

Original

# Conexión de registros sanitarios: base poblacional de salud de Andalucía



Dolores Muñozerro-Muñiz, Juan Antonio Goicoechea-Salazar, Francisco Javier García-León\*, Antonio Laguna-Téllez, Daniel Larrocha-Mata y Manuel Cardero-Rivas

Subdirección Técnica Asesora de Gestión de la Información, Servicio Andaluz de Salud, Sevilla, España

## INFORMACIÓN DEL ARTÍCULO

### Historia del artículo:

Recibido el 26 de diciembre de 2018

Aceptado el 18 de marzo de 2019

On-line el 25 de mayo de 2019

### Palabras clave:

Sistemas de ayuda a decisiones

Conexión de registros médicos

Historia clínica electrónica

Almacenamiento de datos

Estadística

Sistemas de información en salud

Organización y administración

Investigación

## R E S U M E N

**Objetivo:** Describir el desarrollo de un sistema de información que conecta datos procedentes de múltiples registros, sanitarios y otros, para su uso con fines asistenciales, de administración, gestión, evaluación, inspección, investigación y salud pública.

**Método:** Conexión determinística de datos pseudonimizados de una población de 8,5 millones de habitantes, procedentes de Base de datos de usuarios, Historia clínica electrónica DIRAYA, Conjunto mínimo básico de datos (hospitalización, cirugía mayor ambulatoria, urgencias hospitalarias y hospital de día médico) y sistemas de información de salud mental, pruebas de imagen, pruebas analíticas, vacunas, pacientes renales y farmacia. Se utilizó un codificador automático para los diagnósticos clínicos y se definieron 80 enfermedades crónicas para su seguimiento. La arquitectura del sistema de información constó de tres capas: datos (base de datos Oracle 11 g), aplicaciones (MicroStrategy BI) y presentación (MicroStrategy Web, librerías JavaScript, HTML 5 y hojas de estilo CSS). Se implantaron medidas para la gobernanza del sistema.

**Resultados:** Se incluyeron datos de 12,5 millones de personas que fueron usuarias entre los años 2001 y 2017, con 435,5 millones de diagnósticos. El 88,7% de estos diagnósticos fueron generados por el codificador automático. Los datos se presentan mediante informes predefinidos o consultas dinámicas, ambos exportables a ficheros CSV para su tratamiento fuera del sistema. Analistas expertos pueden acceder directamente a las bases de datos y realizar extracciones mediante SQL o tratar directamente los datos con herramientas externas.

**Conclusión:** El trabajo ha mostrado cómo la conexión de registros sanitarios abre nuevas posibilidades en el análisis de datos.

© 2019 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Health record linkage: Andalusian health population database

### A B S T R A C T

### Keywords:

Decision support systems

Medical record linkage

Electronic health records

Data warehouse

Data analysis

Health information systems

Organization and administration

Research

**Objective:** To describe the development of an information system that connects data from multiple health records to improve assistance to patients, health services administration, management, evaluation, and inspection, as well as public health and research.

**Method:** Deterministic connection of pseudonymized data from a population of 8.5 million inhabitants provided by: a users database, DIRAYA electronic medical records, minimum basic data sets (inpatients, outpatient mayor surgery, hospital emergencies and medical day hospital), mental health information systems, analytical and image tests, vaccines, renal patients, and pharmacy. An automatic coder was used to code clinical diagnoses and 80 chronic pathologies were identified to follow-up. The architecture of the information system consisted of three layers: data (Oracle Database 11 g), applications (MicroStrategy BI) and presentation (MicroStrategy Web, JavaScript libraries, HTML 5 and CSS style sheets). Measures for the governance of the system were implemented.

**Results:** Data from 12.5 million health system users between 2001 and 2017 were gathered, including 435.5 million diagnoses, 88.7% of which were generated by the automatic coder. Data can be accessed through predefined reports or dynamic queries, both exportable to CSV files for processing outside the system. Expert analysts can directly access the databases and perform queries using SQL or directly treat the data with external tools.

\* Autor para correspondencia.

Correo electrónico: [fjavier.garcia.leon@juntadeandalucia.es](mailto:fjavier.garcia.leon@juntadeandalucia.es) (F.J. García-León).

**Conclusion:** The work has shown that the connection of health records opens new possibilities for data analysis.

© 2019 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introducción

La digitalización de los servicios sanitarios está generando un volumen de datos sin precedentes, tanto para la atención sanitaria como para su uso secundario<sup>1,2</sup>.

En el Sistema Sanitario Público de Andalucía (SSPA), la historia clínica electrónica y otros sistemas de información generan datos cuya explotación se realiza en módulos específicos de tratamiento, creándose silos independientes de datos; esta misma evolución ha sido observada en otras organizaciones<sup>3</sup>.

Tal fragmentación de datos crea duplicidades, propicia su infrautilización y afecta a su calidad y costes de gestión. Esta situación está evolucionando hacia modelos de análisis más globales, gracias a la conexión de registros y la interoperabilidad de los sistemas de información<sup>4</sup>. La integración de datos produce sinergias que aumentan su valor, y genera un conocimiento superior al obtenido a partir del análisis aislado e independiente de cada sistema de información.

La conexión de registros es habitual en proyectos de investigación específicos, pero lo es menos como actividad continuada, sistemática e institucional, para el uso secundario de datos administrativos. Estas iniciativas institucionales suelen tener como finalidad la evaluación de políticas públicas<sup>5</sup> o el desarrollo de infraestructuras para investigación social<sup>6</sup>, si bien tendrán cada vez mayor relevancia como apoyo a estrategias nacionales de inteligencia artificial<sup>7</sup>.

La integración de datos con fines gerenciales (*business intelligence*) en el sector de la salud es limitada<sup>4,8</sup>, situación que previsiblemente se verá modificada por el interés creciente en el uso secundario de los datos sanitarios<sup>9</sup>. En este contexto, el SSPA puso en marcha un proyecto de «gobernanza de datos»<sup>10</sup> para superar las dificultades que se encuentran en la gestión de datos dispersos en los múltiples sistemas de información y contribuir a una mejora de la atención sanitaria, al ayudar a la toma de decisiones en un entorno cada vez más complejo.

El objetivo de este trabajo es describir el desarrollo del sistema de información que conecta datos procedentes de diferentes silos de información del SSPA y de otras fuentes externas, para su uso con fines asistenciales, de administración, gestión, evaluación, inspección, investigación y salud pública<sup>11</sup>.

