

UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO



**ARMONIZACIÓN Y ANÁLISIS DE DATOS CLÍNICOS BASADO EN
EL OMOP COMMON DATA MODEL: UN CASO DE USO CON
ALERTA CLAVE AZUL EN EL HOSPITAL CLÍNICO UNIVERSIDAD
DE CHILE**

Alonso Carvajal Moreno

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN INFORMÁTICA MÉDICA

Director de Tesis: Prof. Dr. Steffen Härtel

2023

UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO



**ARMONIZACIÓN Y ANÁLISIS DE DATOS CLÍNICOS BASADO EN
EL OMOP COMMON DATA MODEL: UN CASO DE USO CON
ALERTA CLAVE AZUL EN EL HOSPITAL CLÍNICO UNIVERSIDAD
DE CHILE**

Alonso Carvajal Moreno

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN INFORMÁTICA MÉDICA

Director de Tesis: Prof. Dr. Steffen Härtel

2023

Dedico este trabajo a la convergencia de las disciplinas de la salud y la tecnología, donde la búsqueda del bienestar humano se encuentra con el potencial transformador de la innovación tecnológica, organizacional y humana. Consciente de la responsabilidad que cada individuo tiene sobre su propia salud, aspiramos a un futuro donde la tecnología eleve no solo el conocimiento, sino también la humanidad. Con gratitud a todos los que me acompañaron en este camino de exploración y descubrimiento, y con la firme convicción de que mejorar la salud es un deber compartido por todos.

Agradecimientos

A mi querida novia, por acompañarme en esta travesía llena de retos y desafíos. En momentos en los que la vida me absorbía, tú estabas allí, recordándome la importancia de dedicar tiempo y cariño a nuestra relación. Tu presencia ha sido una luz en las sombras y un recordatorio constante del valor de compartir la vida con alguien especial.

A mi madre, por ser mi inagotable fuente de apoyo y aliento. Siempre has creído en mí y me has brindado el soporte necesario para alcanzar mis metas.

A mi mentor, el Dr. Steffen Härtel, por su invaluable sabiduría, consejos y paciencia a lo largo de este proceso. Su guía experta y apoyo incondicional fueron fundamentales para el éxito de este trabajo. Estoy enormemente agradecido por las oportunidades que me brindó para desarrollarme en el ámbito de la informática médica. Su compromiso y dedicación han sido una inspiración en mi formación académica y profesional.

Agradezco al Hospital Clínico Universidad de Chile y a los médicos del Servicio de Medicina Interna, en particular a Gonzalo Díaz, Fabian Miranda, Francisco Gonzalez, y al Dr. Alfredo Aguila, quien es el jefe de la Unidad Grupo Relacionado al Diagnóstico y Ficha Clínica Electrónica, por su apoyo en este trabajo. Su iniciativa al proponer el tema y generar la atención de los demás clínicos será fundamental para instaurar la informática médica dentro del hospital.

Agradezco a la comunidad OHDSI por su invaluable capital humano, donde siempre encuentro respuestas y apoyo frente a cualquier pregunta o duda. Destacar especialmente a Jose Posada, líder del grupo de trabajo LATAM en OHDSI, quien con su experiencia ha sido clave en los inicios de este proyecto al proporcionar valiosa orientación y apoyo.

Quiero reconocer de manera especial a mis compañeros del Magíster en Informática Médica por compartir espacios y momentos de discusión sobre nuestros proyectos. En particular, a mi compañera y colega, Macarena Molina, por acompañarme en este emocionante viaje hacia el mundo de la salud digital.

A todos ellos, mi más profundo agradecimiento por ser parte fundamental de este camino de aprendizaje y crecimiento profesional.

Este trabajo fue financiado por la Agencia Nacional de Investigación y Desarrollo (ANID) y el German Academic Exchange Service a través de los proyectos FONDECYT 1211988 y DAAD 57519605.

Índice

1. Introducción.....	10
2. Hipótesis y Objetivos.....	16
2.1. Hipótesis.....	16
2.2. Objetivo General.....	16
2.3. Objetivos Específicos.....	16
3. Metodología.....	17
3.1. Diseño del Estudio.....	17
3.2. Fuente de Datos.....	17
3.3. Datos Clínicos.....	18
3.4. Infraestructura Tecnológica para la Gestión de Datos.....	19
3.5. Transformación de los Datos al Estándar OMOP CDM.....	19
3.6. Caracterización de la Base de Datos y Evaluación de la Calidad de Datos.....	21
3.7. Caso de Uso Análisis de Datos.....	22
3.7.1. Elaboración del Protocolo de Estudio.....	22
3.7.2. Elaboración y Validación Interna del Modelo.....	24
3.7.3. Validación Externa.....	24
3.7.4. Ciencia Abierta e Intercambio de Evidencia.....	24
3.8. Privacidad y Anonimización de Datos.....	25
4. Resultados.....	27
4.1. Extracción de los Datos desde la Fuente y Creación de la Base de Datos Clínica.....	27
4.1.1. Extracción Exitosa de los Datos Clave Azul para su Análisis.....	27
4.1.2. Perfilamiento de los Datos para la Base de Datos Clínica.....	27
4.1.3. Mapeo Sintáctico de los Datos Clave Azul al Formato OMOP CDM.....	28
4.1.4. Mapeo Semántico de los Datos para su Interpretación en el Contexto OMOP CDM.....	30
4.1.5. Implementación Exitosa del Proceso de Extracción, Transformación y Carga en los Datos Clave Azul.....	31
4.2. Caracterización y Evaluación de la Calidad de Datos en la base de datos OMOP: Un Análisis Preliminar.....	33
4.2.1. Caracterización Preliminar de la Base de Datos OMOP.....	33
4.2.2. Una Evaluación Armonizada de la Calidad de Datos: Un Enfoque Basado en el Marco de Trabajo de Kahn et al.....	34
4.3. Implementación Estándar y Accesible del Caso de Uso de Análisis de Datos en la Base de Datos OMOP.....	37
4.3.1. Elaboración del Protocolo de Estudio.....	37
4.3.2. Elaboración y Validación Interna del Modelo.....	37
5. Discusión.....	42
5.1. Extracción, Transformación y Carga de Datos Clínicos Relevantes: Transformación Exitosa de la Fuente de Datos al Estándar OMOP CDM.....	42
5.1.1. Experiencia en la Extracción y Transformación de Datos Clínicos: Utilizando	

el Modelo OMOP CDM.....	42
5.1.2. Perfilamiento de Datos con WhiteRabbit: Identificación de Campos y Consistencia con el Diccionario de Datos Planificado.....	42
5.1.3. Mapeo Sintáctico a OMOP CDM: Logros y Desafíos en el Análisis del Procedimiento Clave Azul.....	43
5.1.4. Mapeo Semántico en OMOP CDM: Uso de Vocabularios Estandarizados y Colaboración con Expertos Clínicos.....	44
5.1.5. Integración Eficiente de Datos en OMOP CDM: Implementación del ETL y Desafíos Encontrados.....	45
5.2. Caracterización y Evaluación de la Calidad de la Base de Datos Transformada a OMOP: Utilización de ATLAS y DQD.....	46
5.2.1. Caracterización Detallada de la Base de Datos Transformada a OMOP CDM: Análisis y Resultados con ATLAS.....	46
5.2.2. Evaluación Exitosa de la Calidad de Datos en la Base de Datos OMOP CDM: Resultados y Recomendaciones con DQD.....	47
5.3. Análisis de Datos en la Base de Datos Transformada a OMOP: Un Caso de Uso de Pacientes Clave Azul en un Hospital Universitario.....	48
5.3.1. Enfoque Metodológico Propuesto por OHDSI para el Análisis de Datos.....	48
5.3.2. Contribuciones del Marco Conceptual Propuesto.....	48
5.3.3. Exploración de Covariables Influyentes en el Modelo.....	48
5.3.4. Evaluación de Métricas de Rendimiento y Capacidad del Modelo.....	49
6. Conclusión.....	50
7. Perspectivas Futuras.....	51
8. Bibliografía.....	52
9. ANEXO A: Tablas de Información.....	55
10. ANEXO B: Figuras.....	59

Resumen

Introducción: Este trabajo se enfoca en realizar un análisis de datos de Predicción a Nivel de Pacientes, para lo cual se llevó a cabo la armonización y estructuración de los datos clínicos del Hospital Clínico Universidad de Chile (HCUCH) utilizando el estándar Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). El objetivo es evaluar el riesgo de mortalidad en pacientes hospitalizados tras un procedimiento llamado Clave Azul utilizando sus datos clínicos.

Metodología: Se realiza un análisis de Predicción a Nivel de Pacientes utilizando el estándar OMOP CDM en un estudio observacional retrospectivo de pacientes con alerta Clave Azul en el HCUCH entre Enero 2016 y Diciembre 2020. Para preservar la privacidad, se aplica un proceso de de-identificación. Se recopilan 31 variables en seis categorías: Pacientes, Hospitalizaciones, Laboratorio, Signos Vitales, Clave Azul y Defunciones, seleccionadas por relevancia clínica y sugerencias del equipo de Medicina Interna. El mapeo sintáctico y semántico se efectúa con Rabbit-in-a-Hat y vocabularios como Logical Observation Identifiers Names and Codes (LOINC) y Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT). La calidad de la base de datos se evalúa con DataQualityDashboard (DQD) en las categorías Conformancia, Completitud y Plausibilidad. Se aplica el algoritmo Gradient Boosting Machine (GBM) para predecir la mortalidad intrahospitalaria y se evalúa con métricas como Sensibilidad, Valor Predictivo Positivo y el Área bajo la curva ROC (AUC, Area Under the ROC Curve).

Resultados: El mapeo de datos clínicos a OMOP CDM fue exitoso, transformando el 100% de los registros de pacientes hospitalizados, salvo algunos campos con datos faltantes. Se mapean de forma precisa el 100% de los 46 conceptos identificados, incluyendo exámenes de laboratorio, signos vitales y aspectos clave de la Clave Azul. La calidad de datos en la base OMOP fue alta, con un cumplimiento del 98% en 2334 chequeos. Se detectaron errores menores, validados por expertos clínicos. El análisis de datos involucra entrenamiento y validación de un modelo de clasificación según las mejores prácticas de OHDSI. El modelo muestra un rendimiento con Sensibilidad del 84.6%, Valor Predictivo Positivo del 61.1% y un AUC de 0.7.

Conclusión: El estudio logra sus objetivos al emplear el enfoque metodológico basado en OMOP CDM y las herramientas de OHDSI. Los resultados obtenidos respaldan la calidad y precisión de los análisis realizados, proporcionando una sólida base para futuras investigaciones y la toma de decisiones clínicas.

Abstract

Introduction: This study focuses on conducting a Patient-Level Prediction data analysis. The harmonization and structuring of clinical data from the Hospital Clínico Universidad de Chile (HCUCH) are carried out using the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) standard. The aim is to evaluate the risk of mortality in hospitalized patients following a procedure known as a Code Blue using their clinical data.

Methodology: A Patient-Level Prediction analysis is performed using the OMOP CDM standard in a retrospective observational study of patients with a Code Blue alert at HCUCH between January 2016 and December 2020. To preserve privacy, a de-identification process is applied. Thirty-one variables in six categories are collected: Patients, Hospitalizations, Laboratory, Vital Signs, Code Blue, and Deaths, selected for clinical relevance and based on suggestions from the Internal Medicine team. Syntactic and semantic mapping is performed using Rabbit-in-a-Hat and vocabularies such as Logical Observation Identifiers Names and Codes (LOINC) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT). Data quality is assessed with DataQualityDashboard (DQD) in the categories of Conformance, Completeness, and Plausibility. The Gradient Boosting Machine (GBM) algorithm is applied to predict in-hospital mortality and evaluated with metrics like Sensitivity, Positive Predictive Value, and the Area under the ROC Curve (AUC).

Results: The mapping of clinical data to OMOP CDM is successful, transforming 100% of the records of hospitalized patients, except for some fields with missing data. One hundred percent of the 46 identified concepts, including laboratory tests, vital signs, and key aspects of Code Blue, are accurately mapped. Data quality in the OMOP database is high, with a 98% compliance rate in 2334 checks. Minor errors are detected and validated by clinical experts. The data analysis involves training and validating a classification model following OHDSI's best practices. The model demonstrates performance with a Sensitivity of 84.6%, a Positive Predictive Value of 61.1%, and an AUC of 0.7.

Conclusion: This study achieves its objectives by using the methodological approach based on OMOP CDM and OHDSI tools. The results obtained support the quality and precision of the analyses, providing a solid foundation for future research and clinical decision-making.

1. Introducción

Durante la última década, la investigación biomédica experimenta una revolución en la generación y almacenamiento de datos clínicos. Con avances tecnológicos y la adopción generalizada de registros electrónicos de salud, se observa un crecimiento exponencial en la cantidad de información disponible. Por ejemplo, en el HCUCH, se almacenan miles de registros médicos electrónicos de pacientes, que abarcan desde datos narrativos y resultados de pruebas de laboratorio hasta imágenes médicas y mediciones numéricas de signos vitales. Este vasto conjunto de datos plantean desafíos únicos para los investigadores y profesionales de la salud en la gestión y análisis efectivo de la información clínica relevante.

En el ámbito de la investigación clínica, se utilizan diferentes términos para describir la amplia gama de datos disponibles. Algunos investigadores se refieren a ellos como *dato biomédico*, mientras que otros los clasifican como *dato clínico*. Según Weber et al. [1], los datos biomédicos abarcan información relacionada con la biología y la salud de los pacientes, como registros médicos, ensayos clínicos, registros electrónicos de salud e imágenes médicas. Por otro lado, Shortliffe & Barnett [2] usan otro concepto, el de dato clínico, y lo define “como cualquier observación individual de un paciente, por ejemplo, una lectura de temperatura, un recuento de glóbulos rojos, un historial previo de rubéola o una lectura de presión arterial”. Shortliffe & Barnett añaden también, que los tipos de datos clínicos “abarcan desde información narrativa y textual hasta mediciones numéricas, información genética, señales registradas, dibujos y fotografías u otras imágenes” [2].

Diversos profesionales de la salud recopilan estos datos y los utilizan para evaluar el estado de los pacientes, planificar tratamientos y mejorar la atención médica. En este trabajo, utilizamos indistintamente los términos dato biomédico o dato clínico, ya que ambas perspectivas comparten datos en común.

Cabe destacar que la generación y almacenamiento de datos biomédicos alcanzan volúmenes significativos en la actualidad. Se estima que el 30% de los datos almacenados en el mundo están relacionados con la salud, y un solo paciente puede generar hasta 80 megabytes al año en datos biomédicos [3]. Estos datos constituyen un soporte fundamental para la investigación y desarrollo en el ámbito biomédico, y se pronostica que su crecimiento siga a una tasa anual compuesta del 36% entre 2018 y 2025 [4].

La diversidad y complejidad de los datos clínicos plantean desafíos en el desarrollo de herramientas informáticas para la investigación reproducible y el análisis de datos. En respuesta a este problema, han surgido herramientas de modelado de datos biomédicos que permiten determinar qué elementos se almacenarán, cómo se almacenarán y sus relaciones y limitaciones. Existen diversos modelos de datos que se adaptan a diferentes contextos de aplicación, entre ellos, el *Informatics for Integrating Biology & the Bedside (i2b2)* [5], el *Sentinel Common Data Model (SCDM)* [6], el *National Patient-Centered Clinical Research Network (PCORnet)* [7], y el *Medical Information Mart for Intensive Care (MIMIC)* [8]. Cada uno de estos modelos tiene sus propias características y ventajas para la gestión y análisis de datos clínicos. La elección del modelo apropiado se vuelve una tarea desafiante debido a la necesidad de equilibrar la complejidad requerida para capturar la diversidad de situaciones clínicas, con la usabilidad y accesibilidad necesarias para su aplicación práctica [9]. En este sentido, los modelos de datos clínicos desempeñan un papel fundamental, ya que cada uno tiene sus fortalezas y debilidades, permitiendo adaptarse a diferentes necesidades y contextos específicos.

En este contexto nace OHDSI, una colaboración internacional que se dedica a utilizar datos de salud observacionales para generar evidencia clínica. OHDSI es una comunidad de ciencia abierta que busca mejorar la salud mediante la generación colaborativa de evidencia para decisiones y atención médica. Realiza investigaciones metodológicas, desarrolla herramientas analíticas de código abierto y aplica estas prácticas a preguntas clínicas [10]. Cuenta con más de 2,000 colaboradores en 74 países, y tiene acceso a aproximadamente 810 millones de registros de pacientes únicos de todo el mundo [11]. Ser colaborador de OHDSI implica contribuir al logro de una comprensión completa de la salud y enfermedad a través de la ciencia abierta.

OHDSI ha creado y mantiene el estándar OMOP CDM, un modelo de datos estandarizado utilizado en investigaciones de salud para el intercambio de datos clínicos. Contiene más de 30 tablas que cubren diversos dominios de datos, emplea estándares de vocabulario para codificar conceptos médicos y es altamente escalable. Su adopción ha facilitado la colaboración global en investigación médica y ha impulsado importantes avances en la generación de evidencia clínica [12]. En el ámbito médico, los vocabularios estandarizados son esenciales para organizar y analizar datos de pacientes. OMOP CDM utiliza conceptos para representar eventos médicos y los clasifica por dominios para facilitar su gestión. Además, establece relaciones y jerarquías entre los conceptos para un

análisis más profundo. Con la ayuda de las tablas internas de referencia, se logran abordar casos especiales y situaciones particulares mediante la inclusión de vocabularios adicionales y la colaboración con grupos de trabajo dedicados a la mejora y expansión del modelo de datos [13]. El proceso de Extracción, Transformación y Carga (*ETL, Extract Transform and Load*) es esencial para transformar y estandarizar los datos médicos en el modelo OMOP CDM, permitiendo análisis e investigaciones médicas a gran escala. Cada estudio se enfrenta a desafíos específicos debido a la diversidad de formatos de datos en las diferentes instituciones de salud. Esto dificulta la reproducción del método analítico y conduce a una multiplicidad de métodos escritos en diversos lenguajes de programación. Esta situación conlleva problemas de escalabilidad, transparencia y accesibilidad, además de implicar altos costos y tiempos prolongados, donde los usuarios no expertos se ven limitados en su capacidad para reproducir los resultados de manera efectiva [14]. Para abordar este desafío, OHDSI proporciona herramientas y pautas que simplifican el proceso de ETL, garantizando que se mantenga actualizado y se adapte a cambios en los datos de origen, vocabularios y el propio modelo OMOP CDM. Una regla clave que se enfatiza es la del 80/20, lo que significa priorizar el mapeo de los conceptos más relevantes y aceptar la pérdida de datos de baja calidad. Dicho de otra manera, se prioriza la transformación y mapeo de los conceptos más críticos y esenciales, en lugar de intentar mapear cada concepto individual y detallado presente en la fuente de datos. Para obtener apoyo y asistencia, se recomienda utilizar el foro de implementadores de OHDSI, donde se pueden compartir ejemplos y seguir convenciones acordadas para agilizar el proceso [15].

En el ámbito del análisis de datos, OHDSI se enfoca en casos de uso categorizados en tres grupos principales: *Caracterización*, *Estimación a Nivel Poblacional* y *Predicción a Nivel de Pacientes*. La *Caracterización* implica comprender las características de cohortes o bases de datos completas y responder preguntas sobre las prácticas de atención médica y cómo estas cambian con el tiempo. La *Estimación a Nivel Poblacional* busca establecer inferencias causales sobre los efectos de las intervenciones médicas. La *Predicción a Nivel de Pacientes* se centra en hacer predicciones sobre eventos de salud futuros para pacientes individuales.

