

## EVALUACION DE LA CONSISTENCIA EN LA INDIZACION DEL REPERTORIO DOCUMENTACION MEDICA ESPAÑOLA

M. F. Abad García\*, I. Abad Pérez\* y R. Aleixandre Benavent\*

**Resumen:** Entre los aspectos más delicados del funcionamiento de un SRI están la selección de los trabajos que determinan su cobertura temática y su indización. La adecuación de estos procesos es difícil de medir ya que no hay sólo una forma correcta de llevarlos a cabo debido a la existencia de variaciones individuales. Un método para evaluar la magnitud de estas variaciones es medir la consistencia en la realización de estas tareas cuando las hacen distintas personas (consistencia inter-indizador) o la misma en tiempos distintos (consistencia intra-indizador). Se trata de evaluar la consistencia inter-indizador en la selección de los trabajos del tercer volumen del repertorio *Documentación Médica Española (1996)* y en la asignación de categoría principal y de analizar las causas de las discrepancias. La re-selección y reindización han sido llevadas a cabo de forma independiente por un indizador externo experto en la materia. El material reselectionado han sido los 370 artículos preseleccionados para su inclusión en el repertorio. Se han reindizado los artículos en los que ambos indizadores coincidían que eran relevantes. El indicador utilizado como medida de la consistencia ha sido la proporción de acuerdo simple entre los dos indizadores. El acuerdo en la selección de trabajos ha sido del 86 % y en la asignación de clase y categoría principal del 78,6 % para el conjunto de categorías. El análisis de discrepancias en el proceso de selección ha revelado un alto grado de acuerdo para los trabajos centrales de la disciplina. En el proceso de clasificación, salvo un conjunto de trabajos de contenido similar para los que los indizadores utilizan criterios distintos, el origen de las discrepancias se debe a errores de ambos indizadores.

**Palabras clave:** evaluación; consistencia en la indización; sistemas de información bibliográficos; documentación médica española.

**Abstract:** The selection of the documents to be included in the original input and their indexing are two of the most difficult tasks in running one IRS. The adequacy of the performance of these tasks is difficult to measure because there is no correct way to do them due to individual differences. One way to measure these differences is to calculate the degree of consistence with which the same and different indexers carry out these processes. The aim is to evaluate the inter-indexer consistency in the selection of documents to be included in the third volume of the repertoire *Documentación Médica Española (1996)* and in the asignment of the main category of the classification and to analyze the causes of discrepancies. The new selection and re-indexing of the material has been

---

\* Instituto de Estudios Documentales e Históricos sobre la Ciencia (Universidad de Valencia-CSIC). Blasco Ibáñez, 19. 46010 Valencia.

Recibido: 15-11-97; segunda versión 26-5-98.

carried out by an external indexer with expertise in the field. 370 documents have been reselected. The re-indexed documents have been those where both indexers agree that were relevant. To measure consistency the agreement proportion between indexers was calculated. The agreement proportion for the selection of the material was 86 % and for the assignation of the main category of the classification was 78.6 %. The analysis of the discrepancies has showed an high degree of consistency for documents that deal with central topics of the field. Indexing discrepancies are mainly due to errors of both indexers, exemption made of a small group of documents of similar contents.

**Key words:** evaluation; indexing consistency; bibliographic information retrieval systems; Spanish medical documentantion.

## 1 Introducción

*Documentación Médica Española (1994-)* (1) es un repertorio de resúmenes, de aparición anual, que proporciona información de referencia sobre publicaciones aparecidas en más de 200 revistas españolas biomédicas o de documentación cuyo contenido está relacionado con la documentación médica. Esta fuente, actualmente editada en soporte papel, tiene prevista su difusión en Internet al final del presente año.

El repertorio ha sido diseñado con el propósito primordial de servir de herramienta de trabajo al colectivo de médicos documentalistas, cuya producción científica no ha estado favorecida, hasta fechas muy recientes, por la existencia de una revista especializada ni por la disponibilidad de fuentes secundarias específicas.

La acogida de esta fuente por sus potenciales usuarios ha sido buena<sup>1</sup>, pero su aceptación e inclusión dentro de su rutina de trabajo va a depender, en gran medida, de que comprendan el funcionamiento de este sistema. Este entendimiento no sólo sirve de ayuda a los usuarios de estos sistemas de información, sino también a sus productores, a quienes les permite buscar métodos para su mejora. En concreto, uno de los aspectos que se consideran cruciales en relación con la adopción del repertorio *Documentación Médica Española* y con su correcto uso es que los usuarios conozcan los criterios y procedimientos relacionados con la consecución y selección de los documentos y datos que entran en el sistema y que conforman su cobertura temática, así como los que afectan a los procesos de clasificación e indización del material seleccionado.

