

Sistemas informáticos de soporte a la decisión clínica

Julio Bonis^a, Juan J. Sancho^{a,b} y Ferran Sanz^a

^aGrup de Recerca en Informàtica Biomèdica (GRIB). Institut Municipal d'investigació Mèdica. Universitat Pompeu i Fabra. Barcelona.

^bServei de Cirurgia. Hospital del Mar. Institut Municipal d'Assistència Sanitària. Barcelona. España.

La presente revisión presenta la evolución histórica y el fundamento metodológico de los sistemas informáticos de ayuda a la decisión clínica. También se revisan las metodologías de verificación, validación y evaluación del impacto clínico de dichos sistemas. Entre otros factores, la proactividad y la usabilidad de los sistemas condicionan su aceptación por parte de los profesionales sanitarios. Las metodologías de evaluación del impacto clínico se describen con especial detalle y se actualiza la revisión de los estudios de impacto clínico llevada a cabo por Hunt et al en 1998.

Palabras clave: Sistemas de soporte a la decisión clínica. Revisión. Estudios de evaluación.

Computer-assisted clinical decision support systems

This review deals with the historical evolution and methodological basis of computerized medical decision support tools. It also reviews the verification, validation and impact evaluation methods of these systems. The acceptability of them from the health professionals depends on the pro-activity and usefulness of the systems, between other factors. The clinical impact evaluation methods are described in deep detail and the Hunt et al review of 1998 is updated.

Key words: Clinical decision support systems. Review. Evaluation studies.

Las decisiones clínicas pueden tomarse mediante razonamiento deductivo a partir del conocimiento de la fisiopatología humana. Más frecuentemente, se toman basándose en datos inciertos mediante estimaciones de probabilidades. Los clínicos realizan esta estimación de probabilidades de manera no formal y sin ser enteramente conscientes del proceso que realizan^{1,2}.

La toma de decisiones es un proceso complejo que requiere el manejo de mucha información. La tasa de errores en la toma de decisiones clínicas es lo suficientemente elevada como para ser objeto de preocupación. Algunos informes hablan de que los errores médicos podrían ser la octava causa de muerte en los países industrializados³. Aunque existen numerosas causas de error médico, más allá del error en la toma de decisiones (como el error de ejecución), el informe «To Err is Human»³ cita un estudio en el que se detectó que hasta un 45,8% de los pacientes de una unidad de cuidados intensivos presentaron un «episodio adverso», definido como «situación en la cual una decisión inapropiada fue adoptada, cuando, al mismo tiempo, una alternativa apropiada podría haberse escogido».

Casi desde la aparición de los primeros ordenadores se ha pro-

puesto que las tecnologías de la información podrían ayudar al diagnóstico y tratamiento médicos. Inicialmente se escribieron programas que pretendían emular las capacidades diagnósticas de los clínicos, y se denominaron «sistemas expertos», con la pretensión de que iban a ser capaces de suplir la actividad deductiva de los clínicos. El estrepitoso fracaso de su introducción en la práctica, incluso cuando demostraban ser capaces de generar hipótesis correctas, ha desvelado muchos aspectos, tanto formales como prácticos, que condicionan la toma de decisiones en el ejercicio médico. Actualmente, los denominados sistemas de soporte a la decisión clínica (SSDC) son menos ambiciosos pero más efectivos. Un SSDC se define como cualquier sistema o programa informático diseñado para ayudar a los profesionales sanitarios a tomar decisiones clínicas, ya sean preventivas, diagnósticas o terapéuticas.

Breve recorrido histórico

Ya a mediados del siglo pasado, se pensó en utilizar los ordenadores para el diagnóstico clínico⁴. Los primeros sistemas diseñados a finales de los años cincuenta trataban toscamente de realizar diagnósticos a partir de los datos introducidos por los clínicos en el ordenador. Se basaban en distintas estrategias, de las que la más inmediata es la de los algoritmos lógicos (del estilo «si A, hacer B y si no A, hacer C»). Un ejemplo clásico es el sistema de diagnóstico de alteraciones electrolíticas de Bleich⁴. Otra estrategia, cuyos pioneros fueron Warner et al⁵, se basa en la realización de razonamientos probabilísticos mediante el uso intensivo del teorema de Bayes. En dicho marco, De Dombal desarrolló un sistema de soporte a la decisión diagnóstica sobre apendicitis aguda, que ha sido fuertemente validado y que aún se usa en la práctica clínica⁶.

A finales de los setenta y durante los años ochenta, se desarrollaron diversos programas que obviaban la asunción de independencia de los sistemas bayesianos simples mediante el uso de redes bayesianas. Un ejemplo es el de Ben-Bassat et al⁷, integrado en el sistema MEDAS y orientado al diagnóstico de múltiples enfermedades.