La caracterización de la base de datos en OMOP CDM es un proceso que utiliza estadísticas descriptivas para entender las variaciones en poblaciones basadas en diversas características. El uso de herramientas como *ATLAS* y *R* es común en este proceso para realizar análisis y visualización de datos de salud de manera efectiva [16, 17]. En el

contexto de Predicción a Nivel de Pacientes, OHDSI aborda el desafío de anticipar eventos de salud futuros mediante el uso de datos médicos observacionales. Este proceso involucra la extracción cuidadosa de datos relevantes, seguido del ajuste del modelo mediante diversos algoritmos de aprendizaje automático disponibles en el paquete *PatientLevelPrediction*¹ (PLP). Sin embargo, la investigación con bases de datos OHDSI tiene limitaciones, como la incapacidad de determinar efectos causales de intervenciones en comparación con placebos y desafíos con datos faltantes o erróneos. Para abordar estos aspectos y garantizar la calidad de los casos de uso de Predicción a Nivel de Pacientes, se siguen pautas de desarrollo, validación externa y se proporciona una descripción detallada de los modelos para que los resultados de los estudios se difundan de manera transparente, beneficiando la atención médica personalizada y la toma de decisiones clínicas [18].

Otro punto importante en el análisis de datos en salud es la evaluación de la *calidad de datos*, aspecto que resulta esencial para la toma de decisiones en la atención médica, ya que datos incompletos, inconsistentes o poco confiables pueden llevar a diagnósticos incorrectos, tratamientos inapropiados y riesgos para la seguridad del paciente. Según Roebuck (2011), la calidad de los datos se define como “El estado de completitud, validez, consistencia, puntualidad y precisión que hace que los datos sean apropiados para un uso específico”. A esto hay que sumar que la mayoría de los datos gestionados por sistemas de información en salud no fueron originalmente almacenados con fines de investigación, lo que plantea inquietudes acerca de su calidad. Los problemas en la calidad de datos pueden surgir en distintas etapas del ciclo de vida de los datos, que abarcan desde la recopilación, integración y almacenamiento hasta el análisis y la publicación de la información [19]. Por tanto, es fundamental contar con métodos robustos de gestión de la calidad de los datos [20]. En este contexto, OHDSI ofrece una valiosa herramienta llamada *DQD* que contribuye a abordar estos desafíos [21]. El uso de un modelo de datos común y herramientas estandarizadas, como *DQD*, ayuda a disminuir la variabilidad en la evaluación de la calidad de datos, garantizando que los datos clínicos sean confiables y apropiados para su uso en investigaciones y toma de decisiones clínicas. Y para evaluar si los datos clínicos usados en este trabajo son aptos para su uso, se utiliza la definición genérica de Kahn et al. [22], que consta de tres componentes: *Conformancia* (adherencia a estándares y formatos), *Completitud* (presencia de variables y valores registrados) y *Plausibilidad* (credibilidad de los valores registrados). Cada componente puede evaluarse

¹ <https://ohdsi.github.io/PatientLevelPrediction>

mediante *verificación* (evaluación de restricciones de datos y conocimiento local) o *validación* (alineación de datos con referencia externa).

Numerosos estudios adoptan actualmente el modelo OMOP CDM para realizar investigaciones observacionales a gran escala y analizar datos clínicos de manera reproducible. Por ejemplo, en Papez et al. (2022), se utiliza OMOP CDM con datos del Biobanco del Reino Unido con 500,000 participantes, facilitando análisis federados internacionales e informando sobre COVID-19. De manera similar, en el estudio de Lamer et al. (2021), OMOP CDM se emplea para facilitar la explotación y compartición de datos de anestesia. Además, el trabajo de Boyce et al. (2016) implementa OMOP CDM utilizando datos de instalaciones de enfermería especializada para mejorar la disponibilidad de datos electrónicos para la investigación en hogares de cuidado de enfermería. Estos son solo algunos ejemplos de la creciente adopción de OMOP CDM en la comunidad de investigación médica, y muchos otros estudios utilizan con éxito este modelo para abordar diversas preguntas clínicas [23–33].

En este trabajo de investigación, se aplica la transformación a OMOP CDM en un caso de uso específico: el análisis de datos de los procedimientos Clave Azul en el HCUCH. Estos procedimientos representan una medida de evaluación y se utilizan como indicador para pacientes hospitalizados que presentan signos tempranos de deterioro clínico (Figura 1). El propósito del análisis es desarrollar un modelo de Predicción a Nivel de Pacientes para evaluar la mortalidad de los pacientes que han sido sometidos a los procedimientos Clave Azul. Antes del mapeo a OMOP CDM, se realiza un levantamiento detallado del proceso de este procedimiento para obtener una visión clara de su dinámica y evolución en el tiempo (Figura Anexa B.2). Se lleva a cabo la recopilación de datos clínicos relevantes con fines teóricos y de investigación para conocer la dinámica de los pacientes y su evolución en el tiempo, sin que este modelo tenga aplicación directa en la evaluación o tratamiento real de los pacientes. El desafío principal de este trabajo radica en la transformación y estructuración adecuada de los datos del procedimiento Clave Azul al formato OMOP CDM, teniendo en cuenta su falta de estandarización inicial. Además, se efectúa una evaluación de la calidad de la base de datos transformada y se aplica un análisis estandarizado utilizando las herramientas de OHDSI para desarrollar el modelo predictivo de mortalidad.

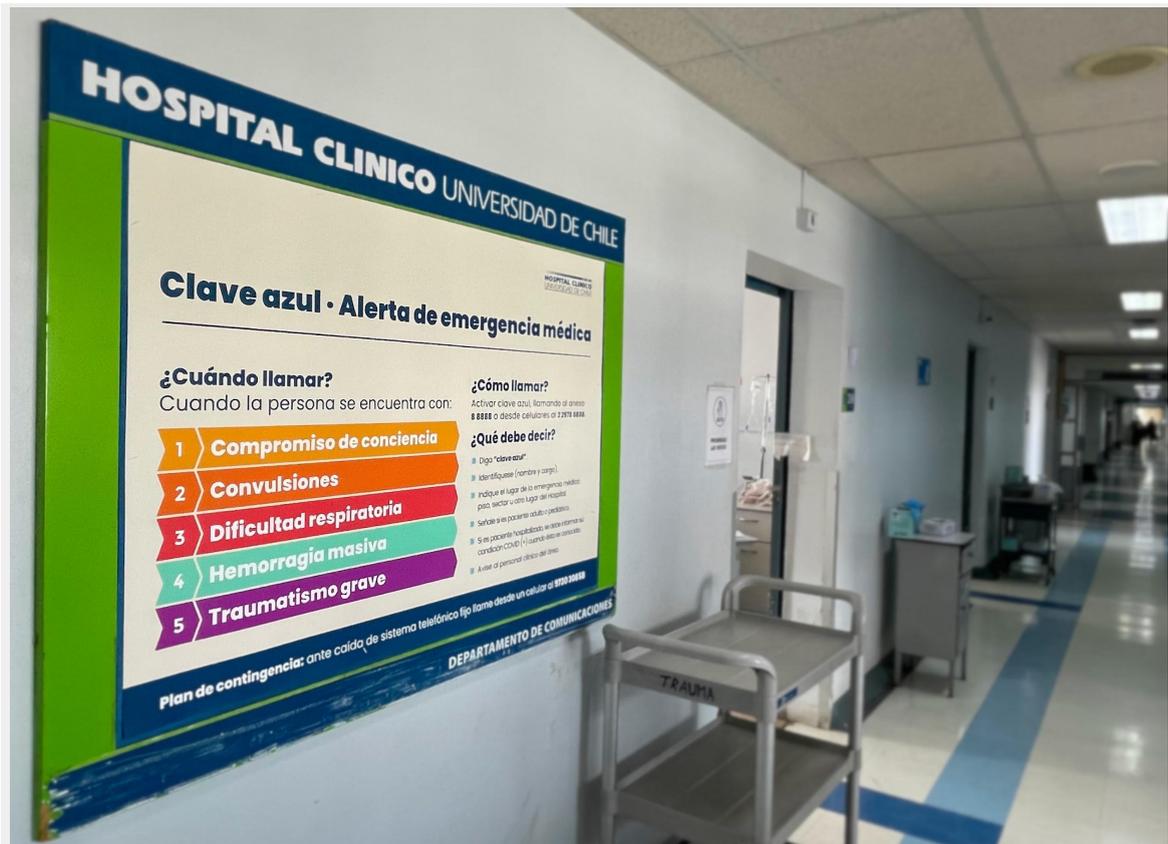


Figura 1. Panel de información de emergencia médica del procedimiento Clave Azul en el Hospital Clínico Universidad de Chile: ¿Cuándo llamar? ¿Cómo llamar? ¿Qué debe decir? La Clave Azul establece un protocolo para actuar en caso de que las personas presenten descompensaciones con riesgo vital o signos tempranos de deterioro clínico (ver detalle del proceso en [Figura Anexa B.2](#)).

2. Hipótesis y Objetivos

2.1. Hipótesis

La armonización de datos biomédicos del HCUCH según el estándar OMOP CDM permitirá desarrollar un modelo predictivo de riesgo de mortalidad para pacientes con alerta Clave Azul. Se espera que el modelo, basado en datos estructurados y estandarizados, identifique patrones y factores de riesgo asociados con el deterioro clínico, contribuyendo a comprender la dinámica de estos pacientes. La calidad y coherencia de los datos transformados bajo el modelo OMOP CDM serán fundamentales para el éxito del modelo predictivo y el avance en la investigación clínica.

2.2. Objetivo General

Desarrollar un modelo predictivo para pacientes hospitalizados en el HCUCH utilizando el estándar OMOP CDM en una base de datos clínicos creada a partir de los datos del hospital. El enfoque se centrará en aplicar el marco de trabajo OHDSI para analizar datos clínicos y asegurar la calidad de los datos con el fin de obtener resultados confiables en el desarrollo de modelos predictivos.

2.3. Objetivos Específicos

1. Transformar los datos del procedimiento Clave Azul al formato OMOP CDM, asegurando una adecuada estructuración y coherencia de los datos.
2. Caracterizar y evaluar la calidad de la base de datos, considerando aspectos como la Conformancia, Completitud y Plausibilidad.
3. Realizar un caso de uso de análisis de datos estandarizado utilizando las herramientas proporcionadas por la comunidad OHDSI, con el fin de realizar un modelo predictivo de mortalidad utilizando la base de datos en formato OMOP CDM.

3. Metodología

3.1. Diseño del Estudio

El presente estudio se trató de una investigación observacional retrospectiva en la que se analizó una muestra de pacientes hospitalizados en el HCUCH entre los años 2016 y 2020 que presentaron una alerta Clave Azul. Esta alerta fue un procedimiento de evaluación y tratamiento inicial de pacientes hospitalizados que presentan signos tempranos de deterioro, ejecutado por un *Equipo de Respuesta Rápida* compuesto por distintos profesionales de la salud, con el objetivo de evitar un mal desenlace. Se buscó evaluar la factibilidad de utilizar el modelo de datos común OMOP CDM para la creación de un caso de uso de análisis de datos de Predicción a Nivel de Pacientes, aplicando un algoritmo de clasificación para predecir el riesgo de mortalidad intrahospitalaria de pacientes con alerta Clave Azul. Este caso de uso fue sugerido por el equipo de Medicina Interna, quienes mostraron un gran interés en el desarrollo de herramientas que permitan mejorar la atención y los resultados en el cuidado de sus pacientes.

3.2. Fuente de Datos

La población de estudio se obtuvo a partir de una planilla proporcionada por el *Servicio de Medicina Interna* del HCUCH en enero de 2022. Esta planilla contenía información de pacientes hospitalizados que tuvieron al menos un procedimiento Clave Azul entre los años 2016 y 2020. Utilizando los identificadores de pacientes, vinculados al *RUN*² y almacenados en la planilla, se realizaron consultas *SQL* para extraer los identificadores únicos correspondientes a cada paciente de la base de datos del *Registro Clínico Electrónico (RCE)*.

Los criterios de inclusión fueron pacientes hospitalizados en el HCUCH con al menos un procedimiento Clave Azul entre 2016 y 2020 que estuvieran registrados en la planilla proporcionada. Aquellos que no cumplieran con este criterio o cuya información no se encontraba en la base de datos del RCE fueron excluidos.

² El Rol Único Nacional (RUN), es el número identificador único e irrepetible que posee todo chileno, residente o no en Chile, y todo extranjero que permanezca, temporal o definitivamente, con una visa distinta a la visa de turista en dicho país. Fuente: Wikipedia.

Para asegurar la correcta interpretación de los datos y facilitar la comunicación con el *Departamento de Informática* del hospital, se elaboró un *Diccionario de Datos*³ detallando el significado de cada variable recopilada. Este diccionario ([Tabla Anexa A.1](#)) se utilizó durante todo el proceso de análisis, garantizando el acceso para todos los miembros del equipo.

La categoría Clave Azul, junto con las variables relacionadas a este procedimiento, fue creada de manera colaborativa utilizando una planilla de cálculo en las dependencias del Servicio de Medicina Interna. Se examinaron las evoluciones de cada paciente en el RCE en busca de la presencia de términos asociados al evento, como “clave azul”, “clave”, “alerta”, “señal” y “código azul”, para determinar la fecha de inicio y fin del evento. Además, se identificaron fechas de traslados a otras unidades, estado de compensación y fecha de fallecimiento. Este proceso permitió tener acceso a la información Clave Azul que no se encuentra disponible de manera confiable en otras secciones del modelo de datos del RCE.

Con el fin de proteger la privacidad de los pacientes, se implementó un procedimiento específico para su de-identificación, el cual se detalla en el apartado [3.8. Consideraciones de Privacidad y Anonimización de Datos](#).

3.3. Datos Clínicos

Se recopilaron 31 variables de referencia para cada paciente, que se agruparon en seis categorías: *Pacientes*, *Hospitalizaciones*, *Laboratorio*, *Signos Vitales*, *Clave Azul*, y *Defunciones*. En la fase de recepción de datos para el estudio, se utilizó un modelo de datos temporal para estructurar los datos en una base de datos temporal previo a la transformación a OMOP. Dentro de las categorías de Laboratorio y Signos Vitales, se seleccionaron algunas variables y pruebas específicas, conocidas por su relevancia clínica en el campo del deterioro clínico y sugeridas por el equipo de Medicina Interna.

La categoría de Laboratorio incluyó las siguientes variables: *Baciliformes*, *Bilirrubina Directa*, *Bilirrubina Total*, *Nitrógeno Ureico*, *Creatinina*, *Exceso Base*, *Fosfatasas Alcalinas*, *GGT*, *GOT*, *GPT*, *Bicarbonato*, *Hemoglobina*, *Protrombina INR*, *Potasio*, *Lactato*, *Lactato Deshidrogenasa*, *Leucocitos*, *PaCO₂*, *PaO₂*, *Proteína C Reactiva*, *pH* y *Plaquetas*.

³ El Diccionario de Datos es un equivalente al conjunto de datos mínimo, y se refiere a la cantidad mínima de información requerida para llevar a cabo este trabajo.

Por otro lado, la categoría de Signos Vitales incluyó las siguientes variables: *Presión Arterial Sistólica, Presión Arterial Diastólica, Presión Arterial Media, Frecuencia Cardíaca, Saturación de Oxígeno y Temperatura Corporal*.

Dentro de la categoría Clave Azul, se incluyeron las siguientes variables específicas relacionadas con el evento: *Identificación del paciente, Identificación de la hospitalización, Fecha y hora del evento, Resultado, Indicador de traslado a otra unidad e Indicador de fallecimiento*.

3.4. Infraestructura Tecnológica para la Gestión de Datos

La arquitectura se basó en una máquina virtual con sistema operativo *GNU/Linux Debian* versión 4.19.171-2, con 188 GB de espacio en disco y 8 GB de memoria RAM, conectada a 50 TB de almacenamiento compartido a una velocidad de conexión de 10 Gb/s gestionada por la *Unidad de Datos del Centro de Informática Médica y Telemedicina de la Universidad de Chile*. Se utilizó el gestor de bases de datos *PostgreSQL* versión 11.10 y el sistema de virtualización *Docker Compose* versión 1.23.1. Para acceder a la base de datos se utilizó *pgAdmin* 5.4 y se implementaron las aplicaciones ATLAS, y DQD mediante contenedores en *Docker* (Figura Anexa B.1). En cuanto a la implementación del modelo de datos OMOP CDM, se siguió el procedimiento descrito en [34], que incluyó la creación de usuarios, asignación de permisos, creación de tablas OMOP CDM y la importación de vocabularios desde *ATHENA*⁴.

3.5. Transformación de los Datos al Estándar OMOP CDM

El proceso de transformación de la fuente de datos del RCE al estándar OMOP CDM se basó en un flujo bien definido de 8 pasos [35], que se describe a continuación (Figura 2):

1. *Extracción de datos*: Se extrajeron los datos definidos en la sección [3.2. Fuente de datos](#).
2. *Perfilado de datos*: Se realizó un análisis detallado de los datos utilizando la herramienta *WhiteRabbit*, que proporcionó, a través de un reporte, información sobre las tablas, campos y distribuciones de frecuencia de valores contenidos en la fuente de datos.

⁴ <https://athena.ohdsi.org>

3. *Mapeo sintáctico*: Utilizando la herramienta *Rabbit-in-a-Hat*, se creó el diseño del proceso de ETL. Esta herramienta permitió documentar y visualizar de manera clara el proceso de mapeo sintáctico entre las tablas de la fuente de datos y OMOP. Para validar el mapeo sintáctico, se realizó una verificación mediante el cálculo de la proporción de registros migrados correctamente desde la fuente de datos a la base de datos OMOP CDM. Esto implicó comparar el conteo de registros de origen con la cantidad de registros correspondientes en la base de datos OMOP CDM. Esta validación aseguró que el mapeo sintáctico se llevó a cabo de manera precisa y que los datos se transformaron correctamente en el proceso ETL.
4. *Mapeo semántico*: En esta etapa, se realizó el mapeo semántico de los conceptos encontrados en la fuente de datos utilizando vocabularios terminológicos específicos. Se empleó LOINC para Signos Vitales y Laboratorio, *Unified Code for Units of Measure (UCUM)* para Unidades, y SNOMED-CT, junto con el vocabulario propio de OMOP, para los demás conceptos relacionados con el contexto de la Clave Azul. Este proceso se llevó a cabo manualmente para garantizar la calidad de la información y asegurar la correcta asociación de los conceptos. La evaluación del mapeo semántico se llevó a cabo mediante el cálculo del porcentaje de conceptos mapeados exitosamente desde la fuente de datos a la base de datos OMOP CDM.
5. *Implementación del ETL*: Se realizó la implementación del proceso ETL utilizando consultas SQL y el lenguaje de programación R. Estas consultas se basaron en el diseño previamente realizado con *Rabbit-in-a-Hat* y permitieron transformar los datos extraídos al formato del OMOP CDM.
6. *Caracterización y evaluación de la calidad de datos*: Se realizó una evaluación de la calidad de los datos transformados al OMOP CDM. En esta etapa, se utilizó la herramienta DQD para evaluar aspectos como la Conformancia, Completitud y Plausibilidad de los datos. El detalle de la metodología utilizada en esta etapa se abordó en la siguiente sección del artículo, donde se realizaron análisis más detallados para evaluar la calidad de los datos mapeados, incluyendo la validación de los mapeos sintáctico y semántico.
7. *Verificación de la calidad de datos*: Si los datos no pasaron la evaluación de calidad, se volvió al paso 5 para corregir y mejorar el proceso de transformación. Si los datos cumplieron con los estándares de calidad establecidos, se procedió al siguiente paso.