La adecuación y calidad de los procesos antes mencionados es difícil de medir. Así, por ejemplo, los juicios de relevancia en los que se basa la inclusión o no de un documento son siempre subjetivos y están sometidos a una gran variabilidad que depende, entre otros factores, de la formación y aptitudes de la persona que los realiza, del tipo de documento que se valora y del momento y condiciones en los que se emite el juicio de valor (2). Del mismo modo, estas variaciones también afectan al proceso de indización ya que no hay una única manera de indizar correctamente un documento (3). El método para evaluar la magnitud de estas variaciones consiste en obtener la medida de la consistencia o concordancia existente en la realización de es-

---

<sup>1</sup> Como demostró la concesión del premio a la mejor aportación científica en el IV Congreso Nacional de Documentación Médica celebrado en Torremolinos en julio de 1994.

tas tareas cuando la llevan a cabo distintos indizadores (concordancia inter-indizador) o la misma persona en tiempos distintos (concordancia intra-indizador) (4).

Para *Documentación Médica Española*, el conocimiento del grado de concordancia alcanzado en los procesos de selección e indización tras sus primeros tres años de existencia se plantea como un paso necesario para superar su período de implantación o prueba y para establecer cuantas recomendaciones puedan mejorar el funcionamiento de esta fuente durante su rutina posterior, habida cuenta, además, de su inminente inclusión en soporte informático para su difusión por Internet.

En definitiva, los objetivos de este trabajo son: a) evaluar la concordancia entre dos indizadores (uno interno y otro externo al sistema) tanto en la selección de los trabajos del tercer volumen del repertorio (año 1996), como en la asignación de su categoría principal y clase; b) analizar las causas de las discrepancias, y c) proponer medios para mejorar el proceso.

## 2 Material y método

La selección de los trabajos de cada volumen del repertorio se realiza en dos etapas. La primera consiste en seleccionar los trabajos potencialmente relevantes. Esta tarea la llevan a cabo cinco indizadores, todos ellos con formación y experiencia práctica en documentación médica. A cada indizador le corresponde revisar los sumarios de un grupo determinado de revistas. La selección se realiza a partir de la lectura de los sumarios y, cuando es necesario, del resumen y del texto completo de los trabajos. Cuando un trabajo es considerado como potencialmente relevante se selecciona y se fotocopia.

La segunda etapa de la selección consiste en decidir la relevancia final de los trabajos para su inclusión en el repertorio. Esta tarea la realiza un indizador distinto a los anteriores (siempre el mismo para todos los trabajos de esta edición y para los de las dos anteriores) al que denominaremos indizador A, quien siempre comenta sus decisiones con cada uno de los responsables de la selección previa.

Los criterios utilizados para considerar un trabajo relevante han tenido en consideración los diversos perfiles profesionales de los médicos documentalistas, tanto aquéllos que trabajan en el área de la documentación médico científica, como los que lo hacen en el ámbito de los servicios hospitalarios de documentación clínica y admisión, en el ámbito de la documentación e información en atención primaria o en el desarrollo y mantenimiento de los sistemas de información concernientes a la salud pública y la administración sanitaria. Estos criterios han quedado reflejados en la clasificación que utiliza el repertorio (ver apéndice 1)

Los materiales seleccionados para su inclusión en el repertorio se indizan mediante dos procedimientos: la asignación de clase y categoría de la mencionada clasificación y la asignación de palabras clave. La indización de los trabajos también la realiza el indizador A. Para ello, previa lectura de las publicaciones, designa a cada una de ellas la categoría de la clasificación que, según su criterio, refleja mejor el contenido central del documento y bajo el que aparecerá el resumen. Si considera que el documento tiene relación con alguna otra materia, establecerá una categoría adicional que aparecerá en el repertorio como referencia cruzada. Además, asigna tantas palabras clave, extraídas del lenguaje natural, como considera necesarias para representar el contenido del documento y facilitar su recuperación a partir del índice de materias.

Para averiguar en qué medida los criterios que utiliza el indizador A en la realización de las tareas de selección y clasificación coinciden con los que utilizaría otro médico documentalista no involucrado en el funcionamiento habitual del repertorio, se procedió a realizar una reelección y una reindización de los trabajos.

Estas tareas las ha llevado a cabo, de una forma totalmente independiente, un médico documentalista con más de 15 años de experiencia profesional y de formación similar al indizador A, al que se ha llamado indizador B. Este indizador ha sido siempre ajeno al mecanismo y a los criterios utilizados en la realización de los procesos que se evalúan.

Para realizar la reelección, al indizador B se le han proporcionado los 370 trabajos manejados por el indizador A para la confección del volumen 3 del repertorio. El indizador B ha ignorado, en todo momento, las decisiones tomadas por el indizador A. Al indizador B se le dio la instrucción de seleccionar los trabajos según sus propios criterios de relevancia y no se le comentaron, en ningún momento, los habitualmente utilizados. A cada trabajo relevante seleccionado, el indizador B le ha asignado la categoría principal y la clase bajo la que, según su criterio, debía aparecer el resumen en el repertorio.