## Método

El proyecto contó con una dirección y un comité director situados al máximo nivel de la organización para garantizar su gobernanza (planificación, seguimiento y gestión); contó además con un comité técnico de seguimiento y grupos de apoyo para aspectos específicos<sup>11</sup>.

Se conectaron de forma determinística registros de los 8,5 millones de personas cuya asistencia sanitaria es cubierta por el SSPA. El identificador de cada persona fue el número único de historia de salud de Andalucía que figura en la base de datos de usuarios del SSPA<sup>12</sup>.

El sistema de información conectó, para cada persona, datos individuales obtenidos de los siguientes sistemas de información corporativos:

- Base de datos de usuarios del SSPA.
- Historia clínica electrónica (DIRAYA).
- Sistema de información de salud mental.
- Conjunto mínimo básico de datos (CMBD): hospitalización, cirugía mayor ambulatoria, urgencias hospitalarias y hospital de día médico.
- Sistema de información de vacunas.
- Sistema de información de pacientes renales.
- Pruebas de imagen.
- Pruebas analíticas.
- Sistema de información de farmacia (parcialmente).

El sistema se desarrolló con una arquitectura en tres capas: capa de datos con servidor de base de datos Oracle 11 g, capa de aplicaciones en MicroStrategy BI y capa de presentación con utilidades de MicroStrategy Web, librerías JavaScript, HTML 5 y hojas de estilo CSS (fig. 1).

En los procesos ETL (*Extract, Transform and Load*), los datos fueron extraídos de sistemas operacionales o de ficheros externos, depositados en un servidor FTP y cargados en la base de datos Oracle, tras controles de recepción y detección de inconsistencias. En el proceso de transformación, los datos fueron enriquecidos para facilitar su explotación, realizando operaciones de codificación de diagnósticos, aplicación de reglas de negocio, agrupación de casos (*Adjusted Clinical Groups*)<sup>13</sup> y grupos de morbilidad ajustados<sup>14</sup>, y cálculo de indicadores. Los datos se cargaron en un modelo de datos en estrella, modelo dimensional, y tablas agregadas que alimentan cubos inteligentes de MicroStrategy para favorecer su explotación y rendimiento.

Los diagnósticos clínicos fueron codificados por documentalistas clínicos para los CMBD de hospitalización, cirugía mayor ambulatoria y hospital de día médico, y mediante un codificador automático de desarrollo propio para urgencias hospitalarias e historia clínica electrónica de atención primaria<sup>15</sup>. Para la codificación se utilizó la modificación clínica de la Clasificación Internacional de Enfermedades (CIE), tanto la novena versión española de la CIE-9-MC de 2014 como las dos primeras versiones CIE-10-ES (2016 y 2018). Los diagnósticos de atención primaria se obtuvieron de DIRAYA, de los apartados «Motivo de consulta», «Problemas de salud», «Programas» y «Procesos».

Se estableció una lista de enfermedades crónicas para su seguimiento («patologías BPS») tomando como referencia los grupos de la clasificación CCS (*Clinical Classifications Software*) elaborados por el HCUP (*Healthcare Cost and Utilization Project*)<sup>16</sup>, que agrupan los códigos CIE en un número de categorías con significado clínico útiles para presentar estadísticas descriptivas. Se utilizó el *Chronic Condition Indicator* de HCUP para establecer si un diagnóstico representaba una enfermedad crónica, y se añadió un atributo propio indicando si la cronicidad era permanente o limitada en el tiempo.

De DIRAYA atención primaria y hospitalaria se obtuvieron diagnósticos de enfermería (NANDA, NIC, NOC) y valoraciones funcionales (Barthel) y cognitivas (Pfeiffer y otras).

Los indicadores de prevalencia, expresados como proporciones, se calcularon para periodos anuales de calendario, utilizando como numerador el número de personas-día que presentaron una condición y como denominador el total de personas-día en seguimiento

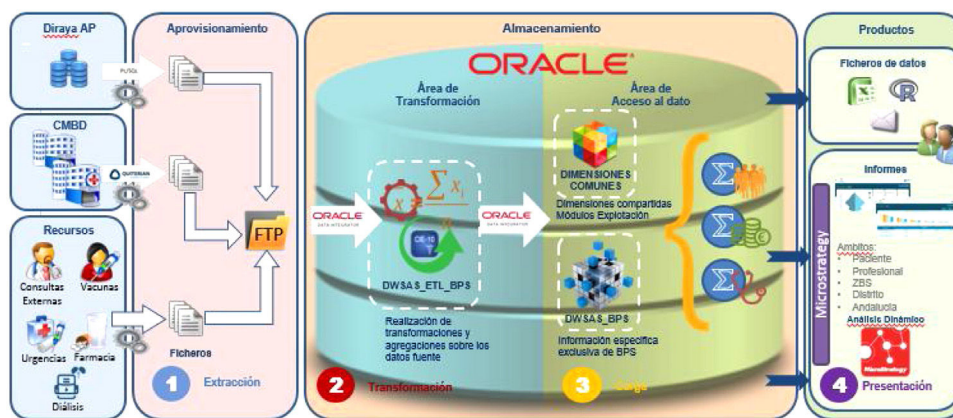


Figura 1. Base poblacional de salud de Andalucía. Arquitectura del sistema.

durante el periodo<sup>17</sup>. Para la comparación de la prevalencia entre distintos ámbitos se calculó la razón estandarizada de prevalencia y su intervalo de confianza.

A cada paciente se le aplicaron los costes unitarios de cada servicio recibido, obtenidos del sistema de contabilidad analítica COAN SSPA<sup>18</sup>.

Los pacientes se asociaron a las claves médicas y de enfermería de los profesionales de atención primaria que los atienden, agrupando los resultados por profesional y centro de atención primaria, unidad de gestión clínica, distrito sanitario y Andalucía.

Se configuró un acceso para las personas autorizadas, con identificación y autenticación segura desde la Red Corporativa de la Junta de Andalucía, limitado a los datos necesarios para el ejercicio de sus funciones específicas. En su desarrollo se tuvieron en cuenta las medidas de seguridad y de protección de datos previstos por la legislación vigente.

## Resultados

### Contenido del sistema

La población incluida en el sistema representaba el 99% de la cifra oficial de población de Andalucía a 31 de diciembre de 2017. De forma acumulada, se incluyeron datos de 12,5 millones de personas que estuvieron dadas de alta en la base de datos de usuarios en algún momento entre los años 2001 y 2017, con un total de 435,5 millones de diagnósticos:

- 347 millones procedentes de atención primaria desde 2001.
- 49 millones procedentes de 8,7 millones de altas del CMBD de hospitalización y de 2,5 millones de altas de cirugía mayor ambulatoria desde 2001.
- 37 millones procedentes de 27 millones de altas del CMBD de urgencias desde 2009.
- 2,5 millones procedentes de consultas externas de salud mental desde 2014.