Una tercera estrategia para el abordaje del diseño de sistemas de soporte a la decisión lo constituye el razonamiento heurístico o simbólico como el sistema HEME, para el diagnóstico de trastornos hematológicos, desarrollado por Lipkin et al⁸. Otros sistemas basados en la aproximación heurística son el INTERNIST-1 de Pople⁹, que abarca unas 600 enfermedades y unos 4.500 signos y síntomas, o el MYCIN de Shortliffe¹⁰, que sugiere terapias antimicrobianas. Otro tipo de sistemas basados en reglas o heurísticas son aquellos que generan avisos que recuerdan al clínico la necesidad de realizar alguna tarea concreta cuando un paciente cumple una serie de criterios. Un ejemplo de estos sistemas es el HELP de Warner et al¹¹.

Si durante los años cincuenta y sesenta se fundaron las bases teóricas para el desarrollo de SSDC, en los setenta se diseñaron e implementaron los primeros sistemas, que se ejecutaban en grandes ordenadores (mainframes). Durante los

Correspondencia: Prof. F. Sanz.
Grup de Recerca en Informàtica Biomèdica. GRIB (IMIM-UPF).
Dr. Aiguader, 80.
08003 Barcelona. España.

Parcialmente financiado por la Comisión Europea mediante el Proyecto SMARTIE (IST-2000-25429).

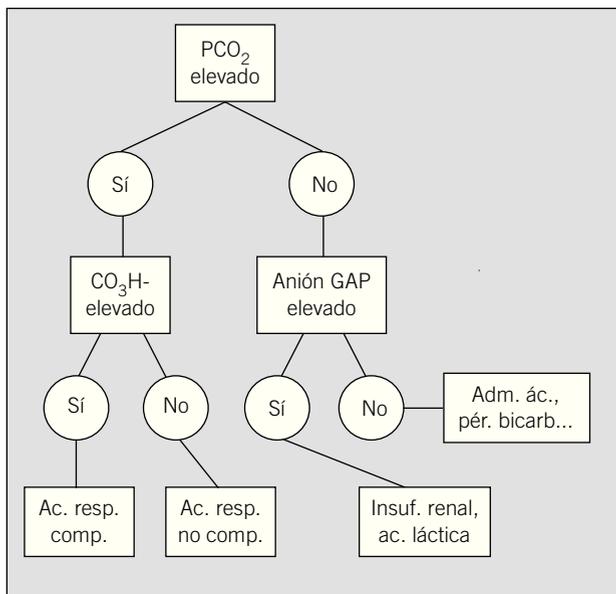


Fig. 1. Algoritmo de decisión en caso de acidosis.

ochenta asistimos a la generalización de los ordenadores personales. Unos años después, la interfaz hombre-máquina mejoró gracias a interfaces gráficas más intuitivas. Desde finales de los noventa se están popularizando los ordenadores de bolsillo y los sistemas embebidos, es decir, la integración de microprocesadores en todo tipo de elementos (relojes de pulsera, esfigmomanómetros e incluso en la ropa).

Aparte de los cambios tecnológicos, se ha producido además un cambio en la filosofía de desarrollo de los SSDC. Hasta finales de los años ochenta el modelo seguido en el desarrollo era el que Miller¹² ha denominado «modelo del Oráculo de Delfos». Según este modelo un SSDC debía ser capaz de, una vez entrados los datos necesarios, elaborar una respuesta correcta y explicar su cadena de razonamiento. El modelo de oráculo de Delfos tiene importantes limitaciones. La principal es que el trabajo requerido para alimentar al sistema con los datos requeridos es muy costoso. Al médico le tomaba más tiempo introducir toda la información que realizar el diagnóstico por sus propios medios. La mayor parte de dichos sistemas no se utilizaban en la práctica real y quedaban como prototipos experimentales. Los modelos más recientes de SSDC tratan de aprovechar sinérgicamente las potencialidades de los sistemas informáticos y de los expertos humanos. Ya no se trata de que el ordenador suplante al médico, sino que en cierto modo colabore con él. Este nuevo enfoque complica los métodos de evaluación pues se hace necesario evaluar también la interacción entre ambos.

Tipología de los sistemas de soporte a la decisión

Los SSDC pueden caracterizarse de acuerdo con múltiples dimensiones, tales como: a) el objetivo que persigue el sistema, b) la forma en que se ofrece la ayuda, o c) el mecanismo de toma de decisión subyacente.

Según su objetivo, los SSDC pueden clasificarse en dos grupos: los que ayudan en decisiones diagnósticas (¿qué tiene el paciente?) y los que ofrecen soporte a decisiones sobre actividades preventivas, diagnósticas (solicitud de pruebas) o terapéuticas que responden a la pregunta «¿qué hacer con el paciente?». En el segundo grupo se debe valorar el coste/beneficio.

