo de 5 semanas.</p> <p>Finalmente los estudiantes aplicaban la información recibida a situaciones reales y describían sus preguntas clínicas, proceso de búsqueda y hallazgos en una discusión moderada por un facultativo médico.</p>	<p>Todos los alumnos contestaban un pre-test sobre sus características demográficas y la percepción sobre sus habilidades clínicas, conocimiento de las funciones de búsqueda en MEDLINE y de los conceptos y metodología de la MBE.</p> <p>Tras el taller, contestaban un post-test en el cual describían su uso de la información y habilidades en MBE, evaluaban dicho taller y aplicaban los principios de la MBE aprendidos sobre un caso clínico.</p> <p>La actividad de búsqueda de los estudiantes fue evaluada de forma cegada e independiente por tres bibliotecarios que las puntuaban en función de la formulación de la pregunta clínica, la estrategia seguida, el uso de términos MeSH, la aplicación de operadores booleanos, la limitación de los resultados a ECAs o meta-análisis y el uso de diversas características y funciones del sistema OVID MEDLINE.</p> <p>Para finalizar los estudiantes debían seleccionar de la búsqueda realizada los dos artículos que considerasen más relevantes. Esta selección fue evaluada por dos de los autores. Se consideraba como excelente la selección de un ECA o un meta-análisis.</p>	<p>Las diferencias en las proporciones fueron evaluadas mediante el test χ^2 y las diferencias en las variables continuas fueron evaluadas usando el test t de Student.</p>	<p>Este estudio muestra como un curso on-line sobre búsquedas en MEDLINE basado en los criterios de la práctica basada en la evidencia puede conseguir literatura de mayor calidad, identificación de evidencia de más alta calidad y mejorar la seguridad en la recuperación de información y habilidades de análisis crítico. También muestra mejora en la aplicación de los principios de la MBE a un caso clínico al observarse como un mayor número de estudiantes del grupo intervención calculó correctamente el NNNT.</p>

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

Autor	Población	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Radack and Valanis, 1986 ⁷¹ .	Médicos de un hospital fueron designados a participar en el programa como grupo intervención (n = 22) y médicos de un hospital distinto fueron utilizados como grupo control (n = 12).	El programa consistía en sesiones semanales de 50 minutos durante 5 semanas realizadas a grupos de 4 a 7 médicos. Se distribuyó entre los asistentes un problema clínico, un artículo relevante y una revisión sobre criterios metodológicos pertinentes al tema. Los participantes valoraban la validez de la información del artículo y su aplicabilidad a la práctica clínica y después interpretaban la adecuación del uso de un test diagnóstico o del manejo terapéutico del tema tratado.	Se utilizó un cuestionario, con respuesta según escala Likert, para valorar la relevancia del curso en la práctica, la satisfacción con el método de enseñanza, el profesorado y la elección de artículos. Además, se empleó un test, antes y después de la intervención, en el cual se pedía evaluar un escenario clínico, la selección de métodos y resultados de un artículo científico y analizar la eficacia del test diagnóstico propuesto mediante el cálculo de su sensibilidad, especificidad y valor predictivo	No se especifican en el artículo los métodos estadísticos utilizados para el análisis de los datos.	Tanto en el grupo control como en el de intervención las puntuaciones del posttest mejoraron pero los diferencias entre ambos grupos no fueron estadísticamente significativas. La mayoría de los participantes calificó el material del curso como altamente relevante. No se encontró relación entre la valoración de la relevancia del curso y la mejora en la puntuación del test.
Bennett et al, 1987 ⁷² .	Estudiantes y tutores de dos hospitales docentes fueron asignados, de forma no aleatoria, al grupo intervención (11 tutores y 49 estudiantes); otros dos hospitales docentes actuaron como grupo control (11 tutores y 43 estudiantes).	Los tutores del grupo experimental recibieron formación (duración total entre 2-5 horas) por epidemiólogos clínicos con experiencia en enseñanza de lectura crítica. En estas sesiones se trabajaba la lectura crítica de artículos sobre diagnóstico y tratamiento y posteriormente desarrollaban paquetes educacionales que contenían: una guía de uso, un escenario clínico que mostraba un problema diagnóstico o terapéutico, un artículo relevante relacionado con el caso, y una guía para la evaluación de la utilidad de un test diagnóstico específico o la eficacia de un tratamiento específico. Se proporcionaba además a cada tutor la crítica del artículo hecha por los investigadores. Los tutores del grupo control recibían formación habitual como tutor. Tras la formación a los tutores los estudiantes recibían sesiones semanales, a grupos de 4 o 5 estudiantes, de una a dos horas de duración e implantadas durante 8 semanas.	Se desarrollaron nueve ejercicios que fueron aleatorizados entre los grupos control e intervención de forma que en el posttest cada estudiante recibiría un problema distinto al del pretest. En base al problema clínico presentado en cada ejercicio, los estudiantes debían tomar una posición, y defenderla por escrito, sobre el uso de un específico test diagnóstico o tratamiento. A los estudiantes del grupo intervención se les preguntaba también si la atención dada durante las tutorías a la lectura crítica había sido baja, moderada o alta. Estos trabajos fueron pretestados en estudiantes graduados en epidemiología clínica y mostraron ser capaces de discriminar entre diferentes niveles de experiencia metodológica. Los ejercicios fueron puntuados por investigadores no facultativos (cegados sobre si era un pre o posttest y sobre el grupo de pertenencia).	la media de cambio entre los pre y posttest se determinó para cada grupo de estudiantes usando el test T de Student no pareado. Se usó el test T pareado para comparar los pre y posttest dentro de cada grupo de estudiantes y el test X ² para comparar la proporción de estudiantes del grupo experimental y control que lograban mejorar sus habilidades en lectura crítica. El acuerdo interobservador e intraobservador en la puntuación de los tests fue calculado usando el índice Kappa.	El grupo experimental demostró un incremento significativo en sus habilidades para la lectura crítica con un incremento mayor en el ejercicio sobre diagnóstico. En el grupo control ambos resultados empeoraron en el posttest. Por tanto, los resultados de este ensayo sugieren que la formación especializada a tutores en el uso de material educativo basado en problemas y haciendo énfasis en la lectura crítica es efectiva para mejorar la habilidad de los estudiantes para evaluar críticamente la evidencia sobre tests diagnósticos e intervenciones terapéuticas.

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

Autor	Población	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Langkamp et al, 1992 ³ .	27 residentes de pediatría de 2 programas distintos: los de una institución participaron como grupo intervención (n = 14) y los de otra institución como grupo control (n = 13).	Al grupo intervención se ofrecieron 2 sesiones didácticas sobre diseño de investigaciones, epidemiología clínica y bioestadística, seguidas de 8 sesiones bibliográficas mensuales. Durante cada sesión bibliográfica 2 residentes presentaban un artículo y discutían sobre su contenido y metodología, bajo la supervisión de dos facultativos con formación en epidemiología y bioestadística.	Ambos grupos contestaba un pretest que recogía información sobre demografía, experiencia previa en epidemiología y bioestadística, y actitudes sobre ellas. El utilizado incluía además 23 preguntas sobre epidemiología clínica y bioestadística. Nueve meses después se administró un posttest (en el del grupo intervención se preguntaba sobre la asistencia a las sesiones bibliográficas y la percepción sobre su utilidad).	Se utilizó el test T de Student para comparar las medias en las puntuaciones del cuestionario. Para comparar las proporciones se utilizó el test χ^2 .	El estudio muestra como las sesiones bibliográficas no son un método efectivo para incrementar los conocimientos en bioestadística y epidemiología clínica al no detectar diferencias estadísticamente significativas entre ambos grupos. El estudio no muestra una correlación significativa entre la asistencia a las sesiones bibliográficas, el interés por la bioestadística y epidemiología clínica y los cambios en la puntuación.
Frasca et al, 1992 ⁴ .	Participaron en el estudio 92 estudiantes de Medicina de tercer año (44 en el grupo control y 48 en el grupo intervención).	Los estudiantes recibieron una sesión didáctica semanal de 1,5 horas de duración durante 10 semanas. Los docentes enseñaban a evaluar la metodología de la investigación y el análisis estadístico de los artículos; los bibliotecarios enseñaban técnicas de búsqueda. Las primeras 2 sesiones eran teóricas y las restantes 8 sesiones prácticas.	Durante la primera y segunda semanas ambos grupos realizaban un examen en el cual, mediante escala de Likert, evaluaban el tiempo dedicado a búsquedas bibliográficas y la frecuencia de evaluación crítica de un artículo. Durante la última semana a todos los estudiantes se le administraba un posttest de 20 preguntas basadas en una situación práctica (13 preguntas estaban centradas en los objetivos de la evaluación crítica y 7 en la destreza en búsqueda bibliográfica).	Las diferencias en la puntuaciones medias de ambos grupos en el pre-test fue evaluada mediante el test T de Student. También se utilizó este test para el análisis de las diferencias en el número de respuestas correctas en el post-test.	Los resultados indican que los integrantes del grupo intervención realizan mejores búsquedas y tienen más habilidad para la evaluación crítica que los del grupo control.
Landry et al, 1994 ⁵ .	Estudiantes de Medicina de 3 ^{er} año (n = 146), distribuidos en grupo control (n = 81) e intervención (n = 65).	La intervención consistió en 2 seminarios interactivos de 90 minutos con una separación semanal entre ellos. El primer seminario se centraba en los tipos de artículos comúnmente usados en la literatura médica y en los tipos de diseño de estudios. El segundo seminario enfatizaba en la lectura crítica.	A los 2 grupos se les administró un test con 10 ítems al inicio y final de la intervención. El test fue desarrollado por consenso por los autores y testado en facultativos con experiencia en MB, aunque en el artículo no se especifica el proceso de validación. El test incluía preguntas sobre diseño de estudios y estadística básica y resúmenes de estudios donde el estudiante debía identificar el diseño del estudio, el gold standard en artículos sobre diagnóstico, el diseño óptimo en artículos de tratamiento y el significado de la sensibilidad. El cuestionario indagaba además en los datos demográficos de los estudiantes y sus actitudes y utilización de la literatura médica.	Mediante una revisión ciega (en cuanto al grupo y el momento de la realización) se valoró: a. número de referencias usadas; b. tipo de literatura referenciada; c. si las referencias estaban también citadas en el cuerpo del relato; d. si se había comentado la calidad del artículo. Para la comparación de las diferencias entre pre y posttest de ambos grupos se usó el test T apareado y el χ^2 para datos categóricos.	Los 2 seminarios interactivos mejoran los conocimientos y actitudes de los estudiantes respecto al grupo control pero no incrementa el uso de la literatura.

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

Autor	Población	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Green and Ellis, 1997 ⁷⁶ .	La población de estudio se constituyó por residentes de 2-3 ^{er} año de medicina interna (grupo intervención = 19; grupo control = 15).	Sesiones tutoriales, semanales, de 1 hora de duración dirigidas por un residente y guiadas por un facultativo. En ellas, basándose en un escenario clínico real, se trabajaban las distintas habilidades de la MBE.	Antes y después del periodo de estudio, ambos grupos completaban un cuestionario sobre comportamientos en MBE, y un autovaloración de competencias en MBE y un test sobre habilidades en MBE.	Los cambios post-pretest fueron valorados por el test T de Students para datos pareados y el test 2. Los cambios en la puntuación media del test sobre habilidades se compararon usando el test T de Student y se desarrolló un análisis estratificado para valorar los efectos de confusión por diferencias en las características basales.	Los participantes del grupo intervención incrementan su uso de artículos originales para contestar preguntas clínicas, el examen de las secciones métodos y resultados de los artículos y su auto-valoración en evaluación del diseño de un estudio, evaluación estadística y aplicación de resultados a pacientes. Mejoran además las puntuaciones del grupo control en el test de habilidades. El grupo control no muestra diferencias.
Fu et al, 1999 ⁷⁷ .	Se constituyó un grupo experimental formado por 12 residentes de Psiquiatría y un grupo control constituido por otros 12 residentes de distinto hospital.	El grupo experimental participó en sesiones bibliográficas semanales de 1.5 horas de duración, durante 12 semanas. En cada sesión bibliográfica los residentes, seleccionaban un artículo de relevancia clínica y posteriormente procedían a su evaluación crítica. El grupo control no participó en una experiencia similar durante este periodo.	a. Cuestionario que evaluaba el uso actual de la MBE y las opiniones sobre su práctica; b. Cuestionario de 19 ítems con respuesta múltiple para evaluar la habilidades en la valoración crítica de un artículo; c. 7 casos clínicos para evaluar los conocimientos en las decisiones terapéuticas, pronóstico de la enfermedad, fisiopatología básica y epidemiología, el residente indicaba la fuente de conocimiento utilizada. Los instrumentos se utilizaban al inicio de las sesiones y a las 12 semanas de finalizarlas.	Para el análisis estadístico de los datos fueron usados el test T de Student para muestras independientes y el test U de Mann-Whitney. Las correlaciones entre la motivación del residente, su rendimiento y la asistencia a las sesiones fue valorada mediante el análisis de correlación de Pearson.	Las sesiones bibliográficas no son un método efectivo para la enseñanza de lectura crítica a los residentes de psiquiatría. Las ganancias en habilidades de lectura crítica son mínimas o inexistentes. Tampoco se aprecia un incremento del uso de la literatura. Por tanto se concluye que esta técnica no es efectiva durante la residencia.
Bazarian et al, 1999 ⁷⁸ .	32 residentes de medicina de emergencias (16 en grupo control y 16 en el de intervención).	Los residentes del grupo intervención participaron en sesiones bibliográficas basadas en un enfoque de MBE y centradas en la lectura crítica, de una hora de duración, mensualmente durante un año. El grupo control fue sometido a tradicionales sesiones bibliográficas también mensuales.	A ambos grupos se les proporcionaba un artículo f1icticio antes y después del periodo de estudio. Este artículo había sido creado a propósito con 17 errores metodológicos (7 considerados errores mayores y 10 menores). Tras la lectura se les preguntaba si usarían el fármaco del que hablaba el artículo para la patología en cuestión y que dieran 5 razones para apoyar su decisión. Del texto libre de los residentes se extralan los posibles errores metodológicos que se puntuaban según fuesen errores mayores o menores. También se analizaba la experiencia investigadora previa, la formación epidemiológica y postdoctoral y los hábitos de lectura.	Las respuestas fueron valoradas por 3 revisores independientes (cegados respecto al autor y el momento de la contestación). Se calculó la media y la mediana de las puntuaciones de los pre y posttest y se compararon los cambios en la medias mediante el test de Wilcoxon.	No se encontraron mejorías significativas en habilidades en literatura crítica, hábitos de lectura y uso de literatura médica en la práctica clínica en el grupo intervención, comparado con el grupo control.

