



1032 - RECOGIDA DE DATOS MEDIANTE METODOLOGÍAS DE REUTILIZACIÓN BASADAS EN ALGORITMOS ETL: UNA HERRAMIENTA PODEROSA EN LA INVESTIGACIÓN CLÍNICA EN GRIPE

Mario Cereza Latre¹, Paula Rubio², Miguel Pedrera², Noelia García², Irene Muñoz¹, Jesús Alonso Carrillo¹, Juan Luis Cruz² y Antonio Lalueza Blanco¹

¹Medicina Interna, Hospital Universitario 12 de Octubre, Madrid, España. ²Informática, Hospital Universitario 12 de Octubre, Madrid, España.

Resumen

Objetivos: La recogida de datos en investigación clínica se realiza habitualmente mediante la introducción manual de información en los formularios electrónicos de reporte de casos (eCRF). Estos datos, en su mayoría, provienen de la historia clínica electrónica (HCE), lo que implica un proceso redundante, propenso a errores y que consume una cantidad considerable de tiempo por parte del personal investigador. La posibilidad de extraer esta información directamente desde la HCE, mediante metodologías de reutilización de datos clínicos basadas en algoritmos de extracción, transformación y carga (ETL), representa una oportunidad para mejorar la eficiencia y la calidad en la captura de los mismos. Estas metodologías automatizadas permitirían reducir los tiempos de recogida, minimizar errores de transcripción y aprovechar de forma más eficaz los recursos ya disponibles en los sistemas de información clínica. El objetivo de este trabajo es demostrar que la aplicación de algoritmos de ETL para la reutilización de datos provenientes de la HCE es una estrategia eficaz para optimizar la recogida de los mismos en investigación clínica garantizando altos estándares de calidad, integridad y fiabilidad de la información.

Métodos: Estudio prospectivo realizado en el Hospital Universitario 12 de Octubre de Madrid en el que se incluyen todos los pacientes mayores de 18 años con infección por gripe demostrada microbiológicamente que precisaron ingreso hospitalario durante la temporada 2022-2023. La obtención de datos se llevó a cabo mediante dos vías alternativas: a) Captura manual de 177 variables clínicas a través de un eCRF en la plataforma RedCap; b) Desarrollo e implementación de una metodología automatizada de reutilización de datos basada en algoritmos de extracción, transformación y carga (ETL) aplicados directamente a la HCE. Para la extracción de los datos se definieron consultas en lenguaje SQL y para el resto de etapas se utilizó el *software* de RStudio. Finalmente, se compararon la completitud y disponibilidad de los datos obtenidos mediante el proceso ETL con los recogidos manualmente, evaluando la eficacia y precisión de la metodología automatizada.

Resultados: Se incluyeron 351 pacientes. Se obtuvieron 177 variables mediante recogida manual de datos en eCRF mientras que las variables cargadas desde la HCE mediante algoritmos ETL fueron 123 (69,5% del total). Los formularios con mayor completitud de datos desde la HCE fueron los de laboratorio debido a la buena estructuración en el origen mientras que los peor registrados son los formularios de hemofagocitosis por no estar recogidos en la HCE o estarlo en texto libre en lugar de como dato estructurado. Para conseguir que la extracción sea de mayor utilidad debe priorizarse que las variables utilizadas tengan formato de dato

estructurado y estandarizado en la HCE en lugar de texto libre para garantizar una reutilización de datos de calidad.

Formulario	Variables totales (N)	Variables reutilizadas (N)
Datos de filiación del paciente	13	9
Enfermedades previas	28	22
Valoración al ingreso	15	11
Analítica al ingreso	15	15
Nadir en hemograma	14	14
Bioquímica	12	12
Hemofagocitosis	17	0
Manejo de la infección	31	27
Complicaciones	22	6
Evolución del episodio	4	4
Asistencia 30 días posgripe	4	3
Comentarios	2	0

Conclusiones: La metodología propuesta (algoritmo ETL) permitiría extraer datos de la HCE con fines de investigación facilitando el trabajo de recogida de datos y minimizando errores. Para optimizar su eficiencia debe priorizarse la utilización de datos estructurados en la HCE en lugar de texto libre.