Según la forma en que los sistemas ofrecen su ayuda, pueden clasificarse en pasivos o activos. En los sistemas pasivos (INTERNIST-1), el médico debe reconocer en algún momento que el SSDC le puede ser útil y consultarlo. Los SSDC más avanzados toman un papel proactivo. Sin necesidad de ser requeridos, detectan determinadas condiciones en un paciente y alertan al clínico (HELP). Los sistemas de recordatorios empleados con éxito en la promoción de medidas preventivas son un ejemplo, así como los SSDC integrados en los sistemas de prescripción electrónica que se activan cuando detectan interacciones medicamentosas potencialmente peligrosas¹³.

Según el método de razonamiento subyacente, en general los SSDC se pueden clasificar en los siguientes grupos:

SSDC basados en algoritmos o en lógica categórica

Consisten en trasladar a un programa informático un algoritmo (fig. 1) de decisión categórico previamente diseñado por clínicos.

Muchos de los SSDC que han demostrado su eficacia en entornos clínicos reales se basan en este modelo. En general son adecuados cuando se trata de abordar un problema muy concreto y donde la toma de decisiones es esencialmente categórica. El sistema puede hacerse tan complejo como se desee, dependiendo del grado de detalle con el que se desarrolle. Cabe destacar aquí uno de los estándares más extendidos de construcción de algoritmos de decisión clínica. Se trata del «Arden Syntax for medical logic modules»¹⁴. Muchos de estos sistemas han sido rechazados por los clínicos, que los consideran demasiado simplistas¹⁵. Además, la utilidad de implementar uno de estos algoritmos en un ordenador es discutible. Los algoritmos puros (sin ningún componente de cálculo numérico complejo) pueden sustituirse por versiones impresas en papel que pueden ser preferibles¹⁶ al permitir una visión global del algoritmo.

SSDC basados en modelos bayesianos simples

Se basan en la formulación básica del teorema de Bayes, una fórmula matemática que permite calcular la probabilidad *a posteriori* de un episodio (tras conocer los resultados de ciertas pruebas diagnósticas) a partir de su probabilidad *a priori* y de probabilidades condicionadas como, por ejemplo, la sensibilidad y especificidad de las pruebas diagnósticas. Sistemas como el de De Dombal et al⁶ se basan en la aplicación repetida del teorema de Bayes.

En su formulación completa, el teorema es difícil de aplicar al requerir una cantidad ingente de información. Por esto se introducen dos aproximaciones:

1. *Exhaustividad y exclusividad en el diagnóstico.* Es decir, se supone que existe un número limitado de diagnósticos que son considerados en su totalidad (exhaustividad) y que no pueden darse dos de dichos diagnósticos simultáneamente en el mismo paciente (exclusividad).

2. *Independencia condicional.* Es decir, que para cada diagnóstico la influencia de un determinado hallazgo es independiente de que se hayan encontrado otros, por lo que la probabilidad de tener un determinado diagnóstico si un hallazgo es positivo es:

$$\text{Prob Posteriori} = \frac{P \cdot S}{P \cdot S + [(1 - P) \cdot (1 - E)]}$$

donde P es prevalencia, S es sensibilidad, E es especificidad.

La exclusividad y exhaustividad hacen que el teorema no sea válido para pacientes con varias enfermedades concomitantes que se relacionan con los síntomas, signos o pruebas diagnósticas considerados. Sin embargo, la limitación más seria es la asunción de independencia, premisa que es difícilmente asumible en muchos casos.

SSDC basados en redes bayesianas y árboles de decisión

Para superar algunas de las limitaciones de los modelos bayesianos simples, durante los ochenta se trabajó en modelos que combinaban la teoría de probabilidades junto con la teoría de grafos (INTERNIST-1).

Una red bayesiana es un grafo acíclico conexo más una distribución de probabilidad sobre sus variables, que cumple la propiedad de separación direccional. La separación direccional es una propiedad matemática que asegura que la probabilidad de una variable x sólo se ve afectada por las variables (nodos) que son sus ascendientes en primer grado (padres directos).

Los nodos representan a las variables que cumplen los criterios de exclusividad y exhaustividad (representarían al modelo probabilístico clásico de Bayes), y los arcos que unen los nodos representan relaciones probabilísticas entre nodos (he aquí la aportación de la teoría de grafos). Mediante los arcos, que son dirigidos (es decir, tienen un sentido de A a B) podemos representar por tanto relaciones de causalidad entre variables (nodos) y decir que A influye en B.

Además de indicar la estructura del grafo (nodos y arcos que los conectan), se debe indicar la distribución condicional de probabilidad para cada nodo. En términos generales, cabe considerar a la distribución condicional de probabilidad como una tabla que indica la probabilidad de que una variable (nodo) tome cada uno de sus posibles valores, dependiendo de la combinación de valores que tomen los nodos padres, es decir, aquellos que están conectados al nodo hijo mediante un arco.

SSDC basados en redes neuronales

Las redes neuronales son sistemas de cálculo basados en arquitecturas modulares, organizadas en redes de numerosos procesadores elementales interconectados, que evalúan localmente una función no lineal. Las redes neuronales disponen de algoritmos de aprendizaje que modifican el valor de sus parámetros (pesos) con el fin de ajustarse lo mejor posible a un conjunto de casos (*training set*). Las redes neuronales deben siempre probarse sobre un conjunto de casos distinto del anterior (*test set*). Resultan apropiadas para la resolución de problemas en los que subyacen interacciones entre los factores que permanecen ocultas al observador y que son difíciles de modelizar.