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

Autor	Población	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Smith et al, 2000. ⁶	Grupo intervención = 18 residentes de 5º año de Medicina Interna; grupo control = 37 residentes.	El curso de MBE consistía en 2 conferencias semanales de 1 hora de duración realizadas durante 7 semanas consecutivas. El curso se centraba en diagnóstico y tratamiento. La primera sesión semanal era un seminario didáctico y en la segunda sesión se aplicaba lo enseñado utilizando un caso clínico. Este curso fue aplicado al grupo intervención al inicio del estudio y a los tres meses lo recibió el grupo control.	Cuestionario que evaluaba habilidades en MBE, actitudes y comportamientos de los residentes. Este cuestionario fue realizado por todos los residentes antes de la intervención educativa, tras ella, y a los 6 y 9 meses del primer curso (3 y 6 meses de la aplicación del curso al grupo control).	El análisis de los datos se realizó mediante el test U de Mann-Whitney, para datos ordinales, y el test X ² para datos dicotómicos. Todos los análisis fueron por intención de tratar. Se utilizó el paquete estadístico SPSS 8.0.	La herramienta de evaluación empleada muestra sensibilidad para detectar el impacto de la intervención educativa: las puntuaciones del grupo experimental mejoran significativamente en formulación de preguntas, búsqueda de literatura y comprensión de términos. Al aplicar el test a los 6 y 3 meses (grupo experimental y control) son mejores las puntuaciones del grupo control. Al administrar el test a los 9 y 6 meses en ambos grupos persisten similares resultados y no hay diferencias entre ellos.
Ghali et al, 2000. ⁷	Estudiantes de tercer año de Medicina distribuidos en grupo intervención (n = 34) y grupo control (n = 26).	Mini-curso de MBE consistente en cuatro sesiones semanales de 90 minutos de duración. En la primera se instruyó sobre como aprovechar los recursos de MEDLINE y en las otras 3 se realizaban búsquedas en relación a preguntas clínicas planteadas. El grupo control recibía sesiones didácticas tradicionales.	Inmediatamente antes y después del curso ambos grupos contestaban un examen con respuestas basadas en escala Likert mediante el cual autocaloraban sus hábitos de lectura/búsqueda de literatura, y sus habilidades y actitudes en relación con la MBE.	Se utilizó el test de Wilcoxon para comparar las puntuaciones de ambos grupos con respecto a su puntuación basal. Se realizó un análisis de regresión múltiple lineal para asegurar que las diferencias entre ambos grupos no eran atribuibles a diferencias basales.	En el grupo intervención se aprecia un incremento significativo en la autovaloración de las habilidades en lectura crítica y en el uso de MEDLINE con respecto al grupo control. También mejora en este grupo la tendencia a usar MEDLINE, y artículos originales para resolver problemas clínicos.
Ross y Verdick, 2003. ⁸⁰	Residentes de Medicina de Familia de dos hospitales distintos, distribuidos en grupo intervención (n = 18) y grupo control (n = 30). La participación en el estudio fue voluntaria.	Taller de 10 sesiones semanales sobre MBE El grupo control no recibía formación especial en MBE.	Se grabó en audio la interacción residente- preceptor durante las consultas externas antes y 6 meses después de la puesta en marcha del plan de estudios (se evaluaban las expresiones relacionadas con búsqueda de literatura, epidemiología clínica y lectura crítica). Se utilizó además un test de 50 ítems de respuesta múltiple que contestaban los residentes a libro abierto antes y después del periodo de estudio.	Se utilizó el test T de Student para comparar las medias en las puntuaciones del cuestionario. Para el análisis de los resultados de la extracción de conceptos y términos los test χ^2 de Pearson y el test exacto de Fisher.	Los resultados pre-intervención fueron similares, n los post-intervención las puntuaciones del grupo experimental mejoraban significativamente. n la cinta de audio del grupo intervención se incrementaba significativamente el uso de términos de MB tanto en los residentes como en los facultativos. Se establece también con el estudio que tanto el test empleado como el análisis de las interacciones son sensibles para detectar cambios tras una estrategia educacional.

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

Autor	Población	Intervención educativa	Instrumento de medida	Métodos estadísticos	Resultados principales
Aki et al, 2004 81.	Grupo intervención = 17 estudiantes y residentes de medicina interna; grupo control = 23 estudiantes y residentes. La asignación a los grupos no se realizó de forma aleatoria pero se utilizó la aleatorización para asignar a los participantes una de las dos series que constituyen el test.	El grupo intervención participaba en una rotación electiva de 2 semanas de duración durante las cuales el residente trabajaban los principios y métodos de MBE y aprendían a transmitir estos conocimientos al resto del equipo (sesiones de lectura crítica basadas en un problema clínico).	Fue utilizada la versión en inglés del cuestionario de Berlin (validado). Este cuestionario fue aplicado, a ambos grupos, antes y después de la intervención educativa.	Se utilizó el test exacto de Fisher para comparar las características basales de ambos grupos; el de Student fue usado para comparar las puntuación media de los tests entre ambos grupos; el análisis de la varianza se usó para comparar la puntuación media de ambos grupo ajustada a otras variables (orden de administración del cuestionario, nivel de formación y habilidades en MBE en el pretest); se realizó análisis de regresión múltiple (puntuaciones de pre y posttest como variables dependientes y las características basales como variables independientes) para determinar si alguna de las características basales afectaba al rendimiento del pre o posttest.	En el grupo intervención, los conocimientos mejoraron ligeramente en el posttest pero no significativamente. La puntuación media en el grupo control decreció en el posttest con respecto al pretest. La diferencia en los cambios de puntuación entre los dos grupos fue significativa (mejoría de conocimientos en el grupo intervención) incluso después de ajustarla por covariantes.
Grad et al, 2005 82.	37 residentes de 5º año en 4 emplazamientos distintos. Los emplazamientos A y B se constituyeron como grupo intervención (n = 20); los emplazamientos C y D se consideraron grupo control (n = 17).	Los integrantes del grupo intervención recibieron una PDA (Personal Digital Assistant), un curso de MBE y tres software sobre MBE para la PDA (incluyendo InfoRetriever: sistema de apoyo para la toma de decisiones y recuperación de información clínica en bases de datos). Los residentes del grupo intervención también recibían formación en el uso correcto de InfoRetriever. Los residentes del grupo control recibieron las mismas herramientas excepto InfoRetriever y el curso de MBE.	Se desarrolló un test de conocimiento constituido por 30 preguntas sobre hipertensión, diabetes, extraídas a partir de una revisión de la literatura. Incluía además 5 preguntas en las que se valoraba la habilidad para estimar la probabilidad de faringitis estreptocócica y enfermedad coronaria a partir de la información proporcionada en un caso clínico y el uso de InfoRetriever. La validez fue calculada a partir de la habilidad del test para discriminar entre 3 grupos de médicos con diferentes niveles de conocimiento (médicos de familia = 105, residentes de familia de 5º año = 37 y estudiantes de Medicina de 3º año = 24). La fiabilidad fue estimada mediante el coeficiente alpha de Cronbach. Ambos grupos respondían el test antes de la intervención y 8 meses después. Tras el post-test, los residentes del grupo intervención volvían a contestar el test usando InfoRetriever para intentar mejorar la puntuación.	Se usó el test T de Student para comparar las puntuaciones medias entre el pre-test y el primer post-test. Las diferencias entre las puntuaciones medias del primer post-test entre el grupo control y el intervención se expresaron en diferencias absolutas. También se utilizó el test T de Student para comparar el número de estimaciones del riesgo correctas entre el primer y segundo post-test.	No se encuentra efecto de la intervención al comparar el pre-test con el primer post-test en ambos grupos pero sí hay un aumento significativo de la puntuación del segundo post-test (cuando se usa la PDA como apoyo) respecto al primero.

TABLA 3. ENSAYOS CLÍNICOS CONTROLADOS NO ALEATORIZADOS.

<p>Autor Gruppen et al, 2005⁸³.</p>	<p>Población 92 estudiantes de 4º año de Medicina distribuidos en un grupo intervención (n = 34) y un grupo control (n = 58).</p>	<p>Intervención educativa Intervención educativa de 2 horas sobre técnicas basadas en la MBE para buscar la evidencia en MEDLINE a partir de problemas clínicos.</p>	<p>Instrumento de medida Los autores realizaron una valoración pre-intervención de las habilidades de búsqueda y una nueva evaluación un mes después de la intervención. La calidad de las búsquedas fue analizada y puntuada por bibliotecarios médicos.</p>	<p>Métodos estadísticos</p>	<p>Resultados principales Los estudiantes del grupo intervención tuvieron menos errores de búsqueda y búsquedas de calidad más alta que los integrantes del grupo control. Los errores de búsqueda más comunes fueron la falta de términos MeSH, falta de límites apropiados, fracaso en la búsqueda de mejor evidencia, e inapropiada combinación de todos los conceptos de búsqueda. Por tanto, una breve sesión formativa en habilidades de búsqueda puede tener marcados efectos beneficiosos en la calidad de las búsquedas realizadas con posterioridad.</p>
---	--	---	--	------------------------------------	---

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validad	Fiabilidad
Cuestionario de Linzer et al, 1988 ⁵² .	<p>44 residentes de Medicina Interna fueron asignados aleatoriamente a una sesión bibliográfica de medicina general (grupo intervención, n = 22) o a conferencia estándar sobre temas de medicina ambulatoria (grupo control, n = 22). Para la validación del cuestionario un panel de expertos en epidemiología y bioestadística y una cohorte de residentes de medicina de familia.</p>	<p>El cuestionario contaba con 4 partes: 1. Preguntas sobre características previas (formación en investigación, lectura crítica...); 2. Hábitos de lectura y autovaloración de las habilidades para leer críticamente un artículo; 3. Conocimientos de epidemiología y bioestadística. Esta sección incluía 15 preguntas que fueron seleccionadas del banco de preguntas sobre epidemiología y bioestadística de la Escuela de Medicina Monte Sinai (basándose en el trabajo de Emerson y Coiditz); 4. Habilidades en lectura crítica (a partir de la lectura crítica de un artículo y puntuando en base a unos criterios gold standard desarrollados por consenso). El cuestionario fue administrado antes y después de las sesiones educativas.</p>	<p>La validez de contenido fue valorada por un panel de expertos. Se utilizó el método Delphi para conseguir un consenso sobre el formato inicial.</p> <p>La validez de constructo fue valorada a partir de la administración del cuestionario al panel de expertos y la cohorte de residentes. Fue encontrada adecuada discriminación entre los diferentes niveles de experiencia metodológica. El cuestionario también muestra sensibilidad para detectar cambios antes y después de la intervención educativa.</p>	<p>Para medir la variabilidad (fiabilidad) interobservador, una fracción de los cuestionarios fue puntuada por otros evaluadores. Para medir la fiabilidad intraobservador, los evaluadores puntuaban varios cuestionarios mas de una vez.</p>
Cuestionario de Stern et al, 1995 ⁵⁴ .	<p>62 residentes de medicina interna (contestaron el cuestionario 28 residentes, 45%; panel de expertos compuesto por 5 médicos con experiencia en evaluación de artículos.</p>	<p>El cuestionario estaba compuesto por 11 ítems con respuestas según escala tipo Likert de 6 puntos y fue diseñado en base a los criterios de McMaster para evaluación de un artículo de tratamiento. Junto con el cuestionario se distribuyó un único artículo sobre un nuevo tratamiento omitiendo las secciones de resumen, introducción y discusión. Las respuestas correctas fueron establecidas tras administrar el mismo artículo y preguntas a un panel de expertos compuesto por 5 médicos con experiencia en la evaluación crítica. El instrumento analizaba la autorvaloración de los residentes de su competencia en evaluación de la investigación a través de la comparación de sus respuestas con las del panel de expertos. Incluyó además preguntas sobre el año de formación y la experiencia previa en lectura crítica.</p>	<p>No se encuentran comentarios sobre la valoración de la validez de criterio</p>	<p>No se realiza medición de la fiabilidad.</p>

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validez	Fiabilidad
Cuestionario de Green y Ellis, 1997 ⁶ .	19 residentes de medicina interna de 2º y 3º año que recibieron un programa de formación de 7 sesiones semanales sobre MBE; 15 residentes de otra rotación en el mismo periodo de tiempo participaron como grupo control.	El cuestionario estaba constituido por 25 ítems que evaluaban 5 áreas: MBE, formación previa en lectura crítica, autovaloración de las competencias en MBE, hábitos de lectura médica y preferencia de fuentes de información. La parte que evaluaba las habilidades en MBE se extrajeron del cuestionario de Stern et al (103) y requería una respuesta libre a una pregunta basada en un caso clínico (fiabilidad en la aplicación de la literatura médica en la toma de decisiones) y un artículo relacionado.	En este estudio no fue valorada la validez del instrumento de evaluación. El cuestionario es sensible para detectar el impacto de la intervención educativa.	El test de habilidades en MBE muestra una buena fiabilidad (inter-observador) medida a partir de las concordancias entre las puntuaciones de los dos evaluadores (el grado de acuerdo fue de 2 puntos en el 98% de los test y de 1 punto en el 67% de los test). El coeficiente de correlación entre ambos evaluadores fue 0.87.
Test de McAlister, 1999 ⁸⁵	Se utilizó muestra de 521 médicos. Se excluyeron aquellos médicos que dedicaban todo su tiempo a la administración/ investigación. Respondieron 296 de los 495 seleccionados (de la muestra inicial 26 médicos habían fallecido) y, tras excluir 2 cuestionarios incompletos, el tamaño final de la muestra fue de 294.	El test se desarrolló partiendo de una revisión de cuestionarios previamente publicados en guías de práctica clínica. El pilotaje (pretest cognitivo) se realizó con 5 internistas que determinaron la aceptabilidad y claridad del cuestionario y confirmaron su validez facial. La herramienta final, además de preguntas sobre características demográficas, incluía preguntas que valoraban el uso de la MBE en la práctica clínica, las opiniones sobre la MBE y sobre las limitaciones de su uso y el uso de fuentes de información. Las respuestas fueron diseñadas según una escala tipo Likert de 5 puntos.	Validez facial: medida durante la fase de pretest cognitivo.	No se realiza medición de la fiabilidad.
			No se especifica evaluación de la validez de constructo.	Se evaluó la validez de criterio a través de las asociaciones esperadas entre las respuestas del cuestionario y otras variables (cálculo de coeficientes de correlación). Por ejemplo, la autopercepción del uso de MBE estaba fuertemente asociada al uso de artículos de investigación primarios y la autoconfianza en lectura crítica.