Para medir la coincidencia en los criterios utilizados por los indizadores A y B, en la realización de los procesos de selección y de clasificación, se ha calculado la proporción de acuerdo simple (5). Esta proporción es la razón entre el número de veces en las que los indizadores A y B coinciden en los resultados de la realización de una tarea respecto del total de ocasiones en que se ha realizado tal comparación, expresada en tanto por ciento. Así, por ejemplo, para medir la concordancia en la clasificación de los trabajos se obtendría la razón entre el número de veces en las que los indizadores A y B coinciden en la asignación de categoría principal a un documento respecto del total de pares de trabajos que han sido clasificados, expresada en tanto por ciento.

El análisis de las discrepancias en el proceso de selección se ha realizado, para cada una de ellas, mediante la puesta en común de los criterios utilizados por cada uno de los indizadores, la revisión de los contenidos del documento y la discusión de la adecuación de cada criterio aplicado, llegando a un acuerdo cuando esto era posible.

El análisis de las discrepancias en el proceso de clasificación se ha realizado, de modo análogo a lo anteriormente expuesto, analizando los criterios utilizados por cada indizador para la clasificación de los trabajos. Posteriormente, las discrepancias se han clasificado en dos grupos: aquéllas en las que los indizadores coincidían en el criterio para su clasificación pero aparecía una discrepancia porque se consideraba que uno de ellos había cometido un error (en la aplicación de los criterios que consideraban adecuados o en el reconocimiento de las materias centrales de las que trata el documento) y aquéllas cuyo origen era la utilización de distintos criterios.

### 3 Resultados

En relación con la emisión de los juicios de relevancia, el indizador A ha considerado relevantes 291 de los 370 trabajos juzgados y no relevantes 79. Por su parte, el indizador B ha considerado relevantes 293 y no relevantes 77.

Cuando se comparan las decisiones de uno y otro indizador al juzgar la relevancia de cada trabajo (figura 1) se observa que ambos indicadores coinciden al considerar las trabajos relevantes en 266 ocasiones y no relevantes en 52. Esta situación supone que están de acuerdo en sus juicios de relevancia en 318 ocasiones, lo que representa una proporción de acuerdo simple del 86 % ( $318/370 \times 100$ ).

**Figura 1****Juicio de relevancia****Investigador B**

<b>Investigador A</b>	<b>Investigador B</b>		
	<b>Relevante</b>	<b>No relevante</b>	
<b>Relevante</b>	<b>266</b> a	<b>25</b> b	<b>291</b> (a + b)
<b>No relevante</b>	<b>27</b> c	<b>52</b> d	<b>79</b> (c + d)
	<b>293</b> (a + c)	<b>77</b> (b + d)	<b>370</b> (a + b + c + d)

En cuanto a las discrepancias, en 27 ocasiones el indizador A ha considerado no relevantes trabajos relevantes para el indizador B. Alternativamente, en 25 ocasiones el indizador B ha considerado no relevantes trabajos relevantes para el indizador A.

La discusión de los criterios que habían llevado a la emisión de juicios de relevancia discordantes tuvo como consecuencia que, en ocasiones, uno u otro indizador modificaran los juicios inicialmente emitidos. De este modo, el indizador A estuvo de acuerdo con el B en que 19 de los 25 trabajos que originalmente había considerado relevantes en realidad no lo eran, ya que los temas tratados en los trabajos no trataban de materias propias de la documentación médica. Concretamente eran artículos que abordaban temas de gestión económica y de personal de las instituciones sanitarias o que presentaban resultados de la explotación epidemiológica de registros ya tradicionales. En los seis trabajos restantes ambos indicadores coincidieron que eran realmente relevantes aunque cuatro se consideraron de difícil clasificación.

El análisis de las discrepancias de los 27 trabajos no relevantes para el indizador A y relevantes para el B concluyó con el acuerdo de que 13 trabajos eran realmente relevantes, prevaleciendo de este modo el criterio del indizador B. En estos casos tampoco podía decirse que los temas tratados se atenían a lo que podría entenderse como documentación médica, ya que muchas veces se centraban en el uso (o mal uso) de determinados términos clínicos o en la exposición de definiciones operativas relativas

a determinadas unidades asistenciales, pero se consideró que constituían una información de referencia muy interesante para el médico documentalista

Del resto de los 27 trabajos, 11 quedaron como no relevantes y finalmente 3 trabajos se consideraron de difícil clasificación. Los 11 trabajos que quedan como no relevantes trataban de la explotación epidemiológica de registros ya tradicionales y sobre aspectos de la evaluación de la calidad asistencial (como por ejemplo, la satisfacción de los pacientes).