Las redes neuronales actúan como cajas negras, es decir, ningún observador externo puede comprender intrínsecamente cómo una red neuronal llega a una conclusión determinada. Esto ha limitado enormemente la aceptación por parte de los clínicos del uso de redes neuronales en los SSDC.

SSDC basados en conocimiento o representación simbólica

Estos sistemas no trabajan exclusivamente con valores cuantitativos, como en el caso de los sistemas probabilísticos, sino que codifican simbólicamente los conocimientos de expertos mediante reglas.

Evaluación de sistemas de soporte a la decisión

Una parte fundamental en el desarrollo e implantación de SSDC es su evaluación. Aún no existe una metodología de evaluación universalmente aceptada y el tema es objeto de investigación y debate académico¹⁷.

Existen múltiples aspectos de los SSDC que pueden ser objeto de evaluación. La metodología ideal debería abarcar aspectos tales como la verificación de los algoritmos y del correspondiente código informático, la validación de los resultados generados por el sistema, la usabilidad y aceptación por parte de los usuarios, el impacto clínico, social y económico, y el análisis de las consecuencias ético-legales del uso de los SSDC.

La verificación analiza la consistencia interna del SSDC. Es decir, son actividades dirigidas a comprobar que las reglas y los modelos de inferencia que utiliza el sistema son correctos y que están bien programados. La validación constituye un análisis del funcionamiento del sistema centrándose en la calidad de los resultados que genera. Muchas veces se lleva a cabo mediante la comparación de las predicciones generadas por el SSDC para casos retrospectivos de los que se conoce el resultado correcto. El análisis de la usabilidad y aceptabilidad trata de resolver una de las mayores fuentes de fracaso de los SSDC. Los aspectos ético-legales de la implantación de SSDC deben tenerse en cuenta en todo estudio de evaluación, aunque no serán tratados aquí.

Verificación y validación: ¿produce el sistema resultados correctos?

El primer paso es la verificación del SSDC. Los métodos utilizados son variados, incluyendo metodologías propias de ingeniería de software. Muy a menudo son procedimientos experimentales en los que se modifican variables y se observan las consecuencias^{18,19}. Los SSDC cuyas reglas de inferencia son transparentes (como los basados en algoritmos, sistemas o redes bayesianas o modelos de representación simbólica) pueden ser verificados por expertos clínicos. En cambio, en las redes neuronales no es posible pues funcionan como «cajas negras».

Las actividades de validación consisten a menudo en comparar, sobre un conjunto de datos (reales o hipotéticos), los resultados obtenidos por parte del SSDC con los obtenidos mediante el juicio del clínico o directamente con los resultados clínicos reales observados.

Un problema importante aparece cuando no existe un *gold estándar* con el que comparar los resultados. Dos sistemas expertos desarrollados por investigadores españoles y que formaban parte de esta categoría fueron validados comparando la «opinión» del SSDC con la de un amplio grupo de clínicos con distintos niveles de experiencia, al suministrar a todos el mismo conjunto de datos de una serie de pacientes^{20,21}. Para comparar dichas opiniones se analizó el grado de acuerdo entre las opiniones mediante un análisis de conglomerados sobre «distancias» diagnósticas, así como con una modificación del índice kappa para varios observadores. Se consideró que dichos SSDC se comportaban adecuadamente al no resultar sus opiniones más separadas de los expertos clínicos que la de ellos entre sí.

En algunas ocasiones los investigadores desarrollan un SSDC utilizando un conjunto de datos de muy difícil obtención y, por esto, la validación acostumbra realizarse sobre el mismo conjunto de datos usado para la construcción del SSDC. En esta situación se produce un sesgo por el que el SSDC presenta mejor capacidad predictiva sobre dichos casos que sobre otros no usados para el desarrollo del SSDC. Para abordar este problema se suelen utilizar estrategias de

validación cruzada como la «*leave-one-out*» o «dejar uno fuera». Este método consiste en elaborar un modelo predictivo utilizando todos los casos disponibles excepto uno, y utilizar este caso excluido para validar el modelo. Si repetimos este proceso a lo largo de todos los casos disponibles, obtendremos predicciones no sesgadas llevadas a cabo con n modelos ligeramente distintos construidos con $n-1$ casos y muy parecidos al que pretendemos validar (construido con todos los n casos).

Usabilidad: ¿se puede utilizar el sistema en el mundo real?

Tras haber comprobado que el SSDC ofrece los resultados esperados, el siguiente paso es evaluar si será utilizado por sus usuarios finales. Frecuentemente el fracaso de los SSDC reside en problemas de usabilidad²². Por esto, es conveniente evaluar la aceptación por parte de los usuarios antes de la implantación de un SSDC en entornos reales. Esta estrategia de evaluación paso a paso evita malgastar recursos en la implantación de un sistema que luego no sería utilizado²³. Puede realizarse con metodologías cuantitativas o cualitativas.

