

## Análisis del control de calidad del Conjunto Mínimo de Datos Básicos de Andalucía de los años 2000 a 2003

Rodríguez del Águila MM, Perea-Milla E, Librero J, Buzón Barrera ML, Rivas Ruiz F, por el Grupo de Variaciones en la Práctica Médica de la Red IRYSS (Grupo VPM-IRYSS)

**RESUMEN:** El Conjunto Mínimo de Datos Básicos (CMBD) de Andalucía es la base de datos administrativa en la que se recoge información de los episodios de hospitalización registrados en los hospitales andaluces. Se ha diseñado un programa que evalúa la calidad de la información contenida en dichas bases de datos, comprobando coherencias e incompatibilidades entre los campos contenidos en las mismas. El objetivo de este estudio fue analizar los errores detectados en los CMBDs de los años 2000 a 2003, ambos inclusive, mediante la aplicación de dicho programa. Se observa en general un porcentaje aceptable de errores, siendo la residencia, fecha de intervención y codificación del diagnóstico principal los campos con mayor número de errores de codificación.

**PALABRAS CLAVE:** errores de codificación, bases de datos, calidad.

### Introducción

El Conjunto Mínimo de Datos Básicos (CMBD) es una base de datos en la que se recogen características tanto administrativas como clínicas de todos los episodios de hospitalización producidos en un centro hospitalario en un periodo de tiempo concreto<sup>1</sup>. Cada hospital la remite a los servicios centrales de las distintas Comunidades Autónomas, donde es utilizada con fines informativos para determinar la casuística de manera periódica<sup>2</sup>. Estas bases de datos son también utilizadas en investigación, previa solicitud formal de la misma, y en concreto son la fuente de información imprescindible a partir de la cual se desarrollan los Atlas de Variaciones de la Práctica Médica.

Como todo sistema de información, el CMBD no está exento de errores, aunque estos pueden ser depurados y controlados para que los análisis posteriores realizados con estos datos tengan mayor fiabilidad. Con esta idea se diseñó desde los tres nodos andaluces adscritos a la Red de Investigación en Resultados en Salud y Servicios Sanitarios (Agencia de Evaluación de Tecnologías Sanitarias de Sevilla, Hospital Costa del Sol de Marbella y Hospital Virgen de las Nieves de Granada), un programa informático que lleva a cabo un control de calidad sobre el CMBD andaluz, desarrollado en el lenguaje de programación Visual Basic y en el entorno de Microsoft Access. Inicialmente se hizo el diseño adaptado a los campos definidos en el CMBD andaluz y sólo es aplicable a esta base de datos, pero el hecho de estar escrito en este entorno y este lenguaje de programación permite que, al no ser un programa compilado, se pueda adaptar a cualquier estructura de CMBD e incluso incluir indicadores de calidad nuevos.

El control de calidad del CMBD andaluz consiste en llevar a cabo una serie de análisis en la base de datos que comprueben la coherencia de los campos o variables, detectando posibles errores de transcripción, codificación, campos no cumplimentados, etc. Se realiza en diversos niveles en función del contenido de dichos campos, teniendo así controles de calidad de tipo administrativo, de tipo clínico, de coherencia intraepisodio e interepisodio y de especificidad de la información clínica.

El objetivo de este estudio fue analizar los errores detectados en los CMBDs andaluces de los años 2000 a 2003, a través del programa del control de calidad.

### Métodos

Los CMBDs andaluces de los años de estudio se solicitaron formalmente a la Subdirección de Asistencia Especializada del Servicio Andaluz de Salud. Se aplicó el programa de control

de calidad a cada año, pasando los 7 módulos de que consta para determinar la coherencia de diferentes grupos de variables, según una adaptación realizada sobre la clasificación realizada por Librero et al<sup>3</sup>. Se analizaron por separado cada una de las tablas de errores generadas, llevando a cabo un estudio descriptivo. Los 7 análisis realizados fueron:

1. Análisis de calidad intracampo de datos administrativos (comprueba la validez de los campos relacionados con los datos administrativos).
2. Análisis de calidad intracampo de datos clínicos (comprueba la validez de datos relacionados con la clínica del episodio).
3. Análisis de calidad entre variables del mismo episodio (verifica si son compatibles los campos de un mismo episodio entre sí).
4. Análisis de calidad entre episodios (analiza la diferencia de los datos administrativos entre diferentes episodios de un mismo paciente).
5. Inespecificidad según diagnósticos (porcentaje de diagnósticos inespecíficos de los 24 grupos de diagnósticos que figuran en el cuadro 1).
6. Especificidad de la información clínica (comprueba si la información contenida en los datos clínicos del episodio es válida).
7. Volumen de la información clínica (resumen del número de registros del CMBD en que falta uno o varios diagnósticos).