En cuanto a la asignación de clase y categoría a los trabajos considerados relevantes por ambos indizadores (266 en total) se estuvo de acuerdo en 209 ocasiones, cifra que representa una proporción de acuerdo simple para el conjunto de categorías del 78,6 %.

La distribución por categorías de la concordancia inter-indizador se muestra en la tabla I, en la que los títulos de las columnas han recibido las iniciales a,b,c,d que

**Tabla I**  
**Proporción de acuerdo simple entre indizadores por categorías**

<i>Lista de categorías</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>% Conc.</i>
1.1.1 Métodos y técnicas del trabajo científico	44	4	20	198	91,0
1.1.2 Bibliometría	36	1	3	226	98,5
1.1.3 Planificación y organiz. de sistemas de información	0	3	0	263	98,9
1.1.4 Evaluación de sistemas de información bibliográficos	2	2	1	261	98,9
1.2 Fuentes de información científica	6	8	5	247	91,1
1.3 Terminología y lenguajes documentales	0	13	0	253	95,1
1.4 Aplicaciones de la informática a la documentación	13	1	3	249	98,5
1.5 Miscelánea	0	0	1	265	99,6
2.1.1 Historia clínica hospitalaria y sus documentos	1	2	0	263	99,2
2.1.2 Sistemas de información de los servicios hospitalarios	2	2	1	261	98,9
2.1.3 Servicio de admisión y documentación clínica	11	1	5	249	97,7
2.1.4 Evaluación de la información clínica hospitalaria	6	4	2	254	97,7
2.1.5 Información para la gestión hospitalaria	5	3	5	253	97,0
2.1.6 Miscelánea	0	0	1	265	99,6
2.2.1 Historia clínica de atención primaria	0	1	1	264	92,2
2.2.2 Sistemas de información de atención primaria	3	1	1	261	99,2
2.2.3 Evaluación de la información de atención primaria	10	1	0	255	99,6
2.2.4 Información para la gestión de atención primaria	0	0	1	265	99,6
2.2.5 Miscelánea	0	0	0	266	100,0
3.1 Fuentes de información sobre mortalidad	2	0	0	264	100,0
3.2.1 Registros de enfermedades	8	1	1	256	99,2
3.2.2 Encuesta de Morbilidad Hospitalaria	0	1	1	264	99,2
3.2.3 Otros sistemas de información	2	1	1	262	99,2
3.4 Evaluación de sistemas de información sanitaria	1	0	0	265	100,0
3.5 Miscelánea	0	0	0	266	100,0
4.1 Terminología y leng. documentales en Doc. Clínica	5	2	0	259	99,2
4.2 Sistemas de clasificación de pacientes	13	1	1	251	99,2
4.3 Aplicaciones de la informática a la document. clínic.	12	4	3	247	97,4
4.4 Miscelánea	0	0	0	266	100,0
5 Documentación en enfermería	26	0	0	240	100,0
6 Documentación en odontología	1	0	0	265	100,0

habitualmente se utilizan para la denominación de las celdillas resultantes de la construcción de tablas de concordancia semejantes a la empleada en la figura 1. En la tabla, *a* significa el número de trabajos para los que ambos indizadores coinciden en su asignación a una determinada categoría (llámese X); *b* hace referencia a los trabajos clasificados por el indizador A en la categoría X y por el B en cualquiera otra distinta; *c* a los trabajos clasificados por el indizador A en una categoría cualquiera y por el B en la categoría X; y *d* número de trabajos en los que los indizadores A y B coinciden en no asignar a la categoría X.

Por categorías, los valores de las proporciones de acuerdo varían desde el 91 % obtenido para la de métodos y técnicas del trabajo científico hasta el 100 % que corresponde a varias de ellas (proporciones que se hallan como  $a + d/a + b + c + d$ ). Entre estas últimas se encuentran tanto aquéllas para las que los indizadores concuerdan al no clasificar en las mismas ningún trabajo (son las categorías de miscelánea de los epígrafes de documentación clínica hospitalaria, documentación en atención primaria y documentación sanitaria) como otras en las que ha habido acuerdo en la clasificación de todos los trabajos (documentación en enfermería, documentación en odontología, fuentes de información sobre mortalidad y registros de enfermedades).

En la tabla II, que es una matriz de concordancia, se puede observar la adscripción a las distintas categorías de los 266 documentos juzgados relevantes por ambos indizadores, mostrándose las discrepancias entre ellos. Así, por ejemplo, entre los resultados más significativos cabe destacar que 20 de los 64 trabajos que habían sido clasificados por el indizador A en la categoría de métodos y técnicas del trabajo científico han sido reclasificados por el indizador B en otras categorías, entre las que destacan la de terminología y lenguajes documentales en documentación médico-científica, con 13 trabajos, y la de fuentes de información científica, con 5.