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validad	Fiabilidad
Cuestionario de Khan et al, 1999. ⁹	Ocho clínicos participaron en el programa educativo. El cuestionario se aplicó al inicio y 4 meses después de concluir la intervención educativa.	El cuestionario, que valoraba conocimientos, incluía 20 ítems que cubrían diferentes aspectos de la MBE (valoración del diseño de un estudio, evaluación de sesgos, evaluación de test estadísticos, valoración del valor general y la aplicación clínica de un artículo). Había además 5 ítems control que no se cubrían en el programa educativo.	No se especifica evaluación de validez de constructo o de criterio.	Se comprobó la fiabilidad del cuestionario mediante el método test-retest (coeficiente de correlación de Pearson = 0.9)
Cuestionario de Taylor et al, 2001. ⁸⁶	20 profesionales de la salud participaron en el pilotaje (pretest cognitivo); 152 profesionales sanitarios agrupados en 4 grupos, en base a su formación previa en MBE, intervinieron en la validación.	Para la construcción del cuestionario los ítems fueron identificados a partir de una revisión bibliográfica sistemática y un cuestionario previo sobre MBE no validado. Se obtuvieron 11 preguntas de respuesta múltiple sobre conocimientos (utilizaban como respuesta las categorías verdadero, falso y no sabe no contextual) y 14 cuestiones para la valoración de actitudes (con respuestas diseñadas según una escala tipo Likert de 5 puntos). Posteriormente, 20 profesionales con distintos grados de experiencia en MBE, trabajaron sobre el formato global del cuestionario y sobre preguntas específicas que podrían resultar problemáticas. El cuestionario definitivo contó con 7 preguntas sobre actitudes y 6 sobre conocimientos.	<p>Validez de constructo:</p> <ul style="list-style-type: none"> - análisis de la validez discriminante: comparación de la media de puntuación de conocimientos y actitudes entre el grupo sin experiencia en MBE y el grupo de expertos en MBE. El análisis estadístico muestra diferencias significativas entre ambos grupos. - sensibilidad (o capacidad del instrumento para detectar el impacto de una intervención educativa) se mide por medio de la comparación de los resultados obtenidos en el grupo sin experiencia en MBE y aquellos que han recibido formación en MBE. Los resultados muestran de moderados a buenos niveles de sensibilidad para conocimientos (0.48, 0.69) y actitudes (0.72, 0.43). 	Para su determinación se midió la consistencia interna, mediante el coeficiente de Cronbach, y el coeficiente de correlación de Spearman. Se obtuvo un valor satisfactorio de alpha de Cronbach para las preguntas relacionadas con el conocimiento y actitudes (0.72 y 0.64 respectivamente). El coeficiente de correlación de Spearman para cada pregunta individual frente a la puntuación total de las preguntas osciló entre 0.12 y 0.66.

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Questionario	Participantes	Características y desarrollo del test	La validez facial fue conseguida mediante preguntas inequívocas, sin ambigüedad. La validez de contenido fue lograda a través de la inclusión de preguntas que incorporaban conceptos clave de la MBE y epidemiología a clínica.	No se hace referencia en el artículo a la evaluación de validez de constructo o criterio.	Validez	Fiabilidad
Questionario de Cramer, 2001 ¹¹	Residentes de medicina de familia (n = 35), de 2 y 3º año que participaron en un conjunto de Journal Clubs (sesiones bibliográficas reestructuradas para introducir, reforzar y reevaluar el conocimiento de los residentes de los conceptos de la MBE.	Los test (utilizados en cada Journal Club) consistían un conjunto de 10-12 preguntas centradas en los principios de la MBE y la epidemiología clínica así como en los contenidos clave de los artículos. Los ítems fueron preparados por el director del Journal Club y aparecían tanto en el pre como en el posttest.	La validez facial fue conseguida mediante preguntas inequívocas, sin ambigüedad. La validez de contenido fue lograda a través de la inclusión de preguntas que incorporaban conceptos clave de la MBE y epidemiología a clínica.	No se hace referencia en el artículo a la evaluación de validez de constructo o criterio.	Validez	La fiabilidad fue equiparada a la repetibilidad o estabilidad de las respuestas en repetidas administraciones (método test-retest).
Questionario de Berlín, 2002 ⁴	En el estudio participaron 43 expertos en MBE y 20 estudiantes de 3º año de medicina (grupo control) y 203 estudiantes de medicina que realizaron 3 cursos sobre MBE que tuvieron lugar desde 1999 a 2001/02 estudiantes curso A, 50 curso B y 71 curso C).	Cinco profesores de MBE redactaron una primera serie de preguntas múltiples. Las preguntas se construyeron partiendo de escenarios clínicos típicos y enlazándolos con los estudios de investigación publicados. A través del método Delphi modificado se revisaron las preguntas y las discrepancias se resolvieron por consenso. Posteriormente las preguntas se pilotaron (pretest cognitivo) en un grupo de 5 médicos voluntarios no familiarizados con MBE. Al final de la primera etapa el instrumento consistía en dos series de 15 preguntas de respuesta múltiple, con similares contenidos aplicados a diferentes casos. Las preguntas del cuestionario evaluaban conocimientos en la interpretación de la evidencia y habilidades en el diseño de preguntas clínicas, en el de estrategias de búsqueda y en el uso de la información para resolver un problema clínico.	La validez facial se midió durante la fase del pretest cognitivo.	Validez de constructo: se valoró la capacidad del test para discriminar entre grupos con diferente nivel de conocimiento: expertos versus participantes del curso (antes de este) versus estudiantes y participantes de los diferentes cursos antes de estos entre ellos (análisis de la varianza con método Scheffe para comparaciones post hoc). El cuestionario también se muestra sensible para detectar el impacto de la intervención educativa.	No se especifica evaluación de la validez de criterio.	La equivalencia entre las dos series fue determinada a través del cálculo del coeficiente de correlación (tras aplicarlas al grupo control). Por medio del coeficiente de correlación de Pearson se estableció la correlación entre un ítem individual y el instrumento como un todo omitiendo ese ítem (correlación ítem-total) para identificar ítems que contribúan a una baja fiabilidad. Para la determinación de la consistencia interna se calculó el coeficiente de Cronbach. Los resultados obtenidos fueron 0.75 para la primera serie y 0.82 para la segunda.

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validez	Fiabilidad
Cuestionario de Beasley 87	Contestaron el cuestionario 177 médicos con actividad asistencial y 22 facultativos con actividad únicamente docente; 7 médicos de familia participaron en el pilotaje del test.	El cuestionario utilizado fue diseñado por los autores 2 años antes para evaluar el rendimiento de estudiantes de medicina tras un curso de MBE. Contaba con 4 secciones principales: una primera sección destinada a describir las características personales y profesionales; una segunda donde se preguntaba sobre la importancia de diferentes conceptos y términos sobre MBE en la práctica diaria y se evaluaba el entendimiento de una lista de conceptos y términos sobre MBE (escala de Likert de 5 puntos); en la tercera sección se analizaba la incorporación de la MBE en la formación médica continuada y actividades docentes; y en la última sección se evaluaban los conocimientos en MBE. El cuestionario fue pilotado en un grupo de 7 médicos de familia y modificado en función de sus sugerencias.	En la fase de pretest por los autores se evaluó por medio de análisis factorial. La validez de constructo se evaluó por medio de análisis factorial. No se especifica la medición la validez de criterio	Se usó el coeficiente de Cronbach para cuantificar la fiabilidad de cada sección y valorar la contribución de cada pregunta a la fiabilidad global de cada sección.
Cuestionario de Green et al, 2002 88.	El cuestionario fue pilotado en primer lugar en una muestra de 112 médicos de familia. Una segunda versión que fue testada en 1217 clínicos (234 médicos de familia, 848 internistas y el resto compuesto por facultativos de diferentes especialidades) y los resultados de esta segunda versión guiaron la preparación de una tercera que fue probada en 64 médicos de familia.	Para elaborar el instrumento los investigadores elaboraron un pool de preguntas que fue refinado por médicos en activo del comité de la Academia Americana de Medicina de Familia. El instrumento final contaba con 17 ítems que hacían referencia a tres factores subyacentes: - evidencia científica, más que experiencia clínica o autoridad, es percibida como la mejor fuente de conocimiento para una buena práctica. - grado de conformidad con prácticas clínicas alejadas de las prácticas locales o recomendaciones de líderes. - importancia del manejo de la carga asistencial mientras se mantiene la satisfacción del paciente (practicabilidad)	El valor de Cronbach fue calculado para cada uno de los 3 factores de la escala y en cada una de las tres interacciones. El valor de Cronbach osciló entre 0.48 y 0.79. No se especifica evaluación de validez del cuestionario.	

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validez de constructo: la capacidad discriminativa del test se evaluó mediante la comparación entre las puntuaciones medias de novatos y expertos mediante el test t de student. También se utilizó el test 'X' para comparar el porcentaje de ítems individuales que pasaron los expertos en relación con los novatos.	Validez	Fiabilidad
Test de Fresno, 2003 ⁵	43 residentes de medicina familiar y miembros de la universidad; 53 voluntarios; 19 profesores de medicina de familia.	Representación de 2 escenarios clínicos seguidos de preguntas abiertas que requieren que el alumno formule una pregunta precisa, identifique el más apropiado diseño para responder tal pregunta, muestre conocer como buscar en bases de datos electrónicas, identifique si un determinado artículo es relevante y válido y discuta la magnitud e importancia de los hallazgos de la búsqueda. Las preguntas se puntuaban en base a la presencia de elementos que un experto en MBE decidía debían estar presentes en una respuesta ideal.	Validez de contenido (facial): se distribuyó un borrador del test entre profesores de MBE y posteriorment e se modificó la escala en función de sus sugerencias.	No se especifica evaluación de la validez de criterio.	- Fiabilidad inter-ratio (estabilidad del test): entre 0.72 y 0.96 para ítems individuales, puntuación total 0.97. - Consistencia interna (α de Cronbach): 0.88 - Correlación ítem-total: entre 0.47 y 0.75 para ítems individuales.
Test de Hong Kong, 2003 ⁸⁵	Cohorte de estudiantes de 5º año (n = 159), reclutados para el pilotaje; grupo combinado de estudiantes de 2º y 3º curso (n = 293), reclutados para la validación cuestionario.	Para la elaboración del cuestionario se parte de una revisión de la literatura, de las aportaciones de un grupo focal compuesto por estudiantes de 5º año de medicina (n = 10), y de la valoración de un panel de expertos internacional y de un panel local. Como resultado final se obtuvo un cuestionario de 43 ítems agrupados en 4 subescalas (conocimientos sobre MBE, actitudes hacia MBE, práctica y uso actual de MBE y uso futuro de MBE). Para dar respuesta a los distintos ítems se utilizaron escalas de Likert de 5 o 6 puntos...	Validez de contenido : valorada por medio de la correlación de los resultados obtenidos tras la aplicación del test a los estudiantes de 2º, 3º y 4º año a principio del curso y los obtenidos por los estudiantes de 2º año tras finalizar el mismo. El análisis estadístico T de Students mostró diferencias significativas. También se calcularon los coeficientes de correlación de las puntuaciones medias de cada estudiante en cada una de las 4 subescalas. No se obtuvo una correlación significativa entre las 4 subescalas lo que confirma su ortogonalidad.	Validez de criterio: se realizó un análisis de correlación entre la puntuación del cuestionario y otras actividades basadas en la evidencia realizadas por los estudiantes obteniéndose resultados positivos.	Se calculó el α de Cronbach para cada subescala tanto en el cuestionario preliminar (0.75-0.88) como en el definitivo (0.71-0.88)

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	La validez facial se mide durante la fase de pretest cognitivo.	No se especifica la medición de la validez de constructo.	Validez	Fiabilidad
Cuestionario de Holloway, 2004 ⁹⁰	100 estudiantes que iniciaron el primer curso de Medicina en agosto del 2000 y al los que se siguió durante los cursos 1 ^o y 2 ^o hasta marzo del 2002 (durante este periodo de tiempo los estudiantes recibieron formación sobre la práctica de MBE); 46 estudiantes de 3 ^o curso de medicina (sin formación previa en MBE) participaron en el pilotaje del test.	Se desarrolló un instrumento de evaluación que pretendía valorar los 4 pasos básicos de la MBE (construcción adecuada de preguntas clínicas, búsqueda de la mejor evidencia externa, valoración crítica de la evidencia y aplicación de la evidencia a la práctica clínica) y además incluía una auto-evaluación. Tras la experiencia piloto, el instrumento definitivo quedó constituido por dos módulos que presentaban cada uno un caso clínico a partir del cual se evaluaban las habilidades de los estudiantes en la práctica de la MBE.			La validez de criterio fue evaluada determinando los coeficientes de correlación entre cada paso dentro del módulo 1, así como dentro del módulo 2.	La fiabilidad test-retest fue determinada usando el coeficiente de correlación de Pearson entre la puntuación de cada paso y la global de los 2 módulos del test administrados a cada estudiante individual con un separación de 3 meses.
Cuestionario de MacRae et al, 2004 ⁹¹	Participaron en la validación 44 residentes de cirugía general.	El examen consistía en una serie de preguntas cortas y respuestas según escala de Likert de 7 que valoraban la calidad del estudio descrito en cada uno de los artículos. El test quedó constituido por un total de 55 ítems para los tres artículos que se utilizaron (15 a 21 preguntas por artículo). Las preguntas estaban basadas en el libro "Evidence based medicine: how to practice and teach EBM".	El artículo comenta que el instrumento diseñado mostraba validez facial y de contenido sin entrar en detalles.	La validez de constructo fue valorada mediante la comparación de la puntuación de los residentes que habían participado en sesiones bibliográficas sobre lectura crítica guiadas por epidemiólogos y los que no habían participado en una actividad similar (validez de constructo discriminativa). Se usó para dicha comparación el test T de Student (56.6% versus 49.3%, t ₃₅ = 2.31, p = 0.02).	No se encuentran comentarios sobre la valoración de la validez de criterio.	La consistencia interna del examen como un todo (55 ítems) fue valorada usando el coeficiente alfa de Cronbach (0.77). Para valorar fiabilidad interobservador se comparó, usando el coeficiente de correlación de Pearson, la puntuación generada por un epidemiólogo clínico al corregir el examen en relación a un artículo con la puntuación obtenida por dos adicionales correctores (clínico no epidemiólogo y enfermera). La correlación entre el epidemiólogo clínico y el clínico no epidemiólogo fue de 0.91 y, la correlación entre el epidemiólogo y la enfermera fue de 0.78.

TABLA 4. HERRAMIENTAS VALIDADAS PARA EVALUAR ACCIONES FORMATIVAS.