El 88,7% de estos diagnósticos fueron generados por el codificador automático, y el resto por los servicios de documentación hospitalarios.

También se incorporaron:

- 911.000 valoraciones funcionales y 536.000 valoraciones cognitivas desde el año 2014.
- 12 millones de diagnósticos de enfermería desde 2004.
- 35,5 millones de actos vacunales desde 2000.

- 12 millones de pruebas radiológicas (excepto radiología simple) desde 2014.
- 35 millones de resultados de pruebas analíticas desde 2014.
- 3 millones de constantes (peso, talla, índice de masa corporal, presión arterial) desde 2016.

Al aplicar la contabilidad analítica a las prestaciones recibidas por cada persona incluida en nuestro trabajo se explicó el 91,28% del presupuesto del SAS.

En esta fase inicial, la periodicidad de carga de las fuentes de información fue anual y la unidad temporal de presentación de datos fue el año-calendario.

El sistema ha alcanzado la cifra de más de 300 usuarios con perfil de gestión y salud pública, que forman parte de una comunidad de usuarios en la Red Profesional de la Junta de Andalucía. Para estos usuarios, los datos identificativos de los pacientes están enmascarados. Se ha comenzado la apertura del sistema a los profesionales asistenciales de atención primaria con acceso a los datos nominales de sus pacientes como apoyo a la gestión asistencial.

### Utilidades de acceso a los datos

#### 1) Informes predefinidos

Se elaboraron informes predefinidos para cada nivel de usuarios (médico o enfermera, centro de atención primaria, unidad de gestión clínica, distrito y Andalucía).

El sistema permite a los usuarios visualizar los informes propios de su ámbito y navegar, en su caso, a los de niveles inferiores. Los profesionales asistenciales pueden acceder a una ficha que incluye toda la información disponible de cada uno de sus pacientes (fig. 2).

Los listados predefinidos están agrupados en varias dimensiones que incluyen, entre otras:

- Características de la población: pirámide por edad y sexo, características demográficas, contacto con los servicios sanitarios, valoraciones funcionales, patologías crónicas o polimedicación (fig. 3).
- Prevalencia de enfermedades crónicas, su evolución temporal, razón estandarizada e intervalo de confianza (fig. 4).
- Uso de recursos, incluyendo la evolución del número de consultas y urgencias, los costes totales y los de servicios específicos por grupos de edad y sexo (fig. 5).

#### 2) Consultas dinámicas

Los profesionales sanitarios, según su perfil, pueden realizar mediante MicroStrategy determinadas consultas dinámicas. Ello les



## Ficha del usuario



## INFORMACIÓN DEL PACIENTE

XXXXXXXXXX XXXXXXXX XXXXXXXX ♂

NUHSA - ANXXXXXXXXXX

Activo en BDU ✓

Residente en ZNTS<sup>a)</sup>

Fallecido

71 AÑOS / XX-XX-1946

Días activo en BDU: 365

Residente en Andalucía ✓

Exportar a PDF

Ver Estructura

## Nº de patologías BPS: 8

Categoría	Patología	Fecha inicio	Edad inicio	Duración años	Fuentes <sup>b)</sup>
Endocrino	Diabetes	08-08-2007	60	11,33	3
Endocrino	Dislipemia	08-03-2010	63	8,75	3
Endocrino	Obesidad	11-02-2013	66	5,83	1
Salud Mental	Trastorno de ansiedad	25-06-2014	67	4,50	1
Sistema Nervioso	Enfermedad de Parkinson	28-09-2000	54	18,25	4
Circulatorio	Hipertensión	28-09-2000	54	18,25	5
Circulatorio	Insuficiencia cardiaca	14-05-2013	66	5,58	1
Osteomuscular	Artrosis, espondilosis	21-12-2004	58	14,00	2

## Paciente polimedcado con 15 medicamentos

Índice BARTHEL

Sin valoración Barthel

Índice PFEIFFER

Sin valoración Pfeiffer

Índice NORTON

Sin valoración Norton

Índice ESFUERZO

Sin valoración Esfuerzo Cuidador

Ha contactado con el SSPA

Atención Primaria	2016	2017	Atención Hospitalaria	2016	2017
Consultas Atención Primaria	10	17	Número de Consultas Atención Hospitalaria	7	12
Consultas de Enfermería	2	4	Procedentes de Atención Primaria	0	1
Centro	2	4	Total urgencias	1	1
Domicilio	0	0	Altas hospitalización	0	0
Consultas Médicas	7	13	Estancias hospitalización	0	0
Centro	6	13	Sesiones diálisis	0	0
Domicilio	1	0			
Otras Consultas	1	0	Farmacia	2016	2017
Centro	1	0	Dispensación Farmacia Atención Hospitalaria	962,19€	715,43€
Domicilio	0	0	Dispensación Farmacia Atención Primaria	4.092,15€	4.085,46€
Urgencias	0	0	Farmacia entrega Hospital	0,00€	0,00€

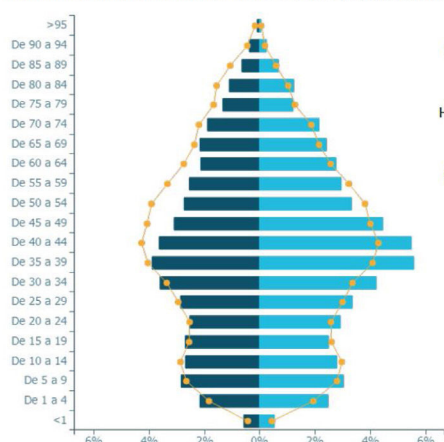
**Figura 2.** Base poblacional de salud de Andalucía. Informe resumen de un paciente. <sup>a</sup> Zona de transformación social. <sup>b</sup> Número de fuentes de información distintas en las que aparece este diagnóstico.