8. *Generación de la base de datos OMOP CDM*: Una vez que los datos pasaron la evaluación de calidad, se generó la base de datos OMOP CDM, que sirvió como el repositorio final de los datos transformados.

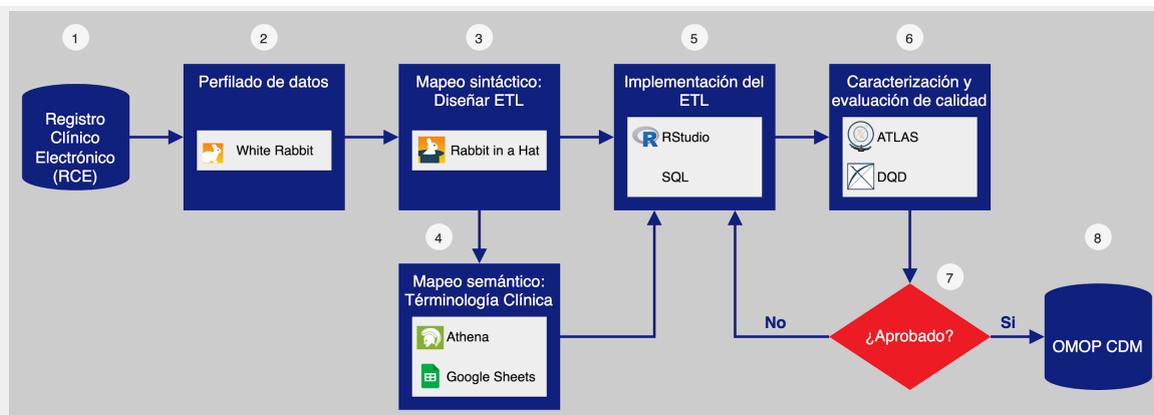


Figura 2. Flujo del proceso de ocho pasos para el mapeo de la fuente de datos del Registro Clínico Electrónico (RCE) al OMOP CDM usando las herramientas OHDSI. Modificado de [35].

3.6. Caracterización de la Base de Datos y Evaluación de la Calidad de Datos

En esta fase, que correspondió a la etapa 6 del flujo mencionado en la [3.5. Transformación de los Datos al Estándar OMOP CDM](#), se caracterizó la base de datos y se evaluó la calidad de los datos transformados. Estas actividades fueron cruciales para comprender la estructura y la calidad de la base de datos OMOP CDM.

Se utilizó la herramienta ATLAS para explorar y caracterizar la base de datos OMOP. ATLAS proporcionó una interfaz para acceder a informes descriptivos y estandarizados para la fuente de datos configurada en este trabajo.

La evaluación de la calidad de los datos transformados a OMOP fue realizada mediante la herramienta DQD, teniendo en cuenta las dimensiones de Conformancia, Completitud y Plausibilidad propuestas en [22]. El paquete de DQD ejecutó sistemáticamente 2334 comprobaciones aproximadamente y evaluó los resultados contra un umbral por defecto en una instancia de OMOP CDM, en el caso de este trabajo, la versión v5.4. Posteriormente se creó un objeto *JSON* que se leyó en una aplicación R *Shiny*⁵ para visualizar los resultados.

⁵ <https://shiny.posit.co/>

3.7. Caso de Uso Análisis de Datos

Para la implementación del caso de uso de análisis de datos, se utilizó el marco conceptual y las herramientas proporcionadas por la comunidad OHDSI, específicamente a través de la elaboración de un *Prediction Study*⁶, propuesto por [35] para estandarizar la implementación del proceso de análisis y asegurar la validación adecuada de los modelos predictivos resultantes. La **Figura 3** muestra el marco de trabajo propuesto, el cual se utilizó para guiar la definición de la población de estudio, selección de variables predictoras, construcción y validación del modelo.



Figura 3. La metodología se divide en dos fases: la primera fase en color azul llamada *Data Harmonization and Quality Control*, que implica la transformación semántica y sintáctica a OMOP CDM. La segunda fase en color rojo llamada *Stages of the study*, que consiste en cuatro etapas correspondientes a (a) protocolo del estudio, (b) configurar el modelo, ejecutarlo y validarlo, (c) validación externa con datos de otra institución, y (d) ciencia abierta y compartición de la evidencia.

3.7.1. Elaboración del Protocolo de Estudio

Esta fase, que marcó el inicio de la metodología, constó de cuatro pasos que definieron de la siguiente manera:

1. *Especificación del problema predictivo*: Se estudió el siguiente problema predictivo (**Tabla 1**): “Entre los pacientes que se observó la ocurrencia de un procedimiento Clave Azul, qué pacientes desarrollarán resultado de fallecimiento por descompensación o fallecimiento por causa desconocida durante el intervalo de tiempo de 0 a 365 días posterior al inicio del procedimiento Clave Azul”.
2. *Generación y evaluación de fenotipos*: En la **Tabla 2** se definieron los fenotipos utilizados para identificar a los “pacientes hospitalizados con clave azul activada al menos una vez” y “fallecimiento”.
3. *Evaluación de la idoneidad de las fuentes de datos*: La base de datos seleccionada contenía la información necesaria sobre el fenotipo de interés (Clave Azul) y cumplía

⁶ Un *Prediction Study* (Estudio Predictivo) es una investigación que se basa en el análisis de datos para predecir resultados o eventos futuros.

con los criterios de idoneidad establecidos, evitando así la necesidad de usar la herramienta *CohortDiagnostics* para evaluar múltiples fuentes de datos.

4. *Definición de la configuración para el desarrollo del modelo:* Se utilizó el paquete *FeatureExtraction* de R para generar características o covariables para pacientes que cumplieron con los parámetros de inclusión de la [Tabla 2](#). Se aplicó un conjunto predefinido de covariables que incluyen género, edad y una variable binaria de edad agrupada en intervalos de 5 años (15-19, 20-24, 25-29, ..., 90-94). Además, se incluyeron los valores de los signos vitales y exámenes de laboratorio registrados entre 5 y 10 días previos a la *fecha índice*⁷ como medidas. Estas covariables predefinidas fueron utilizadas como entrada para construir el modelo predictivo del evento de interés. Se realizó una partición (aleatoria y estratificada según los resultados de los datos) en la que se utilizó el 75% de los datos para entrenamiento y el 25% restante para prueba. Luego, se aplicó un proceso de validación cruzada con 3 iteraciones en el conjunto de entrenamiento para seleccionar la configuración óptima de los hiperparámetros en el modelo predictivo supervisado. Para esta prueba de concepto se entrenó un modelo GBM por medio del paquete PLP. El detalle sobre la configuración del modelo se puede ver en las especificaciones del estudio predictivo ([Tabla 3](#)).

Tabla 1. Especificación del problema predictivo.

Component	Specification
Target	Patients hospitalized with Code Blue activated at least once
Outcome	Death
Time-at-risk	0 days to 365 days after activating the Code Blue

Tabla 2. Definición del fenotipo.

Component	Phenotype definition
Target	Procedure occurrences of "Signposting to rapid response team"
Outcome	Death of "Decompensation" or "Unknown problem"

⁷ Se refiere a la fecha específica en la que se produce un evento clínico de interés o el inicio de una observación en un estudio observacional. En el contexto de este trabajo es la activación de la Clave Azul.

Tabla 3. Especificaciones del estudio predictivo.

Definition	Value
Problem Definition	
Target Cohort (T)	A procedure occurrences of Signposting to rapid response team
Outcome Cohort (O)	A death of Unknown problem or death of Decompensation
Time-at-risk (TAR)	0 day till 365 days from cohort start
Population Definition	
Washout Period	0
Enter the target cohort multiple times?	No
Allow prior outcomes?	No
Start of time-at-risk	0 days
End of time-at-risk	365 days
Require a minimum amount of time-at-risk?	No
Model Development	
Algorithm	Gradient Boosting Machine
Hyper-parameters	ntree:5000, max depth:4 or 7 or 10 and learning rate: 0.001 or 0.01 or 0.1 or 0.9
Covariates	Gender, Age, Age Group, Measurement Value (<5, <10)
Data split	75% train, 25% test. Randomly assigned by person

3.7.2. Elaboración y Validación Interna del Modelo

Para el desarrollo y validación interna del modelo, se utilizó el paquete PLP. Se empleó la herramienta ATLAS para el desarrollo de las cohortes *Target* y *Outcome* descrita en la [Tabla 2](#) definiendo los fenotipos mediante la opción *Cohort Definitions* ([Figura 4](#)). La herramienta PLP incluyó la generación de una aplicación R Shiny que visualizó de manera interactiva los resultados de la validación interna del modelo. Se evaluó el rendimiento del modelo mediante métricas como *Sensibilidad*, *Valor Predictivo Positivo* y *AUC*. Se utilizaron los tutoriales oficiales de PLP como referencia durante el proceso [36] .

3.7.3. Validación Externa

Dado que los datos utilizados para desarrollar el modelo provenían de una sola institución, no se llevó a cabo una validación externa. Debido a esta limitación, el modelo fue evaluado y validado internamente utilizando las técnicas y métricas mencionadas anteriormente para asegurar su rendimiento dentro de los datos del hospital universitario.

3.7.4. Ciencia Abierta e Intercambio de Evidencia

Todo el código utilizado en este trabajo fue puesto a disposición en el repositorio público de GitHub: <https://github.com/iamalonso/hcuch-omop>.



Figura 4. Captura de pantalla de la sección Cohort Definitions en la herramienta Atlas de análisis de datos. Navegación: Atlas > Cohort Definitions > Cohort > Export.

3.8. Privacidad y Anonimización de Datos

Antes de procesar los datos para su transformación a OMOP, se realizó la de-identificación de todos los pacientes de la cohorte, con la aprobación del Comité de Ética según el Certificado N° 1179/21 emitido el 16 de marzo de 2021 por la dirección general del HCUCH. Este proceso implicó modificar o eliminar los datos personales que podrían identificar a los individuos en la base de datos, con el objetivo de proteger su privacidad. Se siguió el método utilizado por Lima et al. (2019) para esta de-identificación, donde la Información Protegida de Salud (*PHI, Protected Health Information*) de los pacientes se redujo al mínimo necesario antes de cargarla en la base de datos estandarizada. Los identificadores RUT y Código de Hospitalización se seudonimizaron utilizando una arquitectura que permitía la re-identificación en caso de ser necesario por el equipo clínico (Figura 5). Estos identificadores se reemplazaron por un nuevo identificador (seudónimo) generado mediante la función *RANDOM* de PostgreSQL, dentro de un rango compatible con los conceptos estandarizados OMOP. Los seudónimos se almacenaron en una tabla específica dentro de la base de datos OMOP, que contenía tres campos: el nombre de la tabla a la que pertenecía el seudónimo, el valor original y el valor seudonimizado.

Además, la fecha de nacimiento se anonimizó mediante el proceso de generalización, que implicaba reemplazar la fecha original por una representación menos específica. En este caso, se eliminó el día y el mes, conservando únicamente el año.

Las fechas de los eventos restantes, como la adquisición de signos vitales, la toma de pruebas de laboratorio, la alerta de la Clave Azul, entre otros, no se anonimizaron en esta tesis debido a su complejidad y la necesidad de llevar a cabo tareas adicionales. Las técnicas de anonimización de fechas requerían una evaluación detallada, lo que las hacía inviables en este estudio. Por lo tanto, se decidió centrarse en la anonimización de otros datos relevantes y dejar la aplicación de técnicas más sofisticadas para futuras investigaciones.

A

patients			hospitalizations				
patient_id	birthdate	gender	patient_id	hospitalization_id	admission_date	discharge_date	discharge_code
12345678	1960-07-10	1	12345678-9	987654321	2020-07-10	2020-07-15	1

B

keys		
table	source_id	new_id
patients	12345678	68626129835711
hospitalizations	987654321	152924209735366

C

patients			hospitalizations				
patient_id	birthdate	gender	patient_id	hospitalization_id	admission_date	discharge_date	discharge_code
68626129835711	1960	1	68626129835711	152924209735366	2020-07-10	2020-07-15	1

Campos PHI
 Campos PHI anonimizados o seudonimizados

Figura 5. Almacenamiento, de-identificación y recuperación del identificador del paciente. A. Datos originales de los pacientes provenientes de la fuente de datos (amarillo). B. Tabla de búsqueda de seudónimos que contiene el identificador de la fuente de datos (source_id) y el identificador seudonimizado (new_id) generado aleatoriamente, (rojo). Esto permite la vinculación de información y la recuperación de los valores reales del RUT o Código de Hospitalización a través del nuevo identificador. C. Datos seudonimizados y anonimizados. Se muestra cómo se lee la información seudonimizada de las tablas procesadas utilizando este método. PHI: Protected Health Information.

4. Resultados

4.1. Extracción de los Datos desde la Fuente y Creación de la Base de Datos Clínica

4.1.1. Extracción Exitosa de los Datos Clave Azul para su Análisis

Se logró extraer con éxito los datos clínicos relevantes para el seguimiento del procedimiento Clave Azul de los pacientes. Estos datos fueron obtenidos desde las fuentes del hospital y sometidos a un proceso de transformación y carga en una plataforma de gestión de datos diseñada específicamente para este propósito. Se implementó un *Data Lake* como un repositorio centralizado para facilitar la carga y el procesamiento eficiente de los datos. Además, se utilizó un área de carga temporal conocida como *Staging*, donde los datos se transformaron siguiendo la estructura del Diccionario de Datos ([Tabla Anexa A.1](#)). Este enfoque permitió acelerar los procesos críticos, como la de-identificación, el perfilamiento y el mapeo de los datos al estándar OMOP CDM.

4.1.2. Perfilamiento de los Datos para la Base de Datos Clínica

Se generó correctamente el perfilamiento de la fuente de datos usando WhiteRabbit. Esta herramienta entregó un informe en formato Excel llamado *ScanReport* ([Figura 6](#)). El informe de escaneo generado proporcionó una visión detallada de los campos y tablas presentes en el área de *Staging*. Se identificaron un total de 31 campos distribuidos en 6 tablas, mostrando consistencia con la estructura planificada del Diccionario de Datos. La tabla *VITAL_SIGNS* (67,142 filas) fue la más extensa, seguida por *LABORATORY* (40,219) y *HOSPITALIZATIONS* (416).

Se seleccionaron las hospitalizaciones en las que se activó una Clave Azul, lo que resultó en 143 casos. Se recopiló información sobre laboratorio y signos vitales asociados a estas hospitalizaciones. Para las hospitalizaciones restantes, se retuvo únicamente la información del paciente y la fecha del evento, permitiendo determinar la duración completa de la observación del paciente.

En términos de integridad de los datos, se encontró que la mayoría de los campos tenían una baja fracción de valores vacíos, siendo los campos *vital_sign_value* (2.2%

fracción vacía) y *discharge_code* (1.7%), *discharge_date* (1.2%) y *test_result* (0.2%) los únicos con valores vacíos en un pequeño porcentaje de registros.

Además, se observó que el campo *vital_sign_id* tenía un porcentaje del 100% de valores vacíos, lo que indica que no se registró ningún identificador de signo vital para esos casos.

Table	Field	Descr	Type	Max length	N rows	N rows ch	Fraction empty	N unique	Fraction unique
laboratory	hospitalization_id		character varying	15	40219	40219	0.0%	141	0.4%
laboratory	patient_id		character varying	15	40219	40219	0.0%	141	0.4%
laboratory	service_code		character varying	8	40219	40219	0.0%	26	0.1%
laboratory	sample		character varying	16	40219	40219	0.0%	7	0.0%
laboratory	test_time		timestamp without time zone	19	40219	40219	0.0%	5097	12.7%
laboratory	test		text	14	40219	40219	0.0%	22	0.1%
laboratory	test_result		numeric	7	40219	40219	0.2%	5433	13.5%
patients	patient_id		character varying	15	143	143	0.0%	143	100.0%
patients	birthdate		numeric	4	143	143	0.0%	60	42.0%
patients	gender		smallint	1	143	143	0.0%	2	1.4%
code_blue	hospitalization_id		character varying	15	143	143	0.0%	143	100.0%
code_blue	patient_id		character varying	15	143	143	0.0%	143	100.0%
code_blue	code_blue_date		date	10	143	143	0.0%	131	91.6%
code_blue	code_blue_time		time without time zone	8	143	143	0.0%	106	74.1%
code_blue	outcome		smallint	1	143	143	0.0%	3	2.1%
code_blue	transferred		smallint	1	143	143	0.0%	2	1.4%
code_blue	dead		smallint	1	143	143	0.0%	2	1.4%
deaths	hospitalization_id		character varying	15	54	54	0.0%	54	100.0%
deaths	patient_id		character varying	15	54	54	0.0%	54	100.0%
deaths	date_of_death		date	10	54	54	0.0%	52	96.3%

Figura 6. Captura de pantalla de la pestaña Field Overview en el reporte de perfilamiento de datos generado por la herramienta WhiteRabbit en formato de planilla de cálculo.

4.1.3. Mapeo Sintáctico de los Datos Clave Azul al Formato OMOP CDM

En la fase de mapeo sintáctico, se procedió al diseño del ETL basado en la solución diseñada para modelar el caso de uso de pacientes clave en OMOP. Durante este proceso, se utilizaron varias tablas y se establecieron relaciones entre ellas, como la tabla *PERSON*, *OBSERVATION_PERIOD*, *VISIT_OCCURRENCE*, *VISIT_DETAIL*, *MEASUREMENT_PROCEDURE_OCCURRENCE* y *DEATH*. Estas tablas fueron seleccionadas de manera específica para capturar la información relevante del proceso Clave Azul del hospital y asegurar una representación adecuada en el modelo OMOP CDM. Cada tabla desempeña un papel en el almacenamiento de datos relacionados con los pacientes, las visitas al

hospital, los detalles de cada visita, los procedimientos realizados, las mediciones clínicas y los desenlaces registrados. El diseño del ETL se centró en asegurar que los datos de origen se transformen y carguen correctamente en las tablas correspondientes del modelo OMOP CDM, siguiendo las convenciones y estructuras definidas. Se empleó un diagrama *Business Process Model and Notation (BPMN)* (Figura Anexa B.2) para describir de forma concisa el procedimiento Clave Azul del hospital. Este diagrama facilitó la comprensión del proceso y la implementación de la solución en el contexto de OMOP. Durante esta etapa, se identificó que 18 campos (~58% del total) son necesarios para poblar 35 campos (~33%) en el modelo OMOP CDM (Tabla Anexa A.2). Un diagrama con los resultados del mapeo entre las tablas de la fuente de datos y OMOP CDM describe en la Figura 7.