Cuestionario	Participantes	Características y desarrollo del test	Validez	Fiabilidad
Cuestionario de Bradley y Herrín, 2004 ³⁶ .	Participaron en la validación del cuestionario 55 profesionales de la salud. Además se llevó a cabo una re- validación en la que intervinieron 175 estudiantes de Medicina ³⁷ .	Un conjunto de preguntas presentes en estudios previos fue modificado por un panel de 4 expertos con experiencia en la enseñanza de MBE. Se seleccionaron preguntas relevantes y que incorporasen todos los aspectos claves requeridos en el programa de estudios que se iba a desarrollar y obtuvo un pool de 23 preguntas. Estas preguntas fueron sometidas a una doble traducción (inglés-noruego, noruego-inglés) para controlar inconsistencias. Posteriormente las 23 preguntas se validaron en 55 asistentes a 2 conferencias internacionales sobre MBE.	La validez facial y de contenido del cuestionario se evaluó a través de la evaluación de expertos en el pretest cognitivo.	Se midió la consistencia interna calculando el valor de α de Cronbach y covarianza inter-item. Se obtuvo una α de Cronbach global subóptima en los tres instrumentos (0.64, 0.58 y 0.03, respectivamente). En una última fase, cada instrumento fue revalidado en una cohorte de estudiantes usando únicamente la puntuación post-intervención. En esta fase sólo se valoró la consistencia interna mediante el recálculo de α de Cronbach que también obtuvo valores inferiores al considerado como satisfactorio (en este caso: 0.52, 0.47, 0.61).
Weberschock et al, 2005 ²² .	Participaron en la validación 132 estudiantes de 3º año de Medicina y 11 estudiantes componentes de Grupo de Trabajo de MBE.	A partir de lo publicado en la literatura se obtuvo un pool de preguntas de respuesta múltiple que se repartieron en 5 series de 20 preguntas (5 fáciles, 10 intermedias, 5 difíciles) con similar grado de dificultad (determinado previamente el grado de dificultad de cada pregunta por los componentes del Grupo de Trabajo de MBE). Las preguntas estaban ligadas a escenarios clínicos y evaluaban conocimientos y habilidades en MB	Fue evaluada la validez de constructo discriminativa al comparar las diferencias obtenidas en las puntuaciones de los estudiantes antes y después de la intervención educativa (seminarios). También se demostró la capacidad de los tests para detectar el impacto de la estrategia educacional.	La fiabilidad de cada uno de los exámenes fue testada mediante el cálculo de la consistencia interna; se obtuvo un α de Cronbach superior siempre a 0.7.

1. Dobbie A, Scheneider D.F., Anderson A., and Littlefield J. What evidence supports teaching evidence-based medicine? *Academic Medicine* 2000;75: 1184-85.
2. Hatala R and Guyatt G. Evaluating the teaching of evidence-Based Medicine. *JAMA*288[9], 110-12. 2002.
3. Hutchinson, Linda. Evaluating and researching the effectiveness of educational interventions. *BMJ* 1999; 318:1267-1269.
4. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. *BMJ* 2002; 325: 1338-41.
5. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *BMJ* 2003; 326: 319-21.
6. Smith CA et al. Teaching Residents Evidence-based Medicine Skills. *J Gen Intern Med* 2000; 15 (10): 710-15.
7. Reed et al. Challenges en systematic Reviews of Educational Intervention Studies. *Ann Intern Med* 2005;142:1080-9.
8. Nazareth I et al. Evaluation of a complex intervention for changing professional behaviour: the Evidence Based Out Reach (EBOR) Trial. *J Health Serv Res Policy* 2002;7(4):230-8.
9. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD Efectos de la auditoría y la retroalimentación (feedback) en la práctica profesional y las medidas de resultado de atención sanitaria (Revisión Cochrane traducida). En: *La Biblioteca Cochrane Plus*, 2006 Número 4. Oxford: Update Software Ltd. Disponible en: <http://www.update-software.com>. (Traducida de The Cochrane Library, 2006 Issue 4. Chichester, UK: John Wiley & Sons, Ltd.).
10. Dorsch JL, Aiyer MK, Meyer LE. Impact of an evidence-based medicine curriculum on medical students attitudes and skills. *J Med Libr Assoc* 2004;92(4):397-406.
11. Cramer JS, Mahoney MC. Introducing evidence based medicine to the journal club, using a structured pre and post test: a cohort study. *BMC Medical Education* 2001; 1: 6-8.
12. Grad R, Macaulay AC, Varner M. Teaching Evidence-based Medical Care: Description and Evaluation. *Fam Med* 2001;33(7):602-6.
13. Seeling CB. Affecting Residents' Literature Reading Attitudes, Behaviors an Knowledge through a Journal Club Intervention. *J Gen Intern Med*, 1991;6:330-4.
14. Cartwright CA, Korsen N, Urbach LE. Teaching the teachers: Helping Faculty in a Family Practice residency Improve Their Informatics Skills. *Acad. Med* 2002;77:385-91.
15. Vogel EW, Block KR, Wallingford KT. Finding the evidence: teaching medical residents to search MEDLINE. *J Med Libr Assoc* 2002;90(3):327-30.

16. Straus SE, Ball C, Balcombe N, Sheldon J, AcAlister FA. Teaching Evidence-based medicine Skills Can Change Practice in a Community Hospital. *J Gen Intern Med* 2005;20:340-3.
17. Kellum J, Rieker J, Power M, Powner D. Teaching critical appraisal during critical care fellowship training: A foundation for evidence-based critical care medicine. *Crit Care Med* 2000;28(8):3067-70.
18. Schoenfeld P, Cruess D, Peterson W. Effect of an Evidence-based Medicine Seminar on Participants Interpretations of Clinical Trials: A Pilot Study. *Acad Med* 2000;75(12):1212-14.
19. Khan K et al. Postgraduate journal club as a means of promoting evidence-based obstetrics and gynaecology. *J Obstet Gynaecol* 1999;19(3):231-4.
20. Thom DH, Haugen J, Sommers PS, Lovett P. Description and evaluation of an EBM curriculum using a block rotation. *BMC Med Educ* 2004;4:19.
21. Byrnes JA, Kulick TA, Schwartz DG. Information-seeking behavior changes in community-based teaching practices. *J Med Libr Assoc* 2004;92(3):334-40.
22. Weberschock TB et al. Change in Knowledge and skills of year 3 undergraduates in evidence-based medicine seminars. *Med Educ* 2005;39: 665-71.
23. Dinkevich E, Markinson A, Ahsan S, Lawrence B. Effect of a brief intervention on evidence-based medicine skills of pediatric residents. *BMC Medical Education* 2006;6:1
- 24 Lee A et al. Structure Journal Club as a Tool to Teach and Assess Resident Competence in Practice based Learning and Improvement. *Ophthalmology* 2006;113:497-500.
25. Lucas BP et al. The impact of evidence on physicians' inpatient treatment decisions. *J Gen Intern Med* 2004;19:402-409.
26. Thomas KG, Thomas MR, Dupras DM. Assessment tools for evaluating critical appraisal skills. *Med Educ* 2004;38:569.
27. O'Riordan M, Ni Riain, A. Distance learning: linking CME and quality improvement. *Medical teacher*(2004),26:559-564
28. Fiesci M, Soula G, Giorgi R, Gouvernet J., Fieschi D., Botti G, Volot F, Berland Y. Experimenting with new paradigms for medical education and the emergence of a distance learning degree using the internet : teaching evidence-based medicine. *Med. Inform.*(2002);27:1-11-
29. Fordis M, King JE, Ballantyne CM, Jones PH, Comparison of the instructional efficacy of Internet-Based CME with live Interactive CME with live interactive CME workshops. *JAMA* (2005), 294:1043-52.
30. Stewart M, Marshall JN, Ostbye T, Feightner JW, Brown JB, Harris S, Galajda J. Effectiveness of Case-based on-line learning of evidence-based practice guidelines. *Fam Med* (2005); 37:131-8
31. Bradley P., Otherlot C., Nordhemim L., Bjorndal A. Medical student's and tutor's experiences of directed and self-directed learning programs in evidence-based medicine. A qualitative evaluation accompanying a randomised controlled trial. *Evaluation Reviw*,(2005) 29:149-177.
32. Wilkes M, Bligh J, Evaluating educational interventions. *BMJ* 1999 ; 318:1269-1272
33. Gómez de la Cámara. A. La formación del médico de Familia en la Medicina Basada en evidencias. *Tribuna Docente*. Volumen 3 n.º1. 2002.9-15.

34. Straus SE, Green ML, Bell D S, Badgett R, Davis D, Gerrity M, Ortiz E, Shaneyfelt TM, Whelan Ch, Mangrulkar R and the Society of General Internal Medicine.Evidence-Based Medicine Task Force. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ* 2004;329:1029-1032.
35. Guyatt G, Meade MO, Jaeschke RZ, Cook D, Haynes B, Practitioners of evidence based care.. *BMJ* 320[7240], 954-55. 2000.
36. Wadland WC et al. Training Medical Students in Evidence-based Medicine: A community campus approach. *Fam Med* 1999;31(10):703-8.
37. Khalid S. Khan, Arri Coomarasamy. A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Medical Education* 2006, 6:59
38. Martin Dawes, William Summerskill, Paul Glasziou, Antonino Cartabellotta, Janet Martin, Kevork Hopayian, Franz Porzsolt, Amanda Burls, James Osborne. Sicily statement on evidence-based practice *BMC Medical Education* 2005, 5:1
39. Audet N, Gagnon R, Lasouceur R, Marcil M. L'enseignement de l'analyse critique des publications scientifiques médicales est-il efficace ? Révision des études et de leur qualité méthodologique. *Can Med Assoc J* 1993; 148(6):945-52.
40. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of Computer-based Clinical Decision Support Systems on Clinician Performance and Patient Outcome. A Critical Appraisal of Research. *Ann Intern Med* 1994;120:135-42.
41. Davis D, Thomson M, Oxman A, Haynes B. A systematic Review of the effect of continuing medical education strategies. *JAMA* 1995;274(9):700-5.
42. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *Can Med Assoc* 1998;158(2):177-81.
43. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med* 1999;74(6):686-94.
44. Taylor et al. A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ* 2000;34:120-25.
45. Ebbert JO, Montori VM, Schultz HJ. The journal club in postgraduate medical education: a systematic review. *Med Teach* 2001;23(5):455-61.
46. Parkes J, Hyde C, Deeks J, Milne R. Enseñanza de habilidades de evaluación crítica en ámbitos de atención sanitaria (Revisión Cochrane traducida). En: *La Biblioteca Cochrane Plus*, 2005 Número 1. Oxford: Update Software Ltd.
47. Garg A, Turtlet KM. Effectiveness of training health professionals in literature search skills using electronic health databases-a critical appraisal. *Health Information and Libraries Journal*;20:33-41.
48. Coomarasamy A, Taylor R, Khan K. A systematic review of postgraduate teaching evidence-based medicine and critical appraisal. *Med Teach* 2003;25(1):77-81.
49. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ* 2004; 329: 1017-21.

50. Werb SB, Matear DW. Implementing Evidence-based Practice in Undergraduate teaching clinics: A Systematic Review and Recommendations. *J Dent Educ* 2004;68(9):995-1003.
51. Shaneyfelt T et al. Instruments for Evaluating Education In Evidence-Based Practice. A Systematic Review. *JAMA* 2006;296(9):1116-27.
52. Linzer et al. Impact of a Medical Journal Club on House-Staff Reading Habits, Knowledge, and Critical Appraisal Skills. A Randomized control Trial. *JAMA* 1988;260(17):2537-41.
53. Haynes RB, Johnston ME, McKibbon KA, Walker CJ, Willan AR. A program to enhance clinical use of MEDLINE: a randomized controlled trial. *Online J Curr Clin Trials*. 1993;(No. 56).
54. MacAuley D, McCrum E, Brown C. Randomised controlled trial of the READER method of critical appraisal in general practice. *BMJ* 1998;316:1134-7.
55. Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *JR Coll Physicians Lond*. 1998;32:557-563.
56. MacAuley D, McCrum E. Critical appraisal using de READER method: a workshop based controlled trial. *Fam Pract* 1999; 16:90-3.
57. Stevermer JJ, Chambliss ML, Hoekzema GS. Distilling the literature: a randomized, controlled trial testing an intervention to improve selection of medical articles for reading. *Acad Med* 1999;74(1): 70-2.
58. Neville AJ, Reiter HI, Eva KW, Norman GR. Critical appraisal turkey shoot: linking critical appraisal to clinical decision making. *Acad Med* 2000;75(10):S87-S89.
59. Cabell et al. Resident utilization of information technology. A randomized trial of clinical question formation. *J Gen Intern Med* 2001;16:838-44.
60. Villanueva EV, Burrows EA, Fenecí PA, Rajendran M, Anderson JN. Improving question formulation for use in evidence appraisal in a tertiary care setting: a randomised controlled trial. *BMC Med Inform Decis Mak*. 2001;1:4.
61. Bradley DR et al. Real-time, evidence-based medicine instruction: a randomized controlled trial in a neonatal intensive care unit. *J Med Libr Assoc* 2002;90(2):194-201.
62. Bernal-Delgado E, Galeote-Mayor M, Pradas-Arnal F, Peiro-Moreno S. Evidence based educational outreach visits: effects on prescriptions of non-steroidal anti-inflammatory drugs. *J Epidemiol Community Health*. 2002 Sep;56(9):653-8.
63. Verstappen W et al. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians. A randomized Trial. *JAMA* 2003;289(18):2407-12.
64. Cheng GY. Educational workshop improved information-seeking skills, knowledge, attitudes and the search outcome of hospital clinicians: a randomised controlled trial. *Health Info Libr J*. 2003;20(suppl 1):22-33.
65. MacRae et al. Teaching practicing surgeons critical appraisal skills with an Internet-based journal club: A randomized, controlled trial. *Surgery* 2004;136(3):641-6.
66. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial. *BMC Med Educ* 2004;4(1):30.