Entre las cuantitativas encontramos escalas de satisfacción que evalúan distintas dimensiones de la usabilidad. También se analiza la interacción del clínico con la interfaz del SSDC, por ejemplo cuantificando el número de pantallas que se visualizan o los botones que se pulsan. Algunos SSDC consisten simplemente en sistemas alternativos de visualización de datos, como ciertos paneles de control diseñados para anestesiistas que, en entornos experimentales, mejoraron el tiempo de detección de determinados eventos²⁴.

Los métodos cualitativos han merecido una atención especial en los últimos años²⁵. Aunque poco comunes en el ámbito biomédico, tienen gran tradición en otras áreas como las ciencias sociales, en el marco de las cuales se produjo su desarrollo. Incluyen entrevistas mediante preguntas abiertas, estudios etnográficos de campo, interaccionismo social y otras técnicas. Una de sus ventajas es que permiten descubrir las causas de los problemas de usabilidad. Así, Kaplan²⁶ señala que las posibles causas de la falta de aceptación de SSDC serían: *a*) su interferencia en la relación médico-paciente; *b*) la sensación de detrimento en su estatus profesional; *c*) la alteración en el arte de la práctica médica, y *d*) la limitación de la autonomía profesional.

Impacto clínico: ¿mejora el uso del sistema la salud de los pacientes?

Los estudios de impacto clínico pueden clasificarse en estudios descriptivos, meramente observacionales, y estudios comparativos. En los comparativos, el investigador modifica las condiciones de un grupo respecto a otro para tratar así de ver si ocurren diferencias entre ambos²⁷. La dificultad reside en cómo lograr un grupo control que estando aislado sea comparable. Se han propuesto distintos diseños:

Diseños antes-después. El investigador observa la situación basal antes de la introducción del SSDC. En la segunda fase, el sistema es introducido en la práctica diaria y se evalúa si existe un cambio. Esta metodología ha sido puesta en duda en numerosas ocasiones. Entre otros factores, el personal sanitario puede mostrar mayor diligencia durante la segunda fase²⁸. También se pueden obtener conclusiones sesgadas por la contaminación de otros factores temporales externos, como cambios organizativos.

Diseños aleatorizados controlados. Los distintos sujetos se distribuyen en grupos de manera aleatoria sujetos a distin-

tas intervenciones. En los ensayos clínicos farmacológicos, la aleatorización se realiza generalmente sobre pacientes. En el caso de estudios sobre intervenciones no farmacológicas el método de aleatorización se complica debido al sesgo de contaminación. Para evitarlo se suele realizar la aleatorización por *clusters* o conglomerados, esto es, se selecciona a dos grupos de médicos, servicios u hospitales de manera aleatoria, y mientras que unos utilizan SSDC para todos sus pacientes, otros no lo hacen. Por supuesto, ambos grupos deben ser comparables entre sí.

El cálculo del tamaño muestral en el caso de aleatorización por conglomerados esconde ciertas trampas que en demasiadas ocasiones han sido obviadas por los investigadores²⁹. En dicho cálculo debería aplicarse un factor de corrección denominado factor de diseño³⁰.

Tradicionalmente, los investigadores se han centrado en la evaluación del SSDC mediante estudios clínicos aleatorizados y controlados. La extensa revisión de ensayos clínicos sobre SSDC de Hunt et al³¹ desde 1974 a 1998 muestra que, en general, el impacto clínico demostrado ha sido escaso. Los SSDC centrados en el diagnóstico han mostrado poco éxito: sólo uno de cinco estudios demostró cierto impacto. Los sistemas proactivos, basados en recordatorios para la realización de medidas preventivas, sí han demostrado su impacto en un 74% de los casos (14 sobre 19 estudios). Hunt advierte que la mayoría de los estudios tenía tamaños muestrales demasiado pequeños. Además Chuang et al²⁹ ilustraron el problema de la falta de rigor metodológico en el tratamiento de ensayos basados en conglomerados.

La necesidad de metodologías alternativas

Otros autores¹⁷ han señalado que quizá los ensayos clínicos diseñados al estilo de los usados en la evaluación de fármacos no sean la manera más adecuada (al menos por sí solos) para evaluar el impacto de los SSDC. Junto con ellos, Kaplan²⁶ aboga por el uso de metodologías cualitativas que permitan entender las razones de la falta de impacto. En un estudio sobre el uso de un SSDC en el manejo del asma y la angina de pecho no se pudo constatar un impacto significativo debido a que el SSDC no era utilizado²². Las investigaciones muestran que el nivel de satisfacción con el uso de tecnologías informáticas está más relacionado con las percepciones subjetivas del usuario acerca de los efectos del sistema en la productividad que con su efecto en la calidad de los cuidados médicos³².