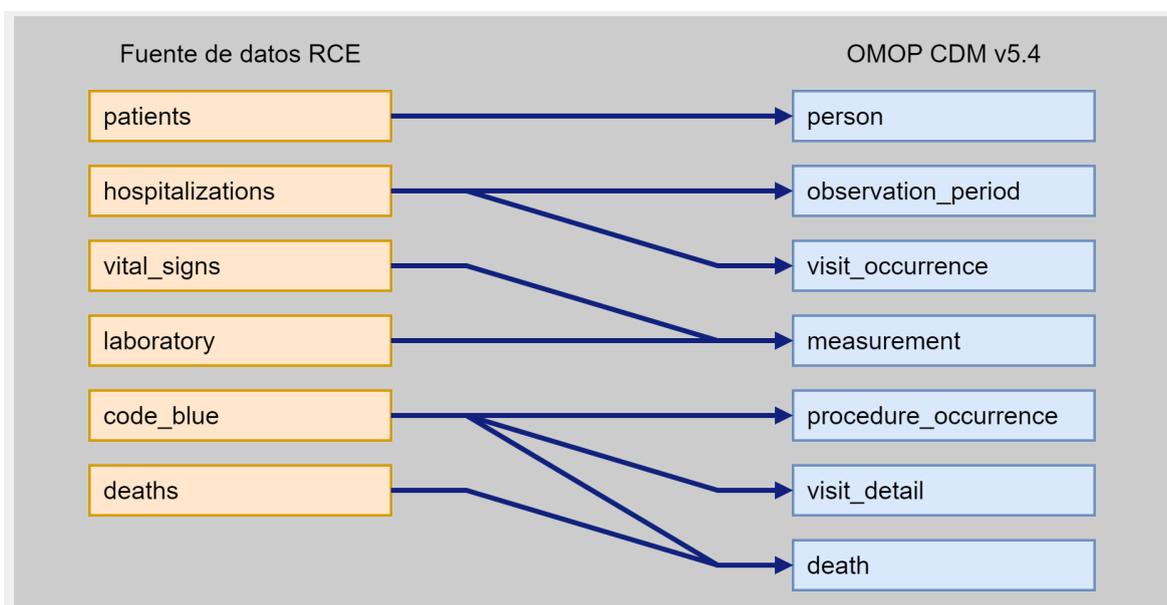


Figura 7. Mapeo sintáctico entre las tablas en la fuente de datos y OMOP CDM. Muestra la correspondencia de las 6 tablas de origen a 7 tablas en el dominio OMOP. Figura adaptada de la documentación generada por la herramienta Rabbit-in-a-Hat.

Se logró exitosamente la transformación de un total de 108,179 registros seudonimizados y anonimizados de 143 pacientes hospitalizados desde el RCE del HCUCH al formato OMOP CDM. En promedio, se logró migrar el 100% de los registros, a excepción de aquellos mapeados a las tablas *VISIT_OCCURRENCE* (98.8%) y *MEASUREMENT* (99.9%) (Tabla 4). Esto se debió a la presencia de algunos campos vacíos en las fechas de egreso de hospitalizaciones y en los resultados de exámenes de Laboratorio, los cuales fueron descartados durante el proceso ETL. Los datos transformados a OMOP presentaron un alto nivel de precisión, ya que no se encontraron diferencias con respecto a las proporciones y estadísticas calculadas sobre las características clínicas y demográficas de la fuente de datos (Tabla Anexa A.3).

Tabla 4. Comparación de registros migrados exitosamente desde la fuente de datos Clave Azul del RCE a las tablas OMOP CDM.

Fuente de datos RCE		OMOP CDM		
Nombre tabla	Registros (N=108,117)	Nombre tabla	Registros (N=108,179)	Proporción migrada
patients	143	person	143	100%
hospitalizations	416	visit_occurrence	411	98.8%
vital_signs	67,142	observation_period	143	100%
laboratory	40,219	measurement	107,240	99.89%
code_blue	143	visit_detail	45	100%
deaths	54	procedure_occurrence	143	100%
—	—	death	54	100%

RCE: Registro Clínico Electrónico.
OMOP CDM: Observational Medical Outcomes Partnership Common Data Model.

4.1.4. Mapeo Semántico de los Datos para su Interpretación en el Contexto OMOP CDM

En el mapeo semántico, se utilizaron conceptos específicos para modelar la Clave Azul en OMOP. Por ejemplo, cada episodio de hospitalización se representó utilizando el concepto *Inpatient Visit* (Visita de Paciente Hospitalizado) del Dominio *Visit* en el vocabulario OMOP, asociado al campo *visit_concept_id* en la tabla *VISIT_OCCURRENCE*. Esto permitió capturar y registrar de manera precisa la información sobre los ingresos de los pacientes al hospital. Además, para representar los procedimientos Clave Azul, se utilizó el concepto *Signposting to rapid response team (procedure)* (Señalización al Equipo de Respuesta Rápida) del vocabulario SNOMED-CT. Estos procedimientos fueron registrados en la tabla *PROCEDURE_OCCURRENCE*, estableciendo su relación mediante los campos *visit_occurrence_id* y *visit_detail_id*. Esta elección permite capturar de manera precisa la ocurrencia de la alerta Clave Azul en el estudio. Para aquellos pacientes cuyo desenlace implicó un traslado a la *Unidad de Paciente Crítico (UPC)*, se utilizó el concepto *Intensive Care* (Cuidados Intensivos) del Dominio *Visit* en el vocabulario OMOP, asociado al campo *discharged_to_concept_id* en la tabla *VISIT_DETAIL*. También se registró cada fallecimiento ocurrido en la tabla *DEATH*, asociando al campo *death_type_concept_id* el concepto *Decompensation* o *Unknown problem* (Descompensación o Problema Desconocido) del dominio *Condition* en SNOMED-CT. La categoría *Decompensation* se utilizó para aquellos pacientes cuyo fallecimiento fue identificado a través de la revisión de las evoluciones del registro clínico, indicando que ocurrió durante el proceso de la Clave Azul. Por otro lado, la categoría *Unknown problem* se asignó a los pacientes que fallecieron después de la ocurrencia de la Clave Azul y donde no se pudo determinar de manera definitiva si la causa directa de su fallecimiento fue el procedimiento, ya que

existían otros factores involucrados. Asimismo, se mapearon diversos parámetros Clave Azul, como los resultados en las categorías Laboratorio y Signos Vitales. Por ejemplo, el nivel de *Hemoglobina* se mapeó utilizando el concepto *Hemoglobin [Mass/volume] in Blood*, la *Presión Arterial Sistólica* utilizando *Systolic blood pressure*, la *Frecuencia Cardíaca* utilizando *Heart rate* y *Temperatura corporal* utilizando *Body temperature*. Estos conceptos se utilizaron junto con la tabla de mapeos para garantizar una representación precisa de los datos en el modelo OMOP CDM.

En el proceso de mapeo semántico, se logró exitosamente mapear el 100% de los 46 conceptos identificados (Tabla 5). Los datos del RCE, que no estaban codificados según estándares terminológicos, fueron mapeados manualmente con ayuda de expertos en Medicina Interna del hospital hacia los estándares LOINC, UCUM, SNOMED-CT y OMOP, garantizando así una representación precisa de los exámenes de Laboratorio, Signos Vitales y otros conceptos relacionados con el procedimiento Clave Azul (Tabla Anexa A.4).

Tabla 5. Comparación del mapeo semántico de los conceptos Clave Azul del RCE hacia conceptos OMOP CDM: Proporciones de mapeo exitoso.

Fuente de datos RCE		OMOP CDM		
Categoría	Nombre de tabla	Conceptos encontrados ¹	Conceptos encontrados	Conceptos mapeados exitosamente
Demográfico	patients	2	2	100%
Clave Azul	code_blue	5	5	100%
Laboratorio	laboratory	22	22	100%
Signos Vitales	vital_signs	6	6	100%
Unidades	vital_signs, laboratory	11	11	100%

OMOP: Observational Medical Outcomes Partnership
RCE: Registro Clínico Electrónico
¹ Mapeados manualmente usando el repositorio OHDSI de vocabularios terminológicos ATHENA.

4.1.5. Implementación Exitosa del Proceso de Extracción, Transformación y Carga en los Datos Clave Azul

Tras una exitosa implementación del proceso ETL, se logró una integración eficiente de los datos en el modelo OMOP CDM. Esta implementación se basó en los recursos y herramientas proporcionados por la comunidad OHDSI a través de su repositorio de GitHub. Como resultado, se poblaron las tablas de los dominios *Standardized clinical data*, *Standardized vocabularies* y *Standardized metadata* (Figura 8).

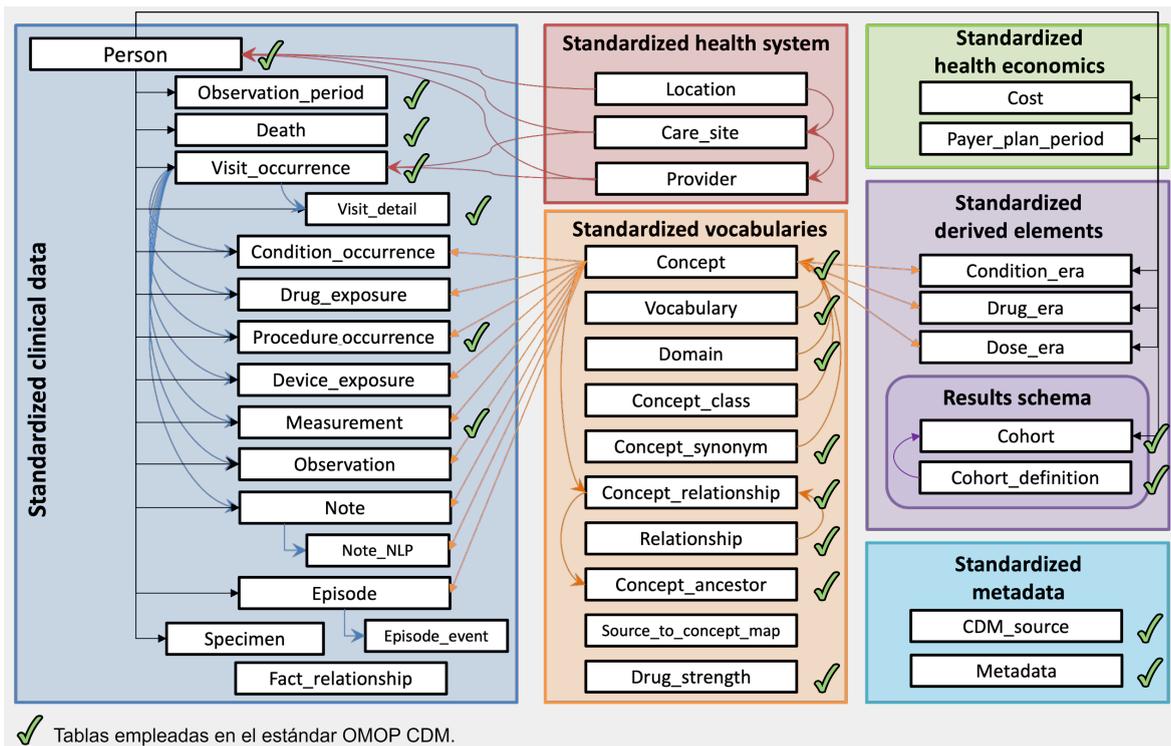


Figura 8. Diagrama del modelo relacional OMOP CDM v5.4. Se presentan los dominios Standardized clinical data, Standardized health system, Standardized vocabularies, Standardized derived elements y Standardized metadata, junto con sus respectivas tablas. Además, se ilustran en gris las tablas que no fueron pobladas en el contexto de este trabajo.

Para el dominio Standardized clinical data, se desarrollaron 7 archivos SQL para cada tabla OMOP utilizada en el mapeo sintáctico (Figura 7), permitiendo la transformación y carga de datos desde R. La ejecución exitosa de estos archivos, como parte del diseño del ETL realizado con la herramienta Rabbit-in-a-Hat, brindó resultados significativos en el mapeo semántico y sintáctico, los cuales se discutieron previamente en las secciones [4.1.3. Mapeo Sintáctico de los Datos Clave Azul al Formato OMOP CDM](#) y [4.1.4. Mapeo Semántico de los Datos para su Interpretación en el Contexto OMOP CDM](#).

En el dominio Standardized vocabularies, se poblaron 8 tablas con las terminologías, vocabularios, clasificaciones y otros datos necesarios para el estándar OMOP CDM. Estos datos fueron obtenidos desde ATHENA y se incorporaron a las tablas correspondientes mediante un proceso de carga utilizando un procedimiento específico que se encuentra detallado en el repositorio de GitHub del presente trabajo.

Se incluyeron varios vocabularios, como SNOMED, ICD-9 (que incluye ICD-9-CM e ICD-9-Proc), CPT-4, LOINC, Gender, Race, ICD-10 (junto con ICD-10-PCS e ICD-10-CM), DRG, Ethnicity, OMOP Extension, entre otros más. Estos vocabularios se utilizaron en su versión V5.0 04-FEB-22.

La implementación del modelo de datos se realizó creando un esquema en el motor PostgreSQL, siguiendo la definición de la [Figura Anexa B.1](#).

4.2. Caracterización y Evaluación de la Calidad de Datos en la base de datos OMOP: Un Análisis Preliminar

4.2.1. Caracterización Preliminar de la Base de Datos OMOP

Utilizando la herramienta ATLAS, se realizó un análisis detallado de la base de datos transformada a OMOP para obtener información relevante sobre los pacientes y los datos. Los informes y visualizaciones generados revelaron detalles significativos sobre la base de datos, permitiendo su caracterización. En particular, se utilizó el reporte *Dashboard* para obtener información clave. Por ejemplo, desde el reporte *Dashboard*, se identificaron 143 pacientes, con una distribución de género de 75 (52%) masculino y 68 (48%) femenino ([Figura Anexa B.3](#)).

El reporte *Measurement* desglosa los conceptos LOINC relacionados con Laboratorio y Signos Vitales, revelando hallazgos relevantes:

1. *Hemoglobina (Hemoglobin [Mass/volume] in Blood, Concept Id 3000963)*: Este concepto tiene la mayor prevalencia, alcanzando un 97.2%. Se registró en 139 personas, con un promedio de 18.8 registros por persona.
2. *Plaquetas (Platelets [# /volume] in Blood by Automated count, Concept Id 3024929)*: Este concepto presenta una prevalencia del 96.5%. Se registró en 138 personas, con un promedio de 17.9 registros por persona.
3. *Leucocitos (Leukocytes [# /volume] in Blood by Automated count, Concept Id 3000905)*: Este concepto también tiene una prevalencia del 96.5%. Se registró en 138 personas, con un promedio de 16.5 registros por persona.

Además, se identificaron tres conceptos con menor prevalencia en la base de datos:

1. *Bicarbonato (Bicarbonate [Moles/volume] in Arterial blood, Concept Id 3008152)*: Este concepto tiene una prevalencia del 21.7%. Se registró en 31 personas, con un promedio de 6 registros por persona.

2. *Temperatura Corporal (Body temperature, Concept Id 3020891)*: Este concepto presenta una prevalencia del 63.6%. Se registró en 91 personas, con un promedio de 74.7 registros por persona.
3. *Presión Arterial Diastólica (Diastolic blood pressure, Concept Id 3012888)*: Este concepto también tiene una prevalencia del 64.3%. Se registró en 92 personas, con un promedio de 131.4 registros por persona.

Se destaca que el concepto *Hemoglobin [Mass/volume] in Blood* tiene la mayor prevalencia, alcanzando un 97.2%. Este concepto se registró en 139 personas, con un promedio de 18.8 registros por persona. Por otro lado, el concepto *Bicarbonate [Moles/volume] in Arterial blood* presenta la menor prevalencia, con un 21.7%. Se registró en 31 personas, con un promedio de 6 registros por persona. Al analizar las prevalencias de los Signos Vitales en la base de datos OMOP, se evidencia una prevalencia promedio del 64.2% (desviación estándar aproximada de 0.003). En relación a los exámenes de Laboratorio, se observa una prevalencia promedio del 80.1% (desviación estándar aproximada de 0.17). Estos datos resaltan las diferencias en la frecuencia de aparición de estos elementos en la base de datos ([Figura Anexa B.4](#)).

Asimismo, al analizar el reporte *Visit* de la base de datos, se encontró que todas las visitas registradas tuvieron una prevalencia del 100%, con un promedio de 2.9 registros por persona. Esto indica que cada persona en la base de datos tuvo en promedio 2.9 visitas registradas ([Figura Anexa B.5](#)). Por su parte, al examinar el reporte *Procedure*, se identificó que los procedimientos Clave Azul también presentaron una prevalencia del 100%, con un único registro por persona. Esto indica que cada persona en la base de datos tuvo un único procedimiento Clave Azul ([Figura Anexa B.6](#)). En cuanto a la edad al momento del fallecimiento, según el reporte *Death*, se determinó una mediana de 70 años para el género femenino y 73 años para el género masculino ([Figura Anexa B.7](#)).

4.2.2. Una Evaluación Armonizada de la Calidad de Datos: Un Enfoque Basado en el Marco de Trabajo de Kahn et al.

La [Figura 9](#) presenta la evaluación obtenida de la calidad de datos en las dimensiones Conformancia y Completitud. En la dimensión de Plausibilidad, se realizaron un total de 1889 chequeos, de los cuales 5 resultaron en errores, lo que representa un cumplimiento del 100%. En la dimensión de Conformancia, se llevaron a cabo 202 chequeos, de los

cuales 22 presentaron fallas, con un cumplimiento del 89%. En la dimensión de Completitud, se efectuaron 123 chequeos, de los cuales 18 arrojaron errores, con un cumplimiento del 85%. En general, la base de datos transformada a OMOP ha obtenido un alto porcentaje de aprobación, con un cumplimiento total del 98% en un total de 2334 chequeos. La visualización ha sido generada utilizando la herramienta DQD, proporcionando una visión precisa del estado de la calidad de los datos.

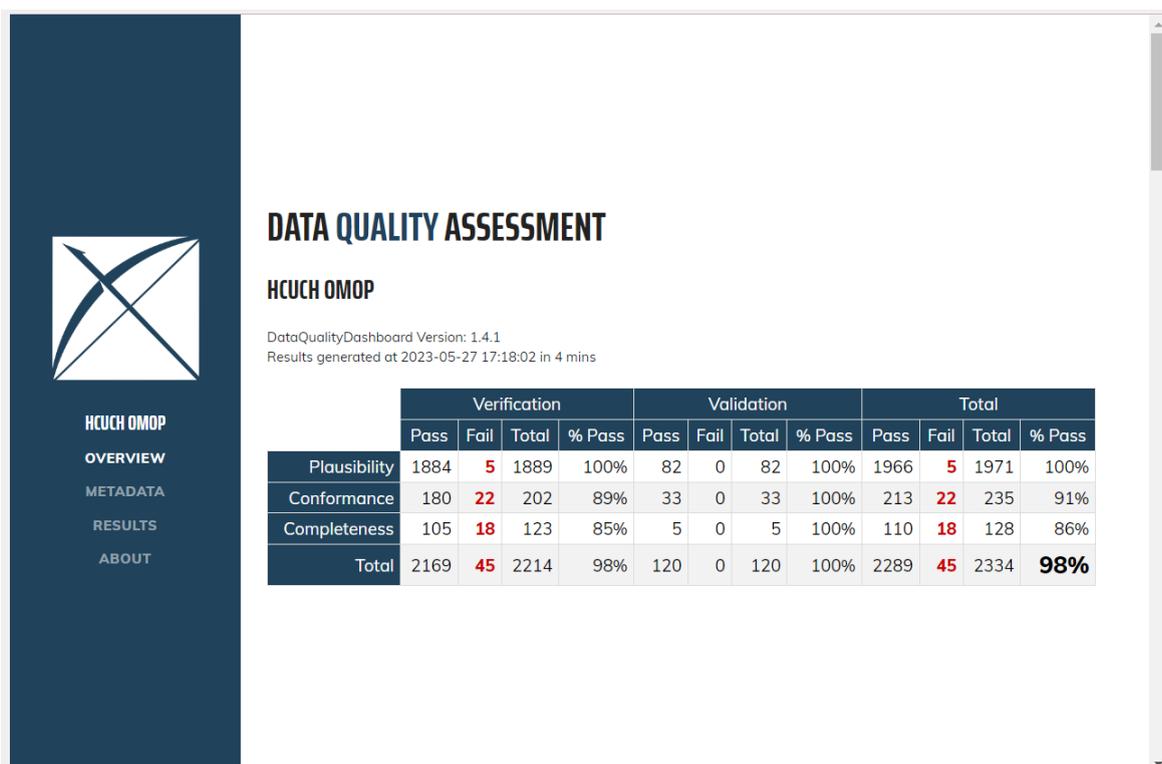


Figura 9. Evaluación detallada de la calidad de datos en las dimensiones Conformance, Completitud y Plausibilidad. Visualización obtenida desde la herramienta OHDSI DataQualityDashboard (DQD) v1.4.1.