- 
67. Johnston JM et al. Evaluation of a handheld clinical decision support tool for evidence-based learning and practice in medical undergraduates. *Med Educ* 2004;38:628-37.
 68. Bradley P et al. Comparison of directed and self-directed learning in evidence-based medicine: a randomised controlled trial. *Med Educ* 2005;39:1027-35.
 69. Mukohara K, Schwartz MD. Electronic delivery of research summaries for academic generalist doctors: a randomised trial of an educational intervention. *Med Educ* 2005; 39:402-9.
 70. Schilling K, Wiecha J, Polineni D, Khalil S. An interactive web-based curriculum on Evidence-based Medicine: design and effectiveness. *Fam Med* 2006;38(2)126-32.
 71. Radack KL, Valanis B. Teaching Critical Appraisal and Application of Medical Literature to clinical Problem-Solving. *J Med Educ* 1986;61:329-31.
 72. Bennett KJ et al. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA* 1987;287(18):2451-54.
 73. Langkamp DL, Pascoe JM, Nelson DB. The effect of a Medical Journal Club on residents' knowledge of clinical epidemiology and biostatistics. *Fam Med* 1992;24:528-30.
 74. Frasca MA, Dorsch JL, Aldag JC, Christiansen RG. A multidisciplinary approach to information management an critical appraisal instruction: a controlled study. *Bull Med Libr Assoc* 1992;80(1):23-28.
 75. Landry FJ et al. A controlled Trial of a Seminar to Improve Medical Student Attitudes toward, Knowledge about, and Use of the Medical Literature. *J Gen Intern Med* 1994;9:436-9.
 76. Green ML, Ellis PJ. Impact of an Evidence-Based Medicine Curriculum Based on Adult Learning Theory. *J Gen Intern med* 1997;12:742-50.
 77. Fu C, Hodges B, Regehr G, Goldbloom D, Garfinkel P. Is a journal club effective for teaching critical appraisal skills? A controlled trial with residents in Psychiatry. *Academic Psychiatry* 1999;23:205-9.
 78. Bazarian JJ et al. Teaching Emergency Medicine Residents Evidence-based Critical Appraisal Skills: A controlled Trial. *Ann Emerg Med* 1999;34(2):148-54.
 79. Ghali WA, Sitz R, Eskew AH, Gupta M, Quan H, Hershman WY. Successful teaching in evidence-based medicine. *Med Educ* 2000;34:18-22.
 80. Ross R, Verdick A. Introducing an Evidence-based Medicine Curriculum into a Family Practice Residency: Is it Effective?. *Acad Med* 2003;78(4):412-417.
 81. Akl EA et al. Integrating an Evidence-Based Medicine Rotation into an Internal Medicine Residency Program. *Acad Med* 2004;79(7):897-904.
 82. Grad RM et al. Effect of a PDA-assisted Evidence-based Medicine course on knowledge of common clinical problems. *Fam Med* 2005;37(10):734-40.
 83. Gruppen LD, Rana GJ, Arndt TS. A controlled comparison study of the efficacy of training medical students in evidence-based medicine literature searching skills. *Acad Med* 2005;80(10):940-4.

84. Stern DT, Linzer M, O'Sullivan PS, Weld L. Evaluating Medical Residents' Literature-appraisal Skills. *Acad Med* 1995;70(2):152-4.
85. McAlister, Graham I, Karr GW, Laupacis A. Evidence-Based Medicine and the Practicing Clinician. *J Gen Intern Med* 1999;14:236-42.
86. Taylor et al. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. *Med Educ* 2001; 35: 544-7.
87. Beasley BW, Woolley DC. Evidence-based Medicine Knowledge, Attitudes, and skills of Community Faculty. *J Gen Intern Med* 2002;17:632-40.
88. Green LA, Gorenflo DW, Wyszewianski L. Validating an instrument form selecting interventions to change physician practice patterns: A Michigan consortium for Family Practice Research study. *J Fam Pract* 2002;51:938-42.
89. Johnston JM, Leung GM, Fielding R, Tin K, Ho L. The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidence-based practice teaching and learning. *Med Educ* 2003; 37: 992-1000.
90. Holloway R, Nesbit K, Bordley D, Noyes K. Teaching and Evaluating first and second year medical students practice of evidence-based medicine. *Med Educ* 2004; 38: 868-78.
91. MacRae HM et al. Assessment of critical appraisal skills. *Am J Surg* 2004;187:120-3.
92. Bradley P, Herrin J. Development and Validation of an Instrument to Measure Knowledge of Evidence-Based Practice and Searching Skills. *Med Educ Online* [serial online] 2004;9:15-19.



Acotado Automático de Términos: Brevemente, consiste en el enlace automático de una palabra o frase (que no es un término MeSH) al término MeSH que mejor lo describe. Una descripción más completa en el apartado correspondiente. Sinónimos (Sin): Automatic Term Mapping, ATM, Mapeo automático de términos.

Aleatorización: En un estudio en el que se comparan dos o más modalidades de tratamiento, es conveniente que los grupos formados sean parecidos en todas las características pronósticas excepto en el tratamiento recibido, de modo que cualquier diferencia en el curso clínico que se registre podrá ser atribuida a los diferentes tratamientos administrados (y sólo a ellos). La asignación aleatoria consiste en distribuir a cada participante de uno de los grupos de tratamiento por un método disciplinado de azar, de modo que cada sujeto tenga exactamente las mismas probabilidades de formar parte de uno u otro grupo de tratamiento. Sin: Randomización.

Alertas: Método usado para la difusión de información mediante mensajes electrónicos a ordenadores, agendas electrónicas e incluso teléfonos móviles. Representa una forma de actualización rápida y eficaz.

Análisis coste/beneficio: Medición de los costes y beneficios de una intervención en términos comensurables, generalmente monetarios. Permite determinar en primer lugar si una intervención concreta ofrece un beneficio global neto de bienestar, y en segundo lugar cómo se compara la ganancia de bienestar obtenida con esta intervención, con la obtenida con intervenciones alternativas. Al valorar todos los costes y beneficios en las mismas unidades, el análisis coste-beneficio compara diversas intervenciones con el criterio del beneficio neto (véase). Análisis económico en el que se consideran los costes de la asistencia sanitaria y la pérdida de ganancias netas debida a muerte o incapacidad, traduciendo todos los costes y beneficios a unidades monetarias.

Análisis coste/efectividad: Modalidad de estudio económico en la que las consecuencias de diferentes intervenciones pueden ser diferentes pero pueden ser medidas en unidades naturales idénticas, Mide los beneficios sanitarios en términos de unidades naturales, como años de vida salvados o mejorías del estado funcional. Dado que los costes no se expresan en las mismas unidades que los beneficios, su razón (ratio coste-efectividad) constituye una herramienta que permite valorar la eficiencia relativa. Análisis económico que compara los costes médicos incrementales y los resultados sobre la salud de programas sanitarios alternativos. Al contrario que el análisis de coste beneficio, los efectos sobre la salud se expresan en diversas unidades, como vida-años ganados, días libres de síntomas, casos evitados, etc., en lugar de en unidades monetarias

Análisis coste-utilidad: Adaptación del análisis coste-efectividad que mide el efecto de una intervención a la vez sobre los aspectos cuantitativos y cualitativos de la salud, mediante el uso de una variable denominada AVAC -Años de Vida Ajustados por Calidad- o QALYs . También evalúa la eficiencia relativa con una razón incremental, que en este caso se denomina ratio coste-utilidad.



Análisis de decisiones (AD): Es la aplicación de métodos explícitos y cuantitativos que cuantifican e integran los efectos del tratamiento, el pronóstico y los valores del paciente para analizar una decisión en condiciones de incertidumbre.

Análisis de minimización de costes: Modalidad de análisis coste-efectividad donde los resultados o beneficios de las alternativas se considera que son equivalentes o intercambiables para el mismo tipo de pacientes. A igualdad de resultados, la atención de los investigadores se centra en la diferencia de costes asociados.

Análisis de sensibilidad: En economía de la salud, técnica por la que se repiten las comparaciones entre recursos necesarios y resultados, pero con sucesivas modificaciones de las suposiciones en las que se basan los cálculos. De este modo, el análisis de sensibilidad pone a prueba la robustez de las conclusiones mediante variaciones de los ingredientes sobre los que existe incertidumbre.

Análisis de regresión: En un conjunto de datos sobre la variable dependiente y sobre una o más variables independientes, x_1, x_2, \dots, x_i , consiste en determinar el modelo matemático más ajustado que describa y como una función de las x o para predecir y a partir de las x . Término general para los métodos de análisis que se ocupan de estimar los parámetros de alguna posible relación entre una variable respuesta y una o más variables explicativas

Análisis de la varianza (ANOVA): Conjunto de técnicas estadísticas para conocer el modo en que el valor medio de una variable es afectado por diferentes tipos de clasificaciones de los datos. Con el análisis de la varianza se pueden ajustar las estimaciones del efecto de un tratamiento según otros factores como sexo, edad, gravedad, etc.

Análisis estratificado: Análisis de los resultados de un ensayo clínico o un estudio observacional, con el fin de examinar si hay diferencias de respuestas asociadas a alguna característica (por ej., edad o sexo).

Análisis factorial: Procedimiento estadístico empleado para la validación de constructo que permite identificar grupos o conglomerados de variables relacionadas.

Análisis multivariante: Cualquier método analítico que permite el estudio simultáneo de dos o más variables dependientes. Término genérico para los métodos que están diseñados para determinar las contribuciones relativas de diferentes causas a un solo suceso o resultado

Análisis por intención de tratar: En un ensayo clínico, análisis de los datos según el tratamiento asignado inicialmente por distribución aleatoria, independientemente de que el paciente no recibiera después este tratamiento.

Árbol de decisiones: Expresión gráfica de un análisis de decisiones.

Artículo científico: informe escrito que comunica por primera vez los resultados de una investigación. Los artículos científicos publicados en miles de revistas científicas componen la literatura primaria de la ciencia

Artículo de revisión: artículos de síntesis que resumen el conocimiento de un tema determinado y durante un periodo de tiempo definido. Junto con los libros, los artículos de revisión constituyen la literatura científica secundaria.

Asiento: Conjunto de información que, en una base de datos, se refiere a un mismo elemento. En Bases de Datos de bibliografía, es el conjunto de información que se refiere a una misma publicación. Sin: Registro, Record.

ATM: Ver Acotado Automático de Términos.

Atributos: Propiedades o características del instrumento que merecen consideraciones separadas e independientes en la evaluación.

Auditoria: Examen o estudio que evalúa e informa sobre la medida en que una condición, proceso o desempeño se ajusta a estándares o criterios predeterminados.

Automatic explosion: Ver Explode.

Automatic Term Mapping: Ver Acotado Automático de Términos.

AVAC: Índice que combina la supervivencia con la calidad de vida. La medida de la calidad de vida no está estandarizada y suele variar de un estudio a otro, según la enfermedad, el tratamiento evaluado y las preferencias de los autores. Para calcular un AVAC se multiplica la duración del estado de salud (en años) por un factor que representa la calidad (“utilidad”) de este estado de salud. Sin: QALY (Quality Adjusted Life Years).

Azar: Se denomina así al proceso que ocurre debido a la casualidad, de forma imprevisible, sin planeamiento ni influencia consciente.

Base de datos: Conjunto de datos homogéneos y ordenados, que se presentan normalmente, de forma legible, por ordenador y que se refieren a una materia determinada. Colección estructurada de datos que está organizada de tal manera que se puede acceder a ella con una amplia variedad de programas de aplicación.

Beneficio neto: Diferencia entre los costes y los beneficios de una intervención.

Bioestadística: Aplicación de la teoría y los métodos de la estadística a las ciencias biológicas. En sentido estricto, rama de la ciencia que aplica los métodos estadísticos a los problemas biológicos, aunque ahora se utiliza la mayoría de las veces para referirse a la estadística aplicada a la medicina y las ciencias de la salud.

Blinding: Ver Enmascaramiento.

Buscadores: Motores de búsqueda de información en Internet que identifican la información contenida en bases de datos formadas a partir de publicaciones previamente indexada. Ver Indización.

Calidad de Vida Relacionada con la Salud (CVRS): Impacto que una enfermedad y el tratamiento consiguiente tienen sobre la percepción del paciente de su satisfacción y bienestar físico, psíquico, social y espiritual.

Campo: Cada una de las unidades de información en que se divide un asiento. Sin: Field.

Carga de administración: Se define como el grado de trabajo y esfuerzo que deben realizar quienes administran un instrumento de medida (cuestionario), y depende del tiempo precisado para su administración, formación requerida del entrevistador y recursos necesarios para su aplicación y puntuación.

Carga de respuestas: Se entiende como el tiempo, esfuerzo y otros aspectos que se exigen a quienes se les administran un instrumento de medida (cuestionario).

CAT: Herramienta de lectura crítica consistente en un resumen breve de las pruebas científicas más relevantes en respuesta a una pregunta clínica concreta.

Causalidad: Indica que las asociaciones encontradas entre las exposiciones y los efectos son realmente de tipo causal y no consecuencia de sesgos ni del efecto de factores de confusión.

Cociente de probabilidad: Véase Razón de probabilidad.

Coefficiente de correlación: Índice que cuantifica la relación lineal entre un par de variables. Medida de asociación que indica el grado en el que dos variables continuas x e y poseen una relación lineal ($y = a + bx$). Su valor puede situarse entre -1 y $+1$. Cuando su valor es 0 no existe ningún grado de correlación. El coeficiente de este tipo que más se utiliza es el de Pearson

Coefficiente de Cronbach: Fórmula utilizada para la comprobación de la consistencia interna de un test que presupone que si varios ítems están midiendo el mismo atributo o dimensión, estos ítems deberían estar intercorrelacionados. Esta correlación la mediremos mediante el coeficiente alfa de Cronbach.

Coefficiente kappa: índice corregido por el azar de la concordancia entre, por ejemplo, juicios o diagnósticos que hacen 2 ó más observadores. El coeficiente adopta desde el valor 1 (cuando hay una concordancia perfecta) hasta el valor 0 cuando la concordancia observada es igual a la debida al azar.

Confusión: Situación en la que la estimación del resultado estudiado está sesgada por algún factor relacionado con la exposición y el resultado.

Consentimiento informado: Procedimiento formal para aplicar el principio de autonomía. Debe reunir por lo menos tres elementos: voluntariedad, información y comprensión. La voluntariedad implica que los sujetos puedan decidir libremente si quieren participar en un estudio, sin que haya persuasión, manipulación ni coerción. La información debe ser comprensible y debe incluir el objetivo del estudio y su procedimiento, los beneficios y riesgos potenciales, las posibles incomodidades derivadas de la participación y la posibilidad de retirarse del estudio en cualquier momento sin que ello de lugar a perjuicio, en este contexto es básica la comprensión de la información, que no sólo depende de esta, sino también de la competencia de quien la recibe, es decir de su capacidad para comprender.

Consistencia interna: Grado en que los ítems de una dimensión se correlacionan entre ellos mismos. Ver: Coeficiente de Cronbach.

Constructo: En psicometría, concepto abstracto no medible directamente que se quiere convertir en variable operativa medible. Generalmente, se utiliza para un concepto que existe en teoría, pero que no se puede observar directamente. Esencialmente, es una variable latente.

Correlación ítem-total:(Correlación total de las preguntas): Verifica la homogeneidad. Proceso mediante el cual comprobamos si cada ítem mide lo mismo que los demás y por lo tanto es sumable en una puntuación total que supuestamente mide el rasgo y que es la que después interpretamos y utilizamos.

Coste: Es el resultado del producto de la cantidad total del recurso consumido y el valor monetario de la unidad de dicho recurso.

Coste oportunidad: En economía de la salud, valor de las oportunidades perdidas por usar recursos en un servicio. Generalmente es cuantificado considerando el beneficio que se obtendría invirtiendo los mismos recursos en la mejor alternativa posible.