Los resultados de la discusión entre el indizador A y el B en relación a los criterios utilizados en la clasificación de los trabajos han puesto de manifiesto el uso de criterios similares y han permitido atribuir la mayoría de las discrepancias detectadas (40 de las 57) a un error en su aplicación en el que han incurrido tanto el indizador A como el B. Sólo en 14 de las 57 ocasiones las discrepancias se debían a la aplicación de criterios distintos o, como ha ocurrido en tres ocasiones, a situaciones de difícil clasificación.

La aplicación de criterios distintos ha afectado siempre a trabajos de contenido similar. El contenido central de estos trabajos versaba sobre el mal uso de determinados términos médicos en las publicaciones médicas y solía acompañarse de una descripción acerca del origen y significado del término en cuestión. En el caso de este tipo de trabajos el indizador A había seguido el criterio de clasificarlos bajo la categoría 1.1.1 (métodos y técnicas del trabajo científico) por cuanto sus contenidos trataban sobre aspectos de la escritura científica. No obstante, el indizador B consideró como más adecuado la clasificación de estos trabajos en la categoría de terminología y lenguajes documentales. Como resultado de la discusión surgida de esta diferencia de criterio, se acordó que el criterio del indizador B parecía el más conveniente, por cuanto es necesario un buen conocimiento de la etimología y significado de los términos para una correcta utilización de los lenguajes documentales. En relación con los trabajos de difícil clasificación, su causa se atribuyó sobre todo a ausencia de una categoría en la que ese tipo de contenido pudiera adaptarse.

**Tabla II**  
**Matriz de concordancia entre indizadores**

Indizador A	Indizador B	1.1.1	1.1.2	1.1.3	1.1.4	1.2	1.3	1.4	1.5	2.1.1	2.1.2	2.1.3	2.1.4	2.1.5	2.1.6	2.2.1	2.2.2	2.2.3	2.2.4	2.2.5	3.1	3.2.1	3.2.2	3.2.3	3.4	3.5	4.1	4.2	4.3	4.4	5	6	TOTAL
1.1.1 Métodos y Técnicas		44	2																														48
1.1.2 Bibliometría		36	1																														37
1.1.3 Planif./org./Des. SIC		1	0	2																													3
1.1.4 Evaluación SI		5	1	2	1																												14
1.2 Fuentes de inf. cient.		13	1	6	2																												14
1.3 Terminolog. y Docum.				0																													13
1.4 Informatización				1																													14
1.5 Miscelánea									0																								0
2.1.1 Historia Clínica										1																							3
2.1.2 Inf./Doc./SIH										2	1																						4
2.1.3 Admisión y D. Clín.											11	1																					12
2.1.4 Evaluación											3	6	3	1																			10
2.1.5 Informac. Gestión												5																					8
2.1.6 Miscelánea																																	0
2.2.1 Historia Clínica										0																							1
2.2.2 Inf./Doc./SIAP											3																						4
2.2.3 Evaluación												10																					11
2.2.4 Informac. Gestión													0																				0
2.2.5 Miscelánea																																	0
3.1 Fuentes mortalidad																																	2
3.2.1 Registros enferm.																					2												9
3.2.2 EMH																					1												1
3.2.3 Otros SI																					1												3
3.4 Evaluación																																	1
3.5 Miscelánea																																	0
4.1 Terminolog./L. Docum.																																	7
4.2 Sist. Clasific. Pacient.																																	14
4.3 Informatización																																	16
4.4 Miscelánea																																	0
5 Doc. Enfermería																																	26
6 Doc. Odontología																																	1
TOTAL		64	39	0	3	11	0	16	1	1	3	16	8	10	1	1	4	10	1	0	2	9	1	3	1	0	5	14	15	0	26	1	266



## 4 Discusión

Los estudios de concordancia o de consistencia son un medio para poner de relieve la forma en la que se llevan a cabo determinadas tareas en un sistema de recuperación de información.

Sus resultados son de utilidad tanto para los productores de la información como para sus usuarios. Para los primeros, sirven como medida para conocer aspectos concretos de la realización de determinadas tareas que plantean dificultades y de esta manera poner medios para su mejora. Para los usuarios es un medio para conocer cómo funciona el sistema de información y así poder utilizarlo del mejor modo posible.