### Descripción del cupo



Periodo 2017, Diciembre / UGC Cuevas de Almanzora



- Mujer**: El cupo medio de su UGC tiene **1315** usuarios frente a los 1368 del cupo medio del SSPA. El **45,50%** de sus usuarios son mujeres frente al 50,92% del SSPA. El **15,55%** de sus usuarios tiene más de 64 años frente al 16,64% del SSPA.
- Hombre**: El **80,29%** de sus usuarios han sido atendidos en AP frente al 82,91% del SSPA. El **0,81%** de sus usuarios acudió a Atención Hospitalaria pero no a Primaria. El **81,73%** de sus usuarios mayores de 65 años no tienen una valoración Barthel.
- Media**: El **82,45%** de sus usuarios mayores de 65 años no tienen una valoración Pfeiffer. Sus usuarios polimedcados consumen una media de **8,04** medicamentos frente a 8,49 del SSPA. El número medio de patologías BPS de sus usuarios es **1,40** frente a 1,53 del SSPA.

	UGC				Distrito 2017	SSPA 2017
	2014	2015	2016	2017		
Usuarios por profesional	1.264	1.270	1.294	1.315	1.180	1.368
Edad media	39	39	39	39	43	41
% Mujeres	43,94%	43,90%	45,67%	45,50%	48,08%	50,92%
% Usuarios contacto AP	83,91%	83,96%	83,86%	80,29%	84,04%	82,91%
% Usuarios contacto solo AH	0,84%	0,82%	1,00%	0,81%	0,96%	1,77%
% Usuarios sin contacto SSPA	15,24%	15,21%	15,13%	14,63%	14,99%	15,31%
Residentes fuera Andalucía	407	375	435	503	3.180	74.882
% Residentes fuera Andalucía	2,93%	2,68%	3,06%	3,48%	2,19%	0,90%
% Copago farmacia medio	33,55%	33,81%	34,21%	34,39%	33,62%	33,65%
Fallecidos	88	62	65	66	999	54.678
Media patologías BPS	1,30	1,35	1,39	1,40	1,53	1,53
Medicación media polimedcados	8,05	8,10	8,06	8,04	8,33	8,49
% > 65 años con Barthel	15,77%	24,01%	20,62%	18,27%	10,72%	12,21%
% > 65 años con Pfeiffer	13%	21,30%	19,65%	17,55%	8,81%	8,36%
<b>Nacido</b>						
Nacido	168	182	174	168	1.246	76.576
% Nacidos mujeres	45,83%	49,45%	50,00%	50,60%	50,80%	48,46%
Tasa bruta natalidad	12,09‰	13,03‰	12,22‰	11,62‰	8,58‰	9,08‰
Tasa global fecundidad	54,37‰	59,28‰	54,72‰	52,11‰	39,32‰	38,57‰
Fallecidos	88	62	65	66	999	54.678
% Fallecidos mujeres	29,55%	38,71%	36,92%	31,82%	45,55%	49,66%
Tasa bruta mortalidad	6,29‰	4,42‰	4,55‰	4,54‰	6,83‰	6,50‰
RE mortalidad	1,08	0,65	0,71	0,76	0,88	1,00
% > 64 años	15,38%	15,42%	15,47%	15,55%	20,50%	16,64%
Índice envejecimiento	0,85	0,85	0,85	0,85	1,27	0,98
Tasa dependencia	0,50	0,50	0,51	0,51	0,58	0,51
Tasa dependencia < 16 años	0,27	0,27	0,28	0,27	0,25	0,26
Tasa dependencia > 64 años	0,23	0,23	0,23	0,23	0,32	0,25

Figura 3. Base poblacional de salud de Andalucía. Características demográfico-sanitarias de la unidad de gestión clínica de Cuevas de Almanzora.