Criterios: Condiciones utilizadas como un estándar para juzgar los atributos.

Criterios de selección: Características sociodemográficas, clínicas, etc. que determina que los pacientes puedan ser o no incluidos en un estudio. Nos van a servir para definir la población de estudio.

Curva de Gauss: En matemáticas, la campana o curva de Gauss es la representación gráfica de la ecuación matemática que corresponde a una Distribución normal. Tiene forma de campana y debe su nombre al matemático alemán Carl Friedrich Gauss.

Curvas ROC: Representación gráfica completa de la capacidad de una prueba para discriminar lo normal de lo que no lo es.

Delimitador de campo: Ver Etiqueta.

DeCS - Descriptores en Ciencias de la Salud: Tesoro que puede considerarse, hasta cierto punto, como una versión en español del MeSH - Medical Subject Headings de la U.S. National Library of Medicine. Fue desarrollado a partir de él con el objetivo de permitir el uso de terminología común para búsqueda en tres idiomas (español, inglés y portugués), proporcionando un medio consistente y único para la recuperación de la información independientemente del idioma. Además de los términos médicos originales del MeSH contiene áreas específicas de Salud Pública y Homeopatía.

Descriptor: Término del lenguaje documental, carente de variabilidad, sinonimia y polisemia, escogido para representar el contenido de un documento, que puede ser utilizado en el análisis y recuperación de la información. Relacionado con Palabra clave.

Descriptor mayor: Ver Mayor Topic.

Descriptores en Ciencias de la Salud: Ver DeCS

Descuento: Procedimiento para calcular los costes o beneficios que ocurren en diferentes fechas a su valor en el momento presente, llamado valor actual.

Dimensiones: Aspectos relevantes que engloba el constructo o concepto que se mide a través del cuestionario.

Diseño: Conjunto de estrategias y procedimientos que se seguirán para dar respuestas al problema y comprobar la hipótesis. Engloba al selección del tipo de estudio, la población a estudiar, la definición de las variables y escalas de medición y los métodos de recogida, procesado y análisis de los datos.

Docencia virtual (“on-line” o “e-learning”): Definida como la utilización de las nuevas tecnologías multimedia y de internet para mejorar la calidad del aprendizaje, facilitando el acceso de recursos y servicios; así como los intercambios y la colaboración a distancia.

Dosis Diaria Definida (DDD): Dosis de mantenimiento promedio en la principal indicación de un fármaco en pacientes adultos. Se usa para cuantificar el consumo de medicamentos, sobre todo cuando se compara el consumo de diversos fármacos de un mismo grupo, o cuando se realizan comparaciones internacionales. No se ve afectada por diferencias de precios o de tamaño de las presentaciones. El consumo se suele expresar en número de DDD por 1.000 habitantes y por día, y esta medida da una idea del número promedio de personas tratadas diariamente con un fármaco o grupo de fármacos en una región o país.

Efectividad: Grado en que una determinada intervención origina un resultado beneficioso en las condiciones de la práctica clínica habitual, sobre una población determinada.

Efecto indeseado: Ver Reacción Adversa.

Eficacia: Grado en que una determinada intervención origina un resultado beneficioso en condiciones ideales de aplicación.

Eficiencia: Relación entre los efectos o resultados finales que se alcanzan con una intervención y coste que supone la misma (recursos humanos, materiales y de tiempo).

Encabezamientos de Materias Médicas: Ver MeSH.

Enmascaramiento: Procedimiento por el que se asegura que los sujetos participantes en un ensayo clínico, los observadores o ambos, no conocen el tratamiento u otra característica que pudiera sesgar los resultados. Se denomina simple ciego cuando el sujeto, pero no el observador ignora cuál de los posibles tratamientos recibe; doble ciego es cuando sujeto y observador ignoran el tratamiento recibido; y triple ciego es cuando el sujeto participante, el investigador observador y el investigador que analiza los datos ignoran el tratamiento recibido.

Ensayo clínico: “Toda investigación efectuada en seres humanos para determinar o confirmar los efectos clínicos, farmacológicos y/o demás efectos farmacodinámicos, y/o de detectar las reacciones adversas, y/o de estudiar la absorción, distribución, metabolismo y excreción de uno o varios medicamentos en investigación con el fin de determinar su seguridad y/o su eficacia” (Real Decreto de 16 de febrero de 2004).

Ensayo clínico controlado: Ensayo clínico en el que los sujetos son asignados a dos grupos: uno (grupo experimental) recibe el tratamiento que se está probando y el otro (grupo de comparación o control) recibe un tratamiento alternativo. En el ensayo clínico controlado aleatorizado la asignación a un tratamiento experimental o control se realiza al azar de forma que ni el sujeto ni el investigador responsable de su selección o tratamiento puedan influir en su asignación.

Entry Term: Relacionado con Palabra clave (véase).

Epidemiología: Estudio de la distribución y los determinantes de los estados o acontecimientos relacionados con la salud de poblaciones. Ciencia que investiga las causas y la historia de las enfermedades y el desarrollo de éstas en relación con determinadas intervenciones o exposiciones.

Epistemología: Doctrina de los fundamentos y métodos del método científico. La función de la estadística en el modelo empírico analítico.

Equivalencia: capacidad del instrumento de proporcionar los mismos valores en las mismas condiciones clínicas cuando las mediciones las hacen distintas personas.

Error aleatorio: Las diferencias de medidas entre la realidad y nuestra muestra y son debidas al azar.

Escala tipo Likert: La respuesta del individuo se da mediante una escala numérica, a la cual se le atribuyen las propiedades de las escalas de intervalo, permitiendo por tanto la realización de análisis estadísticos más complejos. Escalas que se utilizan con frecuencia en estudios de actitudes en los que las puntuaciones crudas se basan en respuestas alternativas graduadas a cada una de una serie de preguntas.

Especificidad: Referida a una prueba diagnóstica, es la proporción de personas verdaderamente no enfermas que son catalogadas como tales por medio de la prueba.

Estabilidad: Ver Reproducibilidad.

Estabilidad del test: Véase Método test-retest.

Estadística: Recopilación y organización de hechos o datos numéricos, sobre todo cuando hay muchos. Conjunto de métodos que ayudan a la toma de decisiones frente a la incertidumbre. La estadística no ofrece pruebas ni certidumbre, sino probabilidades.

Estadístico: Medida estadística (media, mediana, etc...) calculada sobre los datos de una muestra.

Estadístico kappa: Medida del acuerdo, corregida por el azar, aplicada a datos binarios u ordinales. Se calcula como la proporción de acuerdo, aparte del que ya sería de esperar por azar, que ha sido observado entre dos repeticiones del mismo instrumento.

Estudios ambispectivos: Estudios que se inician como los retrospectivos (Ver: Estudios retrospectivos) para el análisis de un fenómeno ya ocurrido y, posteriormente, continuamos con un seguimiento prospectivo hasta la aparición de un efecto a largo plazo (Ver: Estudios prospectivos).

Estudios analíticos: Estudios diseñados para examinar asociaciones, cuyo objeto final suele ser identificar o medir los efectos de factores de riesgo o de intervenciones específicas sobre la salud.

Estudios de caso-control: Estudios en los que la selección de la población a estudiar se hace en función de la presencia o ausencia de una enfermedad o efecto, eligiéndose un grupo de individuos que tienen la enfermedad o efecto (casos), y otro que no los presenta (controles).

Estudios de cohortes: Tipo de estudios en los que los sujetos del estudio se seleccionan a partir presencia o ausencia de la exposición que se trata de analizar y de la que se sospecha es responsable del efecto observado.

Estudio descriptivo: Estudio cuyo objetivo es constatar hechos, observar lo que ocurre en nuestra población. Permiten generar hipótesis.

Estudio experimental: Estudio en el que el investigador asigna y controla el factor de estudio.

Estudios de incidencia: Tipo de estudios en los que seguimos a una población determinada a lo largo del tiempo para analizar la aparición o no de un fenómeno de interés. La unidad de análisis es el individuo.

Estudios longitudinales: Estudios en los que la recogida de los datos de interés hace referencia a dos momentos separados en el tiempo (se analizan los sujetos en varios momentos a lo largo del tiempo).

Estudios observacionales: Estudios en los que el investigador no participa directamente en la intervención o exposición (factor de estudio), se limita a observar (registrar) lo que ocurre en la realidad.

Estudios pragmáticos: Estudios cuyo objetivo principal es la evaluación de la adecuación de la práctica asistencial en condiciones normales.

Estudios pre y post-intervención: Estudios en los que se registran la situación basal de la muestra y los cambios que experimenta tras su exposición a una determinada intervención.

Estudios de prevalencia: Estudios en los que se describe una serie de variables en una población determinada y en un momento determinado. La unidad de análisis es el individuo.

Estudios prospectivos: Tipo de estudios en los que los hechos a estudiar no han tenido lugar aun y los sujetos son seguidos hasta obtener el resultado.

Estudio retrospectivo: Estudio en el cual los hechos a estudiar ya han tenido lugar al iniciar el estudio.

Estudio transversal: Tipo de estudio en el que se estudia la exposición y el desenlace al mismo tiempo.

Etiqueta: Abreviatura del nombre de un campo que puede utilizarse en su lugar para dirigir la búsqueda de información. Sin: Tag, Delimitador de campo.

Evaluación crítica: Véase Lectura crítica.

Evaluación económica: Análisis comparativo de opciones alternativas en términos de sus costes y consecuencias. Los estudios de evaluación económica pueden ser parciales, cuando sólo se describen los costes, los resultados o ambos a la vez pero de una única opción o que comparan dos o más alternativas pero sólo en función de sus costes o de sus consecuencias sin valorar ambos a la vez, o completos, cuando se analizan comparativamente dos o más alternativas mediante la valoración de los costes y de las consecuencias de cada una de ellas.

Evaluación de tecnologías sanitarias: Proceso de análisis e investigación dirigido a estimar el valor y contribución relativos de cada tecnología sanitaria (véase) en la mejora de la salud individual y colectiva, teniendo en cuenta su impacto económico, social e, incluso, implicaciones éticas.

Explode: Obtención de las referencias que contienen el descriptor especificado o cualquiera de los descriptores de su mismo campo semántico pero de significado más restringido. Sin: Automatic explosion, Exploding, Explosion, Explosión, Explosión automática.

Exploding: Ver Explode.

Exploración complementaria: Conjunto de pruebas diagnosticas realizadas con la finalidad de confirmar una hipótesis diagnostica o hacer el seguimiento de la patologia de un paciente.

Explosión: Ver Explode.

Explosión automática: Ver Explode.

Factibilidad: Indica las posibilidades reales y práctica de que un instrumento de medida pueda ser contestado, completado y analizado. Está relacionad con el nivel de aceptabilidad del instrumento y se refleja en la proporción de rechazos, no respuesta y datos perdidos.



Factor de confusión: Variable que está simultánea e independientemente asociada al factor de riesgo estudiado y a la enfermedad estudiada, que puede modificar el resultado del estudio. Tal variable debe ser identificada y se debe evitar su influencia antes de la recogida de datos (mediante el emparejamiento o la restricción) o bien en la fase de análisis, mediante el análisis estratificado y la regresión múltiple.

Factor de impacto: Indicador bibliométrico, elaborado por el ISI (véase) y publicado en el Journal Citation Reports, que mide el uso de las revistas. Suele calcularse dividiendo el número de veces que son citados los artículos publicados en una revista durante un plazo de tiempo (habitualmente dos años), entre el número de artículos publicados por esa misma revista y en ese mismo plazo de tiempo. Puede modificarse para calcular el impacto de autores, instituciones... Sin: Impact Factor.

Factor de riesgo: Factor relacionado con el riesgo de que ocurra un suceso (p. ej. enfermar).

Falso negativo: Referido a una prueba diagnóstica, número de enfermos en que el resultado sería negativo.

Falso positivo: Referido a una prueba diagnóstica, número de sanos en que el resultado sería positivo.

Farmacovigilancia: Programas que proporcionan de forma continuada una evaluación sobre la seguridad de los fármacos. El factor clave en este sistema son los profesionales sanitarios que participan en el programa de notificación espontánea de reacciones adversas.

Fiabilidad: Grado de consistencia con que mide el atributo para el que está diseñado. La consistencia implica la ausencia de variabilidad o desacuerdo en repetidas observaciones o mediciones. Grado en que las mismas mediciones de los individuos que se obtienen en diferentes condiciones dan resultados similares.

Fiabilidad inter-observador (inter-ratio): Mide la variación que ocurre entre diferentes observadores al medir a un mismo individuo con un mismo instrumento .

Field: Ver Campo.

Filtros metodológicos: Son estrategias de búsqueda que nos ayudan a encontrar, en bases de datos como Medline, diferentes tipos de estudio sobre tratamiento, diagnosis, etiología o pronóstico. En Pubmed se denominan "Clinical Queries". Para aplicar uno de estos filtros sólo tenemos que combinar nuestra búsqueda con AND al filtro que nos interesa.

Forest/ Forest Plot: Tipo de gráfico ampliamente utilizado para representar los resultados de los diversos estudios seleccionados para una revisión sistemática. El Forest Plot ayuda a decidir si los resultados de los estudios individuales son suficientemente similares para poder ser combinados en una sola estimación.

Generabilidad: Véase Validez externa.

Gold Standard: Es el criterio que define quién tiene "realmente" la enfermedad (o condición) en estudio.

Grado o nivel de significación: En las pruebas de significación estadística, es el valor de p, el cual, en sentido estricto, en un ensayo clínico debe ser preespecificado en la fase de diseño, el nivel aceptado con mayor frecuencia es de 0.05. Ver: P (probabilidad), Significación clínica y Significación estadística.

Grupo de comparación: distintos grupos que intentamos comparar en un estudio analítico, ya sean estudios caso-control, estudios de cohortes o ensayos clínicos.

Grupo control: En un ensayo clínico, el grupo que recibe placebo o bien un tratamiento activo ya conocido, aceptado y establecido. Sin: Grupo de referencia.

Grupo intervención: En un ensayo clínico, el grupo que recibe el tratamiento que se desea evaluar.

Guía farmacoterapéutica: Documento que contiene una relación limitada de medicamentos recomendados para la prescripción en un ámbito determinado, seleccionada a partir de la oferta farmacéutica en función de unos criterios previamente establecidos, con la participación y el consenso de los profesionales a los que va destinada.

Guía de practica clínica: Conjunto de recomendaciones diseñadas para ayudar a los profesionales sanitarios y a los usuarios a seleccionar las opciones diagnósticas y/o terapéuticas más adecuada en el abordaje de una condición clínica específica.