La consistencia inter e intra-indizador es uno de los aspectos que recaen en lo que Saracevic (6) denomina *investigación de las diferencias individuales en la recuperación de información*. Es una medida de la similitud de las reacciones de seres humanos diferentes cuando procesan la misma información (7). Los estudios más tempranos datan de los años sesenta y principios de los setenta. Desde sus comienzos, han sido múltiples los estudios realizados con el fin de averiguar la consistencia en la indización. El trabajo de Zunde y Dexter (7) que sintetiza los resultados de más de una docena de estudios sobre la materia y el de Leonard (8) que recoge las publicaciones sobre el tema del período 1954-1975, son buena prueba de ello. Algunos autores (6) son de la opinión de que la edad de oro de las publicaciones sobre esta materia fue justo su período inicial y que desde entonces no ha habido un resurgimiento. No obstante, hay que poner de relieve varias publicaciones que han aparecido en fechas posteriores, como la de Sievert y Andrews (9) que presenta los resultados de la evaluación de la base de datos ISA y la de Mai Chan (10) relacionada con la consistencia inter-indizador en la catalogación por materias. También cabe mencionar algunas otras que han abordado aspectos de tipo metodológico como la de Rolling (11) que analiza las ventajas y limitaciones del uso de la consistencia en la indización como medida de la calidad de este proceso y de su repercusión en la eficacia de la recuperación, y la más actual de Soergel (12) que trata el tema de la corrección y la concordancia en la indización en el contexto de un análisis de las características de la indización que pueden afectar a la eficacia de la recuperación.

De fechas relativamente recientes datan algunos estudios comparativos de concordancia realizados para conocer el tratamiento documental que recibe el mismo material por sistemas distintos. Entre ellos se encuentran, por ejemplo, la investigación de Middleton y Di'orio (13) que estudia la indización de varias fuentes del área de educación; la de Starr (14) que puso de relieve variaciones en las políticas de indización en fuentes para la recuperación de información en ciencias marinas, así como la de Konnings (15) y la de de-Jong Hoffman (16) que analizan, respectivamente, las discrepancias en la indización de fuentes del área de informática el primero y de ingeniería el segundo. Entre este tipo de investigaciones también se encuentran las de Sievert y Verbeck (17) que analiza la concordancia en la indización de los trabajos publicados en tres revistas dedicadas a la búsqueda online por dos sistemas diferentes, ERIC y LISA, y la de Barber y colaboradores (18), que analiza la cobertura e indización de 30 artículos de farmacología de los sistemas BioSIS, Casearch, Embase y MEDLINE y los repertorios, *Biological Abstracts*, *Chemical Abstracts*, *Excerpta Medica e Index Medicus*. Tonta (19) recientemente ha publicado un estudio comparativo

de la consistencia en la indización entre la Library of Congress de EE.UU. y la British Library utilizando la *Library of Congress Subject Headings*.

Los estudios que abordan la evaluación de la consistencia de la indización de sistemas de recuperación de información médico-científica no han sido muy abundantes (3). Entre los primeros está el realizado por Lancaster (20) como parte de la evaluación del sistema MEDLARS y la investigación de Leonard (8) que utilizó indizadores de la *National Libray of Medicine* como parte de su investigación sobre la relación entre la consistencia en la indización y la eficacia en la recuperación. Entre los más recientes se encuentran el de Marcetich y Schuyler (21), cuyo propósito fue descubrir si la indización asistida por ordenador podría aumentar la consistencia en la indización, y el de Funk *et al.* (3) que pese a estar publicado hace casi quince años podría considerarse como una de las últimas investigaciones en este campo. En nuestro país no se ha podido localizar evidencia documental alguna sobre la materia.

Los estudios de consistencia o concordancia obedecen genéricamente a la siguiente cuestión: ¿en qué medida personas con una determinada característica en común producen los mismos resultados cuando realizan la misma tarea?, o dicho de otro modo, ¿cuál es la magnitud de la diferencia? (6).

En este trabajo en concreto se ha tratado de conocer la medida en la que los criterios utilizados por los responsables del repertorio para la selección e indización del material son compartidos por sus usuarios. Estos últimos estarían representados por el profesional que actúa como indizador externo al sistema.

En cuanto a los resultados, las cifras revelan la existencia de un buen nivel de acuerdo en la consideración de la relevancia de los trabajos (proporción de acuerdo simple del 86 %). Estas cifras son superiores a los resultados obtenidos en algunos estudios experimentales en los que los valores varían entre el 55 % y 75 %, dándose una mayor concordancia en las situaciones en las que las personas que emiten los juicios de relevancia tienen un nivel elevado de experiencia en la materia, como es el caso que nos ocupa.

Por regla general, las discrepancias no han afectado a ningún trabajo que tratara como materia principal un aspecto relacionado con la documentación médica. Por el contrario, esta situación se ha producido cuando se han valorado publicaciones que trataban estos temas de un modo muy periférico o cuando se abordaban cuestiones relativas a áreas con gran relación de vecindad con la disciplina o con los puestos de trabajo que con frecuencia ocupan los médicos documentalistas. Este es el caso de trabajos que tratan sobre aspectos relativos a la gestión o la epidemiología. Estos resultados han planteado la conveniencia de ampliar el estudio de concordancia en la selección de trabajos a otros especialistas en la materia, con el fin de que sus resultados puedan servir de material para plantear una discusión sistemática que lleve a establecer unos criterios consensuados y operativos de inclusión y no inclusión de trabajos en el repertorio.