Al analizar en detalle los errores, se comenzó por revisar aquellos que se presentaron en la dimensión de Plausibilidad, la cual se refiere a la credibilidad o veracidad de los valores de los datos. Durante la verificación, se identificaron 5 chequeos fallidos en esta dimensión. Estos errores se encontraron en la tabla *MEASUREMENT* de OMOP, donde se almacenaron los datos de Laboratorio y Signos Vitales. A continuación, se presentan los 5 fallos encontrados:

1. Para el concepto *Leucocitos (Leukocytes [#Volume] in Blood by Automated Count, Concept Id 3000905)*, se encontraron registros con un valor >11 (umbral=5%). Este error afectó al 99.9% de los registros correspondientes a este concepto.

2. En el caso del concepto *Proteína C Reactiva (C Reactive Protein [Mass/Volume] in Serum or Plasma, Concept Id 3020460)*, se registraron valores >24 (umbral=5%). Esto afectó al 81.4% de los registros relacionados con este concepto.
3. Asimismo, para el concepto *GGT (Gamma Glutamyl Transferase [Enzymatic Activity/Volume] in Serum or Plasma, Concept Id 3026910)*, se encontraron registros con valores >80 (umbral=5%). Este error afectó al 73.8% de los registros correspondientes.
4. En el caso de las *Plaquetas (Platelets [# /Volume] in Blood by Automated Count, Concept Id 3024929)*, se encontraron registros con valores <100,000 (umbral=5%). Este error afectó al 25.9% de los registros relacionados con este concepto.
5. Finalmente, para el concepto *Bilirubina Directa (Bilirubin.Direct [Mass/Volume] in Serum or Plasma, Concept Id 3027597)*, se registraron valores >4 (umbral=5%). Esto afectó al 13.4% de los registros correspondientes a este concepto.

Es importante destacar que estos chequeos fallidos en la dimensión Plausibilidad fueron admitidos luego de una revisión junto a expertos del área clínica que consideraron como posibles estos valores fuera de rango dado el contexto clínico de la cohorte analizada. Para visualizar un ejemplo de estos errores en la herramienta, se adjunta una captura de pantalla ([Figura Anexa B.8](#)) que muestra la sección *results* del informe de DQD, donde se detallan los errores encontrados en la dimensión de Plausibilidad.

En las dimensiones de Conformancia y Completitud, que se centran en el formato, estructura, relaciones y presencia de atributos de datos, se han identificado errores comunes sobre los datos almacenados en las tablas *VISIT_DETAIL* y *VISIT_OCCURRENCE*. Estos errores están relacionados con cambios de nombre en algunos campos y han afectado tanto la dimensión de Conformancia como la de Completitud. Si un campo no existe, su valor asociado tampoco estará presente, lo que ha generado errores adicionales en la dimensión de Completitud. El problema principal se debe a un cambio en la versión del CDM utilizado por la herramienta DQD para realizar los chequeos. Por ejemplo, en la versión 5.4 del CDM, el campo *admitting_source_concept_id* de la tabla *VISIT_OCCURRENCE*, que estaba presente en la versión 5.3, ha sido modificado y ahora se llama *admitted_from_concept_id*. Esta modificación del nombre del campo ha generado incongruencias con la configuración esperada por la herramienta DQD, lo cual resulta en errores en la dimensión de

Conformancia. Estos errores se propagan a la dimensión de Completitud, ya que se intenta verificar la presencia de un valor no nulo en un campo que no existe en la versión actual del CDM.

En resumen, el cambio en la versión del CDM ha provocado que la herramienta DQD no encuentre el campo esperado, lo que genera errores tanto en la dimensión de Conformancia como en la de Completitud. En la dimensión de Conformancia se encontraron 22 fallos de un total de 235 verificaciones realizadas. Esto representa un porcentaje de aprobación del 91%. En cuanto a la dimensión de Completitud, se identificaron 18 fallos de un total de 128 verificaciones realizadas. Esto representa un porcentaje de aprobación del 86%.

4.3. Implementación Estándar y Accesible del Caso de Uso de Análisis de Datos en la Base de Datos OMOP

4.3.1. Elaboración del Protocolo de Estudio

Las especificaciones del estudio definidas en la [Tabla 3](#) se implementaron correctamente en R. El archivo incluye el código necesario para instalar los paquetes requeridos, cargar los paquetes necesarios, configurar los ajustes de covariables, establecer la conexión a la base de datos y obtener las covariables. Incluye la configuración de la población de estudio, con un tamaño de muestra de 143 pacientes, ya que no se especificaron criterios adicionales de inclusión para la población objetivo, el entrenamiento del modelo de clasificación utilizando una implementación de GBM en R llamada *XGBoost*⁸, la ejecución del análisis PLP y la generación de visualizaciones de los resultados. Con este archivo de código R, se garantiza la correcta ejecución de todos los pasos necesarios para llevar a cabo el estudio de forma consistente y reproducible.

4.3.2. Elaboración y Validación Interna del Modelo

Se implementó la aplicación R Shiny provista por el paquete PLP, una plataforma web interactiva que permitió la visualización y exploración sencilla de los resultados obtenidos del modelo.

⁸ <https://xgboost.readthedocs.io/en/stable/>

En el estudio, se generó una tabla que muestra la importancia de las covariables en las predicciones del modelo. Entre las covariables más influyentes, destacan los siguientes predictores: *Potassium [Mass/volume] in Blood* (10 días antes, valor de importancia: 16.7%), *C reactive protein [Mass/volume] in Serum or Plasma* (5 días antes, valor de importancia: 12.5%) y *Oxygen saturation in Blood* (10 días antes, valor de importancia: 9.8%). Estos valores de importancia indican la relevancia de estas covariables en la predicción del resultado. Es importante tener en cuenta que estos intervalos de días (10 días antes, 5 días antes) fueron configurados previamente en el modelo y forman parte de la metodología utilizada en el estudio. Esta información se extrajo desde la sección *Model Table* de la pestaña *Model* del visualizador (Figura 10).

The screenshot shows a web interface for a 'Model Table'. At the top, there are navigation buttons for 'Previous', '1', and 'Next'. Below this is a 'Model Table' section with a 'Download Model' button and a 'Show 10 entries' dropdown. A search bar is also present. The main table lists 10 covariates with columns for 'Covariate Name', 'Value', 'Count', 'Outcome Mean', 'Non-outcome Mean', and 'Std Mean Diff'. The table is paginated to show 10 of 37 entries.

Covariate Name	Value	Count	Outcome Mean	Non-outcome Mean	Std Mean Diff
1 gender = FEMALE	0.0059	68	0.4074	0.5169	-0.2209
2 measurement value during day -5 through 0 days relative to index: Mean blood pressure (millimeter mercury column)	0.001	87	39.3519	44.0674	-0.1263
3 measurement value during day -5 through 0 days relative to index: Body temperature (degree Celsius)	0.0044	85	20.6667	21.9438	-0.072
4 measurement value during day -5 through 0 days relative to index: Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma (unit per liter)	0.008	109	133.7963	95.7	0.3467
5 measurement value during day -5 through 0 days relative to index: Lactate dehydrogenase [Enzymatic activity/volume] in Serum or Plasma (unit per liter)	0.0198	106	726.9444	237.2584	0.6412
6 measurement value during day -5 through 0 days relative to index: C reactive protein [Mass/volume] in Serum or Plasma (milligram per liter)	0.1246	100	96.6285	47.2309	0.5391
7 measurement value during day -5 through 0 days relative to index: Oxygen saturation in Blood (percent)	0.0015	88	53.3148	56.7528	-0.0775
8 measurement value during day -5 through 0 days relative to index: Urea nitrogen [Mass/volume] in Venous blood (milligram per deciliter)	0.0031	119	35.0741	20.1348	0.7058
9 measurement value during day -5 through 0 days relative to index: Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma (unit per liter)	0.0127	70	105.037	21.6629	0.4948
10 measurement value during day -5 through 0 days relative to index: Gamma glutamyl transferase [Enzymatic activity/volume] in Serum or Plasma (unit per liter)	0.0006	70	80.313	65.1146	0.1265

Figura 10. Captura de pantalla de la pestaña de la sección Model Table de la pestaña Model donde se encuentran los valores de coeficiente para cada covariable incluida en el modelo final.

En la pestaña *Threshold Dependant* (Figura 11), se evaluaron las métricas utilizando un umbral de 0.4, utilizado para clasificar las predicciones del modelo como positivas o negativas, y una tasa de incidencia del 37.8% para el resultado de fallecimiento posterior a un procedimiento Clave Azul. Los resultados mostraron un Valor Predictivo Positivo (*PPV*, *Positive Predictive Value*) fue de 61.1%, lo que significa que el 61.1% de las predicciones positivas fueron correctas. La Especificidad (*Specificity*) se situó en un 68%, indicando la proporción de casos negativos correctamente identificados. La Sensibilidad (*Sensitivity*) alcanzó un 84.6%, representando la capacidad del modelo para detectar correctamente los

casos positivos. Por último, el Valor Predictivo Negativo (*NPV, Negative Predictive Value*) fue de 88.2%, lo que indica la proporción de casos negativos correctamente identificados.

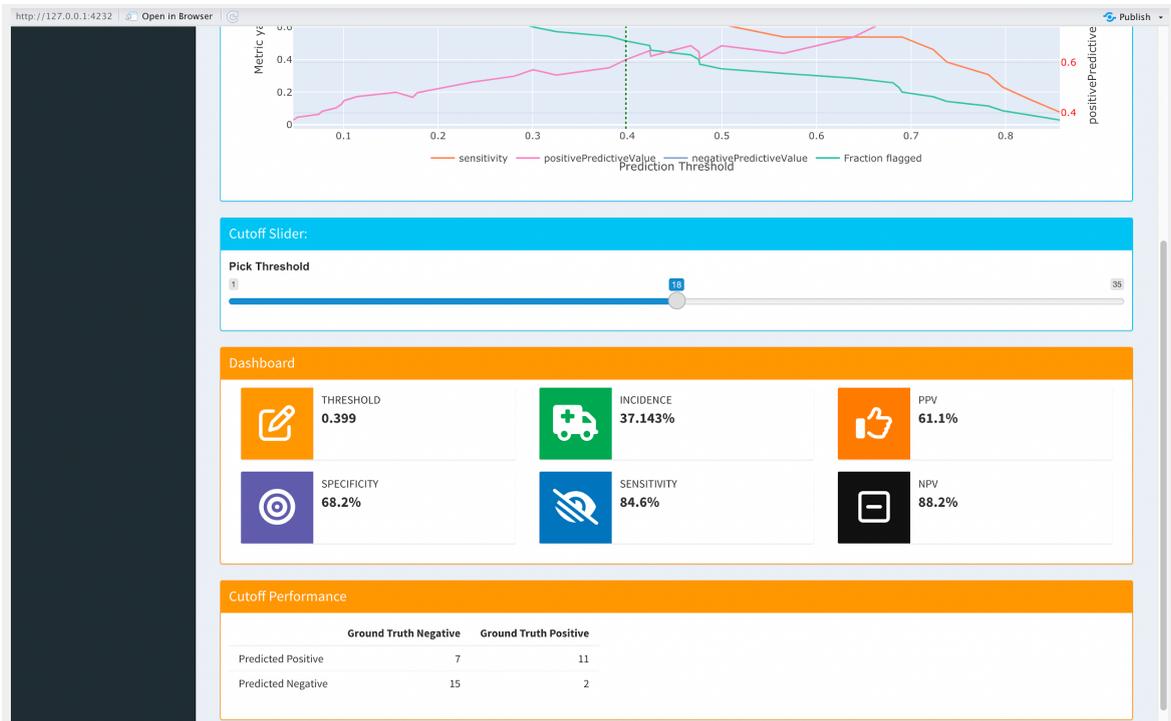


Figura 11. Captura de pantalla de la pestaña Threshold Dependant, donde se encuentran las métricas de PPV, Especificidad, Sensibilidad y NPV. Esta herramienta incluye un control deslizante de umbral que se utiliza para clasificar los resultados del modelo. El ajuste del umbral tiene un impacto significativo en el rendimiento general del modelo, incluyendo el PPV, NPV, Sensibilidad y otras métricas. PPV: Positive Predictive Value. NPV: Negative Predictive Value.

Desde la pestaña *Discrimination* (Figura 12), se evaluó el rendimiento del modelo en la predicción del evento de fallecimiento. Se obtuvo un AUC de 0.7 (*IC* 95%: 0.59-0.8), lo cual indica la capacidad del modelo para clasificar correctamente y distinguir entre los casos de fallecimiento y no fallecimiento. Además, se obtuvo un AUPRC de 0.6, que evalúa la precisión del modelo en la clasificación de las predicciones positivas de fallecimiento.



Figura 12. Captura de pantalla de la pestaña Discrimination, que muestra las métricas de rendimiento del modelo, así como las curvas ROC (Receiver Operating Characteristic) y la curva PR (Precision-Recall).

En la calibración del modelo, se evaluó la precisión de las predicciones en la pestaña *Calibration*. Se encontró que el modelo tiene un $E90$ de 0.1, lo que significa que el 90% de los puntos en la curva de calibración tienen una diferencia absoluta entre la predicción y el riesgo observado menor o igual a 0.1. Además, se obtuvo un E_{max} de 0.1, que representa la mayor diferencia absoluta entre la predicción y el riesgo observado en la curva de calibración. Este valor de 0.1 indica que existe al menos un punto en la curva donde la diferencia entre la predicción y el riesgo observado alcanza ese valor. El gráfico *Calibration Plot* de esta sección muestra esta relación (Figura 13). Estas métricas proporcionan información sobre la precisión y la confiabilidad del modelo en la calibración de las predicciones con los riesgos observados.

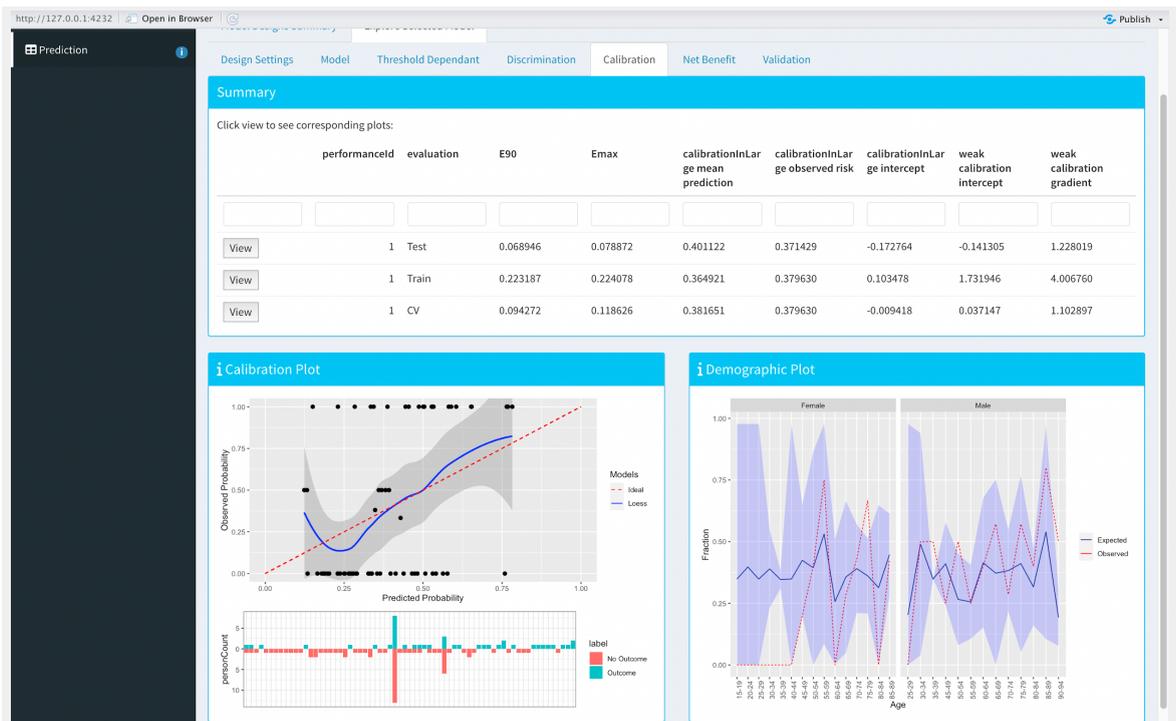


Figura 13. Captura de pantalla de la pestaña Calibration, con métricas y gráficos de calibración del modelo. El gráfico Calibration Plot muestra la relación entre las predicciones del modelo y los riesgos observados. En este gráfico, cada punto representa un grupo de pacientes clasificados en diferentes rangos de riesgo predicho, y la línea diagonal ideal representa una calibración perfecta.

5. Discusión

5.1. Extracción, Transformación y Carga de Datos Clínicos Relevantes: Transformación Exitosa de la Fuente de Datos al Estándar OMOP CDM

5.1.1. Experiencia en la Extracción y Transformación de Datos Clínicos: Utilizando el Modelo OMOP CDM

La extracción y transformación de datos clínicos relevantes para el procedimiento Clave Azul de los pacientes es exitosa en general. Los datos se obtienen de las fuentes del hospital, incluyendo registros digitales y la revisión de fichas del RCE por parte de los profesionales clínicos. El uso del Diccionario de Datos como herramienta de intercambio simplifica la extracción y transformación de los datos al formato estandarizado y final.

Sin embargo, se identifica un punto débil en el proceso: no se llena sistemáticamente el registro de eventos del procedimiento Clave Azul por parte de los profesionales clínicos del hospital. Esto podría afectar la integridad y calidad de los datos disponibles para su transformación.

La infraestructura requerida para llevar a cabo la extracción y carga de los datos en el modelo OMOP CDM puede variar según la complejidad y el volumen de los datos clínicos. En proyectos con conjuntos de datos más pequeños y menos complejos, la implementación puede realizarse de manera efectiva utilizando una infraestructura más sencilla. Sin embargo, a medida que el tamaño y la complejidad de los datos aumentan, puede ser necesario contar con una infraestructura más robusta. La elección de la infraestructura adecuada debe basarse en un análisis de los requisitos del proyecto y en la evaluación de los recursos disponibles.