Hipótesis: Suposición, fundada en observaciones o reflexiones, que puede conducir a predicciones refutables. También, conjetura elaborada de forma que pueda ser comprobada y refutada.

History: Archivo de todas las consultas bibliográficas realizadas a lo largo de una sesión en PubMed. También designa al subgrupo de registros que provienen de artículos sobre historia de la medicina.

Holística: Relativa al "holismo" que es la doctrina que propugna la concepción de cada realidad como un todo distinto de la suma de las partes que lo componen.

Impact Factor: Ver Factor de impacto.

Incremento absoluto del riesgo (IAR): Cuando el tratamiento experimental aumenta la probabilidad de un acontecimiento adverso, diferencia absoluta entre las tasas de acontecimientos adversos.

Incremento relativo del riesgo (IRR): Cuando el tratamiento experimental aumenta la probabilidad de un acontecimiento adverso, aumento proporcional de la tasas de acontecimientos adversos en el grupo experimental relativa a la tasa en el grupo de referencia.

Indicador de adhesión a la Guía Farmacoterapéutica: Indicador genérico de prescripción que indica el porcentaje de Dosis Diaria Definida (DDD, véase) de medicamentos prescritos incluidos en la Guía Farmacoterapéutica respecto del total de DDD de medicamentos prescritos.



Índice Bibliográfico Español en Ciencias de la Salud (IBECS): Base de Datos de la Biblioteca Nacional de Ciencias de la Salud (BNCS) del Instituto de Salud Carlos III (ISCIII), contiene referencias de artículos de revistas científico sanitarias editadas en España e incluye contenidos de las diferentes ramas de las ciencias de la salud tales como Medicina (incluyendo Salud Pública, Epidemiología y Administración Sanitaria), Farmacia, Veterinaria, Psicología, Odontología y Enfermería.

Índice Médico Español (IME): Base de datos referencial bibliográfica, creada y gestionada por el Instituto de Historia de la Ciencia y Documentación (IHCD), centro mixto de la Universidad de Valencia y del Consejo Superior de Investigaciones Científicas (CSIC). Cubre 321 publicaciones periódicas editadas en España especializadas en Biomedicina (Administración sanitaria, Farmacia clínica, Medicina experimental, Microbiología, Psiquiatría, Salud pública).

Indización: Proceso por el que el documentalista asigna unos descriptores que describen el contenido conceptual de documento con el fin de facilitar su recuperación

Index: Listado de todos los términos (unitérminos o compuestos) con significado que están incluidos en la Base de Datos. Esta opción permite hacer una búsqueda seleccionando los términos directamente del listado.

Indexar: Ver indización.

Ineficiencia: Situación en la cual los recursos no son empleados de la forma en que se incrementarían los beneficios producidos.

Inferencia: Proceso de pasar de las observaciones y axiomas a la generalizaciones. En estadística, desarrollo de la generalización a partir de los datos de la muestra, generalmente con un margen de incertidumbre.

Instrumentos de medida de CVRS específicos: Los instrumentos específicos incluyen dimensiones de la CVRS de una determinada afección que se quiere estudiar, capacidad funcional, o de un determinado grupo de pacientes.

Instrumentos de medida de CVRS genéricos: Los instrumentos de calidad de vida genéricos no están relacionados con ningún tipo de enfermedad o especialidad en particular. Son aplicables tanto a la población general como a grupos específicos de pacientes.

Interpretabilidad: grado en el que uno puede asignar un significado comprensible a las puntuaciones cuantitativas del instrumento.

Intervalo de confianza: Margen o recorrido comprendidos entre dos valores extremos (el límite inferior y el superior), entre los que cabe esperar el valor medio real de la población con un margen grado prefijado de seguridad. Ver: Nivel de confianza.

Intervención: Cualquier acción (preventiva, diagnóstica, sintomática o curativa) aplicada con el fin de modificar el curso de la situación previa. Cualquier tratamiento o actuación sanitaria que queramos evaluar

ISI: El Institute for Scientific Information fue creado por Eugene Garfield en 1958 con el objeto de publicar el primer Current Contents que cubría unas 200 revistas. En 1971, cubría más de 4000 revistas, publicaba más de 650 páginas de contenidos a la semana, y era consultado por más de un millón de científicos. En 2001, se desarrolló el ISI Web of KnowledgeSM, que integra literatura científica, tanto revistas (journals) como actas (proceedings), patentes, compuestos químicos, y hasta secuencias genómicas. El ISI fue adquirido por The Thomson Corporation, que lo convirtió en el Thomson Institute for Scientific Information. Es, por tanto, una institución privada que produce diversas Bases de Datos de primer orden en calidad y utilidad en la investigación científica, pero que sobre todo ha obtenido un notable éxito por la difusión de lo que se conoce como Impact Factor o Factor de Impacto.

ISSN: International Standard Serial Number, número internación único que identifica de manera inequívoca cada revista.

Ítem: Cada uno de los elementos, preguntas o cuestiones que constituyen un test.

Key word: Ver Palabra clave.

Lectura crítica: Es el proceso de evaluar e interpretar la evidencia aportada por la literatura científica, considerando sistemáticamente los resultados que se presenten, su validez y su relevancia para el trabajo propio. Sin: Evaluación crítica.

Lenguaje documental: Lenguaje artificial creado específicamente para el análisis y recuperación de la información. Contrapuesto al Lenguaje natural, se diferencia de este especialmente por carecer de variabilidad, sinonimia y polisemia.

Lenguaje natural: Expresión de ideas o de conceptos con los términos y con las reglas propios de un idioma o de un grupo especial de hablantes (p.e. sanitarios). Se contrapone a Lenguaje artificial y, específicamente en documentación, a Lenguaje documental.

Límites: Opción que ofrecen varias bases de datos, como Medline, que permite limitar/restringir una búsqueda de acuerdo a unas características como p.ej. artículos publicados en un idioma determinado, tipo de publicación (revisión, ensayo clínico), o grupo de edad, sexo, etc.

Likehood Ratio: Véase Razón de probabilidad.

Literatura inédita: Literatura que no ha sido publicada.

Mapeo Automático de Términos: Ver Acotado Automático de Términos.

Mayor Topic: Materia principal de un artículo. Sin: Descriptor mayor.

Médical Subject Headings: Ver MeSH

Medicina Basada en la Evidencia (MBE): Práctica de la medicina basada en la integración de la maestría clínica de cada médico (es decir, el dominio del conocimiento y el juicio adquirido por la experiencia) con las mejores pruebas clínica externas disponibles. También ha sido definida como la utilización consciente y juiciosa de las mejores demostraciones provenientes de la investigación clínica para la toma de decisiones en el cuidado de cada paciente.

Medicamentos genéricos (Especialidad Farmacéutica Genérica, EFG): Especialidad con la misma forma farmacéutica e igual composición cualitativa y cuantitativa en sustancias medicinales que otra especialidad de referencia, cuyo perfil de eficacia y seguridad está suficientemente establecido por su continuado uso clínico. La EFG debe demostrar la equivalencia terapéutica con la especialidad de referencia mediante los correspondientes estudios de bioequivalencia.

Medicamentos huérfanos: Cuando un fármaco no es desarrollado porque el detentador de la patente considera que no recuperará los costes de este desarrollo (en el caso de enfermedades raras y otras enfermedades no raras pero que ocurren en países pobres en los que no existe un “mercado” potencial), se lo denomina medicamento huérfano.

MeSH (Médical Subject Headings, Encabezamientos de Materias Médicas): Tesauro (diccionario que permite conocer la equivalencia entre los términos del lenguaje natural y los del lenguaje documental) de Medline y de PubMed.

MeSH Subheading: Subencabezamiento propio de MeSH.

MeSH Term: Descriptor propio de MeSH.

Meta-análisis: Integración estructurada y sistemática de la información obtenida en diferentes estudios sobre un problema determinado. Consiste en identificar y revisar los estudios controlados sobre un determinado problema, con el fin de dar una estimación cuantitativa sintética de todos los estudios disponibles.

Método matricial multirasgo-multimetodo (multitrait-multimethod matriz method, MTMM): procedimiento empleado para la validación de constructo aplicando para ello los conceptos de convergencia (validez convergente) y discriminabilidad (validez discriminante).

Método test-retest: Método utilizado para la comprobación de la fiabilidad de un test. Correlación entre la puntuación de dos administraciones de una prueba a los mismos individuos. De esta manera se mide uno de los conceptos incluidos en la fiabilidad: la estabilidad del test.

Modelo de medida: Se refiere a las escalas y subescalas de un cuestionario y a los procedimientos seguidos para crear las puntuaciones.

Muestra: Fracción seleccionada de una población determinada, idealmente representativa de la población de referencia y de tamaño suficiente.

Muestreo: Selección de los sujetos de una población que tomarán parte en un estudio. El muestreo aleatorio implica que cada participante potencial de la población estudiada tiene las mismas probabilidades de ser incluido en la muestra.

NEAR: Operador booleano utilizado en las búsquedas en Librería Cochrane Plus. Buscará documentos donde las dos palabras que enlaza NEAR estén separadas entre sí por un máximo de seis palabras en cualquier orden.

NEXT: Operador booleano utilizado en las búsquedas en Librería Cochrane Plus. Recupera documentos donde las dos palabras están juntas en ese orden.

Nivel de confianza: Se expresa en términos probabilísticos y se interpreta como la probabilidad de que el verdadero parámetro poblacional esté dentro del intervalo de confianza. Ver: Intervalo de confianza.

Normograma de Fagan: Método de cálculo de la probabilidad posttest mediante la representación gráfica de una tabla. Permite conocer la probabilidad posttest mediante el trazo de una línea que una la probabilidad pretest y la razón de probabilidad, el siguiente punto de intersección con la línea situada a la derecha, señala la probabilidad posttest.

Número necesario para dañar (number needed to harm, NNH): Número de pacientes que deben tomar el tratamiento experimental para que se produzca un caso adicional de un determinado acontecimiento adverso, en comparación con los pacientes que reciben el tratamiento de referencia.

Número necesario para tratar (NNT): Es una medida de la eficacia de un tratamiento. Cuando el tratamiento experimental incrementa la probabilidad de un acontecimiento favorable (o cuando disminuye la de un acontecimiento adverso), número de pacientes que hay que tratar para dar lugar a un paciente más con mejoría (o para prevenir un acontecimiento adverso adicional). Se calcula como $1/RAR$.

Odds: Se llama odds a la razón en la que el numerador es la probabilidad de que ocurra un suceso y el denominador es la probabilidad de que tal suceso no ocurra.

Odds post-test: Equivale a los odds (proporción) de que un paciente tenga la enfermedad una vez que se ha realizado el test diagnóstico.

Odds pre-test: Definida como los odds (proporción) de que un paciente tenga la enfermedad antes de que el test diagnóstico se efectúe.

Odds ratio (OR): Resultado de dividir la odds del grupo de interés (expuesto) por la odds del grupo de referencia (control). Si el $OR = 1$ el riesgo o probabilidad de que ocurra el evento es el mismo para ambos grupos (el efecto del tratamiento experimental es el mismo que el del tratamiento control); si el $OR > 1$ se interpretará como que el tratamiento experimental favorece la aparición del evento y si el $OR < 1$ diremos que protege de la aparición del mismo.

Operadores booleanos: Partículas habitualmente expresadas en inglés (AND, OR, NOT) que indican operaciones lógicas entre conceptos. Las operaciones son las de intersección (AND = recupera los documentos que contengan ambos términos), adición (OR = recupera los documentos que contengan tanto un término como otro, o la suma de los dos) o exclusión (NOT = sirve para excluir documentos que tienen un determinado término).

P: Seguía de la abreviatura n.s. (no significativa) o del símbolo $<$ o una cifra decimal, indica la probabilidad de que la diferencia observada en una muestra haya ocurrido puramente por azar siendo los grupos comparados realmente semejantes, es decir bajo la hipótesis nula.



Palabra clave: Término propio del lenguaje natural, sin otras normas que las del habla de la cual procede, que puede utilizarse en la búsqueda de información. Con frecuencia se utiliza como sinónimo de Descriptor, pero este uso no es correcto y conviene evitarlo. En MeSH las palabras clave son el equivalente a los términos de entrada (Entry Terms) de un diccionario de idiomas. Sin: Key word.

Palabra de texto: PubMed considera “palabras de texto” a todos los números y términos con significado que aparecen, entre otros menos importantes, en los campos título (title [TI]), resumen (abstract [AB]), descriptores (MeSH terms [MH]), subencabezamientos (MeSH Subheadings [SH]), así como los nombres de sustancias químicas (Substance Name [NM]) y los nombres de personas (Personal Name [PS]) cuando unas y otras son la materia del artículo. Sin: Text Word.

Panel de expertos: Reunión de un grupo de profesionales de prestigio en una determinada materia que abordan, exponen, contrastan ... sus opiniones y experiencias sobre la actuación en un tema preocupante de interés para el que se busca la mayor uniformidad de criterios que no es posible obtener por otro tipo de estudios más eficientes.

Perdidas de seguimiento: Aquellos pacientes que son perdidos durante el estudio y en los cuales no podemos valorar los resultados finales.

Población: Grupo individuos, objetos, mediciones, etc., con determinadas características demográficas, de donde se obtiene la muestra y a la que se quiere hacer inferencia de los resultados de un estudio.

Positivismo lógico: También denominado neopositivismo. Movimiento filosófico contemporáneo que destaca en filosofía la importancia del análisis del lenguaje y de la metodología científica.

Pretest cognitivo: Conjunto de estrategias y procedimientos que se utilizan para determinar si el cuestionario funciona tal y como lo ha concebido el investigador. Debe reproducir exactamente las condiciones en las que se responderá el cuestionario definitivo y, por tanto, tiene que realizarse con el cuestionario terminado y empleando el mismo procedimiento de recogida de información, puesto que la única diferencia entre el pretest y el trabajo de campo es el número de cuestionarios contestados.

Prevalencia: Número de casos de una enfermedad en una población y en un momento dados.

Prevalencia del test: Número total de positivos obtenidos al realizar una prueba diagnóstica en relación al total de sujetos que componen la muestra estudiada.

Probabilidad: Es el cociente entre los casos posibles de que ocurra un hecho dividido por el total de casos.

Probabilidad pretest o preprueba: Proporción de pacientes que tienen la enfermedad en la muestra estudiada, y cuyo diagnóstico es fehacientemente establecido por un estándar de referencia. Es la prevalencia de la enfermedad.

Probabilidad postest o posprueba: Probabilidad de tener una enfermedad tras la realización de una prueba diagnóstica.

PubMed: Base de datos de bibliografía sobre ciencias de la salud. Puede obtenerse de ella información acerca de publicaciones de carácter científico y de contenido médico (en su sentido más amplio, incluyendo enfermería, odontología, veterinaria...).

Randomización: Ver Aleatorización.