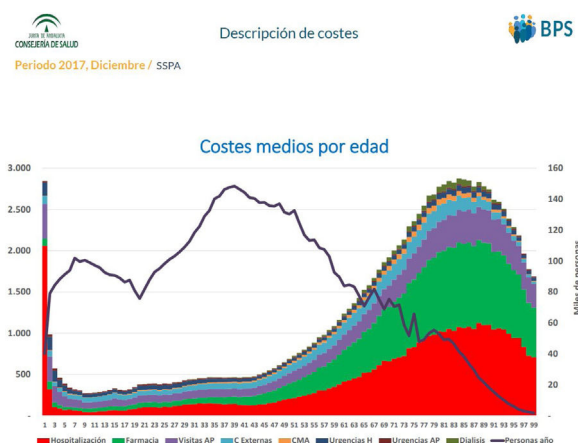
Prevalencia de patologías BPS



Periodo 2017, Diciembre / Distrito Costa del Sol

Patologías BPS	Casos				Distrito			SSPA
	2014	2015	2016	2017	Prevalencia	RE	IC 95%	Prevalencia
					2017			2017
VIH	1.700	1.818	1.911	2.043	3,87%	1,65**	1,57 - 1,72	2,36%
Cáncer de cabeza y cuello	700	733	746	774	1,40%	0,84**	0,78 - 0,91	1,67%
Cáncer de estómago	237	248	246	243	0,41%	0,93	0,81 - 1,07	0,46%
Cáncer colorrectal	1.899	1.983	2.111	2.240	4,03%	0,9**	0,86 - 0,94	4,65%
Cáncer de hígado y vías biliares	273	273	264	276	0,44%	0,98	0,86 - 1,12	0,46%
Cáncer de páncreas	164	164	183	200	0,31%	1,2*	1,01 - 1,4	0,26%
Cáncer de bronquio y pulmón	765	794	818	840	1,25%	1,11*	1,02 - 1,2	1,14%
Cáncer de tiroides	476	524	570	600	1,15%	1,07	0,98 - 1,16	1,08%
Cáncer de hueso y tejidos blandos	181	193	209	215	0,39%	0,82**	0,71 - 0,95	0,48%
Melanoma de piel	772	859	962	1.035	1,90%	1,25**	1,17 - 1,33	1,54%
Cáncer de mama	3.130	3.373	3.643	3.938	7,37%	0,9**	0,87 - 0,93	8,45%
Cáncer de útero	411	446	483	498	0,93%	0,74**	0,67 - 0,81	1,33%
Cáncer de cuello uterino	375	407	444	450	0,85%	0,67**	0,61 - 0,73	1,26%
Cáncer de ovario	234	262	282	300	0,55%	0,9	0,8 - 1,01	0,63%
Cáncer de próstata	2.997	3.061	3.149	3.217	6,05%	1,13**	1,09 - 1,18	5,27%
Cáncer de testículo	129	139	154	171	0,32%	0,92	0,79 - 1,08	0,35%
Cáncer de riñón y pelvis renal	375	410	452	472	0,86%	0,98	0,89 - 1,07	0,90%
Cáncer de vejiga	1.290	1.353	1.392	1.435	2,61%	0,86**	0,81 - 0,9	3,07%
Enfermedad de Hodgkin	178	184	190	206	0,39%	0,9	0,78 - 1,04	0,43%
Linfoma no Hodgkin	603	639	682	724	1,32%	1,01	0,93 - 1,09	1,34%
Leucemia	486	521	526	534	0,98%	1,05	0,96 - 1,14	0,96%
Cáncer inmunoproliferativo	138	152	165	166	0,30%	0,9	0,76 - 1,06	0,34%
Sarcoma de Kaposi	46	48	47	51	0,09%	2,04**	1,5 - 2,71	0,05%
Diabetes	37.731	39.032	40.076	40.941	77,84%	0,9**	0,89 - 0,91	89,23%
Dislipemia	65.662	69.696	72.625	76.414	145,44%	0,75**	0,75 - 0,76	197,41%
Hipotiroidismo	21.978	24.198	26.393	28.573	54,12%	0,87**	0,86 - 0,88	63,03%
Obesidad	16.687	17.923	17.813	18.059	32,59%	0,79**	0,78 - 0,8	41,58%
Demencia	3.585	3.746	3.779	3.941	6,61%	0,69**	0,67 - 0,72	10,67%
Otro trastorno mental orgánico	2.749	3.052	3.021	3.053	4,96%	0,77**	0,74 - 0,8	6,84%
Dependencia alcohol	6.994	8.086	7.771	7.742	12,30%	1,23**	1,2 - 1,26	10,06%
Dependencia tabaco	23.266	23.864	23.541	24.080	43,32%	1,22**	1,2 - 1,24	35,44%
Dependencia otras sustancias	1.697	1.869	1.856	1.865	3,14%	1,02	0,97 - 1,08	3,12%
Trastorno esquizofrénico	2.290	2.616	2.849	2.976	5,62%	0,73**	0,71 - 0,76	7,78%
Trastorno estado animo	11.820	13.647	15.210	16.595	31,12%	0,73**	0,72 - 0,74	43,29%
Trastorno de ansiedad	37.025	39.894	41.140	41.745	74,53%	0,8**	0,79 - 0,81	93,81%
Trastorno conducta alimentaria	571	632	681	669	1,24%	0,85**	0,78 - 0,92	1,44%
Otro trastorno funcional	836	913	879	781	1,32%	1,01	0,93 - 1,09	1,31%
Trastorno personalidad y comportamiento adulto	1.738	2.228	2.577	2.875	5,29%	0,59**	0,57 - 0,61	9,03%
Discapacidad intelectual	812	899	950	995	1,91%	0,5**	0,47 - 0,53	3,85%
Trastorno del espectro autista	440	610	710	809	1,51%	1,16**	1,08 - 1,24	1,29%
Otro trastorno del desarrollo	1.196	1.409	1.555	1.702	2,95%	0,99	0,94 - 1,04	2,90%
Trastorno inicio infancia adolescencia	2.326	2.854	3.122	3.093	5,39%	0,83**	0,8 - 0,86	6,45%
Epilepsia	3.741	3.964	4.098	4.296	8,09%	0,76**	0,74 - 0,79	10,70%
Enfermedad de Parkinson	1.313	1.367	1.380	1.422	2,54%	0,79**	0,75 - 0,84	3,45%
Síndrome extrapiramidal	1.670	1.816	1.935	2.088	3,82%	0,69**	0,66 - 0,72	5,82%
Enfermedad neurológica con déficit motor no ACV	5.641	6.026	6.150	6.323	11,32%	0,84**	0,82 - 0,86	13,75%
Degeneración macular asociada a la edad	1.448	1.535	1.595	1.742	3,17%	0,94*	0,89 - 0,99	3,65%
Glaucoma	7.720	8.171	8.487	8.802	16,69%	0,92**	0,9 - 0,94	18,88%
Retinopatía	4.977	5.395	5.761	6.164	11,51%	0,76**	0,74 - 0,78	15,54%
Enfermedad cerebrovascular aguda	5.461	5.786	6.074	6.245	11,18%	0,8**	0,78 - 0,82	14,66%
Enfermedad cerebrovascular mal definida y otra	820	964	983	1.017	1,61%	0,94	0,88 - 1,01	1,81%
Isquemia cerebral transitoria	1.839	1.922	1.962	2.040	3,45%	0,8**	0,76 - 0,84	4,59%
Oclusión o estenosis arterias precerebrales	889	1.032	977	967	1,61%	1,15**	1,07 - 1,23	1,44%
Secuela enfermedad cerebrovascular	1.100	1.140	1.169	1.181	2,09%	0,81**	0,76 - 0,86	2,72%
Arteriopatía de extremidades	7.163	7.582	7.880	8.150	15,13%	0,85**	0,83 - 0,87	18,33%
Arteriopatía intraabdominal	193	210	232	240	0,43%	0,69**	0,6 - 0,78	0,66%
Aneurisma arterias aorta, periféricas y viscerales	729	800	841	906	1,53%	1	0,93 - 1,08	1,55%
Fibrilación auricular	7.035	7.413	7.684	7.884	13,68%	0,9**	0,88 - 0,92	16,23%
Cardiopatía isquémica	13.206	13.804	14.236	14.665	27,38%	0,96**	0,94 - 0,97	29,76%
Enfermedad valvular adquirida	5.011	5.258	5.152	5.261	8,88%	0,85**	0,82 - 0,87	11,07%
Anomalia congénita cardíaca y circulatoria	695	772	800	819	1,45%	0,72**	0,67 - 0,78	1,98%
Hipertensión	75.152	77.914	80.351	83.094	158,03%	0,81**	0,81 - 0,82	200,63%
Insuficiencia cardíaca	8.920	9.292	9.550	9.911	17,55%	0,82**	0,81 - 0,84	22,82%
EPOC	13.848	14.827	15.423	16.153	29,96%	0,75**	0,73 - 0,76	40,71%
Asma	33.893	35.913	37.456	39.020	75,66%	0,83**	0,82 - 0,84	91,42%
Hepatopatía crónica excepto cirrosis	5.339	5.637	5.713	5.828	10,76%	0,97*	0,94 - 1	11,24%
Esteatosis hepática	4.495	5.011	5.272	5.512	10,19%	0,9**	0,87 - 0,92	11,56%
Cirrosis hepática	1.119	1.177	1.204	1.291	2,27%	0,93*	0,88 - 0,99	2,49%
Enfermedad por reflujo gastroesofágico	7.699	8.388	8.956	9.461	17,30%	0,9**	0,88 - 0,92	19,32%
Enteritis regional y colitis ulcerosa	2.008	2.158	2.318	2.447	4,68%	0,96	0,92 - 1	4,87%
Síndrome de malabsorción e intolerancia alimenticia	4.083	4.511	4.974	5.482	10,33%	0,72**	0,7 - 0,74	14,19%
Psoriasis	5.722	6.123	6.483	6.901	13,24%	0,75**	0,73 - 0,77	17,65%
Dermatitis atópica	6.800	7.127	7.649	7.995	14,76%	0,79**	0,77 - 0,81	18,18%
Osteoporosis	8.897	9.157	9.329	9.544	18,19%	0,77**	0,76 - 0,79	25,53%
Artrosis, espondilosis	42.630	45.739	48.427	50.925	96,93%	0,69**	0,69 - 0,7	145,69%
Gota y otras artropatías por cristales	4.007	4.391	4.743	5.037	9,52%	0,56**	0,54 - 0,57	17,32%
Otra artropatía	7.000	7.900	8.712	9.521	17,92%	0,66**	0,64 - 0,67	27,81%
Artritis reumatoide y enfermedades relacionadas	2.434	2.606	2.753	2.887	5,49%	0,91**	0,87 - 0,94	6,19%
Enfermedad del colágeno y vasculitis	783	843	913	952	1,75%	0,67**	0,63 - 0,72	2,86%
Fibromialgia	3.829	4.082	4.389	4.711	9,01%	0,65**	0,63 - 0,67	13,84%
Insuficiencia renal crónica	5.738	6.175	6.432	6.592	11,81%	1,02	1 - 1,05	12,24%
Litiasis urinaria	10.509	12.192	12.585	13.024	22,81%	0,8**	0,79 - 0,82	28,57%