Fuentes de sesgo

La evaluación del impacto de SSDC es muy compleja debido entre otras razones a los numerosos sesgos potenciales.

En primer lugar está el sesgo del evaluador. En lo posible, el diseño del estudio debe evitar que los evaluadores puedan emitir juicios sesgados. Otra fuente frecuente de sesgo es el sesgo de selección. Así, los clínicos pueden tender a seleccionar a un determinado grupo de pacientes para utilizar con ellos el SSDC. El efecto Hawthorne o sesgo de observación es común en estudios de evaluación de SSDC. En virtud del mismo, los clínicos sujetos a observación y evaluación, por el mero hecho de estar siendo observados, mejorarán sus prácticas y su rendimiento. Otro posible sesgo es la mejora de la toma de decisiones por el simple hecho de llevar a cabo una recogida de datos más estructurada y completa durante la realización del estudio. El efecto por retroalimentación se ha descrito en algunos estudios en los que se observó que al ofrecer a los clínicos información relativa a sus resultados de forma frecuente se les estimulaba a mejorar su rendimiento³³. El sesgo por contaminación

se produce cuando la exposición previa al SSDC influye en la toma de decisiones sobre el grupo control. También debe considerarse el efecto placebo. Se ha argumentado que el hecho de ver que el médico utiliza un sistema informático puede influir en la percepción subjetiva de enfermedad del paciente.

Revisión de los últimos estudios de evaluación de CDSS publicados

La última revisión sistemática de la que tenemos noticia fue realizada por Hunt et al en 1998³¹. Cubre el período comprendido entre enero de 1974 y mayo de 1998. Hemos llevado a cabo una breve revisión de los artículos publicados en revistas indexadas en Medline en el período comprendido entre junio de 1998 y junio de 2003. La metodología de recuperación de artículos no es tan exhaustiva como la de Hunt, pero puede servir como indicativo de la tendencia en los últimos 5 años en este campo.

Hemos identificado 42 estudios, de los cuales 19 eran estudios experimentales de validación interna y 23, estudios de evaluación del impacto clínico. Entre estos últimos se identificaron dos tipologías: a) estudios cualitativos, que constituían un 39% (9)^{23,34-41}, y b) estudios cuantitativos que constituían un 61% (14). En cuanto a los 14 estudios cuantitativos, 8 utilizaban un diseño aleatorizado por conglomerados^{39,42-48} y los otros 6 usaban un diseño del tipo antes/después⁴⁹⁻⁵⁴.

Entre los 6 diseños antes/después, 5 describen resultados globalmente positivos, mientras que uno describe resultados neutros. No se ha identificado ningún estudio antes/después que describa resultados negativos para los SSDC. El tamaño muestral de dichos estudios es muy variable y oscila entre 43 y 1.825 casos, con una media de 719,2. La duración media fue de 28,25 semanas, con un mínimo de 16 y un máximo de 40.

Entre los 8 estudios con diseños aleatorizados por conglomerados, 6 describen resultados globalmente positivos, mientras que en 2 casos se trata de resultados neutros. De nuevo, no se encontró ninguna publicación que describiera resultados negativos para los SSDC. El tamaño muestral de los estudios revisados es nuevamente muy variable, entre 57 y 6.371 casos, con una media de 2.064,6. En cuanto a su duración, la media fue de 39,2 semanas, con un mínimo de 4 y un máximo de 72.

De la revisión bibliográfica se puede concluir que se publican aproximadamente el mismo número de estudios de validación interna que de evaluación del impacto. En los estudios de evaluación, la metodología cuantitativa predomina ligeramente sobre la cualitativa, sin detectar una tendencia clara a favor de un tipo de metodología a lo largo del tiempo. Los estudios de evaluación cuantitativa siguen dos líneas de diseño con parecida prevalencia, estudios de antes/después y estudios aleatorizados por conglomerados. El tamaño muestral y su duración son similares entre ambos tipos de diseño. Existe probablemente un importante sesgo de publicación, ya que no se ha publicado ningún estudio de evaluación con resultados claramente negativos.

En los últimos años, se detecta un interés creciente por la implicación de los pacientes en la toma de decisiones. No es de extrañar que varios SSDC se estén orientando a facilitar la integración de las preferencias del paciente en la toma de decisión, con lo que las opiniones de los pacientes comienzan a tener su importancia en el proceso de evaluación^{23,35,48,50}. Se ha encontrado un grupo importante de SSDC basados en el modelo de Guías de Práctica Clínica. El escenario típico es la implantación de estos SSDC en atención primaria para el manejo de enfermedades crónicas,

como el asma^{34,43}, la diabetes^{23,33,46} u otras^{38,42,45}. Por otra parte, se ha visto que en entornos hospitalarios se prefiere el desarrollo de SSDC más específicos, centrados en la generación de alertas preventivas^{44,49,53} o en el manejo de situaciones muy específicas como ictus⁴¹, petición de pruebas radiológicas⁵¹, cáncer de mama⁵² o próstata⁵⁴, o ventilación mecánica⁴⁷. Un grupo notable lo constituyen los SSDC orientados a unidades de cuidados intensivos^{36,37,47}.