En relación con el proceso de clasificación, los niveles de acuerdo alcanzados por los dos indizadores también son buenos (proporción de acuerdo simple del 78 %), superiores a los obtenidos por otros autores. Así, por ejemplo, el porcentaje de acuerdo que obtienen Funk y colaboradores (3) en la evaluación de la indización de los registros duplicados de la base de datos MEDLINE es del 61 % para la asignación de encabezamientos de materias y del 55 % para los subencabezamientos. Siguiendo el mismo procedimiento, Sievert y Andrews (9) alcanzaron respectivamente un 52 % y

un 45 % de acuerdo para los encabezamientos y subencabezamientos de materias. La obtención en este estudio de cifras más elevadas puede deberse a que la concordancia en las tareas de clasificación u ordenación en clases ya predeterminadas es siempre superior a la que se consigue en la asignación de descriptores o palabras clave (6).

Hay que destacar que lo más provechoso de esta parte del estudio de concordancia ha sido el análisis de las discrepancias y los resultados obtenidos. Se puede desprender de los mismos que, salvo un número de trabajos que afecta a unos contenidos concretos e identificables, ha habido acuerdo entre los criterios que habitualmente utiliza el sistema para la indización y los que utilizaría un usuario tipo representado por el indizador externo.

La mayor parte de las discordancias detectadas en el proceso de clasificación se ha podido atribuir a errores cometidos por uno u otro indizador y su origen se ha atribuido a fallos originados, por ejemplo, por el cansancio, o por la confianza o familiaridad en la realización del proceso. Estos fallos han afectado casi por igual a ambos indizadores. El uso de criterios distintos ha quedado reducido a un pequeño núcleo de trabajos de contenido uniforme.

Del análisis de las discrepancias se ha podido colegir también la existencia de un pequeño núcleo de trabajos que no «encaja» bien en ninguno de los apartados de la clasificación, aspecto que es una contrapartida de la rigidez de estos esquemas y que ha de ser revisado. En este sentido, por ejemplo, se ha planteado la posibilidad de modificar ligeramente el esquema clasificatorio, ya que el utilizado hasta el momento siempre se ha considerado provisional y abierto a las sugerencias de los usuarios y por lo tanto susceptible de modificación.

En general, de los resultados de este trabajo se concluye la utilidad de la puesta en marcha de estudios de esta naturaleza como medio de mejora de los sistemas de información. En concreto, se plantea utilizar la información obtenida con el propósito de elaborar un manual de procedimiento de uso interno que favorezca la aplicación estable de los criterios y la elaboración de unas indicaciones objetivas que queden explícitamente recogidas en la fuente y que permitan un mejor conocimiento de su funcionamiento para los usuarios.

## APENDICE I

### Clasificación

- 1 Documentación Médico-Científica**
  - 1.1 Métodos y Técnicas Documentales
    - 1.1.1 Métodos y Técnicas para la Realización y Publicación del Trabajo Científico
    - 1.1.2 Bibliometría
    - 1.1.3 Planificación, Organización y Desarrollo de Sistemas de Información Científica
    - 1.1.4 Evaluación de Servicios y Sistemas de Información Médico-Científica
  - 1.2 Fuentes de Información Médico-Científica
  - 1.3 Terminología y Lenguajes Documentales en Documentación Médico-Científica
  - 1.4 Aplicaciones de la Informática y de las Nuevas Tecnologías a la Documentación Científica
  - 1.5 Miscelánea
- 2 Documentación Clínica**
  - 2.1 Documentación Clínica Hospitalaria
    - 2.1.1 La Historia Clínica y sus Documentos
    - 2.1.2 Información, Documentos y Sistemas de Información Propios de los Servicios Hospitalarios
    - 2.1.3 Admisión y Documentación Clínica Hospitalaria
    - 2.1.4 Evaluación de la Calidad de la Información, Documentos y Sistemas de Información Clínica Hospitalaria
    - 2.1.5 Información para la Gestión Hospitalaria
    - 2.1.6 Miscelánea
  - 2.2 Documentación Clínica en Atención Primaria
    - 2.2.1 La Historia Clínica en Atención Primaria y sus Documentos
    - 2.2.2 Otra Información y Documentos y Sistemas de Información Propios de la Atención Primaria
    - 2.2.3 Evaluación de la Calidad de la Información, Documentos y Sistemas de Información Clínica en Atención Primaria
    - 2.2.4 Información para la Gestión Asistencial
    - 2.2.5 Miscelánea
- 3 Documentación Sanitaria**
  - 3.1 Fuentes de Información sobre Mortalidad
  - 3.2 Fuentes de Información sobre Morbilidad
    - 3.2.1 Registros de Enfermedades
    - 3.2.2 Encuesta de Morbilidad Hospitalaria
    - 3.2.3 Otros Sistemas de Información
  - 3.3 Fuentes de Información Demográfica
  - 3.4 Evaluación de la Información, Documentos y Sistemas de Información Sanitaria
  - 3.5 Miscelánea
- 4 Aspectos Generales de Documentación Clínica y Sanitaria**
  - 4.1 Terminología y Lenguajes Documentales en Documentación Clínica y Sanitaria
  - 4.2 Sistemas de Clasificación de Pacientes
  - 4.3 Aplicaciones de la Informática y de las Nuevas Tecnologías a la Documentación Clínica y Sanitaria
  - 4.4 Miscelánea
- 5 Documentación en Enfermería**
- 6 Documentación en Odontología**