Figura 4. Base poblacional de salud de Andalucía. Prevalencia de enfermedades crónicas en el distrito sanitario Costa del Sol.



**Figura 5.** Base poblacional de salud de Andalucía. Coste sanitario por persona y año de edad en Andalucía. Año 2017. AP: atención primaria; C: consultas; CMA: cirugía menor ambulatoria; H: hospital.

permite elegir conjuntos de variables para la selección de pacientes (sexo, grupo de edad, número y tipo de patologías BPS, etc.), de los cuales se muestra su identificación y la información clínica y administrativa más relevante disponible en el sistema.

Un grupo más reducido de usuarios de perfil avanzado puede realizar, también mediante MicroStrategy, consultas dinámicas utilizando todas las variables contenidas en el modelo de datos.

### 3) Análisis de datos

Tanto los informes predefinidos como los resultados de las consultas dinámicas son exportables a ficheros CSV para su tratamiento fuera del sistema.

Usuarios expertos en análisis y explotación de datos, en el nivel central del SSPA, pueden acceder directamente a las bases de datos y realizar extracciones mediante SQL o tratar directamente los datos con lenguaje R u otros.

## Discusión

Aunque el proyecto no es novedoso desde la perspectiva tecnológica, supone un avance importante en la gobernanza de datos en un servicio regional de salud. Nos permite adentrarnos en los retos<sup>19–21</sup> y las limitaciones del *Big Data*<sup>22</sup>, que nos facilitará la incorporación de datos desde la web y las redes sociales, los análisis prescriptivos y predictivos, y la toma de decisiones por sistemas expertos<sup>23</sup>. Esta publicación no solo contribuye a aumentar la limitada producción científica<sup>8</sup> en este campo, sino también a la transparencia acerca del uso de los datos sanitarios en las instituciones.

Evolucionaremos, por tanto, a un ecosistema de datos en el que coexistan los entornos *Big Data*<sup>24</sup> con los actuales *Data Warehouse*, sus sistemas de *reporting* y el análisis directo sobre bases de datos relacionales. Estos sistemas tradicionales continuarán teniendo importancia en el futuro, ya que seguirán resolviendo necesidades básicas de la mayoría de los usuarios.

El sistema que hemos implantado permite reconstruir la biografía sanitaria de las personas, y abre posibilidades de análisis longitudinal para la planificación, la gestión y la investigación sanitarias. Además, puede ayudar a la toma de decisiones en los ámbitos estratégico (planes y estrategias), de mesogestión (contratos programa) y microgestión (gestión clínica), con la facilidad de uso de los informes con indicadores predefinidos, y la autonomía del análisis dinámico de datos. Dado que el SSPA presta atención a la mayoría

de la población andaluza, podemos considerar que la cobertura del sistema es poblacional.

Hay que destacar que las fuentes de información sobre actividad asistencial han sido incorporadas al sistema casi en su totalidad, lo que se ha puesto de manifiesto al aplicar la contabilidad analítica a cada una de las personas incluidas en el sistema.

Falta por incorporar al sistema información diagnóstica de consultas externas (salvo salud mental), hospital de día médico, urgencias de atención primaria, anatomía patológica, rehabilitación y farmacia (parcialmente). Igualmente, queda por desarrollar la conexión de BPS con otros registros sanitarios (metabopatías, cáncer, virus de la inmunodeficiencia humana, voluntades anticipadas, etc.).

La incorporación de registros ajenos al SSPA, que no usan el número único de historia de salud de Andalucía como indicador individual, supone un reto importante. En este sentido, se está avanzando en el cruce sistemático de datos entre la base de datos de usuarios y el Registro de Población de Andalucía para incorporar, con fines estadísticos, datos demográficos, sociales y económicos de la población.

El proyecto ha puesto de manifiesto las dificultades en el desarrollo de *Data Warehouse* y *Business Intelligence*<sup>25,26</sup>; por una parte, los procesos ETL desde los sistemas operacionales son caros, complejos y lentos, requieren constantes adaptaciones al cambiar la disponibilidad o la necesidad de datos, y resulta difícil el mantenimiento de las tablas maestras que dan servicio a los sistemas operacionales y de explotación de datos. Además, aunque disponemos de cierto conocimiento acerca de la calidad de los datos del CMDB de hospitalización y urgencias y el proceso de codificación automática, es un aspecto a profundizar<sup>15,27–29</sup>. A medida que profundicemos en el uso de los datos nos podremos encontrar inconsistencias debidas al elevado número de fuentes y profesionales involucrados en la generación de información y a la falta de uniformidad en los criterios de recogida y codificación o en su aplicación.