REFERENCIAS BIBLIOGRÁFICAS

- Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. *Science* 1959;130:9-21.
- Joseph GM, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. *Med Decis Making* 1990;10:31-46.
- Kohn LT, Corrigan JM, Donaldson MS, editors. To err is human: building a safer health system. Washington DC: Committee on Quality of Health Care in America, Institute of Medicine, 2000.
- Bleich HL. Computer evaluation of acid-base disorders. *J Clin Invest* 1969;48:1689-96.
- Warner HR, Toronto AF, Veasey LG, Stephenson RA. Mathematical approach to medical diagnosis. *JAMA* 1961;177:75-81.
- De Dombal FT, Hartley JR, Sleeman DH. A computer-assisted system for learning clinical diagnosis. *Lancet* 1969;1:145-8.
- Ben-Bassat M, Carlson RW, Puri VK, Davenport MD, Schriver JA. Pattern-based interactive diagnosis of multiple disorders: the MEDAS system. *IEEE Trans Pattern Anal Mach Intel* 1980;PAMI-2:148-60.
- Lipkin M, Hardy JD. Differential diagnosis of hematological diseases aided by mechanical correlation of data. *Science* 1957;125:551-2.
- Pople HE. Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics. En: Szolovits P, editor. *Artificial intelligence in medicine*. Boulder: Westview Press, 1982; p. 119-90.
- Shortliffe EH. Computer-based medical consultations: MYCIN Artificial Intelligence Series. New York: Elsevier Computer Science Library, 1976.
- Gardner RM, Pryor TA, Warner HR. The HELP hospital information system: update 1998. *Int J Med Inf* 1999;54:169-82.
- Miller RA. Medical diagnostic decision support systems-past, present, and future: a threaded bibliography and brief commentary. *J Am Med Inform Assoc* 1994;1:8-27.
- Bates DW, Leape LL, Cullen DJ, Laird N, Petersen LA, Teich JM, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998;280:1311-6.
- Hripcsak G. The Arden Syntax for medical logic modules: introduction. *Computers in Biology and Medicine* 1994;24:329-30.
- Grimm RH Jr, Shimoni K, Harlan WR Jr, Estes EH Jr. Evaluation of patient-care protocol use by various providers. *N Engl J Med* 1975;292:507-11.
- Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. *BMJ* 2000;320:686-90.
- Heathfield H, Pitty D, Hanka R. Evaluating information technology in health care: barriers and challenges. *BMJ* 1998;316:1959-61.
- Carbó JM, Sancho JJ, Miravittles MJ, González JC, Sanz F. A new approach to computerized medical knowledge assessment. En: K.-P. Adlansing, Grabner G, Bensgton S, Hansen R, editors. *Medical Informatics Europe 1991. Lecture Notes in Medical Informatics*. Berlín: Springer-Verlag, 1991;45:615-9.
- Kuilboer MM, Van Wijk MA, Mosseveld M, Van der Does E, Ponsioen BP, De Jongste JC, et al. Feasibility of asthma critic, a decision-support system for asthma and COPD which generates patient-specific feedback on routinely recorded data in general practice. *Fam Pract* 2002;19:442-7.
- Verdaguer A, Patak A, Sancho JJ, Sierra C, Sanz F. Validation of the medical expert system PNEUMON-IA. *Comp Biom Res* 1992;25:511-26.
- Hernández C, Sancho JJ, Belmonte MA, Sierra C, Sanz F. Validation of the medical expert system RENOIR. *Comp Biomed Res* 1994;27:456-71.
- Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ* 2002;325:941.
- Carroll C, Marsden P, Soden P, Naylor E, New J, Dornan T. Involving users in the design and usability evaluation of a clinical decision support system. *Comput Methods Programs Biomed* 2002;69:123-35.
- Jungk A, Thull B, Hoefft A, Rau G. Evaluation of two new ecological interface approaches for the anesthesia workplace. *J Clin Monit Comput* 2000;16:243-58.
- Kaplan B. Evaluating informatics applications-clinical decision support systems literature review. *Int J Med Inf* 2001;64:15-37.
- Kaplan B. Evaluating informatics applications-some alternative approaches: theory, social interactionism, and call for methodological pluralism. *Int J Med Inf* 2001;64:39-56.