Razón de probabilidad: Probabilidad de un resultado en los enfermos dividido por la probabilidad de ese mismo resultado en no enfermos. Sin: Cociente de probabilidad, Likelihood Ratio.

Reacciones adversas a medicamentos (RAM): Según la OMS, “reacción nociva y no deseada que se presenta tras la administración de un fármaco, a dosis utilizadas habitualmente en la especie humana, para prevenir, diagnosticar o tratar una enfermedad, o para modificar cualquier función biológica”. Las reacciones adversas se clasifican en dos grandes grupos: RAM tipo A, de tipo cuantitativo o predecibles, y RAM tipo B, de tipo cualitativo o no predecibles. Sin: Efecto indeseado.

Record: Ver Asiento.

Reducción absoluta del riesgo (RAR): Diferencia aritmética absoluta entre la tasa de acontecimientos (riesgo) en el grupo experimental y la tasa de acontecimientos (riesgo) en el grupo control. También se puede denominar: Diferencia de riesgos.

Reducción relativa del riesgo (RRR): Cociente entre la reducción absoluta del riesgo (RAR) y el riesgo en el grupo control. A la reducción relativa del riesgo se le llama también “fracción atribuible” o “diferencia relativa del riesgo”.

Registro: Ver Asiento.

Regresión logística: Tipo de análisis de regresión (véase) en el que la variable dependiente sólo puede tomar dos valores. En el análisis de regresión la variable dependiente puede ser o bien categórica o bien continua (véase).

Repetibilidad: Referida a una prueba, es el grado en que los resultados son idénticos o estrechamente semejantes cada vez que aquella se lleva a cabo por los mismos profesionales y los mismos emplazamientos.

Reproducibilidad: Grado en que los resultados son idénticos o estrechamente semejantes cada vez que una prueba se lleva a cabo por distintos profesionales y es distintos ámbitos. Sin: Estabilidad.

Retroalimentación: Método para mantener la acción o eficacia de un sistema mediante la continua revisión de los elementos del proceso y de sus resultados con el fin de realizar las modificaciones necesarias.

Reunión educacional: Visita individualizada de un educador a un profesional de la salud en su ámbito de trabajo; en ella se suelen combinar técnicas de auditoria y retroalimentación.

Revisión: Se denomina así a todo intento que sintetice los resultados y las conclusiones de dos o más publicaciones relacionadas con un tema dado.

Revisión bibliográfica: recopilación de literatura realizada sobre un tema concreto en un intervalo de tiempo concreto del que se realiza un resumen y en el que no quedan explícitos los criterios de búsqueda y se destacan las conclusiones a criterio del autor.

Revisiónes narrativas: Se basan en el análisis que hace un experto de los estudios originales existentes sobre un determinado tema o problema para posteriormente subrayar aquellos resultados que considera más importantes.

Revisiónes sistemáticas: Revisión en las que la evidencia sobre un tema ha sido sistemáticamente identificada, criticada y resumida de acuerdo a unos criterios predeterminados.

Riesgo: Resultado predecible o incierto desfavorable de una actividad, intervención o exposición. Probabilidad de que ocurra un determinado fenómeno adverso.

Riesgo relativo (RR): Cociente entre el riesgo de padecer un acontecimiento entre los individuos expuestos a un factor de riesgo determinado y el riesgo de padecerlo entre los no expuestos. También se puede denominar: Cociente de riesgos.

Seguimiento: Periodo de duración de un estudio durante el cual se mantienen contactos periódicos con los participantes en el mismo con el fin de evaluar las variables de estudio.

Sensibilidad: Referida a una prueba diagnóstica, es la proporción de personas verdaderamente enfermas que ha sido catalogadas como tales mediante dicha prueba.

Sensibilidad al cambio (de un instrumento): se refiere a la medida en que detecta y mide de manera fiable las variaciones más pequeñas de un atributo, rasgo, dimensión o constructo.

Serie de casos: Tipo de estudio que describe una serie de características poco conocidas de un proceso o tratamiento y facilitan el conocimiento del mismo y su evolución en el tiempo.

Sesgo: Error sistemático producido en la estimación de los resultados, a causa de defectos en el diseño del estudio, recogida de datos o seguimiento.

Sesgo de desgaste: Diferencias sistemáticas que se producen entre los grupos de un ensayo clínico por la pérdida de participantes durante el estudio.

Sesgo de información: Sesgo relacionado con la forma en la que se obtuvieron los datos. Entre ellos se incluyen los errores cometidos en la valoración del estado de los pacientes al inicio, durante o al final del estudio; también los que ocurren cuando los datos se obtienen de fuentes sin que se conozca la calidad de las mismas. Algunos de los sesgos de información serían: el sesgo del recuerdo o de memoria, sesgo de sospecha diagnóstica, sesgo del entrevistador...

Sesgo de interpretación de las pruebas: Error sistemático que surge debido a la realización de las pruebas de forma no independiente y a ciegas con la consecuencia de que el conocer los resultados de una prueba influye en los de la otra.

Sesgo de publicación: Sesgo que surge debido a la tendencia a publicar sólo estudios con resultados positivos.

Sesgo de selección: Error sistemático consecuencia de que las características de los sujetos incluidos en un estudio son diferentes de las características de los no incluidos, de modo que la muestra no es representativa de la población de referencia. Se incluyen en este grupo el sesgo de los voluntarios, el de prevalencia-incidencia, el de los emigrantes, el de selección no aleatorizada, el de Berkson, el de pertenencia a grupo, el del trabajador sano...

Sesgo de sospecha diagnóstica: Sesgo que ocurre cuando la interpretación de una prueba se haga conociendo el resultado de otra.

Sesgo de reproducibilidad: Sesgo debido a la ausencia de información detallada de las condiciones en que se ha aplicado la prueba, así como las definiciones de los resultados positivos y negativos.

Sesgo de resultados no interpretables: Error sistemático consecuencia de no incluir pacientes en diferentes estadios de la enfermedad.

Sesgo de verificación o de confirmación diagnóstica: Sesgo que ocurre cuando el estudio se limita a individuos sobre los que se realizó con anterioridad la prueba "gold Standard", que suelen ser los de mayor probabilidad de padecer enfermedad.

Sesgo en el espectro de la enfermedad: ¿

Sesgo por ausencia de "gold Standard": Error sistemático que se produce cuando entre los criterios de referencia que se emplean en la comparación de la utilidad de una prueba diagnóstica (en ausencia de "gold Standard") se incluye la prueba de estudio.

Sesión bibliográfica: Herramienta docente consistente en reuniones entre profesionales que realizan una lectura crítica de un artículo científico.

Significación clínica: Probabilidad de que una diferencia observada tenga una repercusión sobre el curso del problema o enfermedad tratados que sea relevante para un paciente dado o para un conjunto de pacientes.

Significación estadística: Probabilidad de que una diferencia observada sea resultado de la casualidad y no de los determinantes causales en un estudio. El hallazgo de una significación estadística no implica necesariamente significación clínica.

Síntesis: Fuentes de información que recogen resúmenes estructurados de revisiones sistemáticas de alta calidad o de artículos originales.

Síntesis: Fuentes de información que incluyen revisiones sistemáticas de alta calidad metodológica.

Sistema: Fuente de información que se caracteriza por abarcar una gran variedad de temas, proporcionar un resumen de los resultados de revisiones de alta calidad y ser actualizado con frecuencia.

Subencabezamiento: Término que delimita el significado del descriptor al que acompaña. P.e. “Drug therapy” (tratamiento farmacológico) como subencabezamiento unido al descriptor Hypertension nos informaría que un artículo trata, entre todos los aspectos posibles de la hipertensión, su tratamiento farmacológico. Sin: Subheading.

Subheading: Ver subencabezamiento.

Tag: Ver Etiqueta.

Tamaño de la muestra: Número mínimo necesario de sujetos que participan en un estudio, pero a su vez suficiente para proporcionar una respuesta fiable al objetivo final de dicho estudio. La determinación de dicho tamaño mínimo es de naturaleza estadística y depende de varios factores.

Técnica de grupos conocidos: Procedimiento utilizado para la validación de constructo según la cual el instrumento se aplica a los grupos que, según se presume, diferirán respecto del atributo crítico a causa de una característica conocida.

Tecnología sanitaria: Conjunto de medicamentos, aparatos, procedimientos médicos y quirúrgicos usados en la atención médicos y los sistemas organizativos y de soporte con los que se presta la atención sanitaria.

Teoría estadística de Bayes o Teorema de Bayes: Establecimiento inductiva de la probabilidad de una hipótesis a partir de algunas observaciones. Según el teorema de Bayes, la probabilidad posterior (o a posteriori) de una hipótesis es igual a una constante, multiplicada por la verosimilitud o posibilidad de la hipótesis (definida como la probabilidad de observar unos acontecimientos determinados si la hipótesis considerada fuera de hecho cierta) y por la probabilidad anterior de la hipótesis.

Término MeSH: Ver MeSH Term

Término de entrada: Relacionado con Palabra clave (véase).

Tesaurus: Es un repertorio de términos con expresión de las relaciones semánticas que los unen. Habitualmente consta de dos partes, una sección categorizada, que expresa las relaciones jerárquicas, genérico-específicas, entre los términos normalizados (descriptores), lo que nos permite conocer el campo semántico (significado) que abarcan los distintos términos. La sección alfabética es una ordenación de todos los términos, normalizados (descriptores) o no normalizados (palabras clave), que de forma similar a un diccionario de idiomas, permite conocer las relaciones de equivalencia entre unos y otros términos. Sin: Thesaurus.

Test Chi cuadrado (2): Prueba estadística utilizada para determinar el grado de asociación entre variables categóricas (véase).

Test de McNemar: Prueba de significación estadística para probar la hipótesis nula de inexistencia de cambios en la proporción de sujetos que experimentan un acontecimiento, cuando cada individuo es evaluado dos veces (en condiciones diferentes) y los datos están emparejados.

Test T de Student: Prueba de significación estadística paramétrica (véase) para contrastar la hipótesis nula respecto a la diferencia entre dos medias.

Test U de Mann-Whitney: Prueba de significación estadística no paramétrica (véase) para probar la hipótesis nula de que el parámetro de localización (generalmente la mediana) es el mismo cuando se comparan dos grupos independientes, cualquiera que sea el tipo de distribución de la variable (distribución normal o de otro tipo).

Test de Wilcoxon: El contraste de Wilcoxon es la técnica no paramétrica paralela al test T de Student para muestras apareadas.

Text Word: Ver Palabra de texto.

Thesaurus: Ver Tesoro.

Transferabilidad: Término acuñado en investigación cualitativa sugiere que en la medida en que el investigador ha respetado el criterio de maximizar la diversidad en la obtención de la muestra y realice una detallada descripción del contexto y participantes, los hallazgos pueden ser aplicables en contextos similares.

Triangulación: Técnica empleada en investigación cualitativa para controlar sesgos. Consiste en obtener distintas perspectivas del fenómeno estudiado utilizando diferentes métodos de investigación (cualitativos y cuantitativos). La triangulación puede ser simultánea o secuencial.

Truncado: El término truncar significa en español "cortar el extremo de una cosa". Aplicado a un término de búsqueda consiste en separar del término su terminación y sustituirla por un símbolo comodín (en PubMed es un asterisco *-). El resultado es la selección de todos los términos que contienen la raíz superviviente.

Umbral de decisión: Nivel de probabilidad a partir del cual el beneficio previsible de toma una decisión de intervenir iguala o mejora al beneficio de no tomarla

Validez: Capacidad de un instrumento para medir aquello para lo que ha sido diseñado, denota el grado en que el instrumento mide lo que se supone que debe medir.

Validez aparente o facial: Forma especial de la validez de contenido que se refiere al hecho de que el instrumento parezca, de la impresión a las personas evaluadas, de que es adecuado para medir lo que se pretende medir.

Validez concurrente: Denota la capacidad de un instrumento de distinguir a los individuos que difieren en su condición actual respecto de un criterio.

Validez convergente: Aspecto de la validez de constructo que se refiere al grado en que diferentes formas de medir el mismo constructo (o dimensiones de un constructo) se correlacionan la una con la otra.

Validez de constructo: Grado en que una prueba o instrumento mide un determinado rasgo, característica o constructo.

Validez de contenido: Determina si el instrumento contiene ítems representativos de todas las dimensiones que forman la definición del concepto o constructo en estudio.

Validez de criterio: Concordancia entre los resultados de la medida del instrumento y los resultados de la medida realizada mediante un criterio o patrón de referencia (gold standard) de lo que se quiere medir.

Validez discriminante o divergente: Aspecto de la validez de constructo que demuestra que una medida no se correlaciona con otras medidas cuya finalidad es medir rasgos, dimensiones o constructos diferentes a los que ella mide.

Validez externa: Grado en que las conclusiones obtenidas con la muestra de población que participa en un estudio pueden ser generalizadas a su población de referencia o a otras poblaciones, lugares, momentos e investigaciones.

Validez interna: Grado en que los resultados de un estudio miden la realidad para la muestra de población que participa en un estudio. Grado en que un estudio está libre de sesgos.

Validez predictiva: Se refiere a la idoneidad de un instrumento para diferenciar entre el desempeño o las conductas de los individuos respecto de un criterio futuro.

Valor global de un test: Proporción de sujetos correctamente clasificados por el test en relación al total de sujetos que componen la muestra estudiada.

Valor predictivo negativo: Referido a una prueba diagnóstica, la probabilidad de que una persona con un resultado negativo sea un verdadero negativo.

Valor predictivo positivo: Referido a una prueba diagnóstica, la probabilidad de que una persona con un resultado positivo sea un verdadero positivo.

Variabilidad: Variación entre individuos debido a determinantes genéticos, ambientales, patológicos...(variabilidad interindividual) u oscilación de variables biológicas en un mismo individuo según las circunstancias (variabilidad intraindividual).

Variable: Cualquier atributo, fenómeno o hecho que pueda tener diferentes valores.

Variable categórica: Se dice de la variable que presenta interrupciones. Una variable es categórica si, entre varios valores potencialmente observables, hay un valor que no puede ser observado. Ejemplo: variables binarias.

Variable continua: Se dice de la variable que no presenta interrupciones. Una variable es continua si, entre dos valores observables, siempre existe la posibilidad de que haya otro valor observable. Ejemplos: edad, peso, altura.

Verbatim: Su significado es: Literalmente, o palabra por palabra. Anglicismo utilizado en investigación cualitativa para denominar a la transcripción exacta, en la fase de exposición de resultados, de las expresiones surgidas durante el grupo de discusión de los participantes, es decir conservado la expresión literal grabada.

Verdaderos negativos: Grupo de pacientes sanos en los que el test es negativo.

Verdaderos positivos: Grupo de pacientes con enfermedad y que además son detectados por el test.