Por otra parte, se detectan reglas de negocio escasamente definidas o aplicadas desigualmente en los distintos sistemas operacionales. Estos últimos deben mejorar su diseño y reforzar los controles de calidad en la recogida de datos.

Uno de los retos del sistema es la pronta devolución de la información generada a los usuarios finales. Los tiempos de disponibilidad están siendo acortados, y para algunas fuentes se actualiza trimestralmente. El objetivo es que la periodicidad de generación de la información sea mensual, si bien algunos indicadores de funcionamiento deberían estar disponibles prácticamente en tiempo real, lo que requiere nuevos abordajes metodológicos y tecnológicos.

La reproducibilidad es una limitación importante en los estudios a partir de registros debido a diferencias conceptuales y metodológicas; por ejemplo, para identificar las patologías y su duración. Si bien su efecto puede ser limitado en las comparaciones en el interior del SSPA, puede ser importante al comparar distintos sistemas sanitarios. Así, las prevalencias de enfermedades crónicas obtenidas a partir de nuestros datos son superiores a otras, debido probablemente a que nuestra metodología de extracción de diagnósticos desde las distintas fuentes de información es más exhaustiva.

Respecto a la gobernanza de los datos, nos hemos encontrado con los retos previamente descritos<sup>30–32</sup> acerca de modelos organizativos para la gestión y análisis de datos, seguridad y privacidad, y gestión de tablas maestras y metadatos. Aunque el proyecto se ha desarrollado en una organización con un nivel de gobernanza avanzado, los avances en su gobernanza dependerán de los que se produzcan en el SSPA, reforzándose el liderazgo de la gestión y análisis de datos.

El proyecto ha sido contemporáneo con el Reglamento General de Protección de Datos (RGPD), cuyos principios incorpora, aunque es preciso profundizar en múltiples aspectos, en especial en

los relativos a protección de datos en el diseño del sistema y los procedimientos de tratamiento de datos para su uso por terceros, sobre todo en el ámbito de la investigación.

Con respecto al perfil de los usuarios, está bien identificado el del usuario final de informes o consultas dinámicas, que no requiere formación avanzada en análisis de datos. Parece adecuado, en el marco de equipos multidisciplinares, potenciar el perfil de analista de datos; haciendo una similitud con el perfil en *Population Data Science*<sup>33</sup>, este perfil incorporaría competencias en informática, análisis de datos y conocimiento de la población. En nuestro caso, desde la perspectiva poblacional y sanitaria, necesitaríamos profesionales capaces de combinar y analizar grandes bases de datos de diferente origen, identificar patrones y desarrollar una infraestructura que contribuyera a mejorar la salud individual y poblacional, y diera soporte a la investigación, siendo respetuosa con la privacidad, la seguridad y la ética.

Otros perfiles profesionales de creciente protagonismo en este tipo de proyectos son el delegado de protección de datos establecida en el RGPD, el responsable de gobernanza<sup>31</sup> y el *Chief Data Officer*, que está evolucionando a *Chief Analytics Officer* por la importancia creciente del componente analítico de los datos<sup>34</sup>.

El uso secundario de datos clínicos<sup>2</sup>, las herramientas de *Business Intelligence*<sup>35</sup> y el *Big Data*<sup>20</sup> generan nuevos retos; entre ellos, la gobernanza, los estándares y la calidad de los datos, su volumen, disponibilidad y visualización, los métodos de análisis y el personal cualificado para ello, el impacto clínico, y la seguridad y la privacidad.

Este trabajo ha mostrado cómo en un servicio regional de salud es posible conectar registros de diversos silos de información, lo que abre nuevas posibilidades en el análisis de sus datos.

### ¿Qué se sabe sobre el tema?

La digitalización de los servicios sanitarios está generando un volumen de datos sin precedentes, tanto para la atención sanitaria como para su uso secundario. Estos datos están dispersos, por lo que su disponibilidad es limitada.

### ¿Qué añade el estudio realizado a la literatura?

El estudio muestra que es posible la conexión sistemática de registros sanitarios y ponerlos a disposición de múltiples usuarios para su uso con finalidades diversas; se abren nuevas posibilidades en el uso secundario de los datos.

### Editor responsable del artículo

Carlos Álvarez-Dardet.

### Declaración de transparencia

El autor principal (garante responsable del manuscrito) afirma que este manuscrito es un reporte honesto, preciso y transparente del estudio que se remite a GACETA SANITARIA, que no se han omitido aspectos importantes del estudio, y que las discrepancias del estudio según lo previsto (y, si son relevantes, registradas) se han explicado.

### Contribuciones de autoría

D. Muñozerro-Muñiz y J.A. Goicoechea-Salazar: concepción y diseño del estudio, y revisión crítica. F.J. García-León: análisis e

interpretación de los datos y escritura del artículo. A. Laguna-Téllez: codificación automática de los diagnósticos y revisión crítica. D. Larrocha-Mata y M. Cardero-Rivas: gestión de bases de datos, cruce de registros y revisión crítica. Todas estas personas han aprobado la versión final del artículo.

### Agradecimientos

Agradecemos la colaboración de J. Trancoso Estrada, J. Rodríguez Herrera y M.R. Jiménez Romero por sus aportaciones en cuanto al CMBD y CIE; y a V.D. Canto Casasola por su apoyo estadístico.

### Financiación

Ninguna.

### Conflicto de intereses

Ninguno.