27. Friedman CP, Wyatt JC. Evaluation methods in medical informatics. New York: Springer-Verlag, 1997.
28. Sacks H. Randomised vs historical controls for clinical trials. *Am J Med* 1982;72:233-40.
29. Chuang JH, Hripcsak G, Jenders RA. Considering clustering: a methodological review of clinical decision support system studies. *Proc AMIA Symp* 2000;146-50.
30. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;3:III-92.
31. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998;280:1339-46.
32. Lee F. Implementation of physician order entry: user satisfaction and self-reported usage patterns. *J Am Med Inf Assoc* 1996;3:42-55.
33. Frijling BD, Lobo CM, Hulscher ME, Akkermans RP, Braspenning JC, Prins A, et al. Multifaceted support to improve clinical decision making in diabetes care: a randomised controlled trial in general practice. *Diabet Med* 2002;19:836-42.
34. Rousseau N, McColl E, Newton J, Grimshaw J, Eccles M. Practice based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care. *BMJ* 2003;326:314.
35. Pell I, Dowie J, Clarke A, Kennedy A, Bhavnani V. Development and preliminary evaluation of a clinical guidance programme for the decision about prophylactic oophorectomy in women undergoing a hysterectomy. *Qual Saf Health Care* 2002;11:32-8.
36. Lee TT, Yeh CH, Ho LH. Application of a computerized nursing care plan system in one hospital: experiences of ICU nurses in Taiwan. *J adv Nurs* 2002;39:61-7.
37. Junger A, Michel A, Benson M, Quinzio LA, Hafer J, Hartmann B, et al. Evaluation of the suitability of a patient data management system for ICUs on a general ward. *Int J Med Inf* 2001;64:57-66.
38. Porcelli PJ, Lobach DF. Integration of clinical decision support with on-line encounter documentation for well child care at the point of care. *Proc AMIA Symp* 1999;599-603.
39. Gadd CS, Baskaran P, Lobach DF. Identification of design features to enhance utilization and acceptance of systems for Internet-based decision support at the point of care. *Proc AMIA Symp* 1998;91-5.
40. Farand L, Lafrance JP, Arocha JF. Collaborative problem-solving in telemedicine and evidence interpretation on a complex clinical case. *Int J Med Inf* 1998;51:153-67.
41. Lau F, Penn A, Wilson D, Noseworthy T, Vincent D, Doze S. The diffusion of an evidence-based disease guidance system for managing stroke. *Int J Med Inf* 1998;51:107-16.
42. Jousimaa J, Makela M, Kunnamo I, MacLennan G, Grimshaw JM. Primary care guidelines on consultation practices: the effectiveness of computerized versus paper-based versions. A cluster randomized controlled trial among newly qualified primary care physicians. *Int J Technol Assess Health Care* 2002;18:586-96.
43. McCowan C, Neville RG, Ricketts IW, Warner FC, Hoskins G, Thomas GE. Lessons from a randomized controlled trial designed to evaluate computer decision support software to improve the management of asthma. *Med Inform Internet Med* 2001;26:191-201.
44. Dexter PR, Perkins S, Overhage JM, Maharry K, Kohler RB, McDonald CJ. A computerized reminder system to increase the use of preventive care for hospitalized patients. *N Engl J Med* 2001;345:965-70.
45. Christakis DA, Zimmerman FJ, Wright JA, Garrison MM, Rivara FP, Davis RL. A randomized controlled trial of point-of-care evidence to improve the antibiotic prescribing practices for otitis media in children. *Pediatrics* 2001;107:E15.
46. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H, Furuseth K. Implementing clinical guidelines on the treatment of diabetes mellitus in general practice. Evaluation of effort, process, and patient outcome related to implementation of a computer-based decision support system. *Int J Technol Assess Health Care* 2000;16:210-27.
47. East TD, Heermann LK, Bradshaw RL, Lugo A, Sailors RM, Ershler L, et al. Efficacy of computerized decision support for mechanical ventilation: results of a prospective multi-center randomized trial. *Proc AMIA Symp* 1999;251-5.
48. Flatley-Brennan P. Computer network home care demonstration: a randomized trial in persons living with AIDS. *Comput Biol Med* 1998;28: 489-508.
49. Mekhjian HS, Kumar RR, Kuehn L, Bentley TD, Teater P, Thomas A, et al. Immediate benefits realized following implementation of physician order entry at an academic medical center. *J Am Med Inform Assoc* 2002; 9:529-39.
50. Ruland CM. Handheld technology to improve patient care: evaluating a support system for preference-based care planning at the bedside. *J Am Med Inform Assoc* 2002;9:192-201.
51. Sanders DL, Miller RA. The effects on clinician ordering patterns of a computerized decision support system for neuroradiology imaging studies. *Proc AMIA Symp* 2001;583-7.
52. Bouaud J, Seroussi B, Antoine EC, Zelek L, Spielmann M. A before-after study using OncoDoc, a guideline-based decision support-system on breast cancer management: impact upon physician prescribing behaviour. *Medinfo* 2001;10:420-4.
53. Rocha BH, Christenson JC, Evans RS, Gardner RM. Clinicians' response to computerized detection of infections. *J Am Med Inform Assoc* 2001;8: 117-25.
54. Chang PL, Li YC, Wang TM, Huang ST, Hsieh ML, Tsui KH. Evaluation of a decision-support system for preoperative staging of prostate cancer. *Med Decis Making* 1999;19:419-27.