5.1.2. Perfilamiento de Datos con WhiteRabbit: Identificación de Campos y Consistencia con el Diccionario de Datos Planificado

El perfilamiento de datos realizado con WhiteRabbit revela 31 campos distribuidos en 6 tablas, lo que se alinea con el Diccionario de Datos planificado. La mayoría de los campos tienen pocos valores vacíos, lo que indica consistencia en los datos extraídos. Sin embargo,

se encuentran valores faltantes en diferentes dominios, posiblemente debido a errores de extracción o falta de almacenamiento en los sistemas de información.

Los datos observacionales reflejan el curso normal del cuidado de la salud, por lo que la ausencia de un valor en un campo no significa necesariamente que la información esté perdida, sino que podría indicar la falta de un evento registrado. El uso de herramientas como WhiteRabbit estandariza la presentación de parámetros de datos y proporciona información inicial para comprender mejor los datos que serán transformados al modelo OMOP CDM.

El análisis se basa en los datos extraídos y su correspondencia con el Diccionario de Datos planificado, lo que introduce sesgos o errores en la interpretación de los resultados. Además, la presencia de valores faltantes en algunos dominios afecta la integridad y calidad de los datos disponibles para su transformación.

En conclusión, el perfilamiento de datos mediante WhiteRabbit permite identificar los campos y su consistencia con el Diccionario de Datos planificado. Aunque se encontraron valores faltantes en algunos dominios, es importante abordar estos valores faltantes para mejorar la calidad de los datos y minimizar la pérdida de información. Se recomienda complementar las herramientas de perfilamiento de datos, como WhiteRabbit, con otras que permitan la visualización de los datos para mejorar la comprensión y la discusión en torno a los datos que serán transformados. También se sugiere realizar un análisis más detallado de las causas de los valores faltantes en los dominios identificados, con el objetivo de mejorar la calidad de los datos y minimizar la pérdida de información.

5.1.3. Mapeo Sintáctico a OMOP CDM: Logros y Desafíos en el Análisis del Procedimiento Clave Azul

El mapeo sintáctico de las tablas de la fuente de datos al estándar OMOP CDM se realizó utilizando la herramienta Rabbit-In-a-Hat. Se establecieron relaciones entre varias tablas para garantizar una representación adecuada en el modelo de datos estandarizado.

Se seleccionaron 143 hospitalizaciones en las que se activó una Clave Azul, de un total de 416 hospitalizaciones registradas. Los datos de estos 143 pacientes hospitalizados se transformaron exitosamente desde el RCE del hospital universitario al formato OMOP CDM, con un porcentaje promedio de migración de registros del 99.8%.

Sin embargo, se encontraron algunos campos vacíos en las fechas de egreso de las hospitalizaciones y en los resultados de los exámenes de laboratorio correspondientes a las tablas VISIT_OCCURRENCE y MEASUREMENT del modelo de datos común.

La herramienta Rabbit-In-a-Hat demostró ser altamente satisfactoria en el proceso de transformación debido a su capacidad de visualización gráfica. Además, el uso de recursos estandarizados facilitó una integración eficiente de los datos.

Se recomienda ampliar la base de datos con un mayor número de pacientes para evaluar las métricas del mapeo sintáctico y abordar posibles complejidades que podrían surgir con un conjunto de datos más diverso. Esta ampliación permitiría obtener una evidencia más sólida y confiable en el contexto de este caso de uso. En trabajos anteriores, donde se han transformado bases de datos al estándar OMOP CDM, se han abordado casos con una gran cantidad de pacientes, superando en algunos casos los 20 millones, y con una cantidad significativa de registros, llegando hasta 1,099,195,384 registros clínicos. Estos trabajos previos [23–33, 37–39] han proporcionado una base robusta para el análisis e interpretación de datos clínicos a gran escala.

En conclusión, se logró un exitoso mapeo sintáctico de las tablas de la fuente de datos al estándar OMOP CDM, con un alto porcentaje de migración de registros a nivel de paciente. Sin embargo, se recomienda considerar las limitaciones relacionadas con el tamaño de la base de datos y continuar explorando métricas y ampliaciones en futuros trabajos.

5.1.4. Mapeo Semántico en OMOP CDM: Uso de Vocabularios Estandarizados y Colaboración con Expertos Clínicos

El mapeo semántico de los conceptos relacionados con la Clave Azul se realizó utilizando terminología estándar, superando las limitaciones de la falta de codificación estandarizada en la fuente de datos. Se mapean exitosamente todos los conceptos identificados en las cinco categorías, donde la colaboración con expertos clínicos fue esencial en el mapeo de todas estas categorías.

Durante el mapeo, se destaca la importancia de realizar supuestos para abordar la falta de codificación estandarizada en la fuente de datos. Por ejemplo, en el caso de la Clave Azul, se seleccionó el concepto *Signposting to rapid response team* como una opción

cercana. Además, se establecieron supuestos en relación al fallecimiento de los pacientes, considerando los casos sin mención explícita como *Unknown problem*.

Es importante destacar que durante este estudio no se encontró literatura previa específica que abordará el mapeo semántico de los conceptos relacionados con la Clave Azul usando el estándar OMOP CDM. En cuanto al mapeo de datos en Laboratorio y Signos Vitales, se utiliza el estándar terminológico LOINC, el cual demuestra ser una herramienta intuitiva y efectiva para encontrar y mapear estos parámetros. Asimismo, se logra una representación adecuada de las unidades de medida gracias a la correspondencia entre los códigos LOINC y las unidades estándar utilizadas en la práctica clínica. Se recomienda mejorar el registro del procedimiento Clave Azul, dado que no se llena sistemáticamente por parte de los profesionales clínicos.

En conclusión, se logra una representación efectiva del contexto de los datos no codificados mediante la utilización de conceptos estandarizados en diferentes vocabularios.

5.1.5. Integración Eficiente de Datos en OMOP CDM: Implementación del ETL y Desafíos Encontrados

La implementación del ETL permite una integración eficiente de los datos en el modelo OMOP CDM. Los dominios Standardized clinical data, Standardized vocabularies y Standardized metadata se poblaron siguiendo el diseño del ETL realizado con la herramienta Rabbit-in-a-Hat. El alcance limitado del caso de uso en este trabajo implicó que, en esta etapa, sólo se poblaron parcialmente estos dominios.

El uso del lenguaje de consulta SQL para el mapeo de la fuente de datos a las tablas de OMOP CDM resultó relativamente sencillo gracias a la documentación proporcionada por OHDSI, así como la integración de las distintas consultas al flujo de datos con el lenguaje de programación R.

Se ha establecido una base sólida para futuros proyectos mediante una documentación clara y accesible que facilita la implementación del proceso ETL y el mapeo hacia el modelo OMOP CDM en proyectos similares.

Se identificaron algunas limitaciones en la implementación. En primer lugar, no se realizaron pruebas unitarias del código, lo que podría ocasionar omisión de errores y su paso inadvertido en futuras cargas de datos. Se recomienda abordar esta cuestión mediante la integración de pruebas unitarias utilizando paquetes OHDSI [15]. la carga de datos se

realiza manualmente, lo cual sugiere considerar la integración de una herramienta de orquestación para automatizar las cargas con registros de ejecución y permitir la generación de cargas incrementales de datos en trabajos futuros.

Además, es fundamental destacar la importancia del capital humano cualificado y con experiencia técnica para llevar a cabo con éxito estas tareas. La programación, manejo de bases de datos y análisis de datos son componentes críticos en el proceso de ETL. Contar con profesionales expertos en estas áreas asegura la precisión, eficiencia y fiabilidad de este proceso, así como la adecuada interpretación de los resultados obtenidos.

Finalmente, el trabajo colaborativo y la comunicación efectiva entre expertos en informática y profesionales del ámbito clínico son fundamentales para abordar los desafíos y asegurar la correcta transformación de los datos al modelo OMOP CDM.

5.2. Caracterización y Evaluación de la Calidad de la Base de Datos Transformada a OMOP: Utilización de ATLAS y DQD

5.2.1. Caracterización Detallada de la Base de Datos Transformada a OMOP CDM: Análisis y Resultados con ATLAS

El uso de ATLAS facilita un análisis completo de la base de datos OMOP CDM. Revela detalles sobre los pacientes y sus datos clínicos, incluyendo prevalencias de conceptos estandarizados y características demográficas.

Para los conceptos de Laboratorio y Signos Vitales, se observa una prevalencia promedio del 64.2% para Signos Vitales y 80.1% para Laboratorio. El concepto Hemoglobin [Mass/volume] in Blood destaca con una prevalencia del 97.2% en 139 personas, con un promedio de 18.8 registros por persona. Sin embargo, se nota una menor prevalencia en el concepto Bicarbonate [Moles/volume] in Arterial blood.

Es relevante que todos los campos de los signos vitales se completaron en todos los registros, lo que sugiere una práctica consistente por parte de los profesionales de la salud. No obstante, el desequilibrio en la frecuencia de registro podría tener implicaciones clínicas. Una recomendación importante es realizar análisis más detallados de los conceptos con baja prevalencia, como Bicarbonate [Moles/volume] in Arterial blood en Laboratorio o Systolic blood pressure en Signo Vital, para comprender las razones detrás

de su baja frecuencia en la base de datos. Se recomienda investigar su relevancia estadística. Es importante realizar análisis más detallados de conceptos con baja prevalencia y considerar la inclusión de datos clínicos adicionales para una imagen más completa de los pacientes y su atención médica. La caracterización refleja el estado actual de la base de datos, y su evolución con el tiempo debe considerarse.

Si bien ATLAS ofrece información valiosa, se destaca su limitación. Se recomienda el desarrollo de consultas personalizadas para análisis más específicos. En resumen, ATLAS proporciona información valiosa sobre los pacientes y sus datos clínicos.

5.2.2. Evaluación Exitosa de la Calidad de Datos en la Base de Datos OMOP CDM: Resultados y Recomendaciones con DQD

Esta evaluación confirma la idoneidad de la base de datos para el caso de uso. Se obtuvo un alto porcentaje de aprobación del 98% en 2334 chequeos. Se analizaron las dimensiones clave de calidad de datos: Conformancia, Completitud y Plausibilidad.

En la dimensión de Plausibilidad, se encontraron solo 5 errores en la tabla *MEASUREMENT* relacionados con valores fuera del umbral. Se sugiere validar estos valores independientemente, a pesar de que expertos clínicos hayan considerado posible su existencia en el contexto clínico de la cohorte analizada.

En Conformancia y Completitud, se identificaron errores debido a la incompatibilidad entre la versión del modelo de datos y la herramienta DQD utilizada. Los porcentajes de aprobación en estas dimensiones fueron del 91% y 86% respectivamente, lo que sugiere la existencia de oportunidades de mejora en estas dimensiones de los datos al actualizar a la versión más reciente y compatible de la herramienta DQD.

Para mejorar, se propone incluir variables clínicas más detalladas y evaluar su impacto en los porcentajes de aprobación, teniendo en cuenta la limitada cantidad de datos en comparación con estudios previos de mayor envergadura [40]. La herramienta DQD, aunque es compatible solo con OMOP CDM, fue crucial para evaluar la calidad de datos. Se recomienda mantener una iteración continua y colaboración con expertos clínicos para mantener la calidad de datos.

5.3. Análisis de Datos en la Base de Datos Transformada a OMOP: Un Caso de Uso de Pacientes Clave Azul en un Hospital Universitario

5.3.1. Enfoque Metodológico Propuesto por OHDSI para el Análisis de Datos

El objetivo principal de este estudio es demostrar el valor y eficacia del marco de trabajo de OHDSI. El enfoque metodológico se aplica con éxito en la predicción a nivel de pacientes en un hospital universitario. Se muestra que este enfoque puede guiar y respaldar estudios de manera coherente y confiable. Los resultados respaldan su aplicabilidad y relevancia, promoviendo su adopción por otros investigadores y profesionales de la salud.

5.3.2. Contribuciones del Marco Conceptual Propuesto

El uso de la herramienta PLP y el marco conceptual [35] es fundamental. Garantizan la consistencia y calidad del análisis. La documentación de la comunidad OHDSI⁹ y los repositorios en GitHub¹⁰ son valiosos para la implementación. La participación activa en el foro brinda orientación para mantener la integridad y validez de los resultados.

5.3.3. Exploración de Covariables Influyentes en el Modelo

Se utiliza R Shiny para visualizar la importancia de las covariables en las predicciones del modelo. *Potassium [Mass/volume] in Blood (10 días antes)*, *C reactive protein [Mass/volume] in Serum or Plasma (5 días antes)* y *Oxygen saturation in Blood (10 días antes)* son las covariables más influyentes. Estos hallazgos permiten orientar futuras investigaciones y decisiones clínicas.

Es importante destacar que los modelos de predicción no son modelos causales [41]. La intervención basada en estos predictores no garantiza una reducción en los desenlaces de fallecimiento.

⁹ <https://forums.ohdsi.org/>

¹⁰ <https://github.com/OHDSI/>

5.3.4. Evaluación de Métricas de Rendimiento y Capacidad del Modelo

El paquete PLP y R Shiny permiten evaluar métricas de rendimiento. Estas métricas incluyen el Valor Predictivo Positivo, la Especificidad, la Sensibilidad y el Valor Predictivo Negativo. Además, se calcula el área bajo la curva AUC y el área bajo la curva de precisión-recall AUPRC. Los resultados, con un AUC de 0.7 y un AUPRC de 0.6, demuestran que el modelo tiene un rendimiento comparable a otros modelos predictivos comunes. Según Khalid et al. [35], los modelos clínicos más comúnmente utilizados tienen un AUC que varía entre 0.5 y 0.8. En [35] se citan varias fuentes, incluyendo [42–45], respaldando estos resultados. La calibración del modelo también se evalúa para medir la precisión de las predicciones. En resumen, el estudio destaca la importancia de evaluar y comunicar métricas de rendimiento en la predicción de eventos clínicos y respalda la eficacia del enfoque de OHDSI en el análisis de datos biomédicos.

6. Conclusión

Este estudio logra sus objetivos al armonizar y estructurar los datos clínicos del Hospital Clínico Universidad de Chile bajo el estándar OMOP CDM. Se evalúa la calidad de los datos, asegurando su Conformancia, Completitud y Plausibilidad. Se realiza un análisis de Predicción a Nivel de Pacientes para evaluar el riesgo de mortalidad posterior a un procedimiento Clave Azul en pacientes hospitalizados.

La elección del modelo OMOP CDM demuestra ser acertada, y las herramientas como WhiteRabbit, Rabbit-in-a-Hat, ATLAS, DQD y PLP de OHDSI son fundamentales para obtener resultados consistentes y confiables. El mapeo sintáctico permite una migración exitosa de registros a nivel de pacientes, mientras que en el mapeo semántico se representa de manera detallada los conceptos de la Clave Azul gracias a la colaboración con expertos clínicos. La implementación del ETL integra eficientemente los datos en el modelo OMOP CDM, identificando oportunidades de mejora en la automatización de las cargas y pruebas unitarias del código. La evaluación de la calidad de la base de datos transformada utiliza herramientas como ATLAS y DQD, obteniendo una alta aprobación de la calidad de datos. El análisis de datos con OMOP CDM permite desarrollar un modelo predictivo basado en el algoritmo de clasificación GBM, el cual se entrena y valida siguiendo las mejores prácticas de OHDSI. La herramienta PLP garantiza transparencia y replicabilidad en los análisis, fundamentales en la investigación científica.

En resumen, este estudio cumple exitosamente sus objetivos, evidenciando la eficacia del enfoque metodológico basado en el estándar OMOP CDM y las herramientas de la comunidad OHDSI. Los resultados obtenidos respaldan la calidad y precisión de los análisis realizados, proporcionando una sólida base para futuras investigaciones y la toma de decisiones clínicas.

7. Perspectivas Futuras

El presente trabajo representa un primer paso en el análisis de datos clínicos utilizando el marco de trabajo propuesto por OHDSI y el uso de OMOP CDM para la aplicación de modelos de Predicción a Nivel de Pacientes en el contexto del hospital clínico universitario en Chile. Existen diversas oportunidades para continuar mejorando y ampliando este estudio.

Una perspectiva futura relevante es la aplicación del marco de trabajo OHDSI y la estandarización de datos de OMOP CDM en un contexto interinstitucional. Esto permitiría la colaboración entre diferentes centros de atención médica al analizar datos en un formato común, generando modelos predictivos más robustos y generalizables. Así, se mejoraría la predicción y detección temprana de eventos clínicos adversos, potenciando la colaboración científica y mejorando la calidad de la atención médica.

Además, enriquecer la base de datos con más información clínica relevante permitirá abordar de manera más completa el contexto clínico de los pacientes, no solo en el contexto Clave Azul, sino también en todos los pacientes hospitalizados, reduciendo los posibles sesgos encontrados durante la elaboración de este trabajo. Considerar más información clínica, como diagnósticos o antecedentes sociales, brindará una visión más representativa y generalizable de los casos de uso de Predicción a Nivel de Pacientes.

Asimismo, se podría considerar la implementación de estos modelos predictivos en sistemas de soporte a la decisión clínica, lo que es esencial para mejorar la calidad del cuidado brindado a los pacientes y proporcionar a los profesionales de la salud herramientas adicionales para tomar decisiones informadas y personalizadas en la atención a sus pacientes.