### Bibliografía

- Kostkova P. Grand challenges in digital health. *Front Public Health*. 2015;3. (Consultado el 6/3/2019.) Disponible en: <https://www.frontiersin.org/articles/10.3389/fpubh.2015.00134/full>
- Schlegel DR, Ficheur G. Secondary use of patient data: review of the literature published in 2016. *Yearb Med Inform*. 2017;26:68-70.
- Davidson AJ. Creating value: unifying silos into public health business intelligence. *EGEMS (Wash DC)*. 2015;2:1172.
- Lewer D, Bourne T, George A, et al. Data resource: the Kent Integrated Dataset (KID). *Int J Popul Data Sci*. 2018;3. (Consultado el 6/3/2019.) Disponible en: <https://ijpds.org/article/view/427/389>
- Cohort of 100 million Brazilians. Salvador (BA): Center for Data and Knowledge Integration for Health-CIDACS. (Consultado el 6/3/2019.) Disponible en: <https://cidacs.bahia.fiocruz.br/en/platform/cohort-of-100-million-brazilians/>
- Zhao J, Gibb S, Jackson R, et al. Constructing whole of population cohorts for health and social research using the New Zealand Integrated Data Infrastructure. *Aust N Z J Public Health*. 2018;42:382-8.
- Outline for a German Strategy for Artificial Intelligence. Berlin: Stiftung Neue Verantwortung. (Consultado el 6/2/2019.) Disponible en: [https://www.ip.mpg.de/fileadmin/ipmpg/content/aktuelles/Outline\\_for\\_a\\_German\\_Artificial\\_Intelligence\\_Strategy.pdf](https://www.ip.mpg.de/fileadmin/ipmpg/content/aktuelles/Outline_for_a_German_Artificial_Intelligence_Strategy.pdf)
- Safwan R, Meredith R, Burstein F. Business intelligence (BI) system evolution: a case in a healthcare institution. *J Decis Syst*. 2016;25(s1):463-75.
- Lee ES, Black RA, Harrington RD, et al. Characterizing secondary use of clinical data. *AMIA Jt Summits Transl Sci*. 2015;2015:92-6.
- Liaw S-T, Pearce C, Liyanage H, et al. An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings. *Inform Prim Care*. 2014;21:199-206.
- Resolución 0068/18 de la Dirección Gerencia del Servicio Andaluz de Salud por la que se crea la Base Poblacional de Salud. Sevilla: Servicio Andaluz de Salud. (Consultado el 6/2/2010.) Disponible en: [http://www.sas.junta-andalucia.es/principal/documentosacc.asp?pagina=pr\\_base\\_poblac](http://www.sas.junta-andalucia.es/principal/documentosacc.asp?pagina=pr_base_poblac)
- Orden de 27 de febrero de 2002, por la que se establece la efectividad del carácter individual de la libre elección de médico y su gestión por la base de datos de usuarios del Sistema Sanitario Público de Andalucía. Sevilla: Consejería de Salud. (Consultado el 6/2/2010.) Disponible en: <https://juntadeandalucia.es/boja/2002/28/5>
- Sicras-Mainar A, Navarro-Artieda R. Adjusted clinical groups: a patient classification system through risk adjustment. *Rev Peru Med Exp Salud Pública*. 2013;30:308-14.
- Monterde D, Vela E, Clères B, Grupo Colaborativo Gma. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Atención Primaria*. 2016;48:674-82.
- Goicoechea-Salazar JA, Nieto-García MA, Laguna-Téllez A, et al. Desarrollo de un sistema de codificación automática para recuperar y analizar textos diagnósticos de los registros de servicios de urgencias hospitalarios. *Emergencias*. 2013;25:430-6.
- HCUP tools and software. Rockville (MD): HCUP Healthcare Cost and Utilization Project. (Consultado el 6/2/2010.) Disponible en: [https://www.hcup-us.ahrq.gov/tools\\_software.jsp](https://www.hcup-us.ahrq.gov/tools_software.jsp)
- Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research. Principles and quantitative methods*. Belmont, CA: Lifetime Learning Publications; 1982. p. 560.
- Contabilidad Analítica de Gestión del Servicio Sanitario Público de Andalucía (COAN SSPA). Sevilla: Servicio Andaluz de Salud. (Consultado el 6/2/2010.) Disponible en: [https://ws027.juntadeandalucia.es/principal/documentosacc.asp?pagina=pr\\_coan](https://ws027.juntadeandalucia.es/principal/documentosacc.asp?pagina=pr_coan)



19. Las 7V del Big data: características más importantes. Madrid: Instituto de Ingeniería del Conocimiento (IIC). (Consultado el 6/2/2010.) Disponible en: <http://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/#volumen>
20. Lee CH, Yoon H-J. Medical big data: promise and challenges. *Kidney Res Clin Pract.* 2017;36:3–11.
21. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. *Int J Med Inform.* 2018;114:57–65.
22. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Heal Inf Sci Syst.* 2014;2:3.
23. Islam MS, Hasan MM, Wang X, et al. A systematic review on healthcare analytics: application and theoretical perspective of data mining. *Healthcare (Basel).* 2018;6(2.), pii: E54.
24. Beyond the data warehouse: new data management for analytics. Garnert Webinar Hosted by Rick Greenwald. Stamford (CT): Gartner, Inc. (Consultado el 06/02/2010.). Disponible en: <https://www.gartner.com/webinar/3887381/player?commId=333505&channelId=5502&srclId=1-3931087981&webinarType=UpcomingEvent>
25. Brooks P, El-Gayar O, Sarnikar S. A framework for developing a domain specific business intelligence maturity model: application to healthcare. *Int J Inf Manage.* 2015;35:337–45.
26. Young Lee S. Architecture for business intelligence in the healthcare sector. *IOP Conf. Ser.: Mater. Sci. Eng.* 2018;317. (Consultado el 6/3/2019.) Disponible en: <http://iopscience.iop.org/article/10.1088/1757-899X/317/1/012033/pdf>
27. Martínez Jiménez R, García Benavides F. Estadísticas de morbilidad hospitalaria: exactitud del diagnóstico notificado en el libro de registros de altas. *Med Clin (Barc).* 1991;96:765–8.
28. Peng M, Eastwood C, Boxill A, et al. Coding reliability and agreement of International Classification of Disease, 10th revision (ICD-10) codes in emergency department data. *Int J Popul Data Sci.* 2018;3. (Consultado el 6/3/2019.) Disponible en: <https://ijpds.org/article/view/445/506>
29. Henderson T, Shephard J, Sundararajan V. Quality of diagnosis and procedure coding in ICD-10 administrative data. *Med Care.* 2006;44:1011–9.
30. Elliott TE, Holmes JH, Davidson AJ, et al. Data warehouse governance programs in healthcare settings: a literature review and a call to action. *EGEMS (Wash DC).* 2013;1:1010.
31. Andronis K, Moysey K. Data governance for health care providers. *Stud Health Technol Inform.* 2013;193:299–315.
32. Holmes JH. Privacy, security, and patient engagement: the changing health data governance landscape. *eGEMs (Wash DC).* 2016;4:9.
33. McGrail K, Jones K, Akbari A, et al. A position statement on population data science. *Int J Popul Data Sci.* 2018;3:4.
34. Bugajski J, Popkin J, White A. The rise of data and analytics' roles points to digital business transformation. Stamford (CT): Gartner, Inc. (Consultado el 6/2/2010.). Disponible en: <https://www.gartner.com/doc/3202220?cm.sp=swg--research--tail>
35. Ward MJ, Marsolo KA, Froehle CM. Applications of business analytics in healthcare. *Bus Horiz.* 2014;57:571–82.