Los resultados de cada uno de los 7 módulos fueron analizados en algunos casos desglosados y en otros agregados, fundamentalmente a través de tablas de frecuencias y porcentajes.

### Resultados

El número de episodios o registros anuales que recogió el CMBD desde 2000 a 2003 fue de 577.186, 549.794, 558.114 y 567.063 respectivamente.

#### Análisis de calidad intracampo de datos administrativos

Los campos administrativos con menor cumplimentación son la fecha de intervención y la residencia, siendo mínimos los errores en el resto de campos. En estos años ha ido disminuyendo el porcentaje de errores en el campo de residencia y en 2003 han aumentado los errores en la fecha de intervención y el sexo, esta última de forma insignificante (tabla 1).

#### Análisis de calidad intracampo de datos clínicos

Donde más errores existen es en la cumplimentación del diagnóstico principal, no observándose ninguna tendencia concreta

1. Intoxicación alimentaria e infecciones intestinales
2. Tuberculosis
3. SIDA, complejo relacionado e infección VIH
4. Hepatitis viral
5. Neoplasias malignas
6. Enfermedades del sistema nervioso central
7. Trastornos de ojos y anexos, excepto mastoides
8. Cataratas
9. Trastornos del oído y apófisis mastoides
10. Hipertensión arterial
11. Enfermedades cardíacas isquémicas
12. Enfermedad isquémica cerebro-vascular y aterosclerosis
13. Enfermedad vascular periférica
14. Bronquitis, asma, enfermedad pulmonar obstructiva crónica
15. Úlcera gastrointestinal
16. Gastritis y gastroduodenitis
17. Apendicitis
18. Hernias
19. Nefropatías
20. Cistitis
21. Complicaciones del embarazo, parto y puerperio
22. Osteocondropatías
23. Fracturas
24. Luxaciones y esguinces

Cuadro 1: Listado de diagnósticos de los que se contabilizan los códigos inespecíficos en el módulo 5 del programa de control de calidad del CMBD

en estos años. Cabe destacar un ligero incremento en el porcentaje de errores en los diagnósticos secundarios durante los dos últimos años (tabla 2).

#### Análisis de calidad entre variables del mismo episodio

Los errores detectados en este módulo son inapreciables, a excepción del procedimiento primero sin fecha de intervención (promedio de 18.03%) y la existencia de fecha de intervención sin que haya procedimientos (promedio de 25.32%) (tabla 3).

#### Análisis de calidad entre episodios

Los campos no coincidentes entre reingresos también son prácticamente inapreciables. Cabe destacar que donde más diferencias entre episodios se producen es en la fecha de nacimiento y en la residencia (tabla 4).

#### Inespecificidad según diagnósticos

Se han considerado los 5 diagnósticos principales con inespecificidades más frecuentes dentro del CMBD andaluz. La hipertensión arterial es un diagnóstico que se codifica mayoritariamente como no específico (99,45%). Le siguen, aunque con porcentajes mucho más bajos de inespecificidades, las osteocondropatías (17,4%), enfermedades cardíacas isquémicas (17,04%), fracturas (10,54%) y neoplasias (5,43%) (figura 1). Dentro de los siguientes diagnósticos principales con inespecificidades más frecuentes, cabe destacar la bronquitis, asma y EPOC.