8. Bibliografía

1. Weber GM, Mandl KD, Kohane IS (2014) Finding the missing link for big biomedical data. *JAMA* 311:2479–2480
2. Shortliffe EH, Barnett GO (2006) Biomedical Data: Their Acquisition, Storage, and Use. In: Shortliffe EH, Cimino JJ (eds) *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Springer New York, New York, NY, pp 46–79
3. Suter-Crazzolara C (2018) Better Patient Outcomes Through Mining of Biomedical Big Data. *Frontiers in ICT*. <https://doi.org/10.3389/fict.2018.00030>
4. Reinsel D, Gantz J, Rydning J (2018) The Digitization of the World from Edge to Core. IDC
5. i2b2 Common Data Model. <https://community.i2b2.org/wiki/display/BUN>. Accessed 22 Jul 2021
6. Sentinel Common Data Model. <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model>. Accessed 22 Jul 2021
7. (2020) PCORnet Common Data Model v6.0. In: PCORnet. <https://pcornet.org/data/>. Accessed 22 Jul 2021
8. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG (2016) MIMIC-III, a freely accessible critical care database. *Scientific Data* 3:1–9
9. Kahn MG, Batson D, Schilling LM (2012) Data model considerations for clinical effectiveness researchers. *Med Care* 50 Suppl:S60–7
10. Kostka K (2020) The OHDSI Community. In: *The Book of OHDSI*. pp 3–10
11. Who We Are - OHDSI. In: OHDSI. <https://www.ohdsi.org/who-we-are/>. Accessed 21 Jul 2023
12. Blacketer C (2020) The Common Data Model. In: *The Book of OHDSI*. pp 31–53
13. Reich C, Ostropelets A (2020) Standardized Vocabularies. In: *The Book of OHDSI*. pp 55–74
14. Blacketer C, Voss E, Burrows E, Moinat M (2019) OMOP Common Data Model Extract, Transform & Load Tutorial. *Observational Health Data Sciences and Informatics*
15. Blacketer C, Voss E (2020) Extract Transform Load. In: *The Book of OHDSI*. pp 75–98
16. Schuemie M, DeFalco F (2020) OHDSI Analytics Tools. In: *The Book of OHDSI*. pp 107–120
17. Sena A, Prieto-Alhambra D (2020) Characterization. In: *The Book of OHDSI*. pp 173–200
18. Rijnbeek P, Reys J (2020) Patient-Level Prediction. In: *The Book of OHDSI*. pp 239–284
19. Schuemie M, Huser V, Blacketer C (2020) Data Quality. In: *The Book of OHDSI*. pp 291–306
20. Liaw S-T, Guo JGN, Ansari S, et al (2021) Quality assessment of real-world data repositories across the data life cycle: A literature review. *J Am Med Inform Assoc* 28:1591–1599
21. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR (2021) Increasing trust in real-world evidence

- through evaluation of observational data quality. *J Am Med Inform Assoc* 28:2251–2257
22. Kahn MG, Callahan TJ, Barnard J, et al (2016) A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 4:18
 23. Sathappan SMK, Jeon YS, Dang TK, Lim SC, Shao Y-M, Tai ES, Feng M (2021) Transformation of Electronic Health Records and Questionnaire Data to OMOP CDM: A Feasibility Study Using SG_T2DM Dataset. *Appl Clin Inform* 12:757–767
 24. Haberson A, Rinner C, Schöberl A, Gall W (2019) Feasibility of Mapping Austrian Health Claims Data to the OMOP Common Data Model. *J Med Syst* 43:314
 25. Candore G, Hedenmalm K, Slattery J, Cave A, Kurz X, Arlett P (2020) Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children. *Clin Pharmacol Ther* 107:915–925
 26. Maier C, Lang L, Storf H, et al (2018) Towards Implementation of OMOP in a German University Hospital Consortium. *Appl Clin Inform* 9:54–61
 27. Makadia R, Ryan PB (2014) Transforming the Premier Perspective Hospital Database into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. *EGEMS (Wash DC)* 2:1110
 28. Lamer A, Depas N, Doutreligne M, Parrot A, Verloop D, Defebvre M-M, Ficheur G, Chazard E, Beuscart J-B (2020) Transforming French Electronic Health Records into the Observational Medical Outcome Partnership's Common Data Model: A Feasibility Study. *Appl Clin Inform* 11:13–22
 29. Zhou X, Murugesan S, Bhullar H, Liu Q, Cai B, Wentworth C, Bate A (2013) An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. *Drug Saf* 36:119–134
 30. Papez V, Moinat M, Voss EA, et al (2022) Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. *J Am Med Inform Assoc*. <https://doi.org/10.1093/jamia/ocac203>
 31. Tan HX, Teo DCH, Lee D, et al (2022) Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea. *Healthc Inform Res* 28:112–122
 32. Yu Y, Zong N, Wen A, et al (2022) Developing an ETL Tool for Converting the PCORnet CDM into the OMOP CDM to Facilitate the COVID-19 Data Integration. *J Biomed Inform* 104002
 33. Kim J-W, Kim S, Ryu B, Song W, Lee H-Y, Yoo S (2021) Transforming electronic health record polysomnographic data into the Observational Medical Outcome Partnership's Common Data Model: a pilot feasibility study. *Sci Rep* 11:7013
 34. Lima DM, Rodrigues-Jr JF, Traina AJM, Pires FA, Gutierrez MA (2019) Transforming Two Decades of ePR Data to OMOP CDM for Clinical Research. *Stud Health Technol Inform* 264:233–237
 35. Khalid S, Yang C, Blacketer C, et al (2021) A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data.

36. Rijnbeek P, Williams R, Reys J, Ryan P (2019) OHDSI Tutorial: Patient-level predictive modelling in observational healthcare data.
37. Papez V, Moinat M, Payralbe S, Asselbergs FW, Lumbers RT, Hemingway H, Dobson R, Denaxas S (2021) Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure. *JAMIA Open* 4:ooab001
38. Boyce RD, Handler SM, Karp JF, Perera S, Reynolds CF 3rd (2016) Preparing Nursing Home Data from Multiple Sites for Clinical Research - A Case Study Using Observational Health Data Sciences and Informatics. *EGEMS (Wash DC)* 4:1252
39. Lamer A, Abou-Arab O, Bourgeois A, Parrot A, Popoff B, Beuscart J-B, Tavernier B, Moussa MD (2021) Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study. *J Med Internet Res* 23:e29259
40. Blacketer C, Voss EA, DeFalco F, Hughes N, Schuemie MJ, Moinat M, Rijnbeek PR (2021) Using the Data Quality Dashboard to Improve the EHDEN Network. *NATO Adv Sci Inst Ser E Appl Sci* 11:11920
41. Observational Health Data Sciences and Informatics (2020) *The Book of OHDSI*. 1–470
42. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N (2017) Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 12:e0174944
43. Empana JP, Ducimetière P, Arveiler D, et al (2003) Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. *Eur Heart J* 24:1903–1911
44. Schnabel RB, Sullivan LM, Levy D, et al (2009) Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 373:739–745
45. Deng H, Bai Y, Shantsila A, Fauchier L, Potpara TS, Lip GYH (2017) Clinical scores for outcomes of rhythm control or arrhythmia progression in patients with atrial fibrillation: a systematic review. *Clin Res Cardiol* 106:813–823
46. Hospital Clínico Universidad de Chile (2020) Procedimiento Clave Azul pandemia COVID-19.

9. ANEXO A: Tablas de Información

Tabla Anexa A.1. Diccionario de Datos con 31 variables elaborado para ser usado como referencia por el equipo de Grupo Relacionados al Diagnóstico (GRD) y de Tecnologías de la Información del HCUCH en el proceso de extracción de los datos usados en el presente trabajo. Categoría Pacientes, Hospitalizaciones, Laboratorio, Signos Vitales, Clave Azul, y Defunciones.

Nombre campo	Tipo de dato	Descripción
Pacientes		
patient_id	character varying	Número de identificación seudonimizado del paciente
birthdate	numeric	Año de nacimiento del paciente
gender	smallint	Sexo del paciente
Hospitalizaciones		
hospitalization_id	character varying	Número de identificación seudonimizado de la hospitalización
patient_id	character varying	Número de identificación seudonimizado del paciente
admission_date	date	Fecha de ingreso de la hospitalización que estuvo relacionada con el procedimiento código azul
discharge_date	date	Fecha de alta de la hospitalización que estuvo relacionada con el procedimiento código azul
discharge_code	smallint	Código de estado de alta hospitalaria (estabilizado=1, trasladado=2, fallecido=3)
Laboratorio		
hospitalization_id	character varying	Número de identificación seudonimizado de la hospitalización
patient_id	character varying	Número de identificación seudonimizado del paciente
service_code	character varying	Número de identificación de la prueba
sample	character varying	Tipo de muestra biológica
test_time	timestamp without time zone	Fecha de recogida de la muestra
test	text	Nombre de la prueba
test_result	numeric	Resultado de la prueba
Signos Vitales		
hospitalization_id	character varying	Número de identificación seudonimizado de la hospitalización
patient_id	character varying	Número de identificación seudonimizado del paciente
vital_sign_id	text	Número identificador del signo vital
vital_sign_time	timestamp without time zone	Fecha de la obtención del signo vital
vital_sign	text	Nombre del signo vital
vital_sign_value	numeric	Valor del signo vital
Clave Azul		
hospitalization_id	character varying	Número de identificación seudonimizado de la hospitalización
patient_id	character varying	Número de identificación seudonimizado del paciente
code_blue_date	date	Código de fecha de procedimiento clave azul
code_blue_time	time without time zone	Hora en que ocurrió el procedimiento clave azul
outcome	smallint	Bandera que indica el resultado del procedimiento clave azul (1=estabilizado, 2=transferido, 3=muerto)
transferred	smallint	Bandera que indica si el paciente fue trasladado (1=transferido, 0=no transferido)
dead	smallint	Bandera que indica si el paciente fallece durante el procedimiento clave azul (1=muerto, 0=no muerto)
Defunciones		
hospitalization_id	character varying	Número de identificación seudonimizado de la hospitalización
patient_id	character varying	Número de identificación seudonimizado del paciente
date_of_death	date	Fecha de muerte del paciente

Tabla Anexa A.2. Mapeo sintáctico de la fuente de datos a OMOP CDM. En total se mapean 18 campos origen y 35 campos en destino.

Tabla origen	Campo origen	¿Fue mapeado?	Número de mapeos	Campos mapeados
patients	patient_id	✓	1	person.person_source_value
patients	birthdate	✓	1	person.year_of_birth
patients	gender	✓	2	person.gender_concept_id, person.gender_source_value
hospitalizations	patient_id	—	—	—
hospitalizations	hospitalization_id	✓	1	visit_occurrence.visit_source_value
hospitalizations	admission_date	✓	3	observation_period.observation_period_start_date, visit_occurrence.visit_start_date, visit_occurrence.visit_start_datetime
hospitalizations	discharge_date	✓	3	observation_period.observation_period_end_date, visit_occurrence.visit_end_date, visit_occurrence.visit_end_datetime
hospitalizations	discharge_code	—	—	—
vital_signs	hospitalization_id	—	—	—
vital_signs	vital_sign_id	—	—	—
vital_signs	patient_id	—	—	—
vital_signs	vital_sign	✓	3	measurement.measurement_concept_id, measurement.measurement_source_value, measurement.unit_concept_id
vital_signs	vital_sign_time	✓	2	measurement.measurement_date, measurement.measurement_datetime
vital_signs	vital_sign_value	✓	2	measurement.value_as_number, measurement.value_source_value
laboratory	hospitalization_id	—	—	—
laboratory	patient_id	—	—	—
laboratory	test	✓	2	measurement.measurement_concept_id, measurement.unit_concept_id
laboratory	service_code	✓	1	measurement.measurement_source_value
laboratory	test_time	✓	2	measurement.measurement_date, measurement.measurement_datetime
laboratory	test_result	✓	2	measurement.value_as_number, measurement.value_source_value
laboratory	sample	—	—	—
code_blue	hospitalization_id	—	—	—
code_blue	patient_id	—	—	—
code_blue	code_blue_date	✓	4	visit_detail.visit_detail_start_date, visit_detail.visit_detail_start_datetime, procedure_occurrence.procedure_date, procedure_occurrence.procedure_datetime
code_blue	code_blue_time	✓	2	visit_detail.visit_detail_start_datetime, procedure_occurrence.procedure_datetime
code_blue	outcome	✓	2	visit_detail.discharged_to_source_value, death.cause_concept_id
code_blue	transferred	—	—	—
code_blue	dead	✓	1	death.cause_concept_id
deaths	hospitalization_id	—	—	—
deaths	patient_id	—	—	—
deaths	date_of_death	✓	1	death.death_date

Tabla Anexa A.3. Características demográficas y clínicas de la cohorte relacionada a pacientes Clave Azul presentes en la fuente de datos RCE y en la transformación a OMOP CDM.

Característica	Fuente de datos RCE ¹	OMOP CDM ¹
Demográfico		
Edad	70 (54, 81)	70 (54, 81)
Genero		
Femenino	68 (48%)	68 (48%)
Masculino	75 (52%)	75 (52%)
Clave Azul		
Procedimientos Clave Azul	143 (100%)	143 (100%)
Desenlaces		
Muerte por descompensación	32 (36%)	32 (36%)
Muerte por problema desconocido	22 (24%)	22 (24%)
Transferido a UPC	36 (40%)	36 (40%)
Signos Vitales²		
Body temperature	37.50 (36.25, 38.75)	37.50 (36.25, 38.75)
Diastolic blood pressure	75 (48, 102)	75 (48, 102)
Heart rate	93 (62, 124)	93 (62, 124)
Mean blood pressure	90 (59, 120)	90 (59, 120)
Oxygen saturation in Blood	79 (66, 90)	79 (66, 90)
Systolic blood pressure	133 (89, 177)	133 (89, 177)
Laboratorio²		
Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma	132 (63, 242)	132 (63, 242)
Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma	265 (137, 457)	265 (137, 457)
Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma	134 (62, 250)	134 (62, 250)
Band form neutrophils/100 leukocytes in Blood by Automated count	12 (6, 22)	12 (6, 22)
Base excess in Arterial blood by calculation	-2 (-7, 3)	-2 (-7, 3)
Bicarbonate [Moles/volume] in Arterial blood	23.6 (21.2, 26.2)	23.6 (21.2, 26.2)
Bilirubin.direct [Mass/volume] in Serum or Plasma	2 (1, 10)	2 (1, 10)
Bilirubin.total [Mass/volume] in Serum or Plasma	4 (1, 21)	4 (1, 21)
Carbon dioxide [Partial pressure] in Arterial blood	41 (32, 50)	41 (32, 50)
C reactive protein [Mass/volume] in Serum or Plasma	70 (37, 170)	70 (37, 170)
Creatinine [Mass/volume] in Blood	1.89 (1.02, 3.08)	1.89 (1.02, 3.08)
Gamma glutamyl transferase [Enzymatic activity/volume] in Serum or Plasma	321 (129, 598)	321 (129, 598)
Hemoglobin [Mass/volume] in Blood	10.3 (7.4, 13.1)	10.3 (7.4, 13.1)
INR in Platelet poor plasma by Coagulation assay	2.16 (1.45, 3.31)	2.16 (1.45, 3.31)
Lactate dehydrogenase [Enzymatic activity/volume] in Serum or Plasma	462 (266, 738)	462 (266, 738)
Lactate [Moles/volume] in Arterial blood	4.9 (2.5, 8.9)	4.9 (2.5, 8.9)
Leukocytes [# /volume] in Blood by Automated count	9,830 (6,085, 15,125)	9,830 (6,085, 15,125)
Oxygen [Partial pressure] in Arterial blood	82 (61, 106)	82 (61, 106)
pH of Blood	7.31 (7.18, 7.45)	7.31 (7.18, 7.45)
Platelets [# /volume] in Blood by Automated count	274 (138, 414)	273,500 (137,750, 414,250)
Potassium [Mass/volume] in Blood	5.20 (3.80, 6.60)	5.20 (3.80, 6.60)
Urea nitrogen [Mass/volume] in Venous blood	66 (34, 98)	66 (34, 98)
RCE: Registro Clínico Electrónico		
OMOP CDM: Observational Medical Outcomes Partnership Common Data Model		
UPC: Unidad de Paciente Crítico		
¹ Median (IQR); n (%)		
² La descripción de cada característica está de la manera definida por el vocabulario estándar OMOP.		

Tabla Anexa A.4. Mapeo semántico de los parámetros Clave Azul a los conceptos estándar OMOP CDM. En total se identificaron 46 conceptos no estándar en las categorías: Demográfico (n=2), Clave Azul (n=5), Laboratorio (n=22), Signos Vitales (n=6) y Unidades (n=11). La cantidad de conceptos por vocabulario son: Gender (n=2), LOINC, (n=28), SNOMED-CT (n=3), UCUM (n=11) y Visit (n=2).

Parámetro	Concepto CDM ¹	Dominio ²	Vocabulario	ID Concepto
Demográfico				
Masculino	MALE	Gender	Gender	8507
Femenino	FEMALE	Gender	Gender	8532
Clave Azul				
Fallecimiento por descompensación	Decompensation	Condition	SNOMED	4314553
Fallecimiento por otro motivo	Unknown problem	Condition	SNOMED	44790729
Alerta Clave Azul	Signposting to rapid response team (procedure)	Procedure	SNOMED	37208480
Traslado a unidad de cuidados intensivos	Intensive Care	Visit	Visit	32037
Hospitalización	Inpatient Visit	Visit	Visit	9201
Unidades				
pH	pH	Unit	UCUM	8482
Cuentas por minuto	counts per minute	Unit	UCUM	8483
Por ciento	percent	Unit	UCUM	8554
Unidad por litro	unit per liter	Unit	UCUM	8645
Gramo por decilitro	gram per deciliter	Unit	UCUM	8713
Miligramo por litro	milligram per liter	Unit	UCUM	8751
Milimoles por litro	millimole per liter	Unit	UCUM	8753
Miligramo por decilitro	milligram per deciliter	Unit	UCUM	8840
Mil por microlitro	thousand per microliter	Unit	UCUM	8848
Columna de mercurio milimétrica	millimeter mercury column	Unit	UCUM	8876
Grado Celsius	degree Celsius	Unit	UCUM	586323
Signos Vitales				
Presión Arterial Sistólica	Systolic blood pressure	Measurement	LOINC	3004249
Presión Arterial Diastólica	Diastolic blood pressure	Measurement	LOINC	3012888
Presión Arterial Media	Mean blood pressure	Measurement	LOINC	3027598
Frecuencia Cardíaca	Heart rate	Measurement	LOINC	3027018
Saturación Oxígeno	Oxygen saturation in Blood	Measurement	LOINC	3013502
Temperatura Corporal	Body temperature	Measurement	LOINC	3020891
Laboratorio				
Baciliformes	Band form neutrophils/100 leukocytes in Blood by Automated count	Measurement	LOINC	3035839
Bilirrubina Directa	Bilirubin.direct [Mass/volume] in Serum or Plasma	Measurement	LOINC	3027597
Bilirrubina Total	Bilirubin.total [Mass/volume] in Serum or Plasma	Measurement	LOINC	3024128
Nitrógeno ureico	Urea nitrogen [Mass/volume] in Venous blood	Measurement	LOINC	3027219
Creatinina	Creatinine [Mass/volume] in Blood	Measurement	LOINC	3051825
Exceso Base	Base excess in Arterial blood by calculation	Measurement	LOINC	3003396
Fosfatasa Alcalina	Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma	Measurement	LOINC	3035995
GGT (Gamma glutamil transpeptidasa)	Gamma glutamyl transferase [Enzymatic activity/volume] in Serum or Plasma	Measurement	LOINC	3026910
GOT (Transaminasa glutámico oxalacética)	Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma	Measurement	LOINC	3013721
GPT (Transaminasa glutámico pirúvica)	Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma	Measurement	LOINC	3006923
Bicarbonato (HCO ₃)	Bicarbonate [Moles/volume] in Arterial blood	Measurement	LOINC	3008152
Hemoglobina	Hemoglobin [Mass/volume] in Blood	Measurement	LOINC	3000963
Tiempo de protrombina	INR in Platelet poor plasma by Coagulation assay	Measurement	LOINC	3022217
Potasio (K ⁺)	Potassium [Mass/volume] in Blood	Measurement	LOINC	21490733
Ácido láctico	Lactate [Moles/volume] in Arterial blood	Measurement	LOINC	3018405
Lactato Deshidrogenasa (LDH)	Lactate dehydrogenase [Enzymatic activity/volume] in Serum or Plasma	Measurement	LOINC	3016436
Leucocitos	Leukocytes [# /volume] in Blood by Automated count	Measurement	LOINC	3000905
PCO ₂	Carbon dioxide [Partial pressure] in Arterial blood	Measurement	LOINC	3027946
PaO ₂	Oxygen [Partial pressure] in Arterial blood	Measurement	LOINC	3027801
Proteína C reactiva (PCR)	C reactive protein [Mass/volume] in Serum or Plasma	Measurement	LOINC	3020460
pH (pH)	pH of Blood	Measurement	LOINC	3010421
Recuento de plaquetas	Platelets [# /volume] in Blood by Automated count	Measurement	LOINC	3024929

¹ Traducción literal según su estándar.

² Dominios del OMOP CDM.