## 5 Bibliografía

1. ALEIXANDRE BENAVENT, R.; ABAD GARCIA, M. F. Documentación Médica Española: un instrumento de información profesional. *IV Congreso Nacional de Documentación Médica*, 1995, junio 8-10, Torremolinos.
2. CUADRA, C. A.; KATTER, V. The relevance of relevance assessment. *Proceedings of the American Documentation Institute*, vol 4, 1967. Washington: Thomson Book Company.
3. FUNK, M.; REID, C. A.; MCGOOGAN, L. E. Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association*, 1983, vol. 71, p. 176-183.
4. ABAD GARCIA, M. F. *Investigación evaluativa en documentación: su aplicación a la documentación médica*. Valencia: Servei de Publicacions de la Universitat de València, Col·lecció materials, 1997, n.º 23 .
5. FLEISS, J. L. *Statistical methods for rates and proportions*. Nueva York: Willey, 1981.
6. SARACEVIC, T. Individual differences in organizing, searching and retrieving information. *Proceedings of the 5TH Annual Meeting of the American Society of Information Science*, 1987, p. 82-86.
7. ZUNDE, P.; DEXTER, M. Indexing consistency and quality. *American Documentation*, 1966, vol. 20, p. 259-267.
8. LEONARD, L. E. Inter-indexer consistency studies, 1954-1975. A review of the literature and summary of study results. *University of Illinois Graduate School of Library Science Occasional Papers*, 1977, p. 131.
9. SIEVERT, M. C.; ANDREWS M. J. Indexing consistency in Information Science Abstracts. *Journal of the American Society of Information Science*, 1991, vol. 42, p. 1-6.
10. MAI CHAN, L. Indexer consistency in subject cataloging. *Information Technology and Libraries*, 1989, vol. 7, p. 349-358.
11. ROLLING, L. Indexing consistency, quality and efficiency. *Information Processing and Management*, 1981, vol. 17, p. 69-76.
12. SOERGEL, D. Indexing and retrieval performance: the logical evidence. *Journal of the American Society for Information Science*, 1994, vol. 45, p. 589-599.
13. MIDDLETON, M. R.; DIÒRIO, A. A comparison in indexing consistency and coverage in AEI, ERIC and APAIS databases. *Behavioural and Social Sciences Librarian*, 1984, vol. 3, p. 33-43.
14. STARR, S. Database in the marine sciences. *Online Review*, 1982, vol. 6, p. 109-126.
15. KONNINGS, C. A. G. Comparison and evaluation of nine bibliographies/bibliographic databases in the field of computer science. *Online Review*, 1985, vol. 9, p. 121-133.
16. DE JONG-HOFFMAN, M. W. Comparison of selecting, abstracting and indexing by COMPENDEX, INSPEC and PASCAL and the impact of this on manual and automated retrieval information. *Online Review*, 1981, vol. 5, p. 25-36.
17. SIERVERT, M. E.; VERBEC, A. The indexing of the literature of online searching: a comparison of ERIC and LISA. *Online Review*, 1987; vol. 11, p. 95-104.
18. BARBER, J.; MOFFAT, S.; WOOD, F. Case studies of the indexing and retrieval of pharmacology papers. *Information Processing and Management*, 1988, vol. 24, p. 141-150.
19. TONTA, Y. A study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources and Technical Services*, 1990, vol. 35, p. 177-185.
20. LANCASTER, F. W. MEDLARS: a report on the evaluation of its operating efficiency. *American Documentation*, 1969, vol. 29, p. 119-142.
21. MARCETICH, J.; SCHUYLER, P. The use of AID to promote indexing consistency at the National Library of Medicine. *Eight-first Annual Meeting of the Medical Library Association*, 1981, Montreal, Quebec.