#### Especificidad de la información clínica

Los datos de especificidad de la información clínica corresponden al último año analizado (2003). Se observa un alto porcentaje

Tabla 1: Porcentaje de errores en la cumplimentación de algunos de los datos administrativos contemplados en los CMBDs andaluces durante 2000 a 2003

Indicador	2000	2001	2002	2003
Fecha de nacimiento	<0,01%	0,01%	0	<0,01%
Fecha de ingreso	<0,01%	<0,01%	<0,01%	<0,01%
Fecha de alta	0	<0,01%	0	0
Fecha de intervención	17,44%	16,31%	14,89%	19,94%
Residencia	25,41%	22,71%	22,15%	19,43%
Sexo	0,01%	0,02%	0%	0,08%

Tabla 2: Porcentaje de errores en los diagnósticos y procedimientos

	2000		2001		2002		2003	
	%	N	%	N	%	N	%	N
Dx principal	7,30	577.186	4,70	549.794	6,04	558.114	6,03	567.063
Dx secundarios	1,03	1.199.659	0,88	1.246.196	3,09	1.319.833	3,46	1.425.333
Dx E	1,23	62.590	1,63	67.444	0,52	72.047	0,57	76.650
Procedimientos	0,02	1.066.528	0,02	1.006.718	0,01	1.064.700	0,02	1.119.942

Dx: diagnóstico; N: número de registros

**Tabla 3: Porcentaje de errores entre las variables de un mismo episodio**

	2000	2001	2002	2003
Edad no calculable	0,01%	0,01%	<0,01%	<0,01%
Estancia no calculable	<0,01%	<0,01%	<0,01%	<0,01%
Incoherencia en fechas	0	<0,01%	0	0
Dx principal incompatible	0,04%	0,05%	0,05%	0,04%
Dx secundarios incompatibles	0,04%	0,05%	0,04%	0,04%
Procedimientos incompatibles	0,03%	0,03%	0,02%	0,02%
P1 sin fecha de intervención	23,06%	17,36%	15,95%	15,74%
Fecha de intervención sin P1	24,34%	24,35%	26,37%	26,20%

*Dx: diagnóstico; P1: procedimiento primero*

**Tabla 4: Porcentaje de errores en el análisis de calidad interepisodio**

	2000	2001	2002	2003
Número de reingresos	78.056	73.715	73.680	74.549
Incoherencia en sexo	<0,01%	0,02%	0,01%	0,13%
Fecha de nacimiento diferente	0,15%	0,14%	0,17%	0,38%
Reingreso tras alta por éxitus	0,04%	0,02%	0,02%	0,03%
Residencia diferente	0,82%	0,93%	1,47%	0,95%

**Tabla 5: Especificidad de la información clínica**

Indicador	Porcentaje	Número de errores
Dx o procedimiento duplicado	11,19	63.445
Código E como Dx principal	0,00	0
Falta 4º o 5º dígito en Dx principal	0,12	659
Dx principal inespecífico	3,36	19.055
Falta 4º o 5º dígito en Dx secundarios	5,85	33.167
Dx secundarios inespecíficos	7,50	42.549
Falta 4º dígito en Código E	0,01	42
Falta 4º dígito en procedimientos	0,10	555
Procedimientos inespecíficos	0,66	3.738
Enfermedad crónica no informada*	53,02	16.801

\* sobre 31688 registros;

*Dx: diagnóstico*

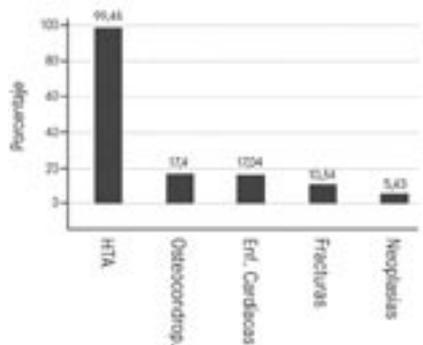


Figura 1: Porcentaje de inespecificidades de los 5 diagnósticos principales más frecuentes del CMBD andaluz. Datos acumulados de 2000 a 2003 ambos inclusive

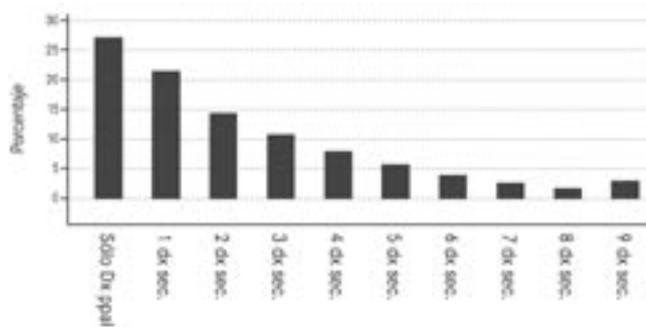


Figura 2: Porcentaje de episodios según el número de diagnósticos secundarios presentes en los episodios registrados en los CMBDs de 2000 a 2003 (acumulado)

de enfermedades crónicas no informadas (53,02%), así como un 11% de diagnósticos o procedimientos duplicados (tabla 5).