10. ANEXO B: Figuras

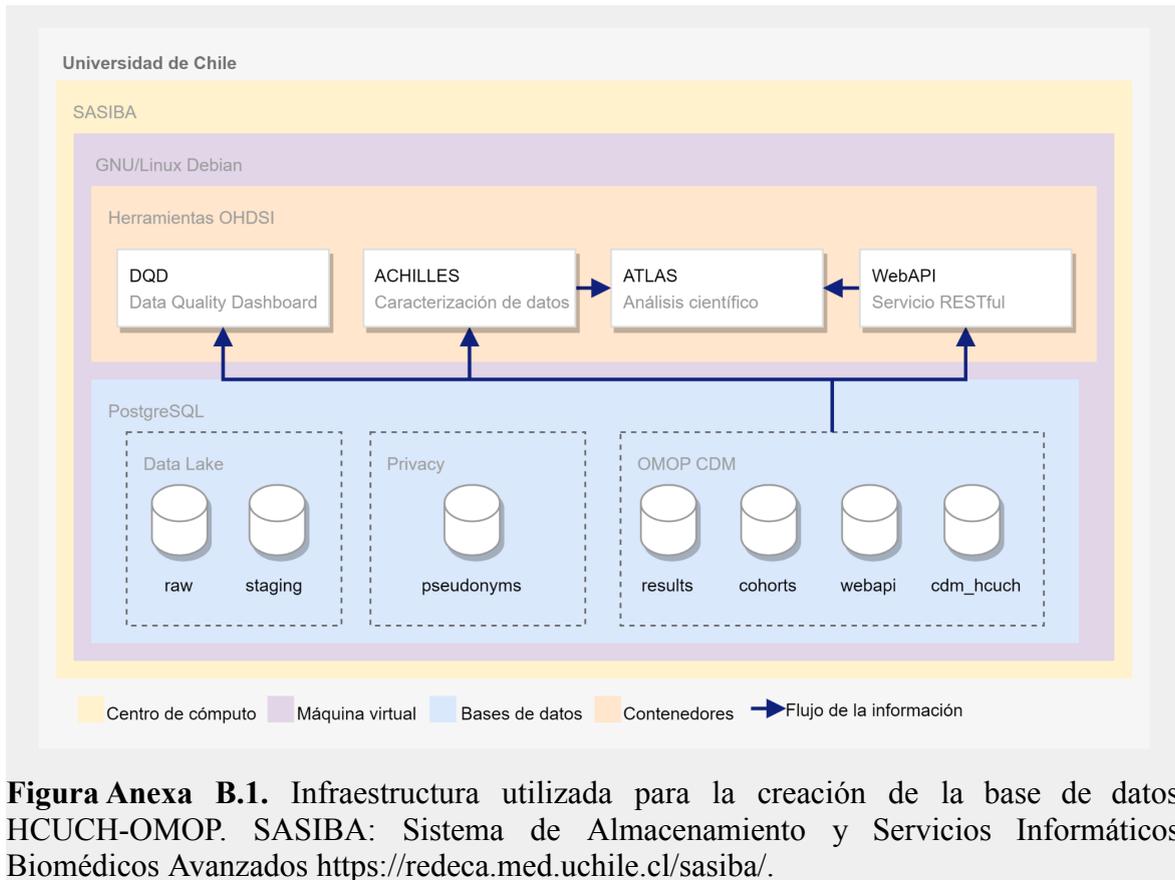


Figura Anexa B.1. Infraestructura utilizada para la creación de la base de datos HCUCH-OMOP. SASIBA: Sistema de Almacenamiento y Servicios Informáticos Biomédicos Avanzados <https://redec.med.uchile.cl/sasiba/>.

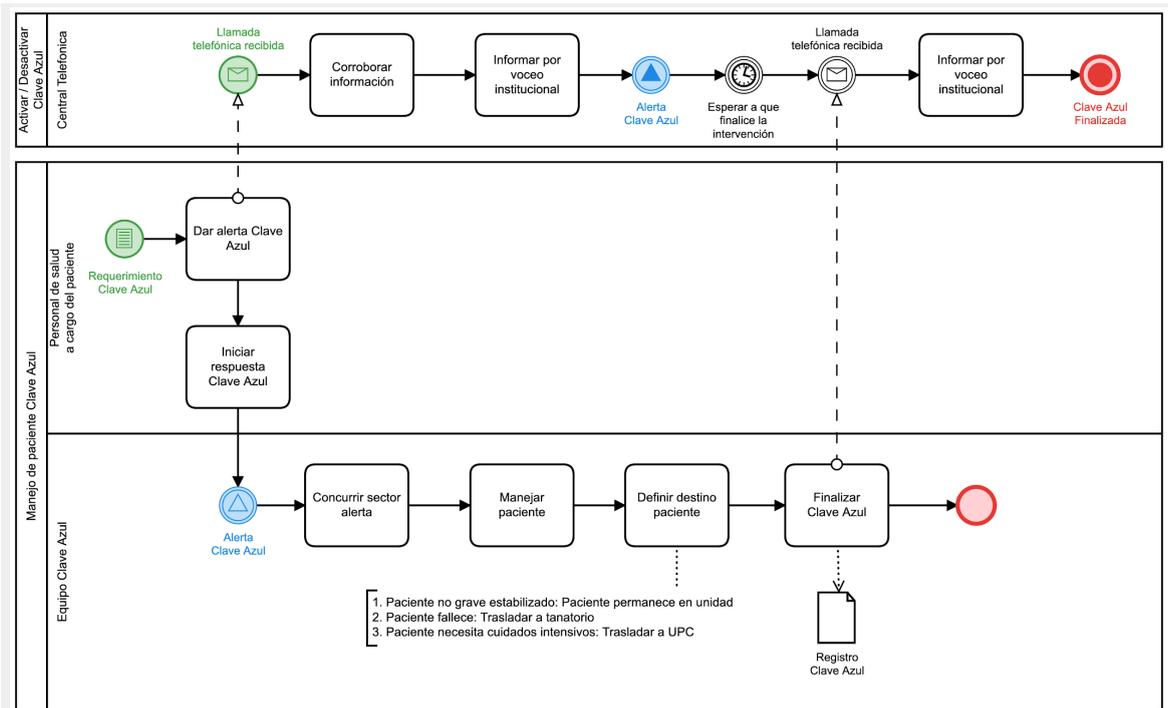


Figura Anexa B.2. Diagrama de procesos usando la notación Business Process Model and Notation (BPMN) Para dar cuenta de cómo es realizado el procedimiento Clave Azul dentro del Hospital Clínico Universidad de Chile (HCUCH). En este procedimiento incurren 3 roles: la central telefónica, el personal de salud a cargo del paciente, y el equipo de respuesta rápida Clave Azul. La elaboración del proceso se realizó en conjunto con el manual de procedimiento Clave Azul elaborado por el hospital [46].

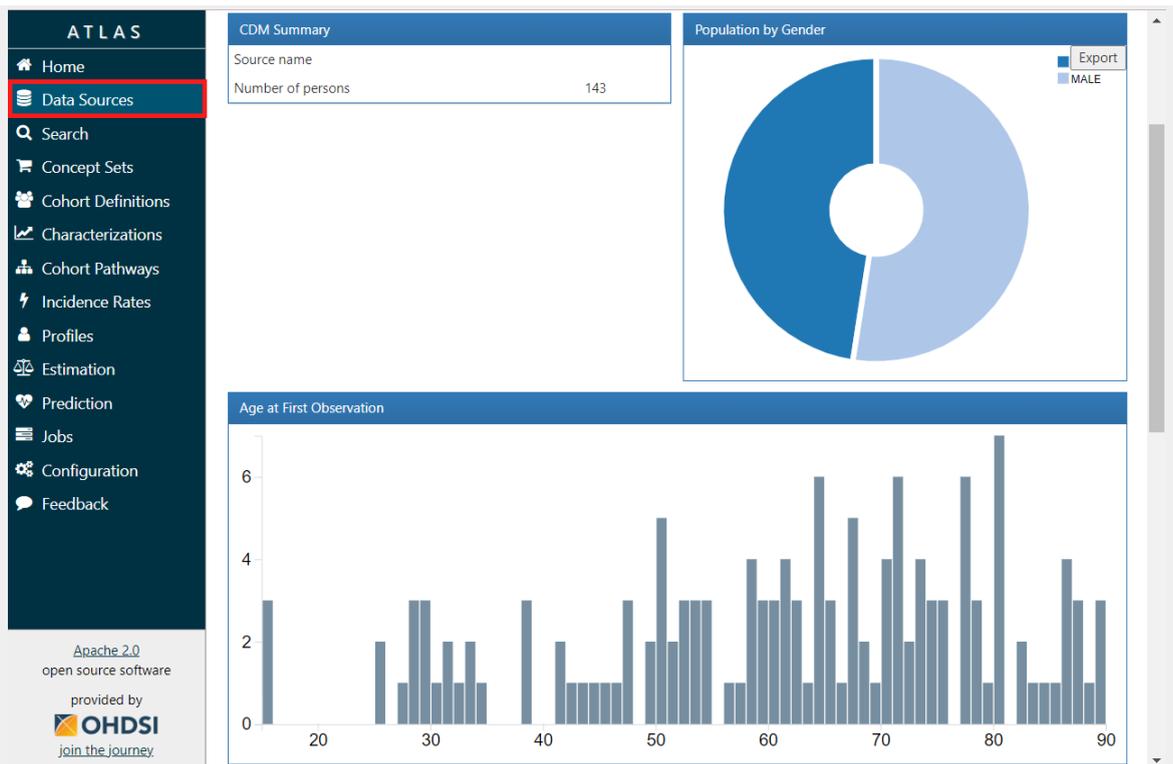


Figura Anexa B.3. Captura de pantalla de la sección de acceso al reporte Dashboard en la herramienta Atlas de análisis de datos. Navegación: Atlas > Data Sources > Report > Dashboard.

The screenshot shows the ATLAS web application interface. The left sidebar contains navigation options: Home, Data Sources (highlighted with a red box), Search, Concept Sets, Cohort Definitions, Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Jobs, Configuration, and Feedback. The main content area is titled 'Data Sources' and shows 'HCUCD CDM' and 'Measurement' selected. Below this is the 'HCUCD CDM Measurement Report' section, which includes a 'Prevalence' tab and a table view. The table displays 28 entries with columns for Concept Id, Name, Person Count, Prevalence, and Records per person. The table is currently showing 1 to 28 of 28 entries.

Concept Id	Name	Person Count	Prevalence	Records per person
3000963	Hemoglobin [Mass/volume] in Blood	139	97.20%	18.83
3035839	Band form neutrophils/100 leukocytes in Blood by Automated count	138	96.50%	16.33
3024929	Platelets [# /volume] in Blood by Automated count	138	96.50%	17.96
3000905	Leukocytes [# /volume] in Blood by Automated count	138	96.50%	16.49
21490733	Potassium [Mass/volume] in Blood	136	95.10%	24.74
3027219	Urea nitrogen [Mass/volume] in Venous blood	133	93.01%	16.96
3051825	Creatinine [Mass/volume] in Blood	131	91.61%	13.00
3035995	Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma	130	90.91%	17.86
3024128	Bilirubin.total [Mass/volume] in Serum or Plasma	130	90.91%	17.93
3013721	Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma	130	90.91%	17.90
3016436	Lactate dehydrogenase [Enzymatic activity/volume] in Serum or Plasma	126	88.11%	12.05

Figura Anexa B.4. Captura de pantalla de la sección de acceso al reporte Measurement en la herramienta Atlas de análisis de datos. Navegación: Atlas > Data Sources > Report > Measurement.

The screenshot shows the ATLAS interface. On the left is a dark sidebar with the ATLAS logo and navigation menu items: Home, Data Sources (highlighted with a red box), Search, Concept Sets, Cohort Definitions, Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Jobs, Configuration, and Feedback. The main content area is titled 'Data Sources' and shows 'HCUCM CDM' and 'Visit' selected in dropdown menus. Below this is the 'HCUCM CDM Visit Report' section, which has a 'Prevalence' header and a 'Table' view selected. The table has columns: Concept Id, Name, Person Count, Prevalence, and Records per person. A single row is displayed for Concept Id 9201, Name Inpatient Visit, with Person Count 143, Prevalence 100.00%, and Records per person 2.87. The interface also includes a search filter, 'Showing 1 to 1 of 1 entries', and navigation buttons like 'Previous' and 'Next'.

Concept Id	Name	Person Count	Prevalence	Records per person
9201	Inpatient Visit	143	100.00%	2.87

Apache 2.0
open source software
provided by
OHDSI
join the journey

Figura Anexa B.5. Captura de pantalla de la sección de acceso al reporte Visit en la herramienta Atlas de análisis de datos. Navegación: Atlas > Data Sources > Report > Visit.

The screenshot shows the ATLAS interface with the 'Data Sources' section selected. The main content area displays the 'HCUCM Procedure Report' for the 'Prevalence' view. The table below shows the data for the selected concept.

Concept Id	Name	Person Count	Prevalence	Records per person
37208480	Systematic Nomenclature of Medicine - Clinical Terms (IHTSDO) 1105741000000103: Signposting to rapid response team	143	100.00%	1.00

Figura Anexa B.6. Captura de pantalla de la sección de acceso al reporte Procedure en la herramienta Atlas de análisis de datos. Navegación: Atlas > Data Sources > Report > Procedure.

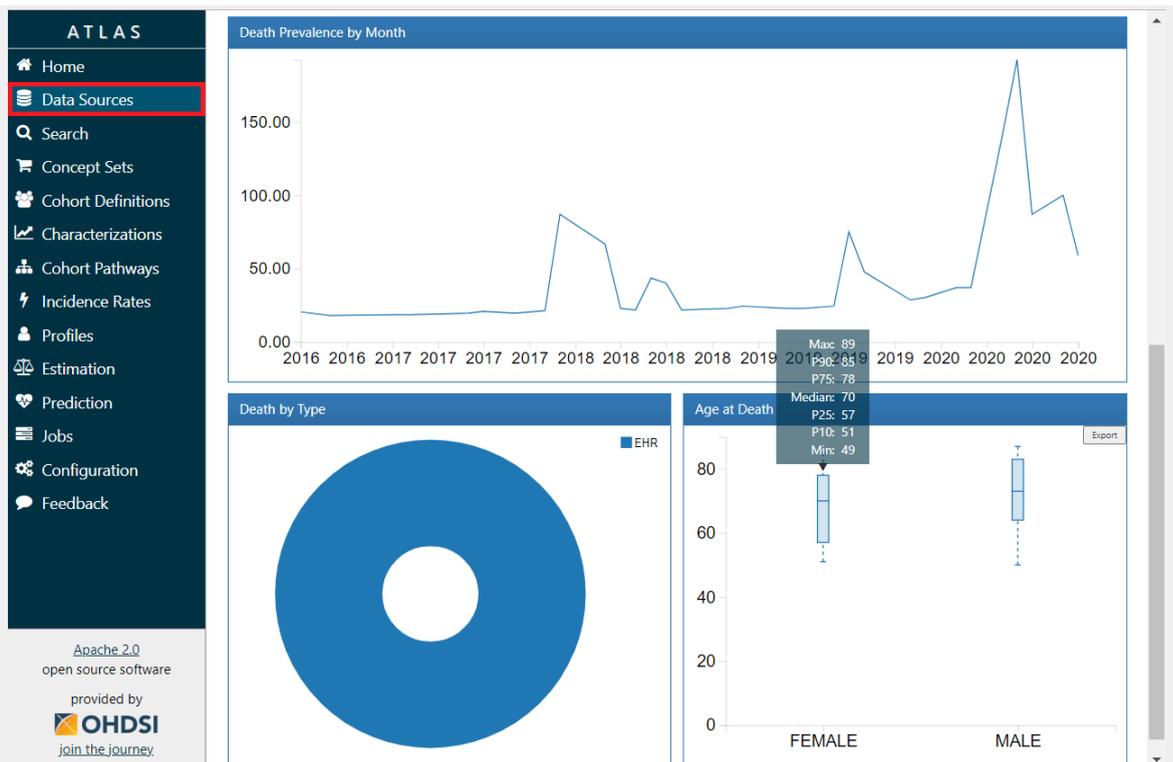


Figura Anexa B.7. Captura de pantalla de la sección de acceso al reporte Death en la herramienta Atlas de análisis de datos. Navegación: Atlas > Data Sources > Report > Death.

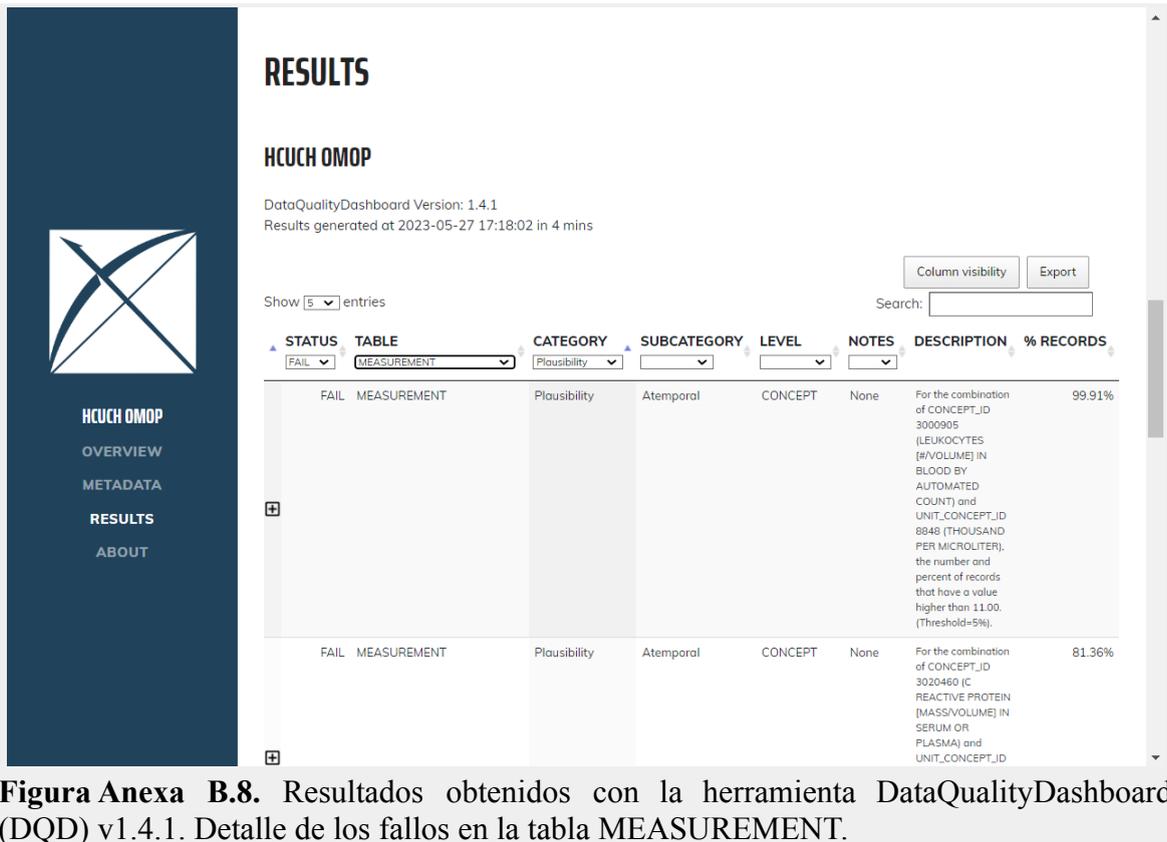


Figura Anexa B.8. Resultados obtenidos con la herramienta DataQualityDashboard (DQD) v1.4.1. Detalle de los fallos en la tabla MEASUREMENT.