### Volumen de información clínica

Los resultados que se muestran de la ejecución de este módulo son datos acumulados de 2000-2003. Hay un alto índice de cumplimentación de los procedimientos principales y el promedio de diagnósticos por episodio que se recogen es en torno a 3 en todos los años (tabla 6). En un 27% de los episodios no se registra ningún diagnóstico secundario y en un 22% se registra un diagnóstico secundario (figura 2).

### Discusión

El control de calidad aplicado sobre los CMBDs andaluces arroja resultados satisfactorios, a excepción de los campos de residencia y fecha de intervención, donde se recoge un volumen importante de errores.

Según Kravets<sup>4</sup> los valores perdidos para los códigos de dirección de los pacientes son más frecuentes en las áreas rurales, mientras que en las áreas urbanas dicha codificación es más precisa. En nuestro estudio no es posible determinar si una mayor precisión en la codificación de la residencia del paciente se da en aquellos episodios asociados al medio rural o urbano. Analizando el CMBD de 2003 se observa que hay un 73% de errores en la codificación de la residencia en hospitales de capitales de provincia (medio fundamentalmente urbano), aunque esta cifra no es representativa ya que muchos de los episodios proceden de poblaciones rurales al ser centros de referencia para éstas.

En los controles de calidad aplicados sobre los años de estudio se ha observado un porcentaje inapreciable de incoherencias en los diagnósticos, salvo en el diagnóstico principal. En dos estudios realizados en sendos servicios de urología en hospitales del Reino Unido<sup>5,6</sup> se comprobó que existían discrepancias de codificación entre la información registrada directamente en el servicio y la recogida en el hospital, sobre todo la referida a los procedimientos y diagnósticos. Los clínicos de los servicios de Documentación Clínica de los hospitales andaluces realizan la codificación de los diagnósticos y procedimientos, comprobando de forma no regular las posibles discrepancias de los datos mediante muestreos de historias clínicas. Sería deseable que estos *audit* se realizaran de forma más o menos periódica verificando además la coherencia con los restantes campos del CMBD y contrastando los resultados obtenidos tras la aplicación del programa.

Los sistemas de información son hoy en día una herramienta de gran utilidad para el almacenamiento y recuperación de información. El CMBD entra dentro del grupo de sistemas de información automáticos y una de las premisas que deben cumplir para que sean precisos es la exhaustividad y una correcta cumplimentación de los registros (además de legibilidad para los sistemas manuales)<sup>7</sup>. El programa de control de calidad descrito asigna a los campos vacíos y erróneos el mismo código (combinaciones de 9), para después considerarlos como perdidos en análisis posteriores. Si se analizan las tablas de errores generadas y las causas se pueden distinguir los campos erróneos de los vacíos. En el CMBD de 2003 los campos vacíos suponían sólo el 0.09% del total de errores (módulo 1),

Tabla 6: Datos referentes al volumen de la información clínica recogida en los CMBDs andaluces (acumulado 2000 a 2003)

	Valor
Registros sin ningún diagnóstico	0
Registros con procedimiento principal	82,68 %
Promedio diagnósticos/episodio	3,31
Promedio procedimientos/episodio	1,94

en el módulo 2 subían hasta un 5%, en el módulo 3 un 11,7% y en el módulo 6 un 9,33%; cifras similares o incluso inferiores a las obtenidas en la revisión de Hogan<sup>8</sup>, en la que analizaron la exhaustividad y adecuación en 20 artículos publicados sobre sistemas informatizados de registros de pacientes (hay que tener en cuenta que un mismo registro puede tener simultáneamente varios errores).

El CMBD es una base de datos administrativa y como tal tiene sus limitaciones, ya que en ella no se recogen datos acerca de la gravedad o proceso de la enfermedad, como en las bases de datos clínicas. Por ello, es difícil comprobar la calidad de los cuidados<sup>9</sup> y hacer un ajuste de riesgos adecuados en la investigación que compara proveedores o analiza resultados.

Por otra parte, sería deseable poder realizar un estudio comparativo con los CMBD de las restantes comunidades autónomas que participan en la Red IRYSS, adaptando el programa a los campos contemplados en las otras bases de datos. Puede que exista cierta dificultad en este proceso, ya que los CMBDs de las distintas comunidades autónomas tienen algunas diferencias en la estructura y nombres de campos. El estudio de la calidad del CMBD de las distintas comunidades autónomas permitiría a los gestores establecer criterios para una mejor codificación y precisión<sup>10</sup>, al menos si no se cumplen unos determinados estándares de calidad de la codificación.

Los códigos de procedimientos quirúrgicos que se han tomado en la ejecución del módulo 3 no son tales, sino que incluyen todos los procedimientos, con lo que los indicadores de errores en fecha de intervención sin procedimiento principal y en procedimiento principal sin fecha de intervención pueden estar sobredimensionados.

El CMBD puede contener otros errores no contemplados en este programa de control de calidad. La ventaja es que al no ser un programa compilado se puede modificar en cualquier momento por programadores que conozcan el lenguaje de programación de Visual Basic y de esta forma incluir módulos adicionales de control de calidad, por ejemplo, análisis del control de calidad de los Grupos Relacionados con el Diagnóstico (GRD).

El hecho de que los campos no coincidentes en los reingresos no se unifiquen puede ser una fuente de variabilidad, pero se observa que los errores derivados de la misma son inferiores al 1%.

En general el control de calidad de los CMBDs andaluces presenta resultados satisfactorios, con la ventaja de que dicho programa puede adaptarse a otros controles de calidad no contemplados aquí. Sería conveniente que las instituciones establecieran mecanismos para reducir el porcentaje de errores en la cumplimentación de los registros de estas bases de datos.

**NOTA:** El software citado en este artículo (Visual Basic y Microsoft Access) corresponde a marcas registradas. El programa y el manual de uso son gratuitos y están disponibles previa petición a la primera autora (mmar.rodiguez.sspa@juntadeandalucia.es).

## Referencias

1. Servicio Andaluz de Salud. Manual de instrucciones del Conjunto Mínimo Básico de Datos al alta hospitalaria de Andalucía 2005. [Internet]. Sevilla: Servicio Andaluz de Salud; 2005. [acceso 1 de junio de 2006]. Disponible en: <http://www.sas.junta-andalucia.es/library/plantillas/externa.asp?pag=../../publicaciones/datos/179/pdf/ManualCMBD2005.pdf>
2. Servicio Andaluz de Salud. Conjunto Mínimo Básico de Datos al alta hospitalaria. Andalucía 2005. [Internet]. Sevilla: Servicio Andaluz de Salud; 2005. [acceso 1 de junio de 2006]. Disponible en: <http://www.sas.junta-andalucia.es/library/plantillas/externa.asp?pag=../../publicaciones/datos/215/pdf/CMBDA2005.pdf>
3. Librero J, Ordiñana R, Peiró S. Análisis automatizado de la calidad del conjunto mínimo de datos básicos. Implicaciones para los sistemas de ajuste de riesgos. *Gac Sanit.* 1998;12:9-21.
4. Kravets N, Hadden WC. The accuracy of address coding and the effects of coding errors. *Health Place.* 2005 [in press].
5. Khwaja HA, Syed H, Cranston DW. Coding errors: a comparative analysis of hospital and prospectively collected departmental data. *BJU International.* 2002;89:178-80.
6. Ballaro A, Oliver S, Emberton M. Do we do what they say we do? Coding errors in urology. *BJU Int* 2000;85:389-91.
7. Ward NS. The accuracy of clinical information system. *J Crit Care.* 2004;19:221-5.
8. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc.* 1997;5: 342-55.
9. Rubinfeld GD. Using computerized medical databases to measure and to improve the quality of intensive care. *J Crit Care.* 2004;19:248-56.
10. Shin E, Dow WH, Kaluzny AD, Park YM, Park K. Disease coding errors by health care organizations: effects of a government quality intervention. *Int J Health Plann Manage.* 2003;18:151-9.

## Agradecimientos

*El programa se ha diseñado en el marco de la Red de Investigación en Resultados en Salud y Servicios Sanitarios (Red IRYSS, Código G03/202).*